Hands-on Introduction to AI Reservoir Modelling Workflows

Short Course

Professor Vasily Demyanov
Farah Rabie, Vitalii Starikov
geodatascience.hw.ac.uk

26 March 2025

Outline for the day

- AM:
  - Well data visualisation with Python (Goggle Colab)
  - Unsupervised learning for clustering wireline data – hands on
  - Supervised classification of lithofacies from wireline data – hands on
- PM:
  - Seismic segmentation with unsupervised learning – demo
  - Unsupervised pattern reignition of PTA well data – hands on
  - Time-lapse PTA monitoring of well dynamic performance – demo

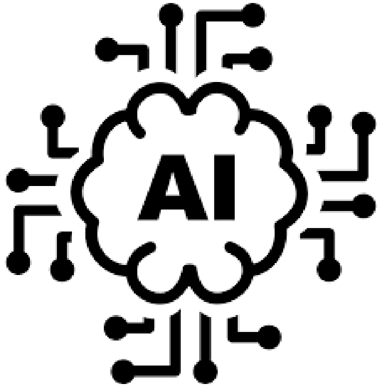
Learning Objectives

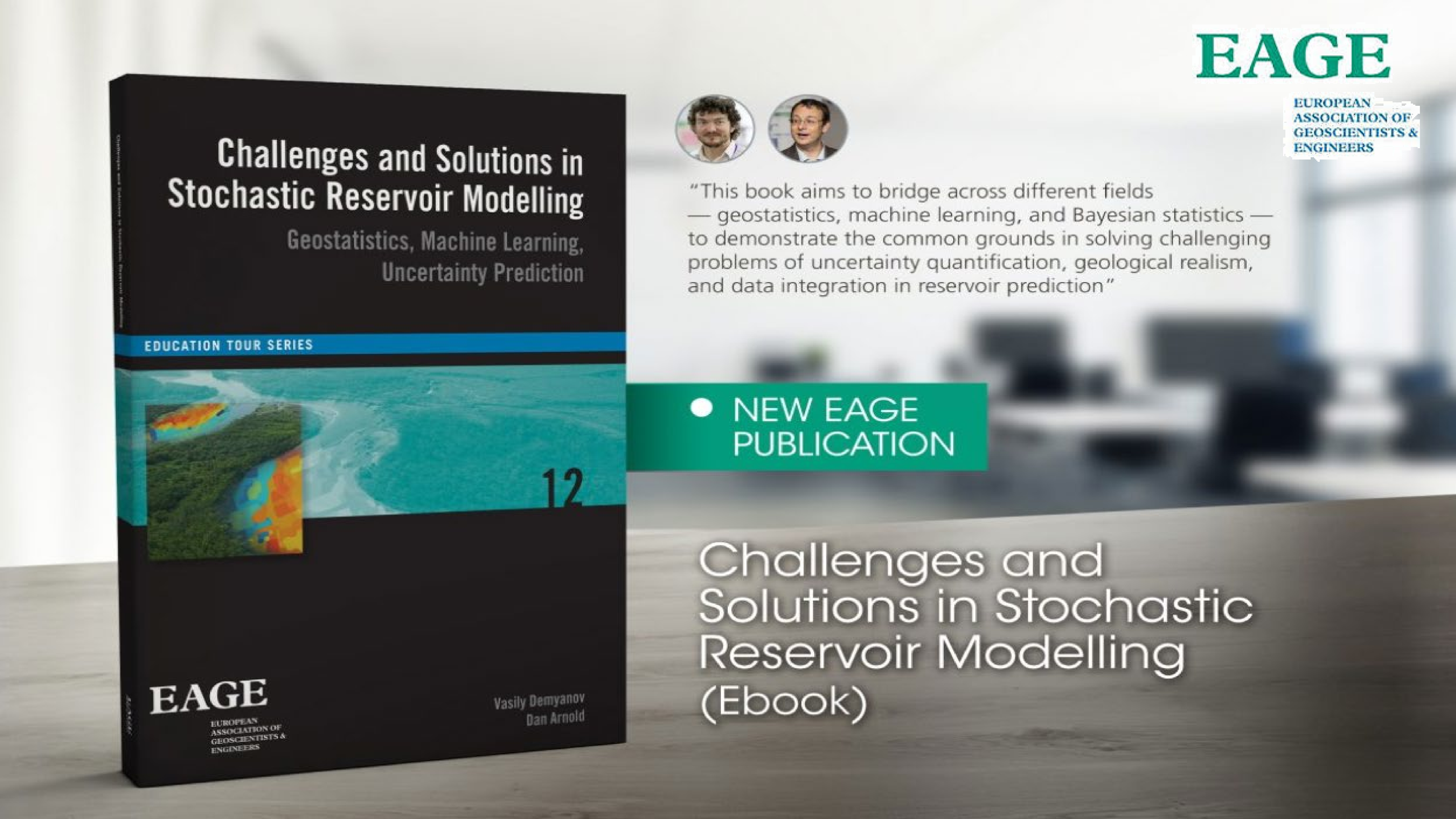
- Get practical introduction of basic concepts of ML – supervised and unsupervised learning.
- Gain a hands on experience with Python workflows in Google Colab using real open access reservoir data (Volve)
- Practice unsupervised clustering with wireline data, compare with interpreted lithologies.
- Practice supervised classification of lithofacies from wireline data, compare different classification methods.
- Practice pattern recognition of well production PTA data for evaluating well productivity characteristics and monitoring well performance.

Geo Data Science & UQ Group

- Use AI to discover patterns in reservoir data to improve our understanding of the subsurface for better management of Earth resources

- Our integrated team implements best in-class data science solutions across all subsurface disciplines to leverage data for reservoir modelling, forecasting and optimisation





EAGE
EUROPEAN ASSOCIATION OF GEOSCIENTISTS & ENGINEERS
Challenges and Solutions in Stochastic Reservoir Modelling
Geostatistics, Machine Learning, Uncertainty Prediction
EDUCATION TOUR SERIES
12
EAGE
EUROPEAN ASSOCIATION OF GEOSCIENTISTS & ENGINEERS
Vasily Demyanov
Dan Arnold
"This book aims to bridge across different fields — geostatistics, machine learning, and Bayesian statistics — to demonstrate the common grounds in solving challenging problems of uncertainty quantification, geological realism, and data integration in reservoir prediction"
NEW EAGE PUBLICATION
Challenges and Solutions in Stochastic Reservoir Modelling
(Ebook)

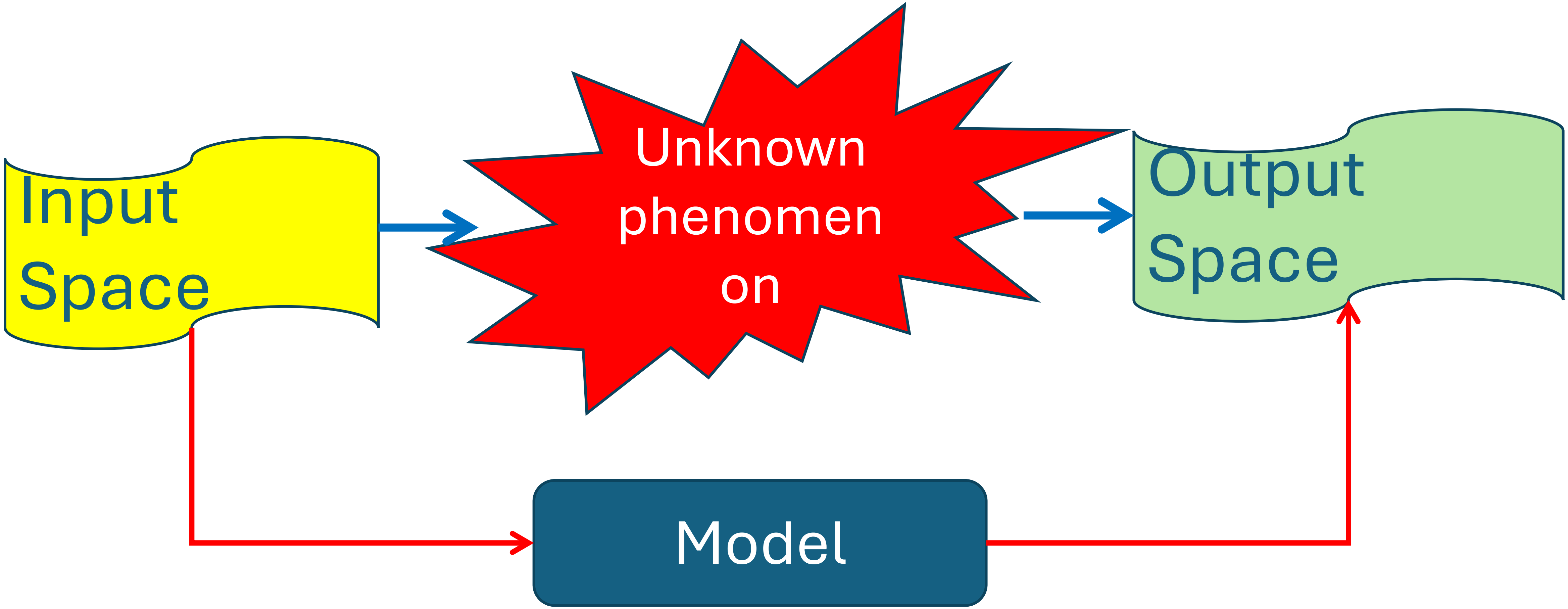
Brief introduction to machine leanring

- Learning from data
- Supervised vs unsupervised
- Overfitting

XXI Century Landscape

- Increasing digital data volumes
- Accumulated domain knowledge expertise in oil & gas
- Fast growing data science technology proved its value in many data rich environments
- Generational change in perception and ethos towards digital environment and AI

Generic Modelling Task – Predictive learning



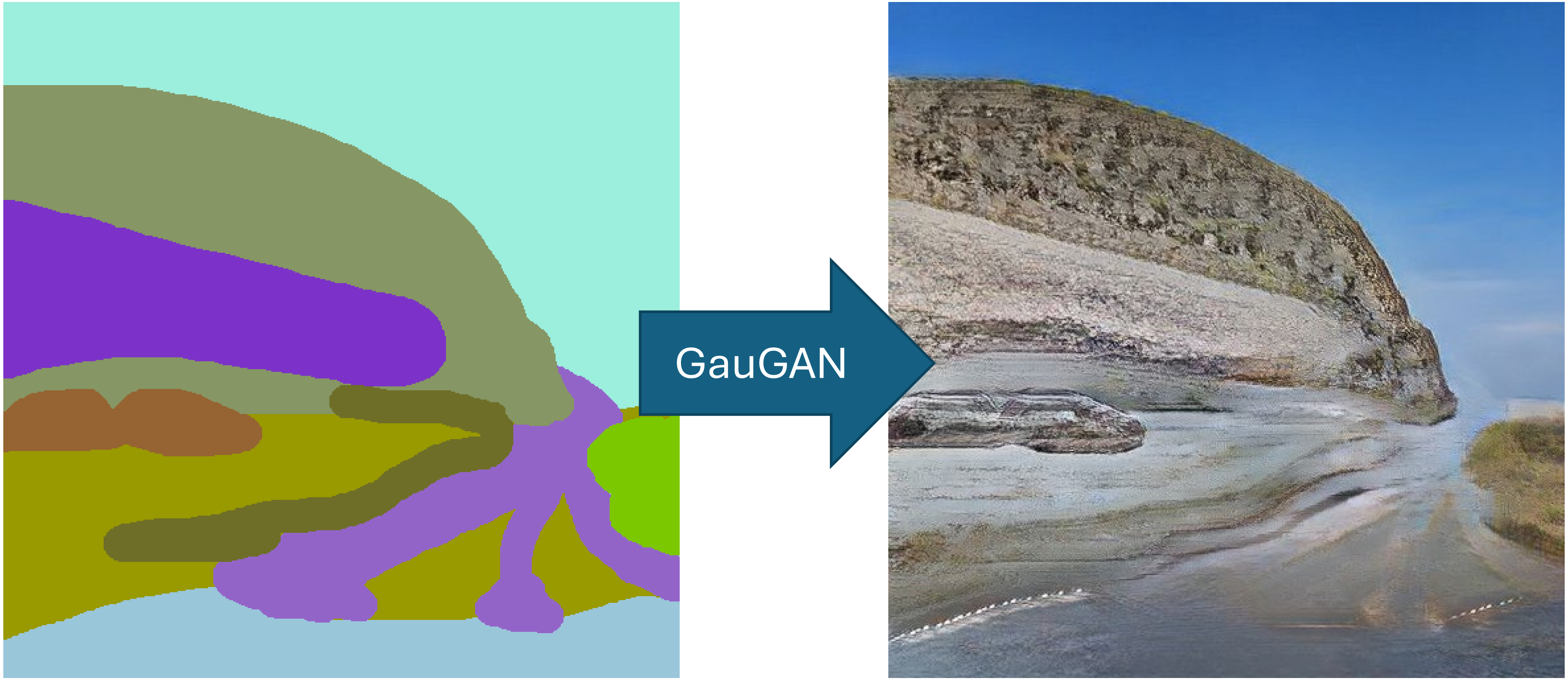
The algorithmic modelling culture, L. Breiman

Why Machine Leaning?

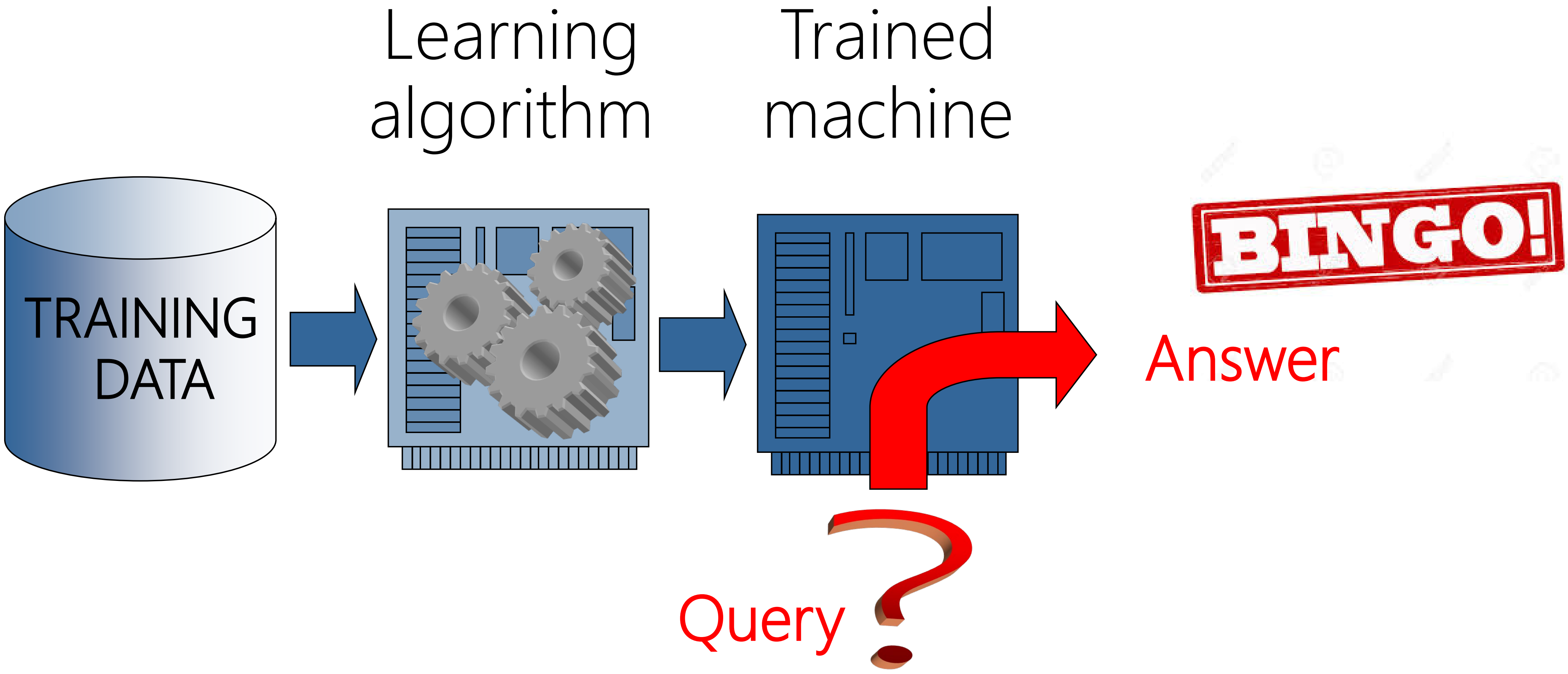
How can we build computer systems
that automatically improve with experience,
and what are the fundamental laws
that govern all learning processes?

Tom Mitchell, Machine Learning, McGraw Hill, 1997

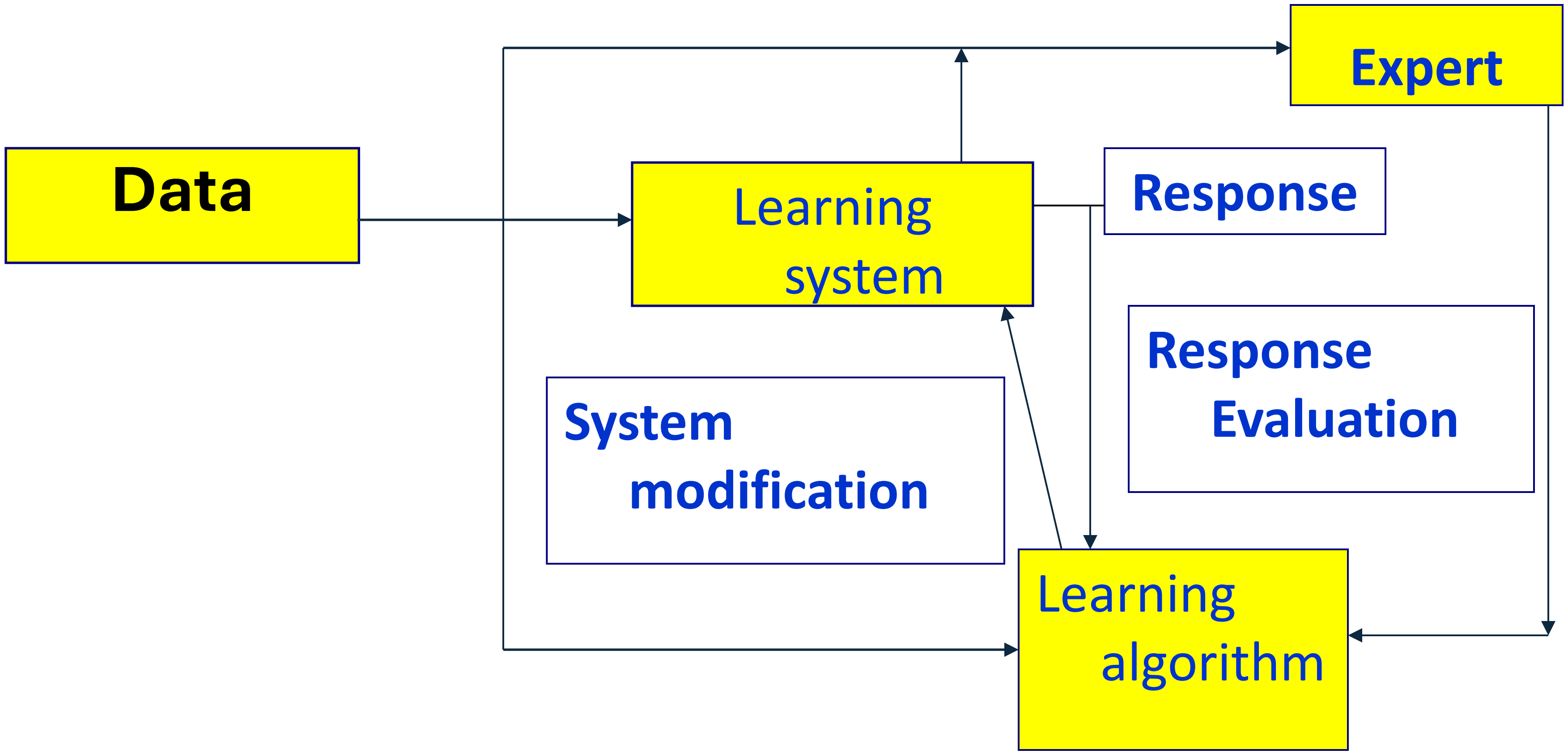
From a concept to a realistic image



What constitutes Machine Learning?

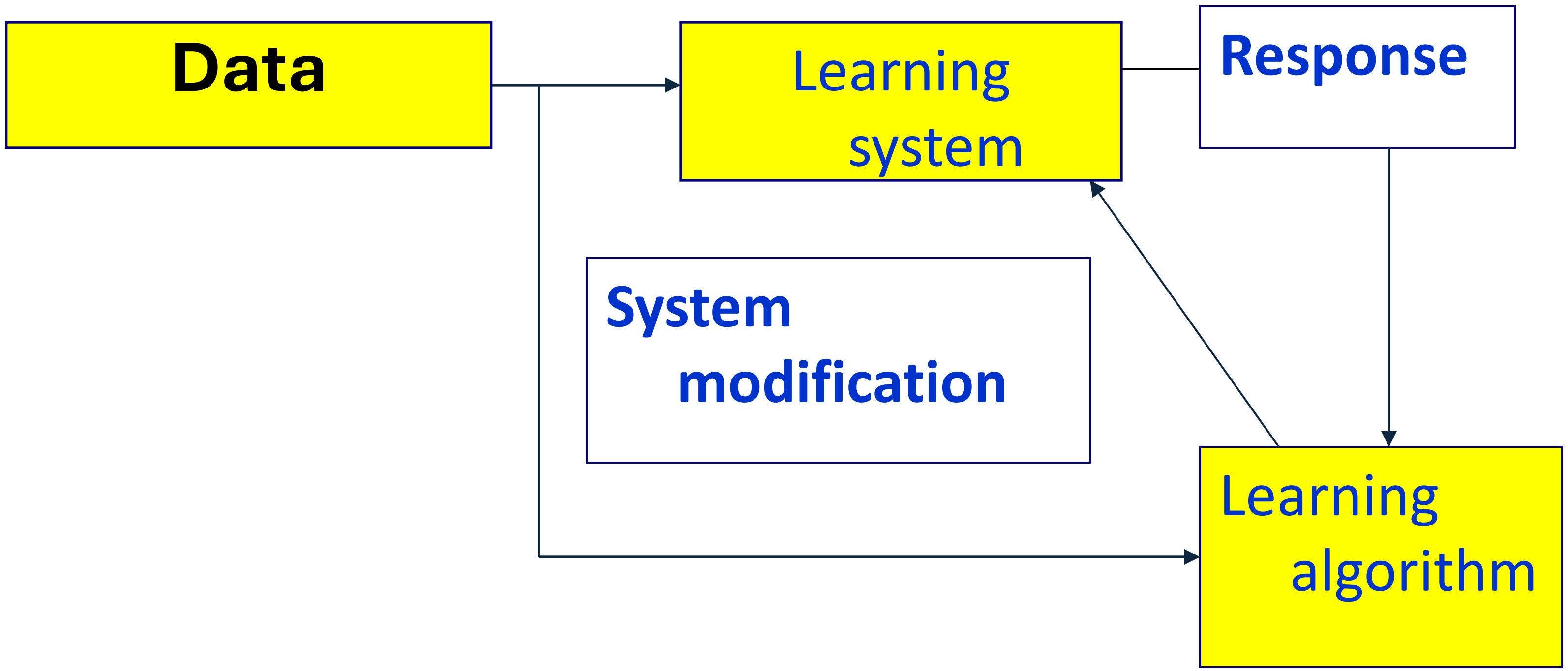


Supervised Learning



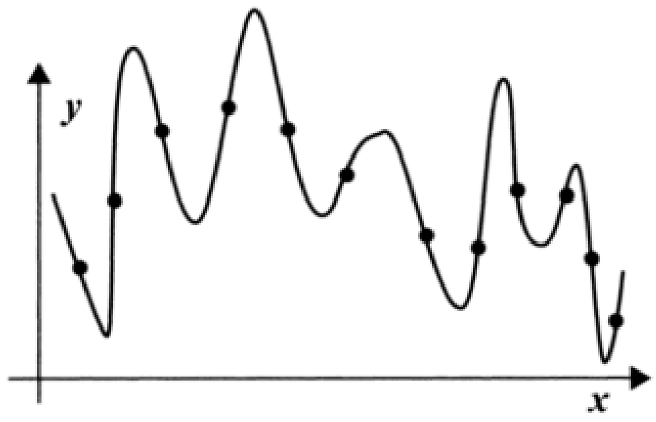
M. Kanevski

Unsupervised Learning



M. Kanevski

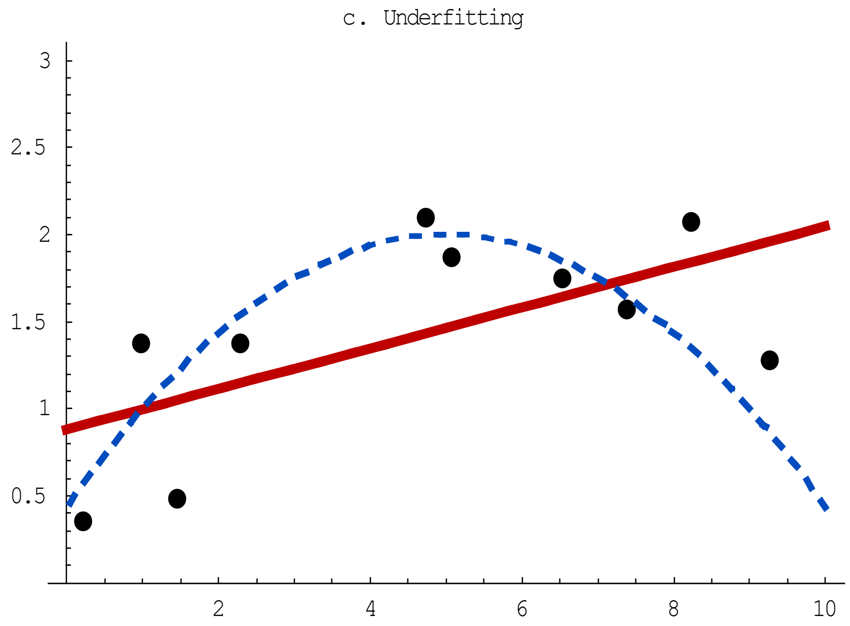
Overfitting

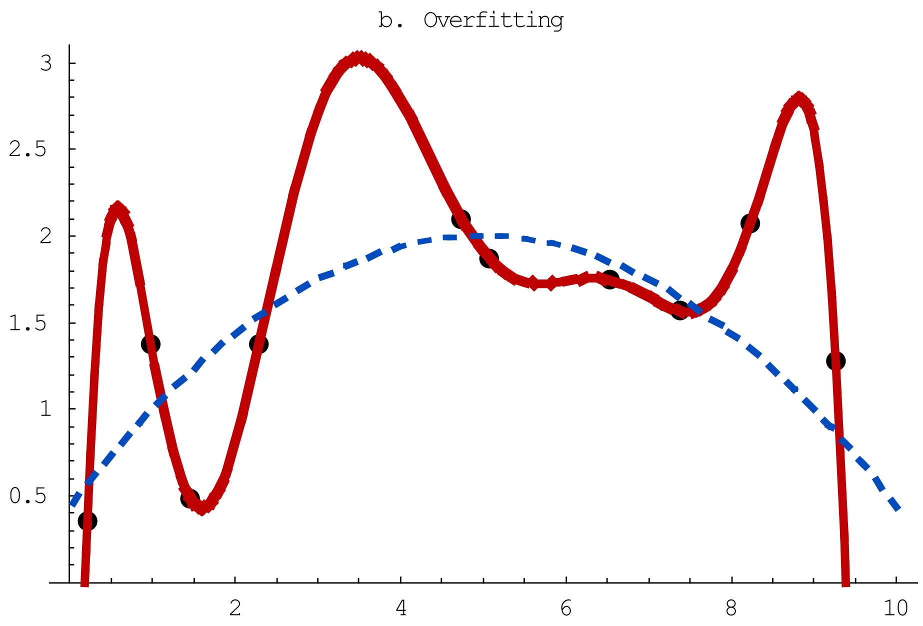


Problem of data-driven modelling: Overfitting

"With four parameters I can fit an elephant
and with five I can make him wiggle his trunk".

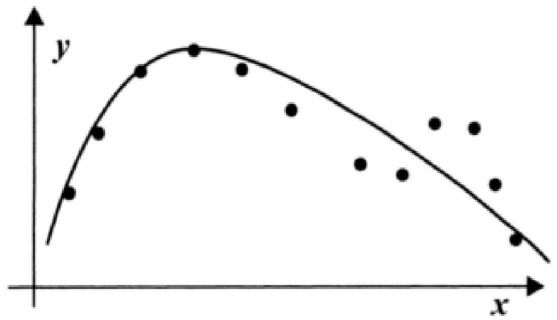
J. von Neumann



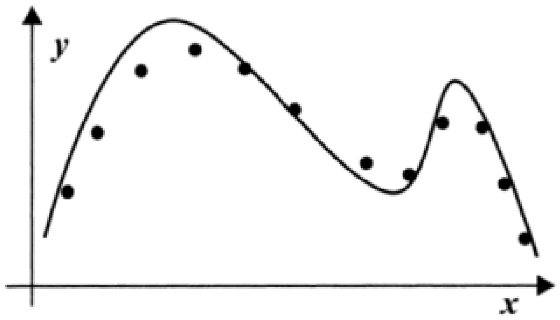


Model Choice in a Regression Problem

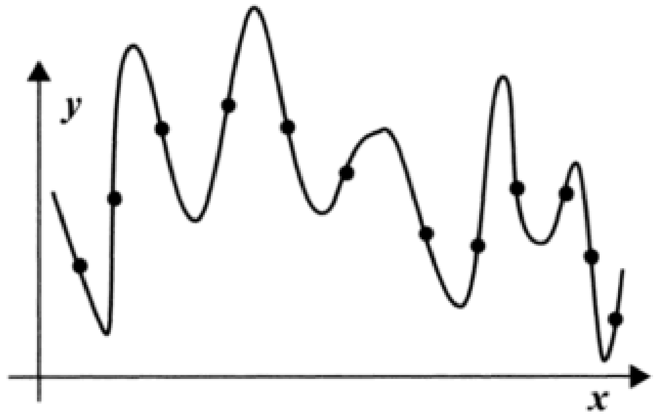
Simple model
- low variability
- oversmoothing (underfitting)



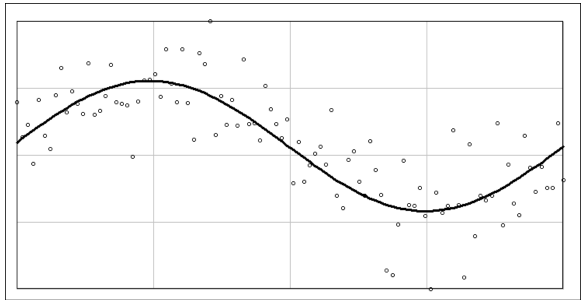
Optimal model
- Balanced complexity and data match



Highly complex model
- Matches the data exactly (overfitting)
- High variability

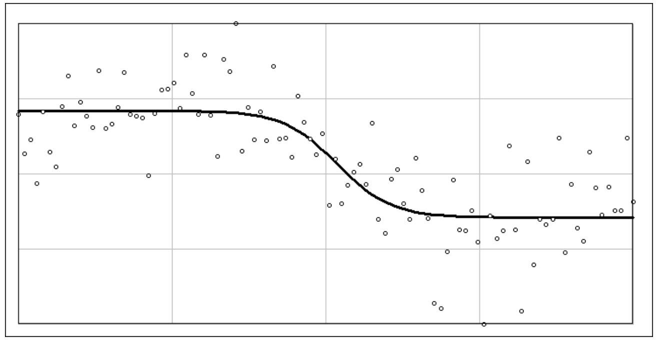


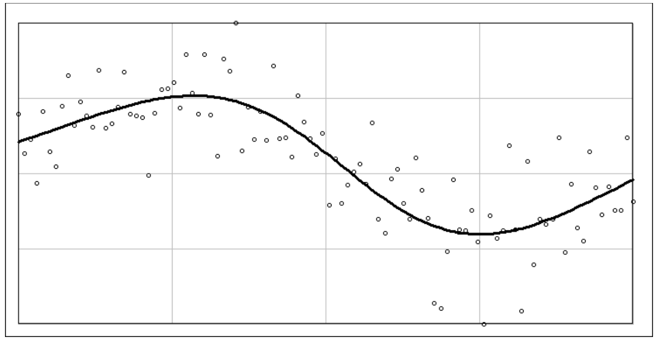
Overfitting

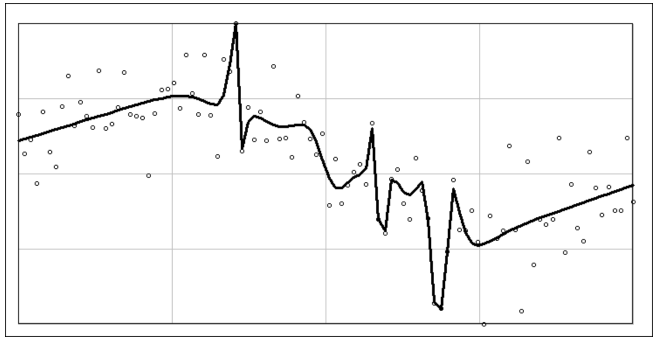


Reference

Underfitting







Overfitting

William Occam (1285 - 1349)

- Occam razor:

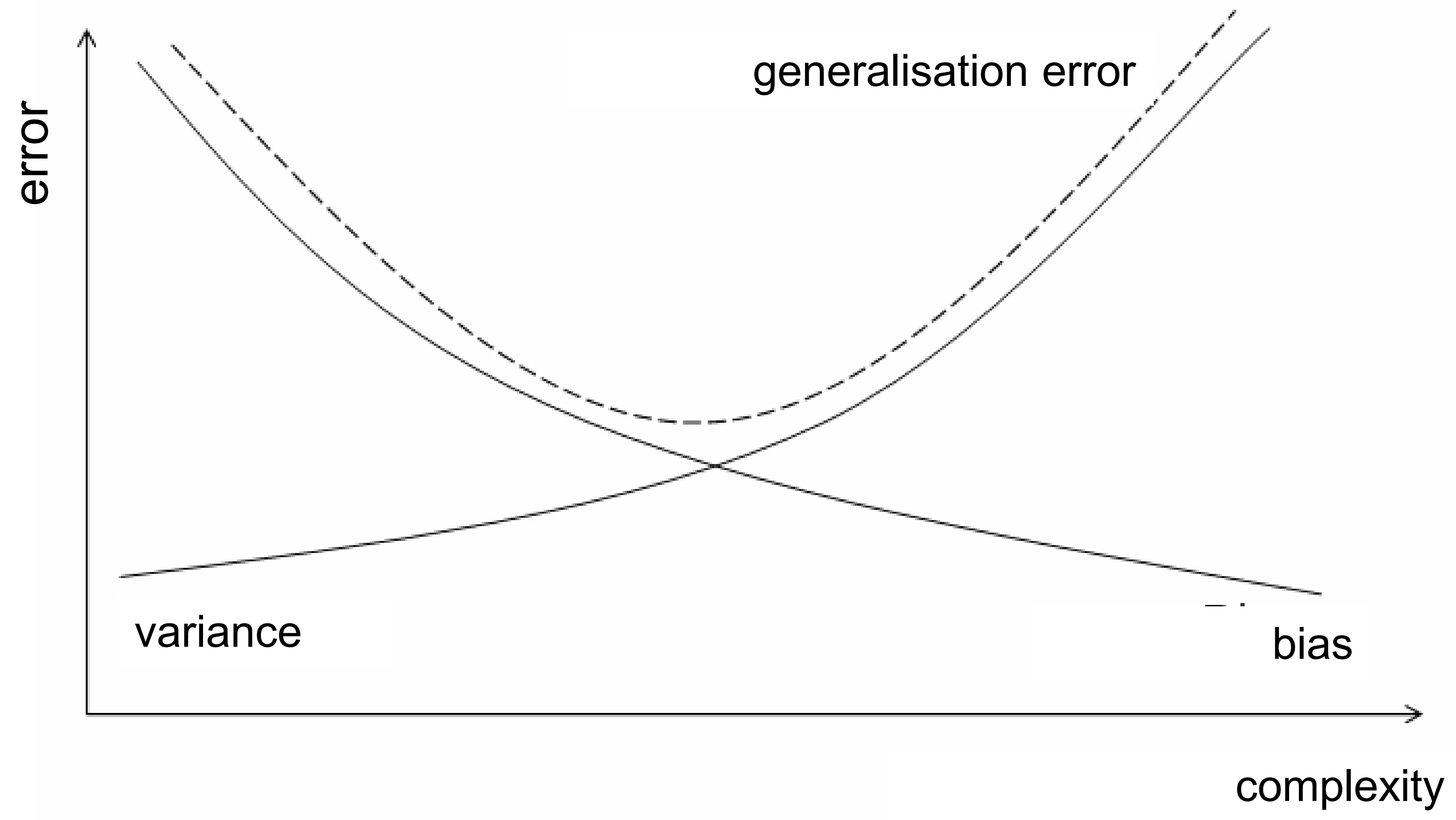
“Pluralitas non est ponenda sine necessitate”

«Plurality must never be positited without necessity»

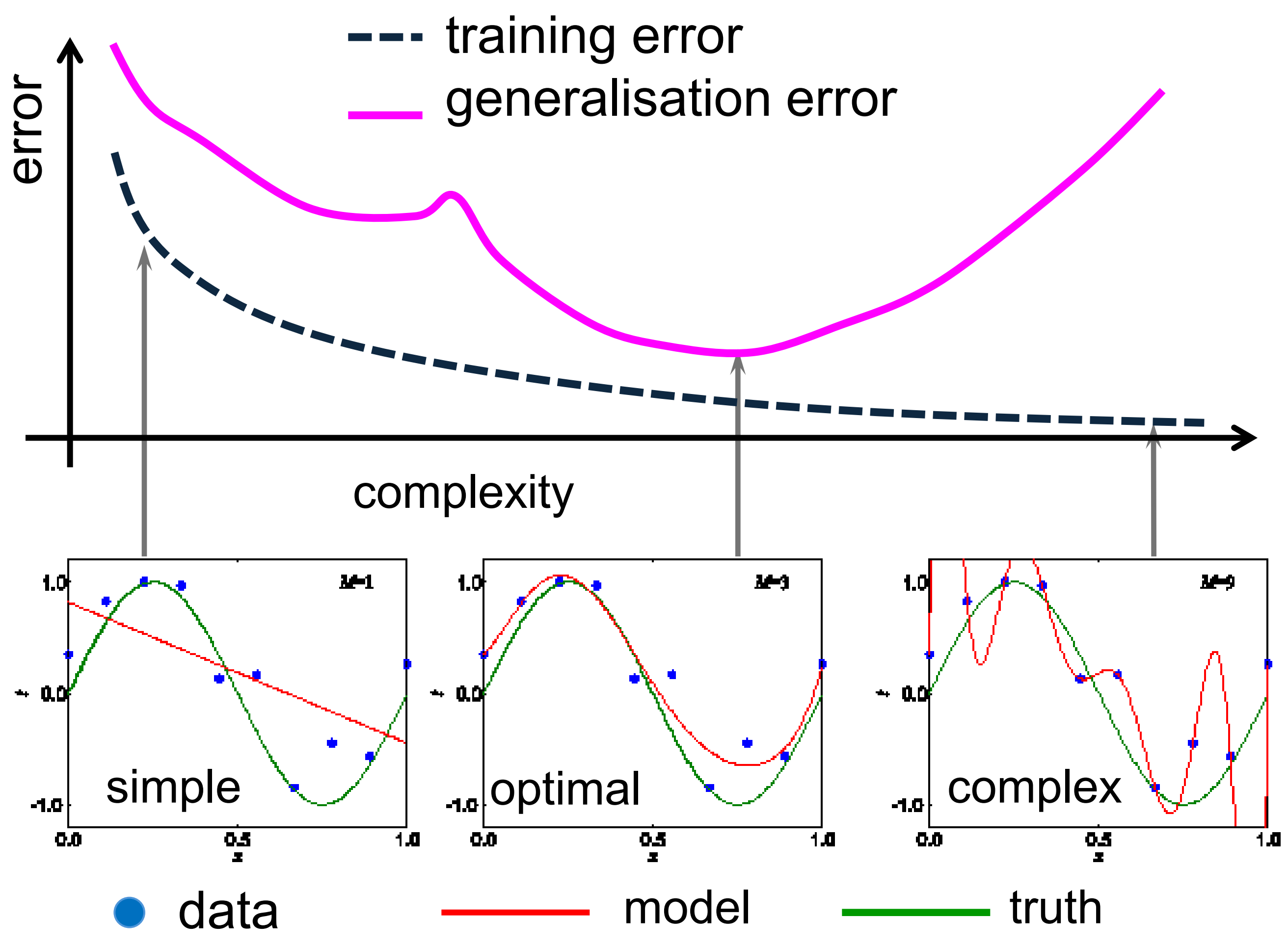
- *Wiki*

“one should proceed to simpler theories until simplicity can be traded for greater explanatory power”

Learning and generalisation



Prediction Errors and Complexity



Over-learning

- Overfitting is one of the most important problems related to learning-based algorithms is
- Estimates well known data used for training
- But is not able to generalise and accurately predict previously unseen data
- Problem of complex over-determined models
- *Solution: balance bias and variance of the estimate*
  - Divide data into training and testing(validation) sets.
  - Test (validation) set is used to control the model's ability to generalise.

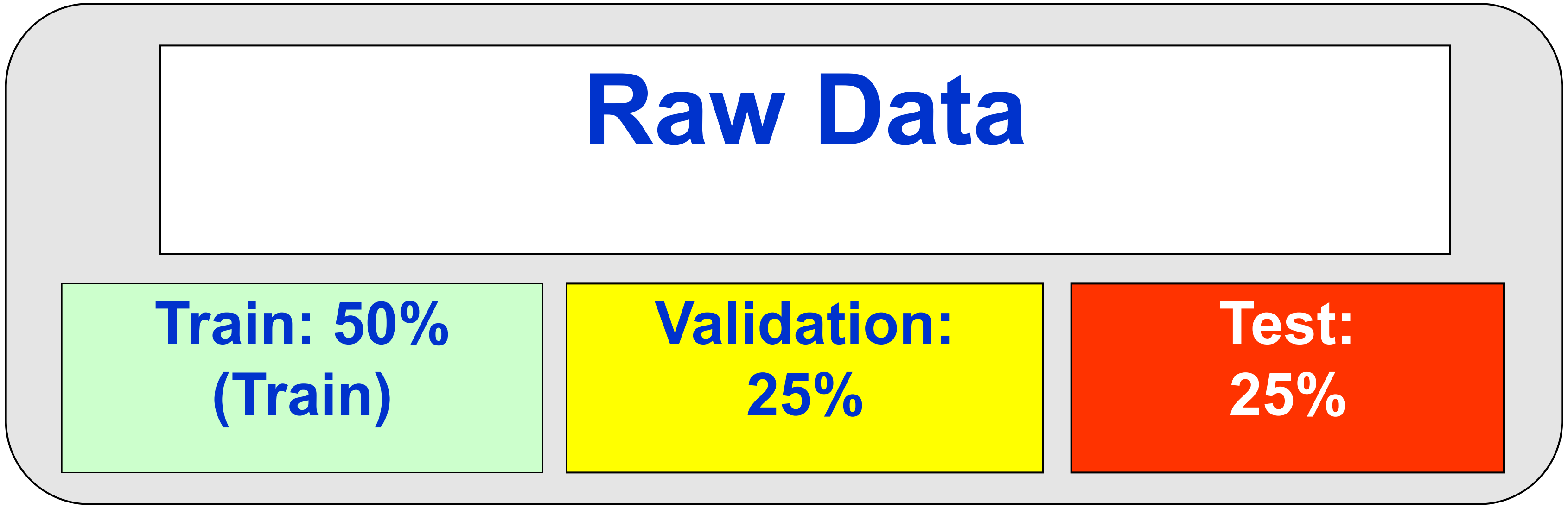
Data sub-set Purpose

- The **training** set is used to fit the models
- The **validation** set is used to estimate prediction error for model selection (tuning hyperparameters)
- The **test** set is used for assessment of the generalization error of the final chosen model

The Elements of Statistical Learning- Hastie, Tibshirani & Friedman 2009

If we are in a data-poor situation – use cross-validation/leave-one-out

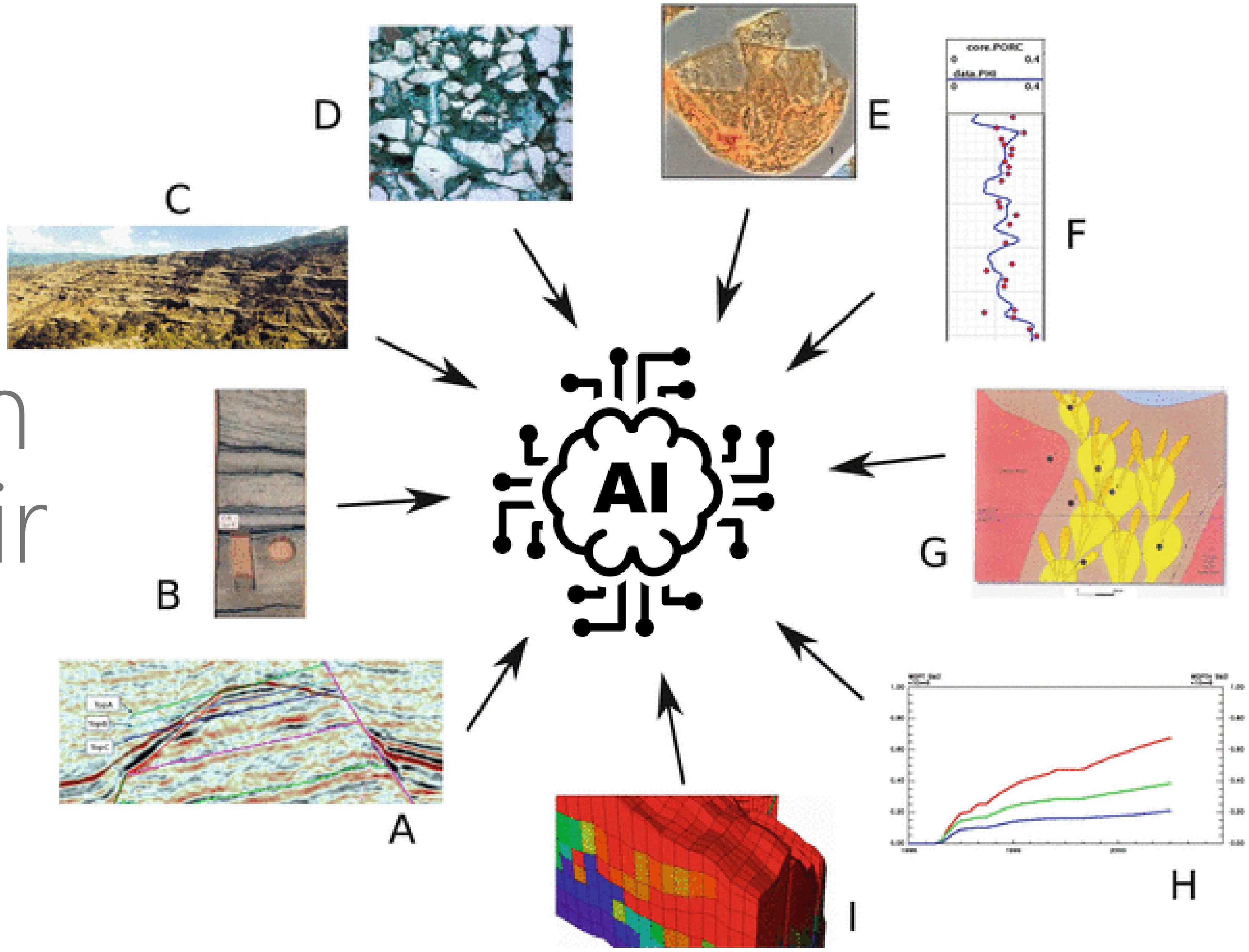
If we are in a data-rich situation – split data randomly (?)



Hands-on AI application to wireline data

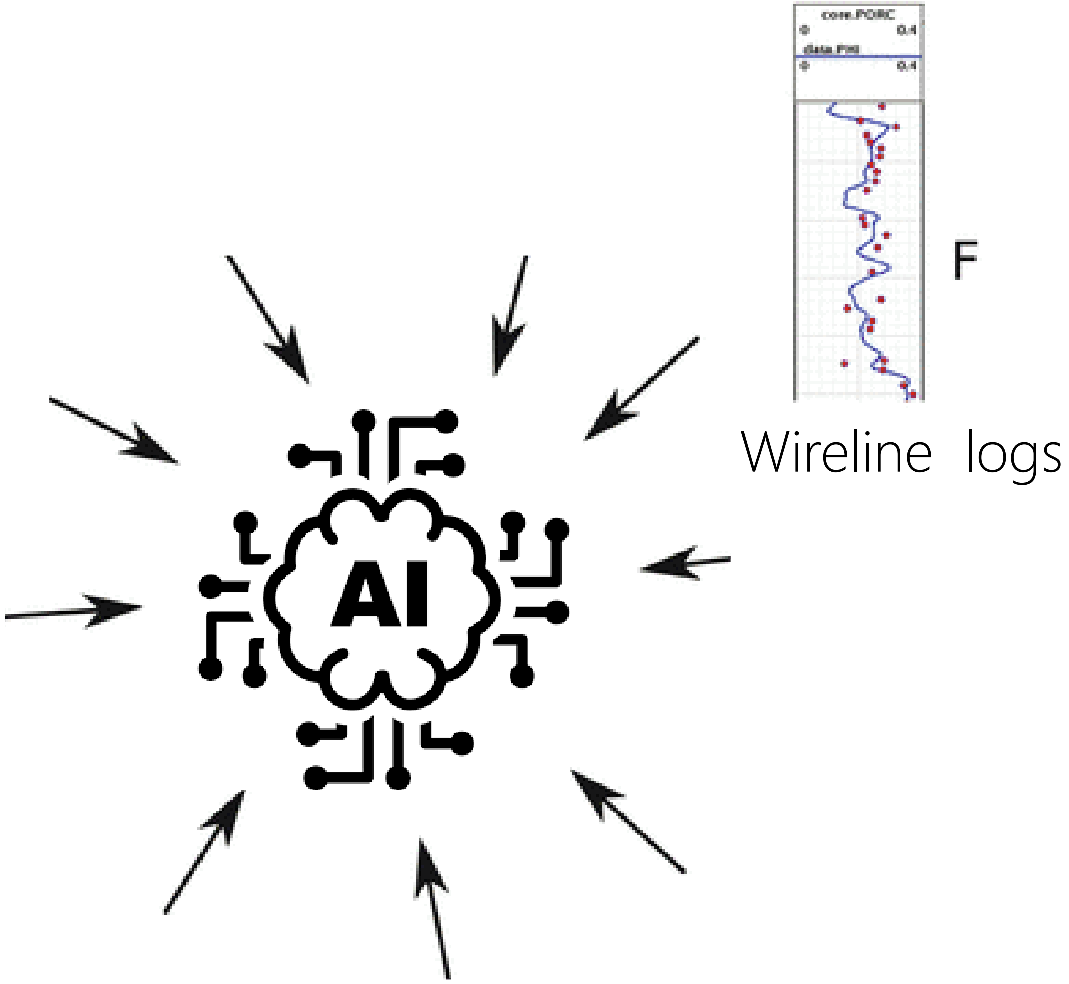
- Volve field
- Unsupervised clustering
- Supervised classification

Data integration in reservoir prediction modelling



Modified from Rivenæs, J.C., Sørhaug, P., Knarud, R. (2015). Introduction to Reservoir Modelling. In: Bjørlykke, K. (eds) Petroleum Geoscience. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-34132-8_22

Data integration in reservoir prediction modelling

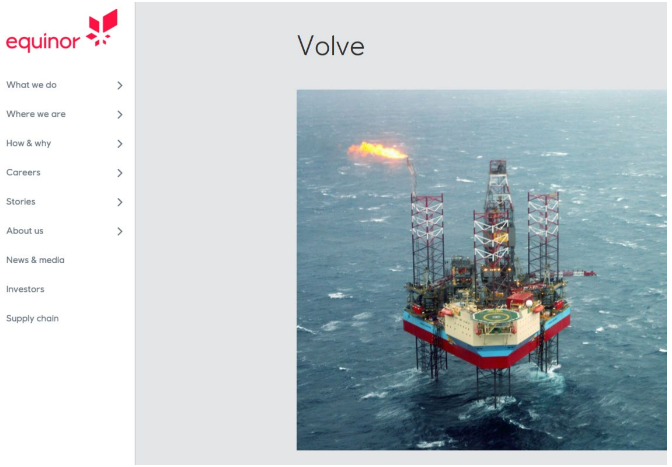


Modified from Rivenæs, J.C., Sørhaug, P., Knarud, R. (2015). Introduction to Reservoir Modelling. In: Bjørlykke, K. (eds) Petroleum Geoscience. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-34132-8_22

Volve Field

- The Volve oil field is located 200 kilometers west of Stavanger at the southern end of the Norwegian sector (central part of the North Sea).

- The reservoir is located at depths ranging from 2750m to 3120 m (water depth 80m).

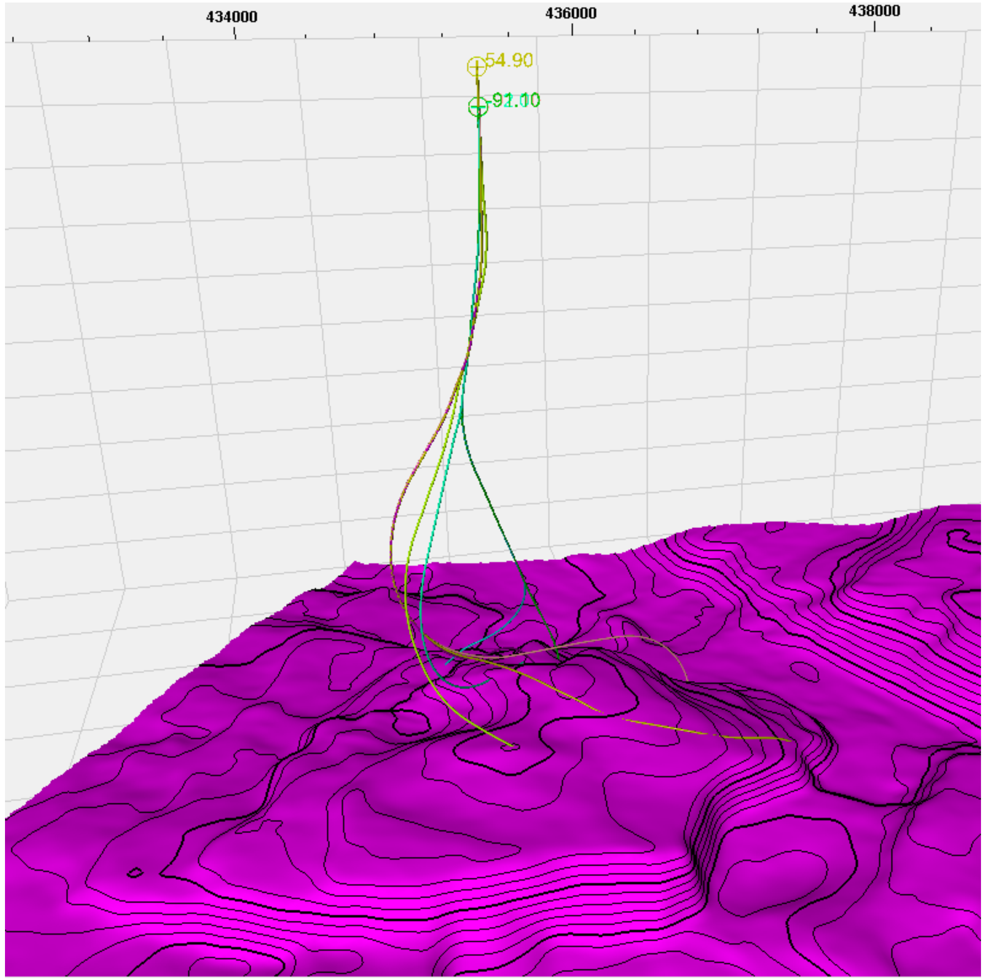


Geological context

- The environment of deposition was interpreted by Equinor to be a shallow marine mouth bar with influence of fluvio-deltaic.
- The oil was produced from sandstone in the Late Middle to Early Upper Jurassic Hugin Formation.

Data description- Volve

- Interpreted Wireline logs
- Core photos
- Seismic
- Simulation model (ECLIPSE)
- Geo-model (RMS)
- Production Data



Well locations

Well Names
- 15/9-F-1 A
- 15/9-F-1 B
- 15/9-F-1 C
- 15/9-F-11 A
- 15/9-F-11 B
- 15/9-F-11 T2
- 15/9-F-12

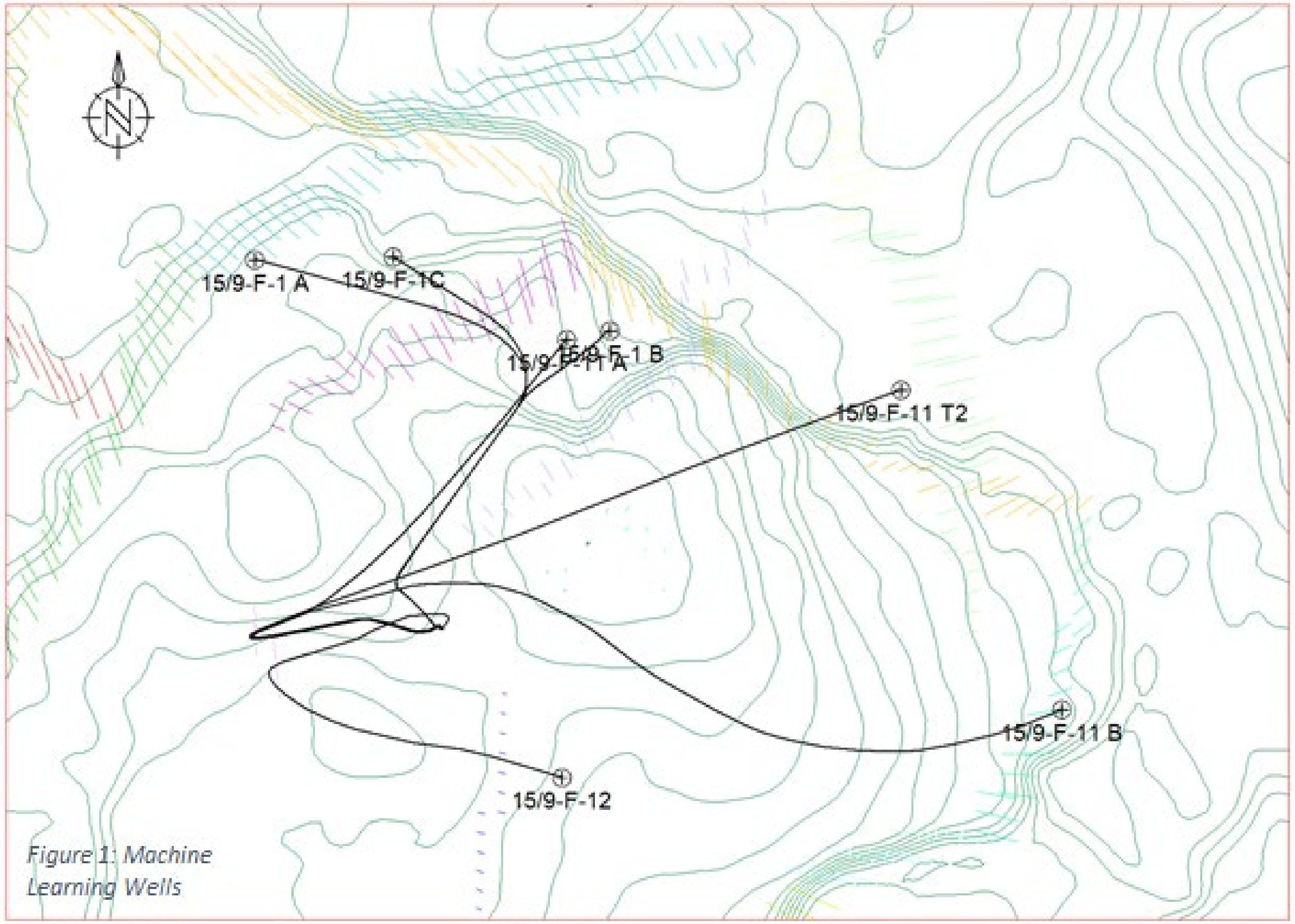
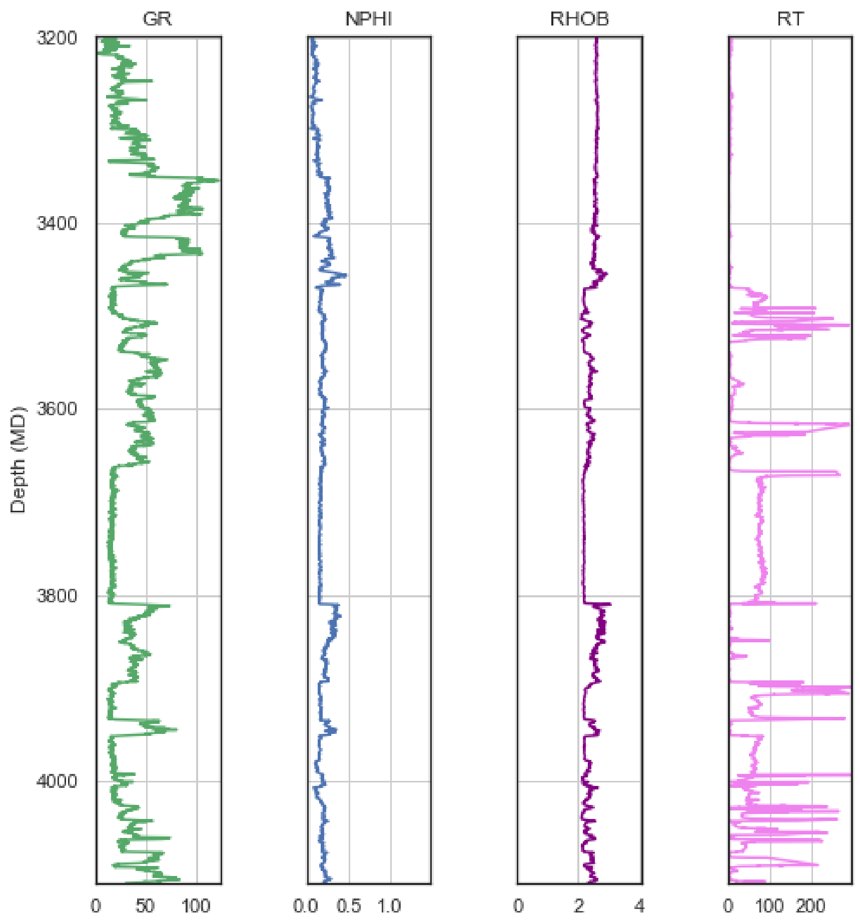
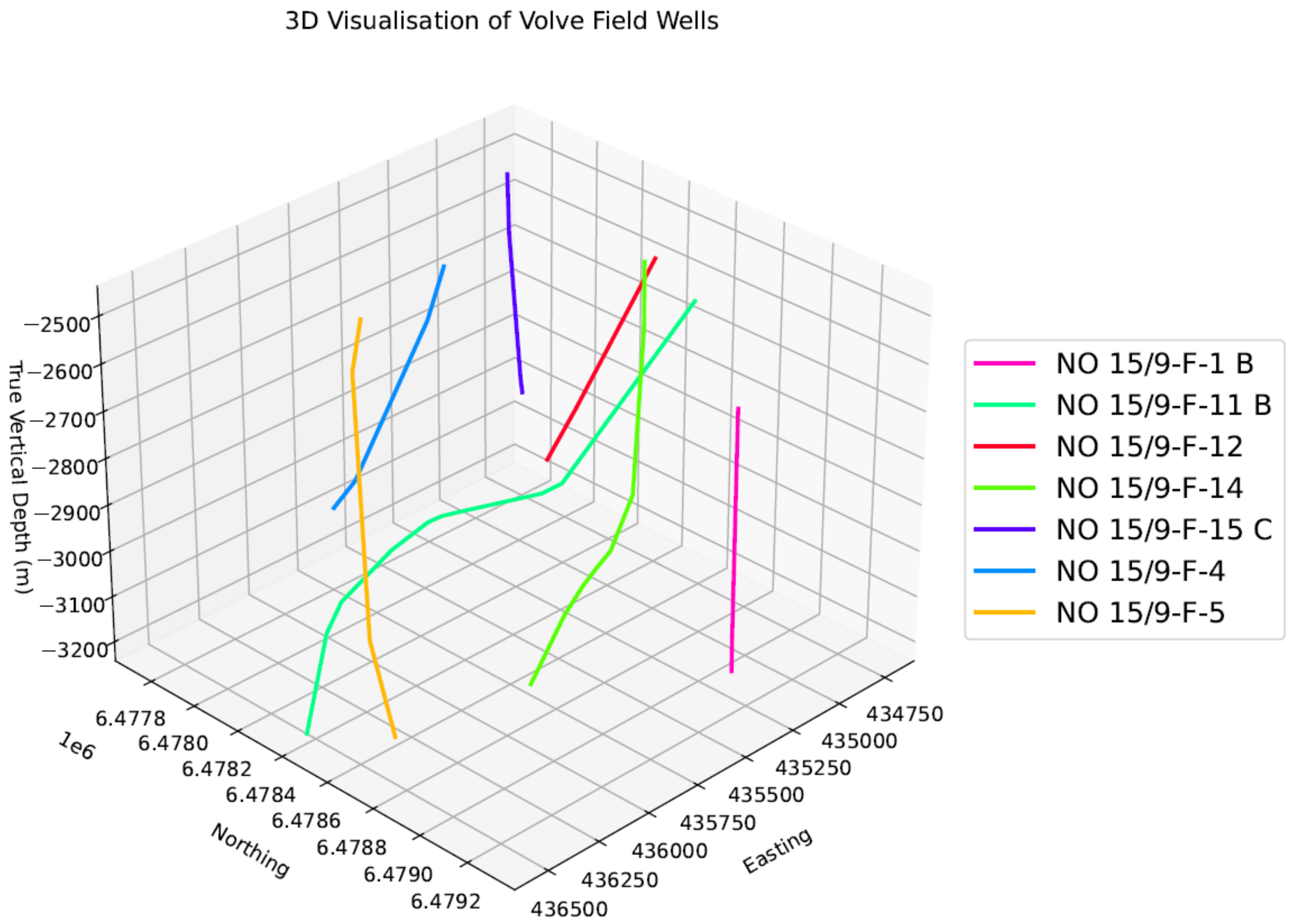


Figure 1: Machine Learning Wells

Wireline logs from six wells

NPHI DT GR PEF RHOB RT BVW VSH KLOGH





Data Visualisation

Hands-on

Data visualisation. Notebook 1

https://colab.research.google.com/drive/1hmBqXKjBpnfH3wIN5t80Zw0RXz98fRk9?usp=sharing

Exercise 1

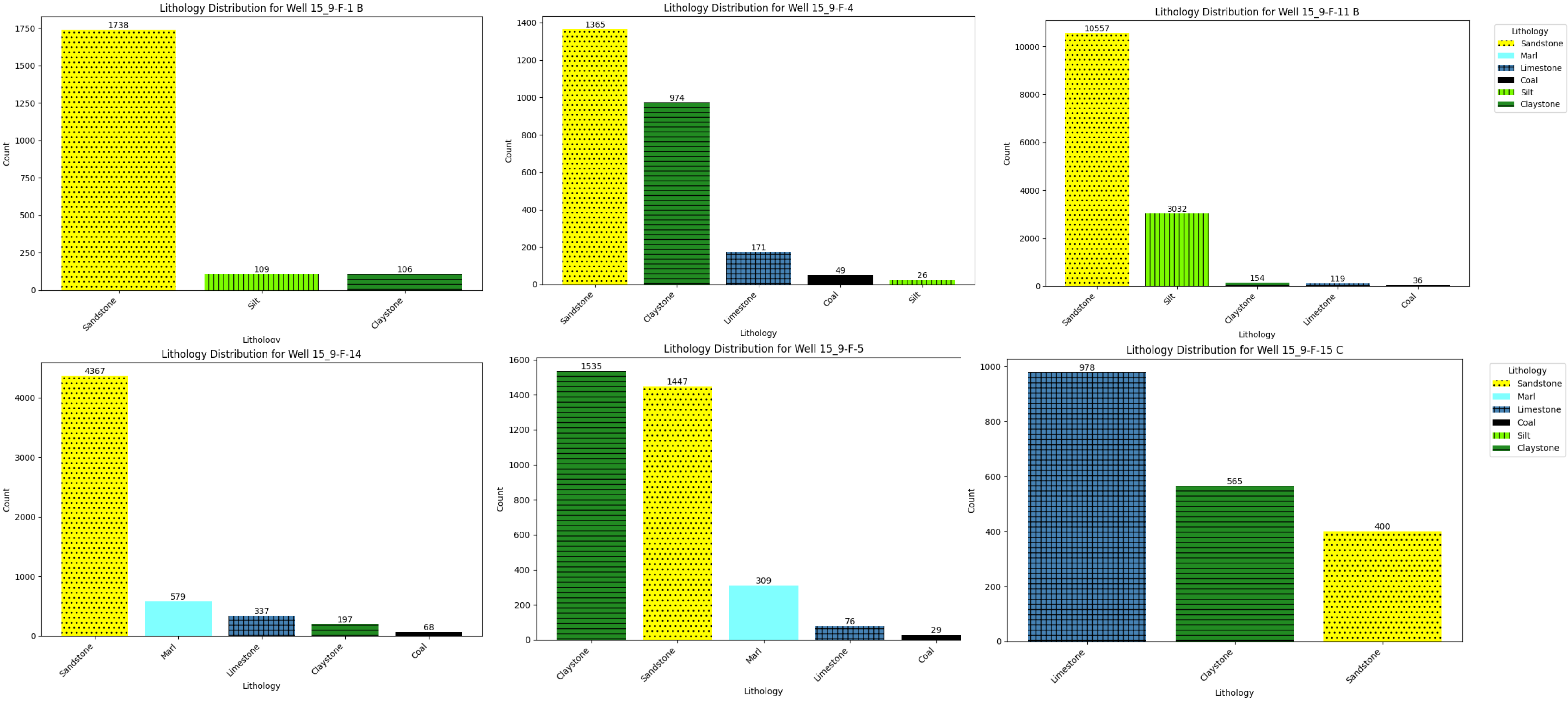
Task 1.1: Run python code to visualise:

- Lithology distribution in a well
- Wireline log panel for a well
- Cross plot wireline (may have to use log scale)
- Save plots in Word/PowerPoint (Copy image/Paste)

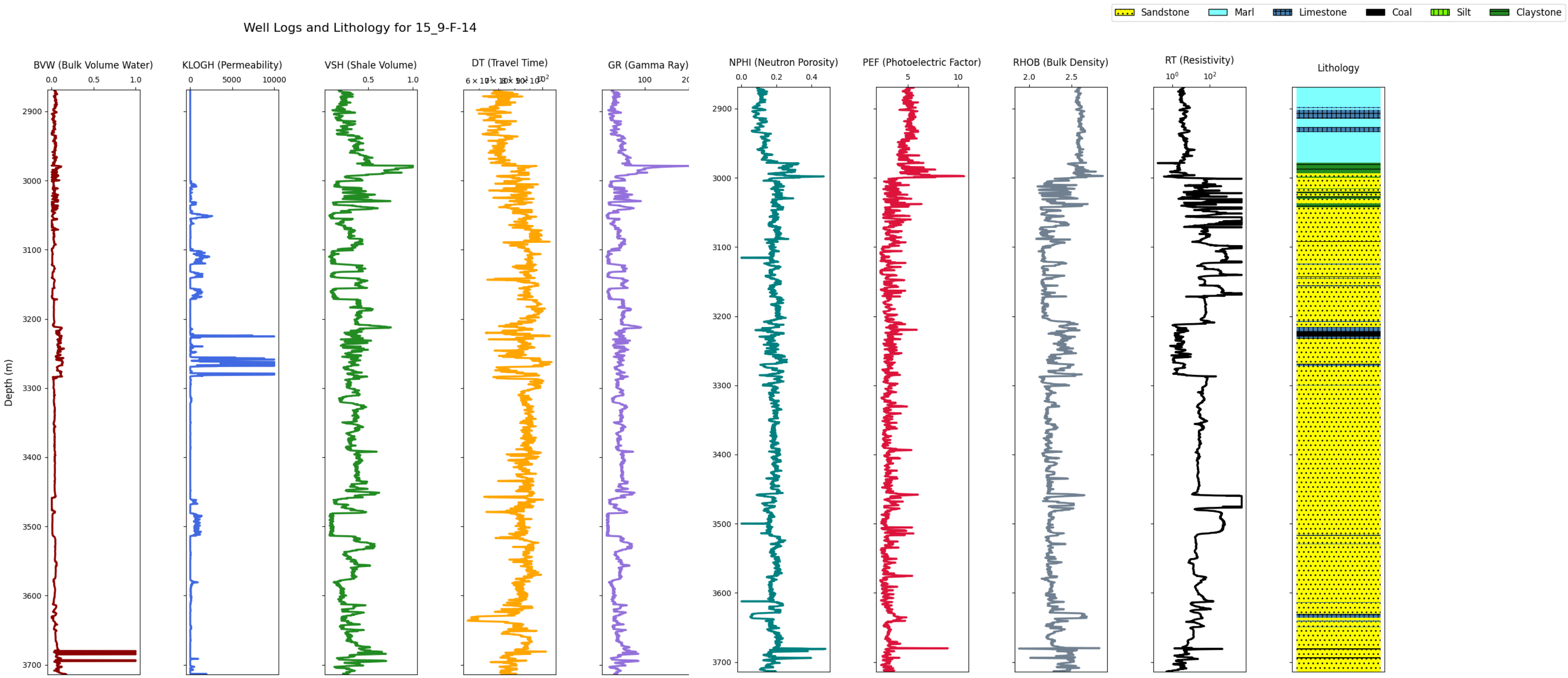
Task 1.2: Select a different well

- Repeat Task 1.1, compare the wells.
- Decide which wells would you select for training/testing/validation for the next exercises.

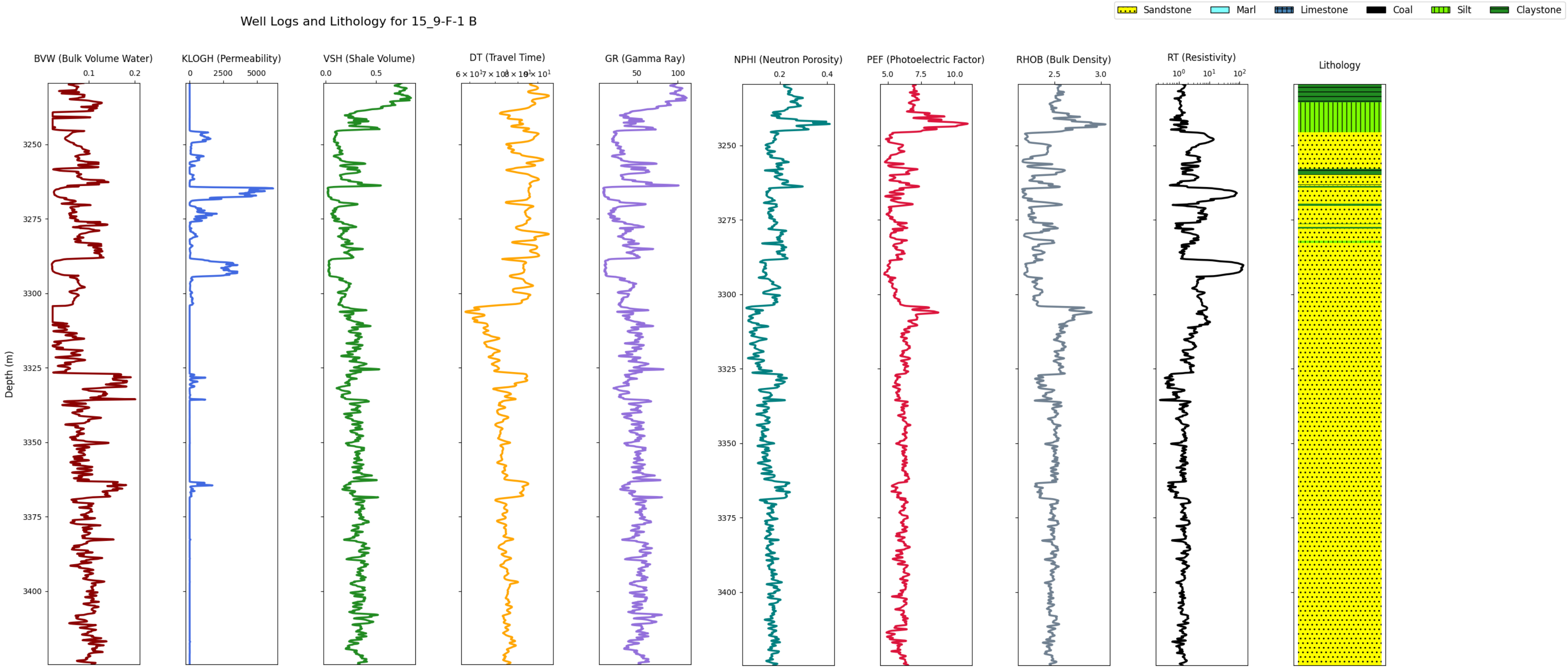
Lithofacies



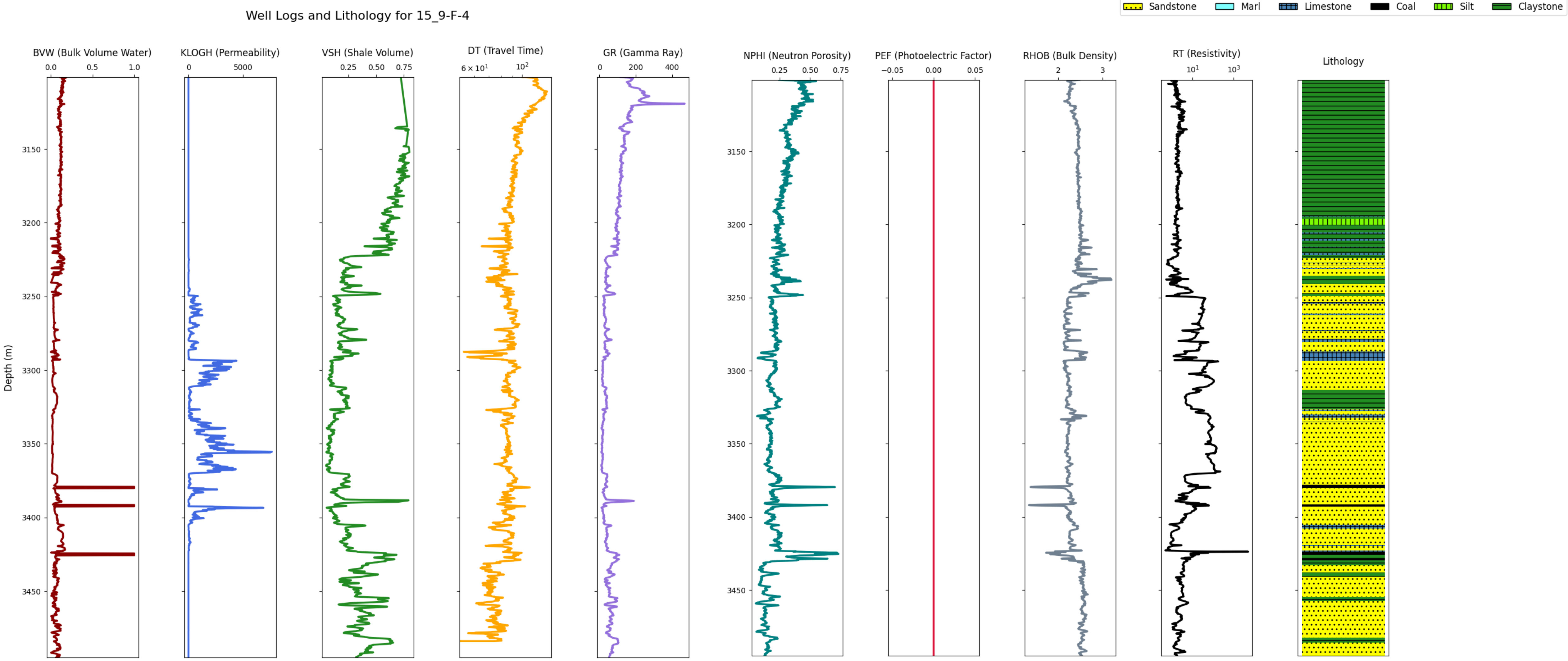
Well data: 15_9-F-14



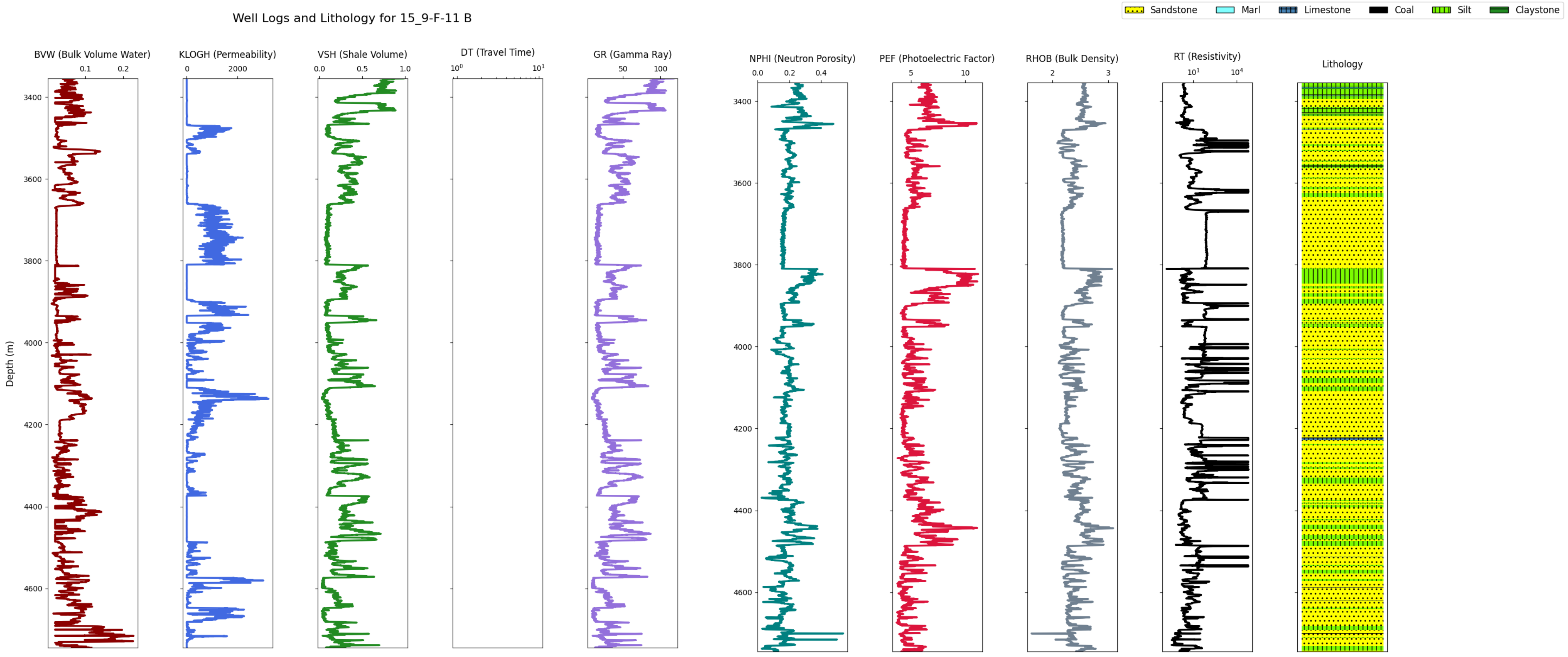
Well data: 15_9-F-1 B



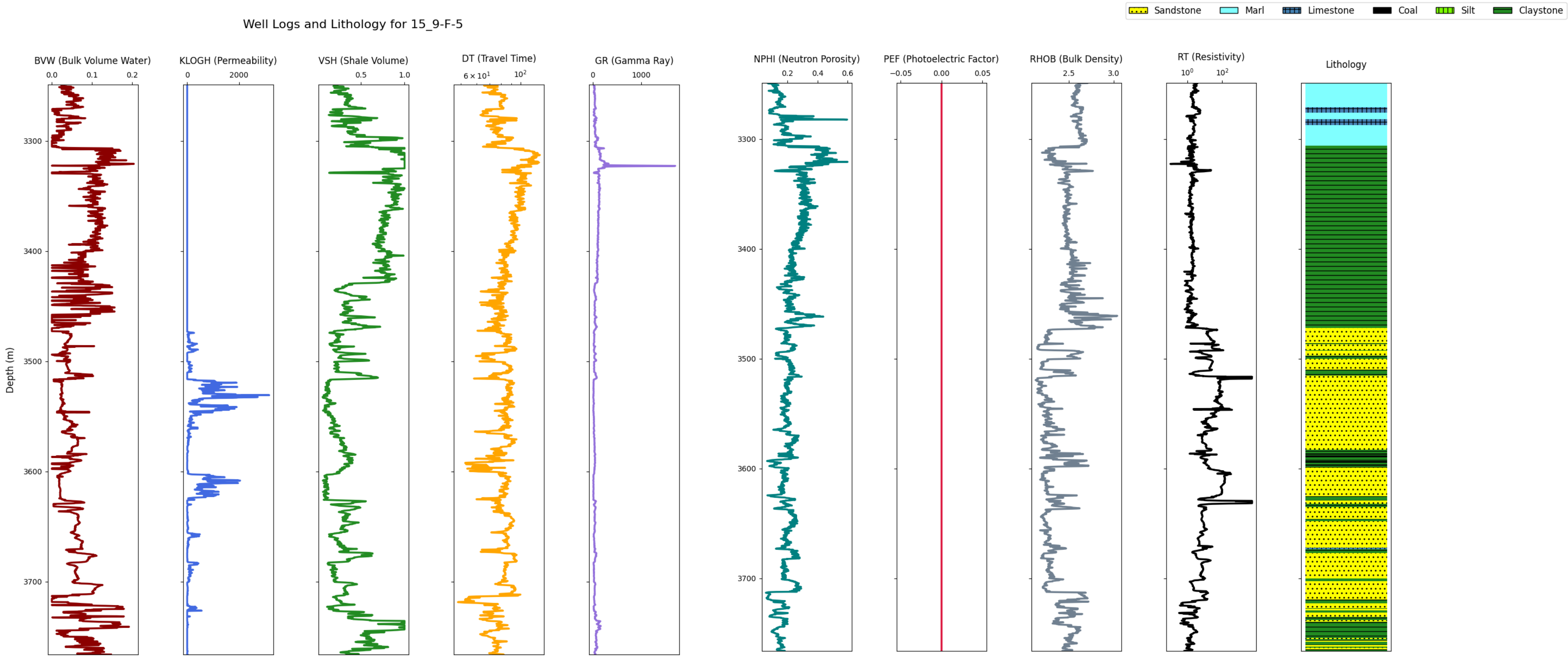
Well data: 15_9-F-4



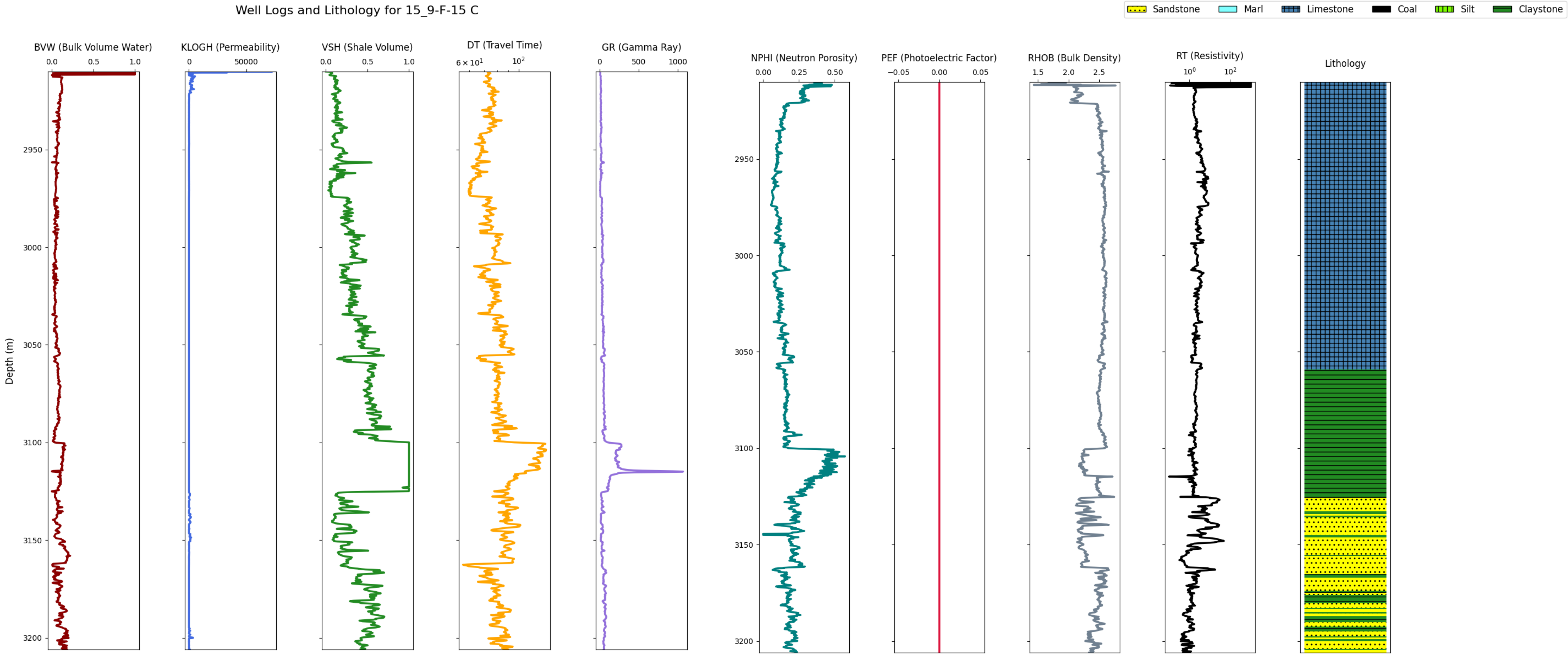
Well data: 15_9-F-11 B



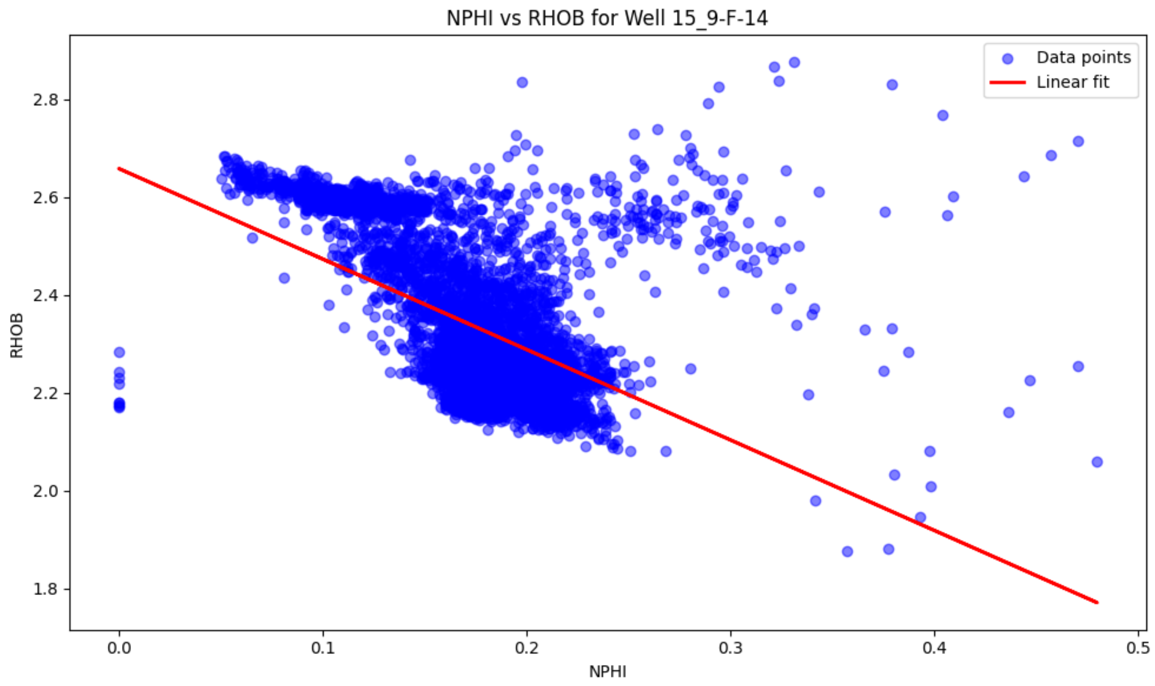
Well data: 15_9-F-5

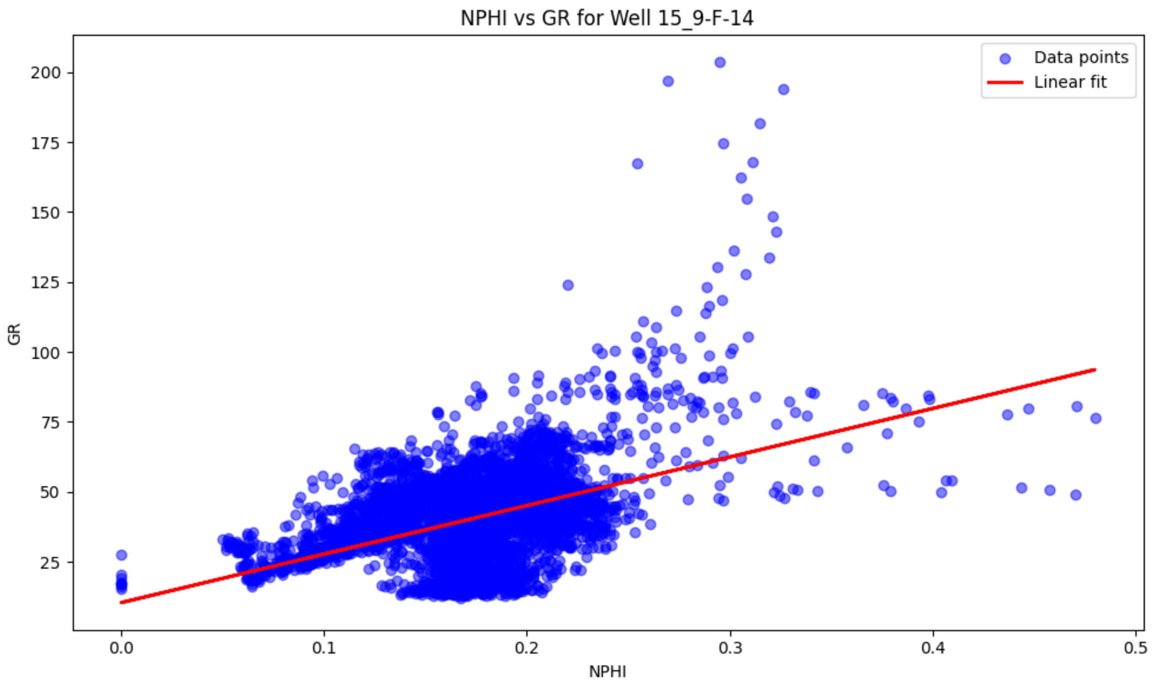


Well data: 15_9-F-15 C



Cross plots


NPHI vs RHOB for Well 15_9-F-14
Data points
Linear fit
RHOB
NPHI
2.8
2.6
2.4
2.2
2.0
1.8
0.0
0.1
0.2
0.3
0.4
0.5


NPHI vs GR for Well 15_9-F-14
Data points
Linear fit
GR
NPHI
200
175
150
125
100
75
50
25
0.0
0.1
0.2
0.3
0.4
0.5

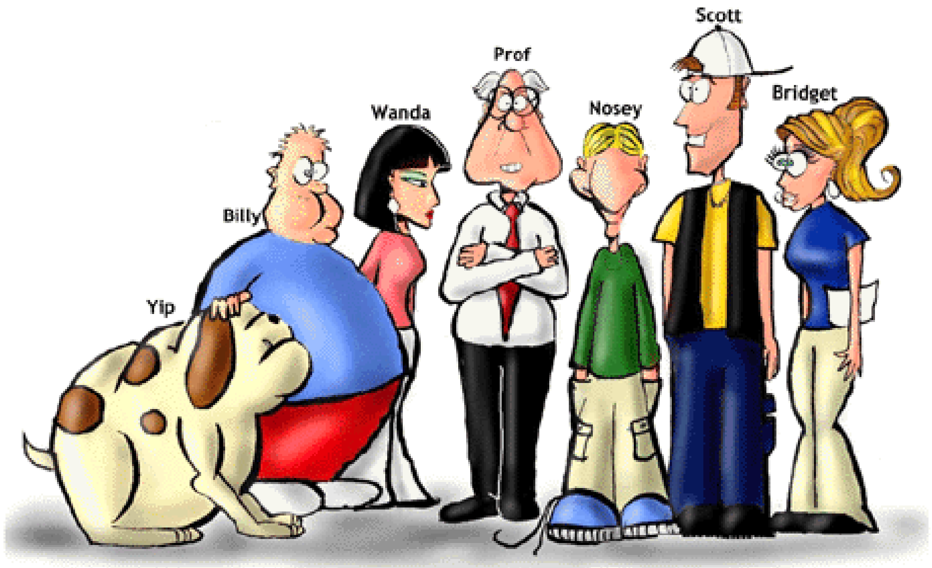
Hands-on AI application to wireline data

- Volve field
- Unsupervised clustering
- Supervised classification

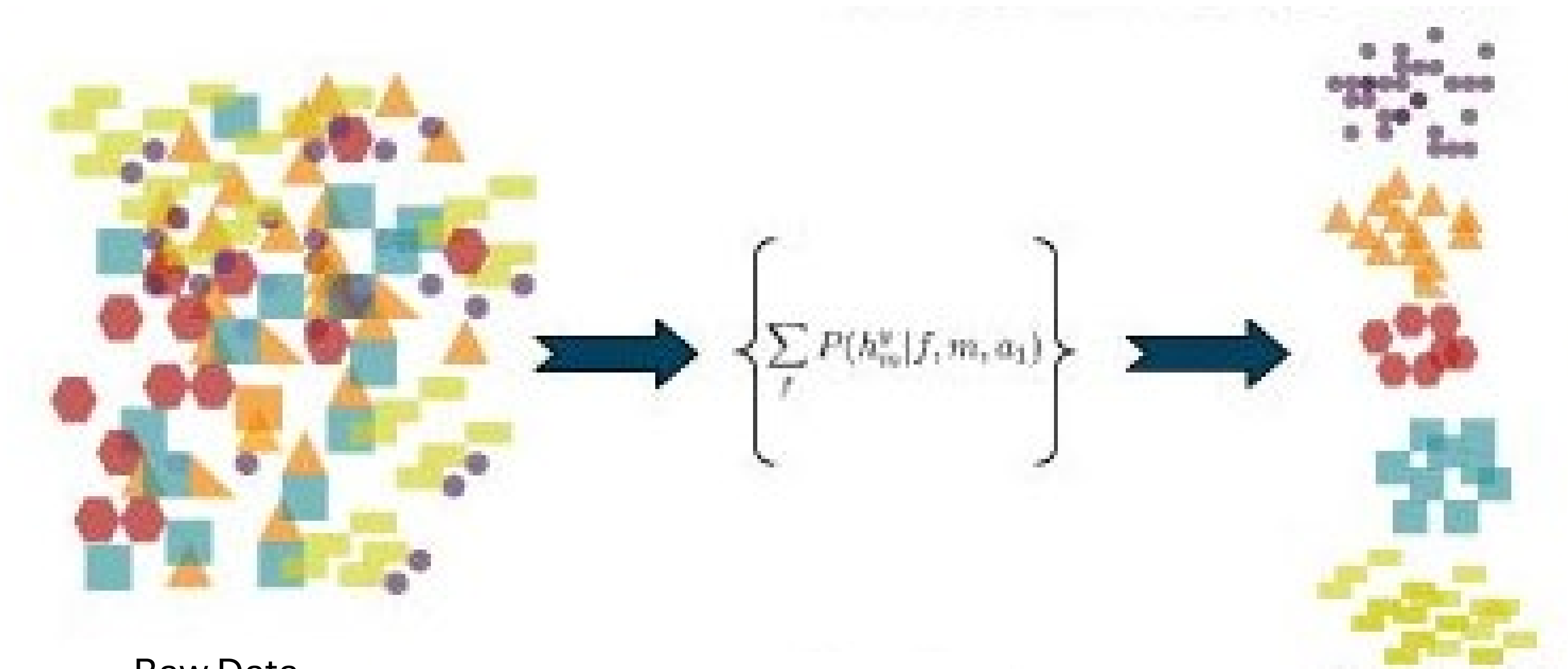
Unsupervised Learning

Training with no answer available

- How fast will they run 100 yards?
- How can we know without a trial?
  - Group them by prior info:
    - height, weight, age, etc.
  - Make limited number of trials with typical group members



Unsupervised Learning



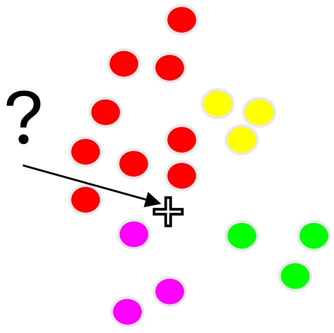
Raw Data

Algorithm

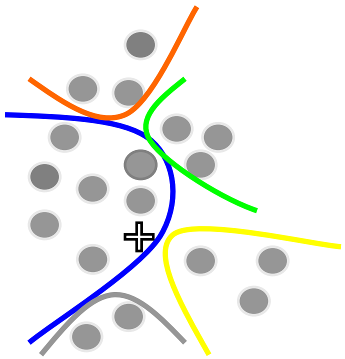
Automated Clusters

Unsupervised learning tasks

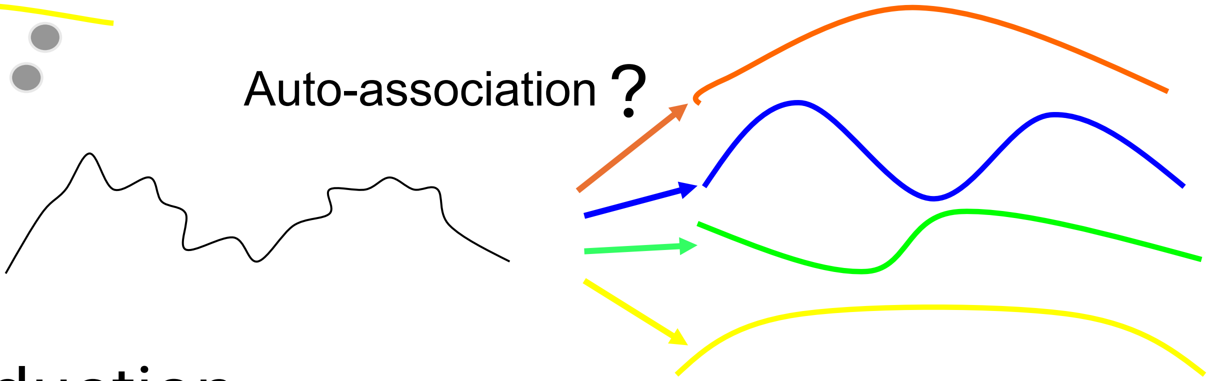
- Pattern classification



- Clustering

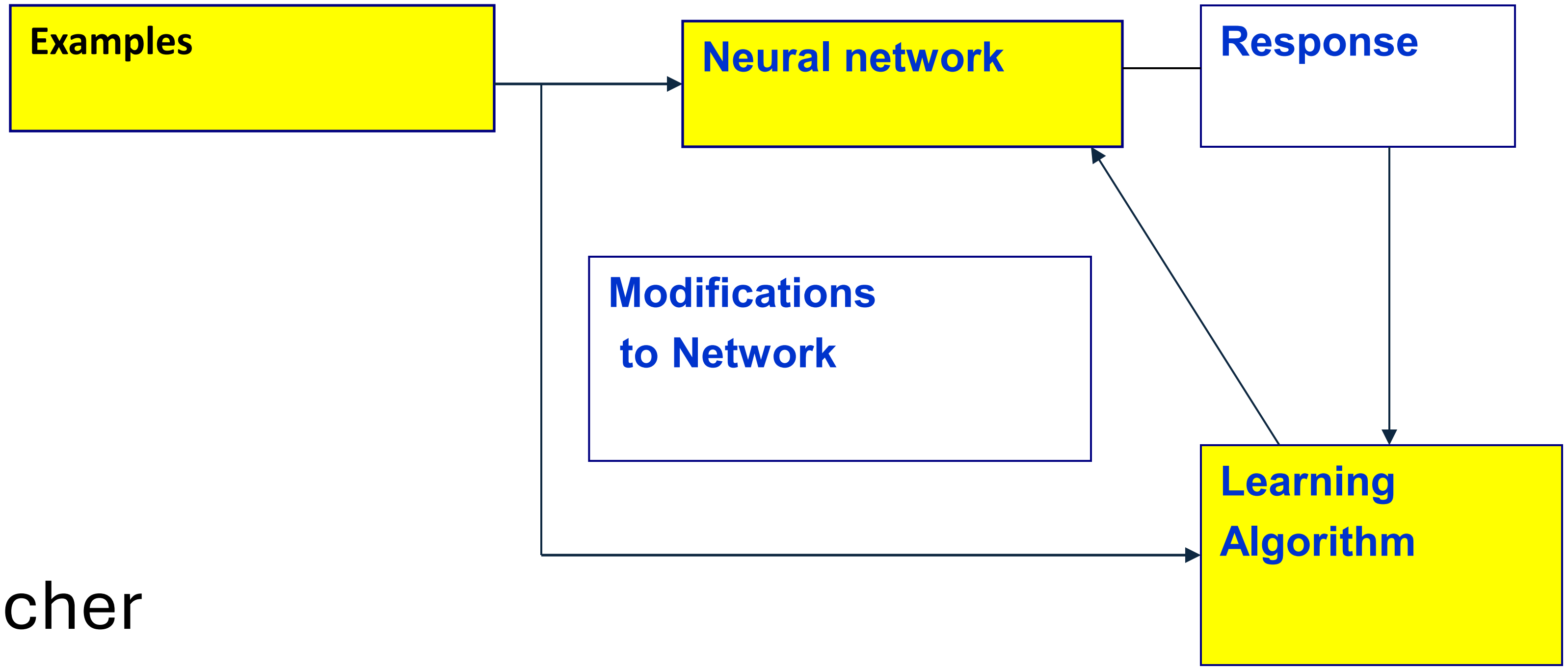


- Auto-association



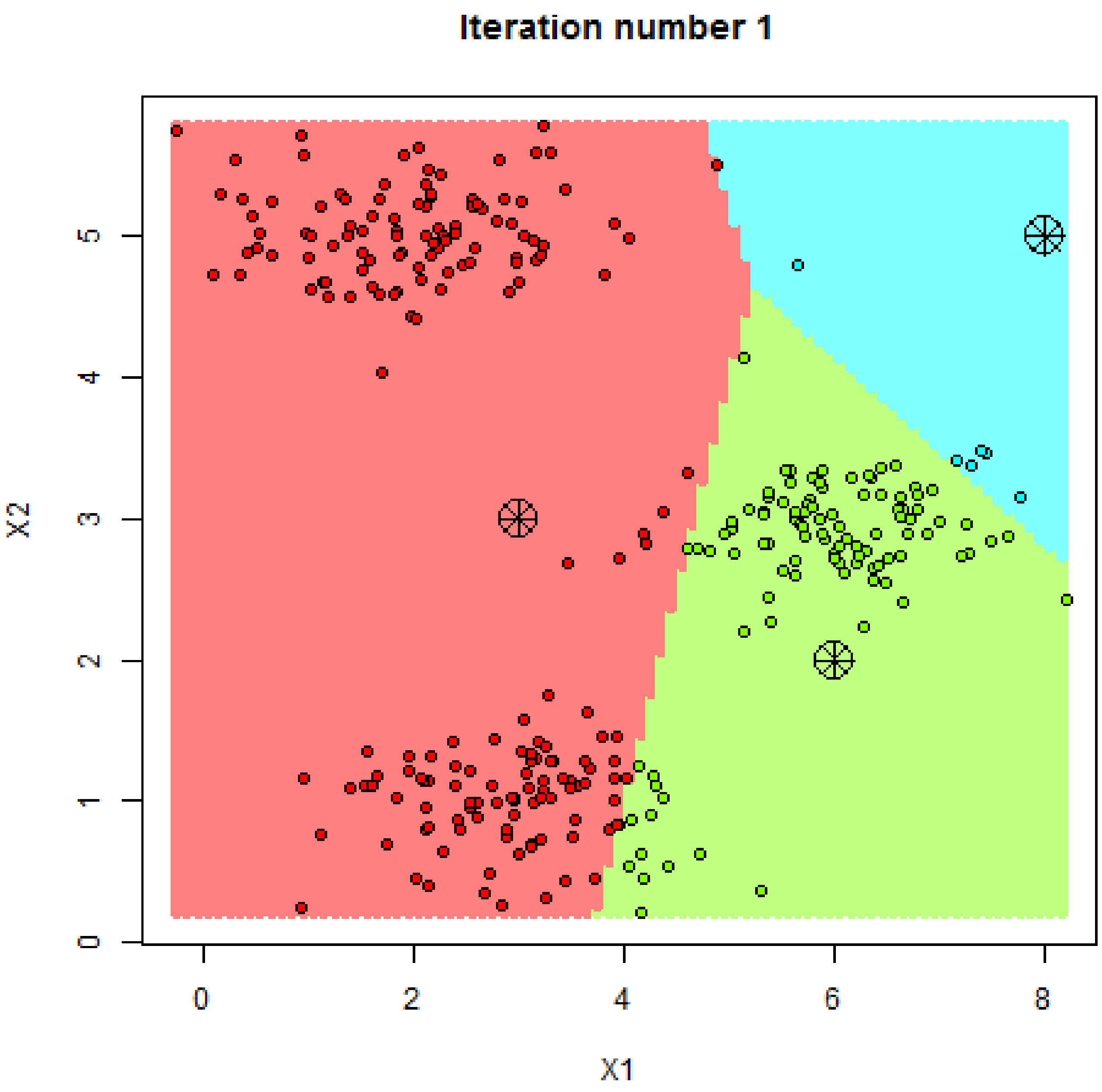
- Dimensionality reduction
- Outliers/anomaly detection

Unsupervised learning



- No external teacher
- Task-independent measure of quality
- Learn features of the input (no outputs)

K-Means clustering



K-Means Algorithm

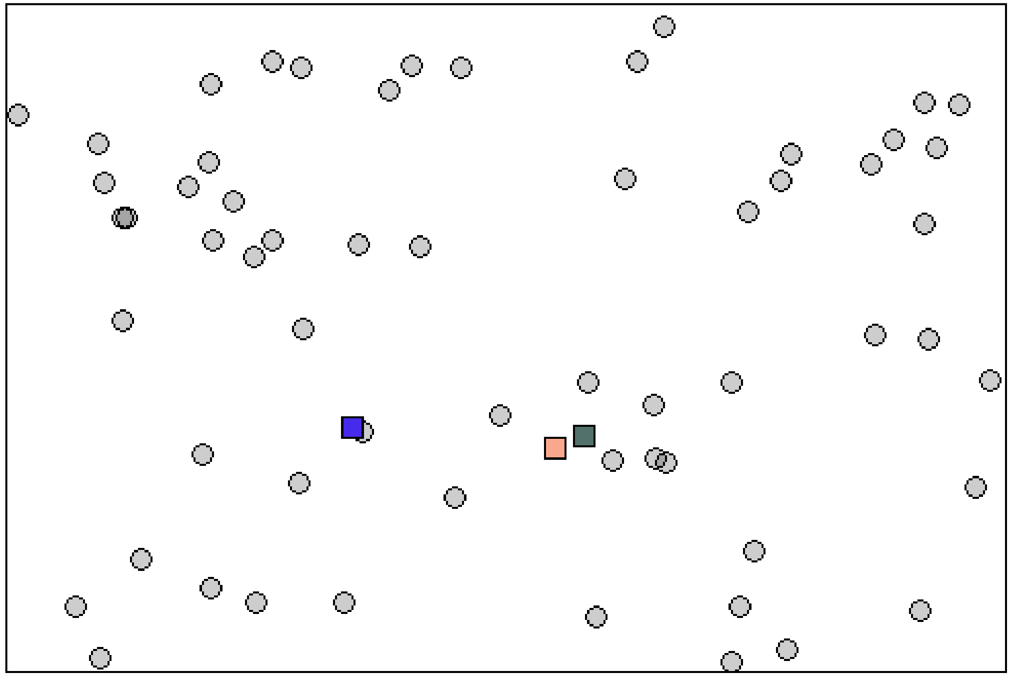
Partitions data according to distance minimisation to cluster centres

$$V = \sum_{i=1}^{k} \sum_{j \in S_i} \left| x_j - \mu_i \right|^2$$

where there are k clusters S_i, i = 1,2,...,k

μ_i is the central point of all the points x_j in the cluster i

Random placement of centres



K-Means Algorithm

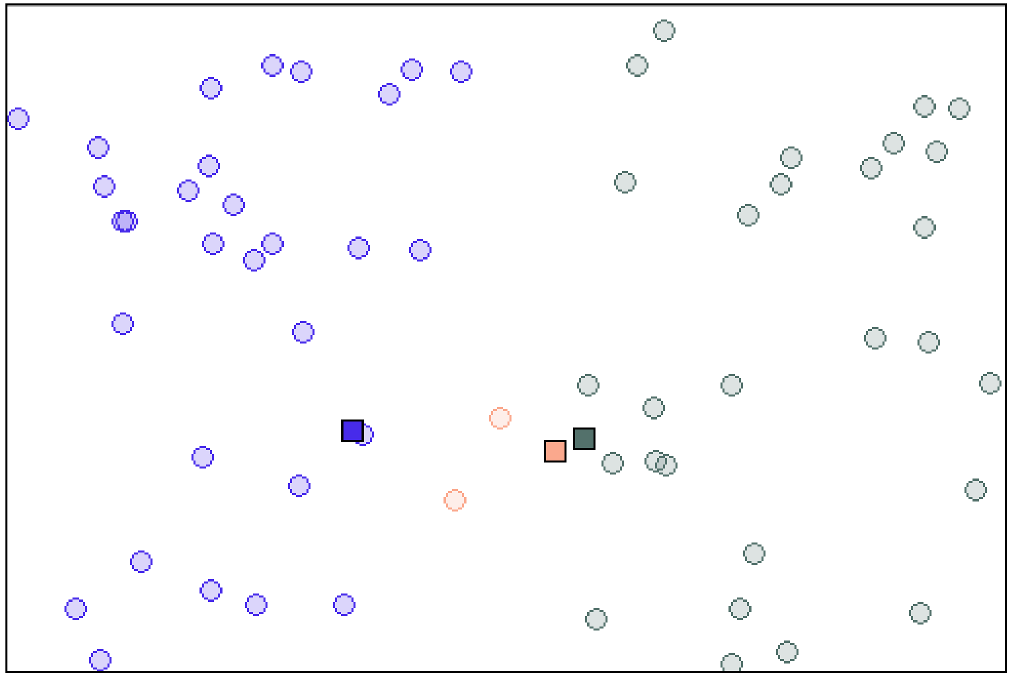
Partitions data according to distance minimisation to cluster centres

$$V = \sum_{i=1}^{k} \sum_{j \in S_i} \left| x_j - \mu_i \right|^2$$

where there are k clusters S_i, i = 1,2,...,k

μ_i is the central point of all the points x_j in the cluster i

Associate points with centres



K-Means Algorithm

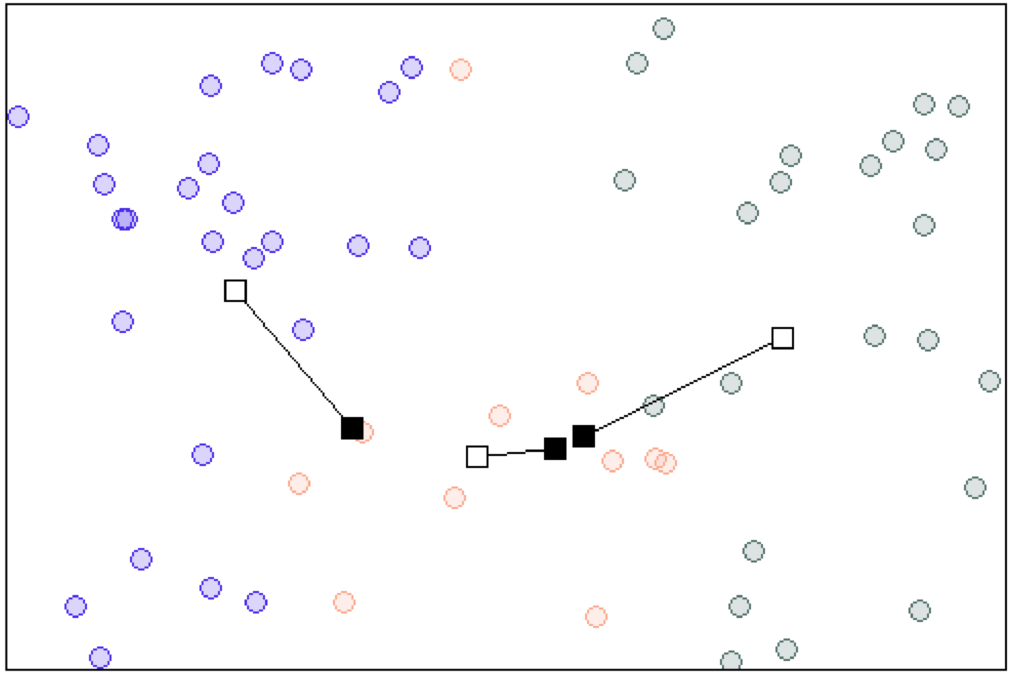
Partitions data according to distance minimisation to cluster centres

$$V = \sum_{i=1}^{k} \sum_{j \in S_i} \left| x_j - \mu_i \right|^2$$

where there are k clusters S_i, i = 1,2,...,k

μ_i is the central point of all the points x_j in the cluster i

Move centres to the middle of each group and re-associate the points with new centres



K-Means Algorithm

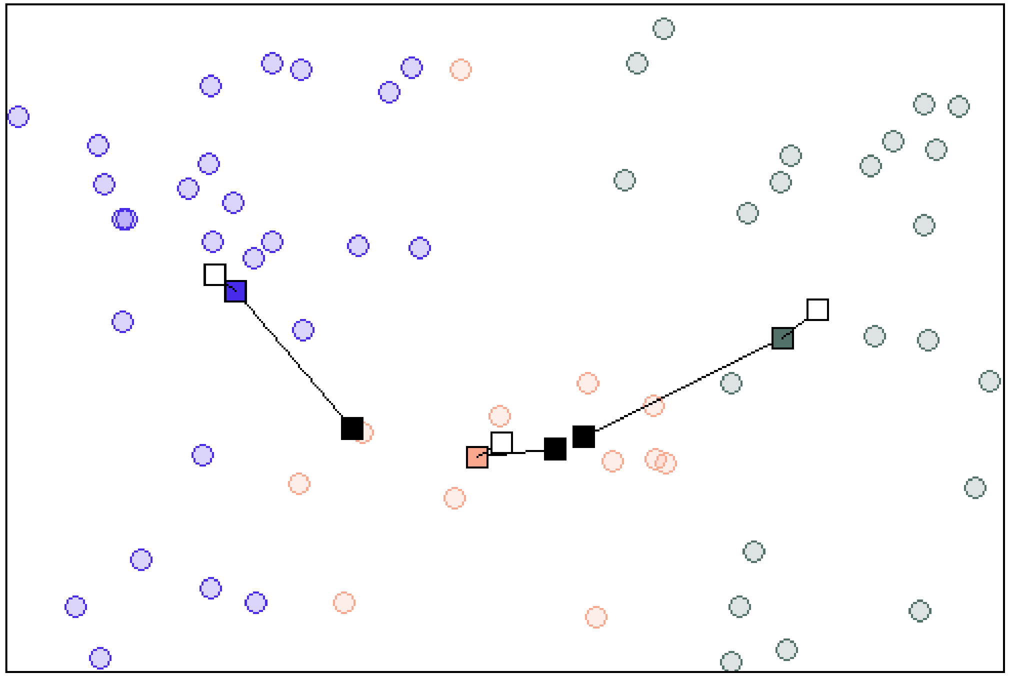
Partitions data according to distance minimisation to cluster centres

$$V = \sum_{i=1}^{k} \sum_{j \in S_i} \left| x_j - \mu_i \right|^2$$

where there are k clusters S_i, $i = 1,2,\ldots,k$

μ_i is the central point of all the points x_j in the cluster i

Iteratively move the centres and re-associate the points minimising V



K-Means Algorithm

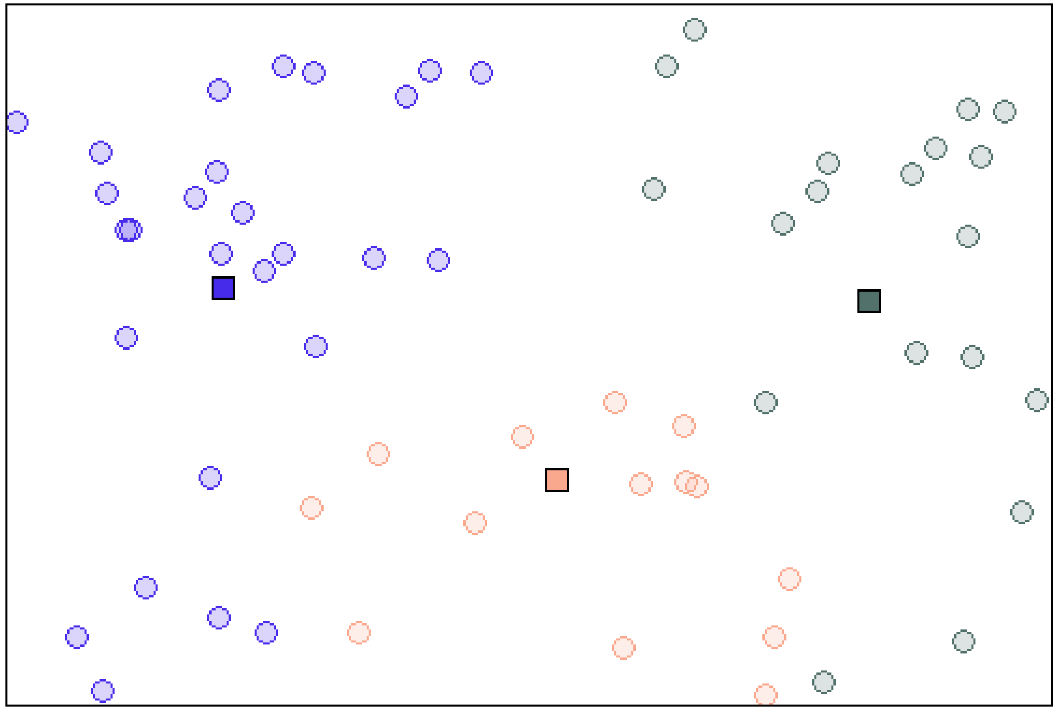
Partitions data according to distance minimisation to cluster centres

$$V = \sum_{i=1}^{k} \sum_{j \in S_i} \left| x_j - \mu_i \right|^2$$

where there are k clusters S_i, $i = 1,2,\ldots,k$

μ_i is the central point of all the points x_j in the cluster i

Converge



Machine Learning Clustering of Reservoir Heterogeneity with Petrophysical and Production Data

by D Konoshonkin, G Shishaev, I Matveev, A Volkova, V Rukavishnikov, V Demyanov, B Belozerov

SPE 200614-MS

SPE-200614-MS

Machine Learning Clustering of Reservoir Heterogeneity with Petrophysical and Production Data

Dmitry Konoshonkin, Gleb Shishaev, Ivan Matveev, Aleksandra Volkova, and Valeriy Rukavishnikov, Tomsk Polytechnic University; Vasily Demyanov, Heriot-Watt University; Boris Belozerov, Gazpromneft STC

This paper was prepared for presentation at the SPE Europec featured at 82nd EAGE Conference and Exhibition originally scheduled to be held in Amsterdam, The Netherlands, 8-11 June 2020. The physical event was postponed until 14-17 June 2021. A virtual SPE Europec was held 1-3 December 2020 for SPE authors to present their papers. The official proceedings were published online on 8 June 2020.

Abstract

Reservoir development decisions strongly depend on our understanding on reservoir heterogeneity, which is often subject to sparse and conflicting data, interpretational bias and constraints imposed by the modelling assumptions. The work tackles a challenging task of accurately and quickly identifying and describing uncertainty in the spatial distribution of reservoir heterogeneity derived from geological well data and with respect to a geological concept. We propose a metric based machine-learning approach to identify and describe spatial trends in reservoir heterogeneity/facies property distribution using wireline and production data.

We demonstrate how the proposed method can help to partition reservoir heterogeneity and discover and verify spatial trends for a real mature producing field in the Western Siberia. The obtained clustering of reservoir facies based on the wireline logs (alpha-SP) demonstrated a good agreement with the reservoir zonation based on manual log interpretation and the geological concept. Clustering based on individual well production profiles has confirmed the reservoir partitioning and matched some of the reservoir features aligned with the prevailing geological concept. The outcome of the proposed method helps to improve the facies distribution model by integrating the discovered spatial trends into a geostatistical model and account for uncertainty in the depositional scenario that is difficult to quantify based on manual interpretation.

Introduction

The interpretation of geological reservoir heterogeneity is one of the major reservoir uncertainty that is usually inferred from different data types: well logs, core data, seismic, production data analysis, etc. as well as based on the geological understanding to derive the conceptual interpretation. Reservoir zonation is commonly used to introduce large scale spatial trends and differentiate between depositional zones according to the derived interpretational concept. This approach is traditionally used to tackle non-stationarity in spatial reservoir property distribution. Reservoir zonation and spatial trends constitute principle uncertainty that is vital to assess in reservoir modellings, simulation and forecasting studies.

Clustering Application

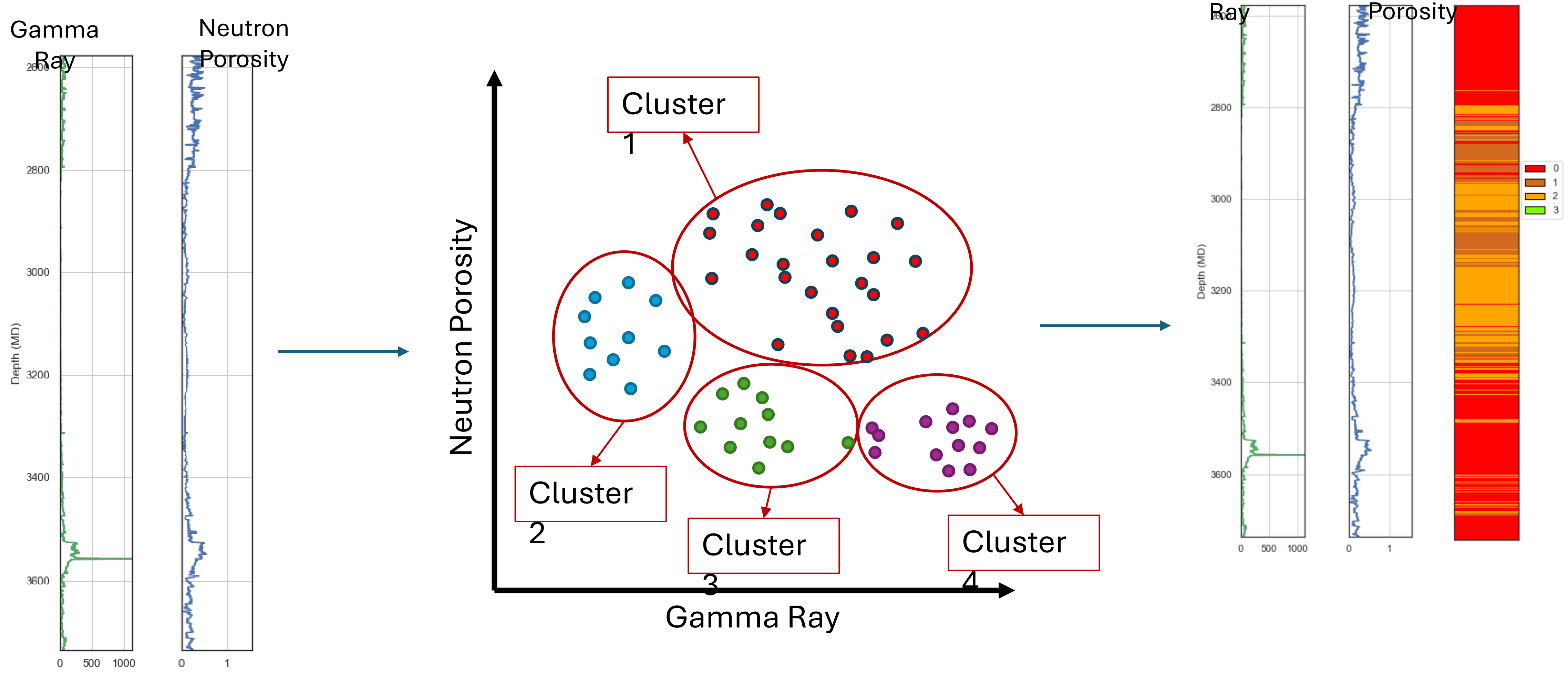
Hands-on

Unsupervised Clustering. Notebook 2

https://colab.research.google.com/drive/1XBw8szzPbncOdTNxsRJgoHswk_BOnDz_?usp=sharing

Objective of the exercise

How **clustering** can help identify lithology classes by similarities in **wireline log** data?



Exercise 2

Task 2.1: Run python code to cluster wireline log input combination:

- Select 15_9-F**4** for **training** and 15_9-F**5** for **testing**
- Select input logs: `BVW, KLOGH, VSH, GR, NPHI, RHOB, RT`
- Select number of k clusters from inertia plot: 4
- Compare clustered groups with the interpreted lithofacies
- Save plots in Word/PowerPoint (Copy image/Paste)

Task 2.2: Select a different number of **k** clusters

- Repeat Task 2.1

Task 2.3: Select a different combination of input logs

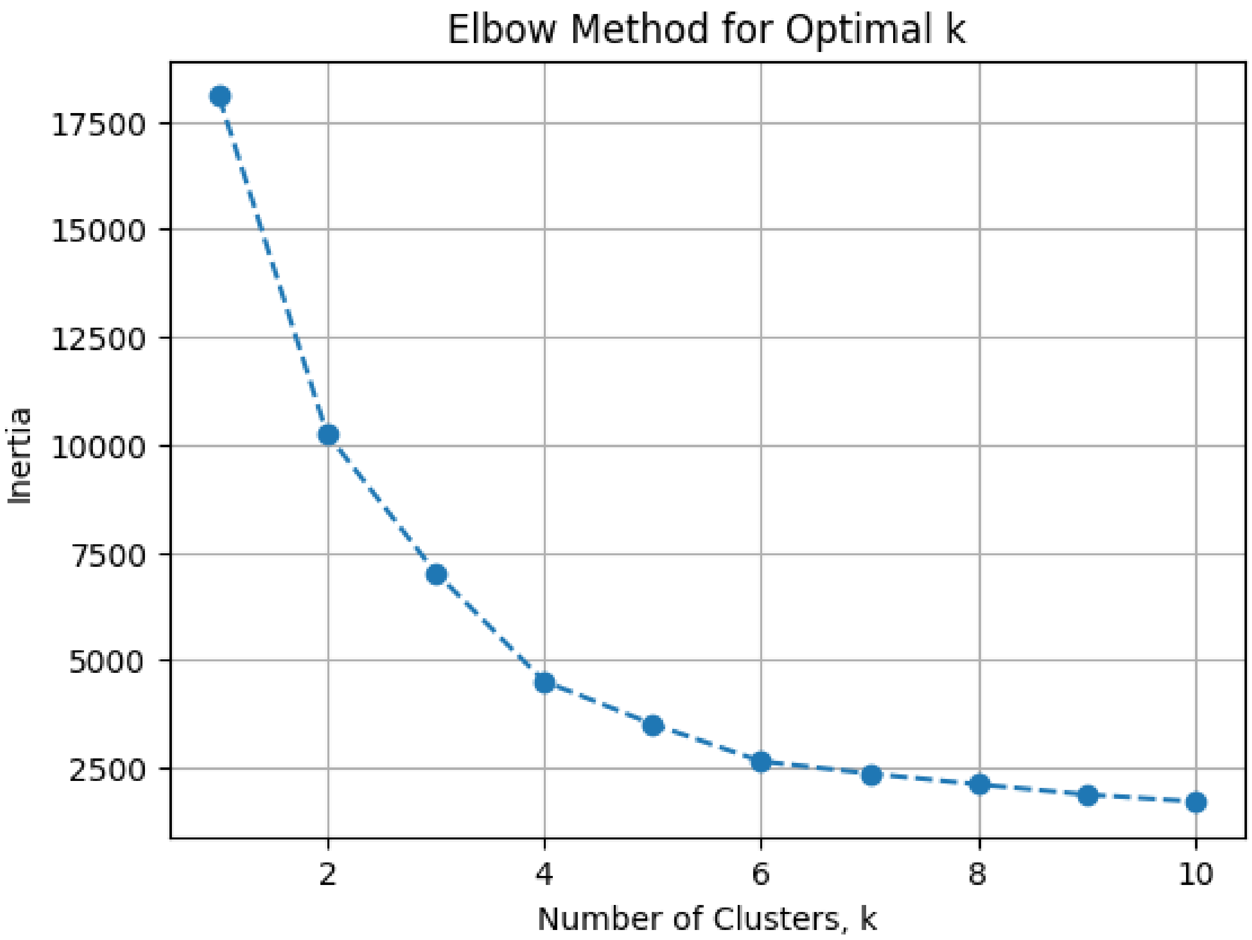
- Repeat Tasks 2.1 and 2.2

Task 2.3: Select different wells for training/testing

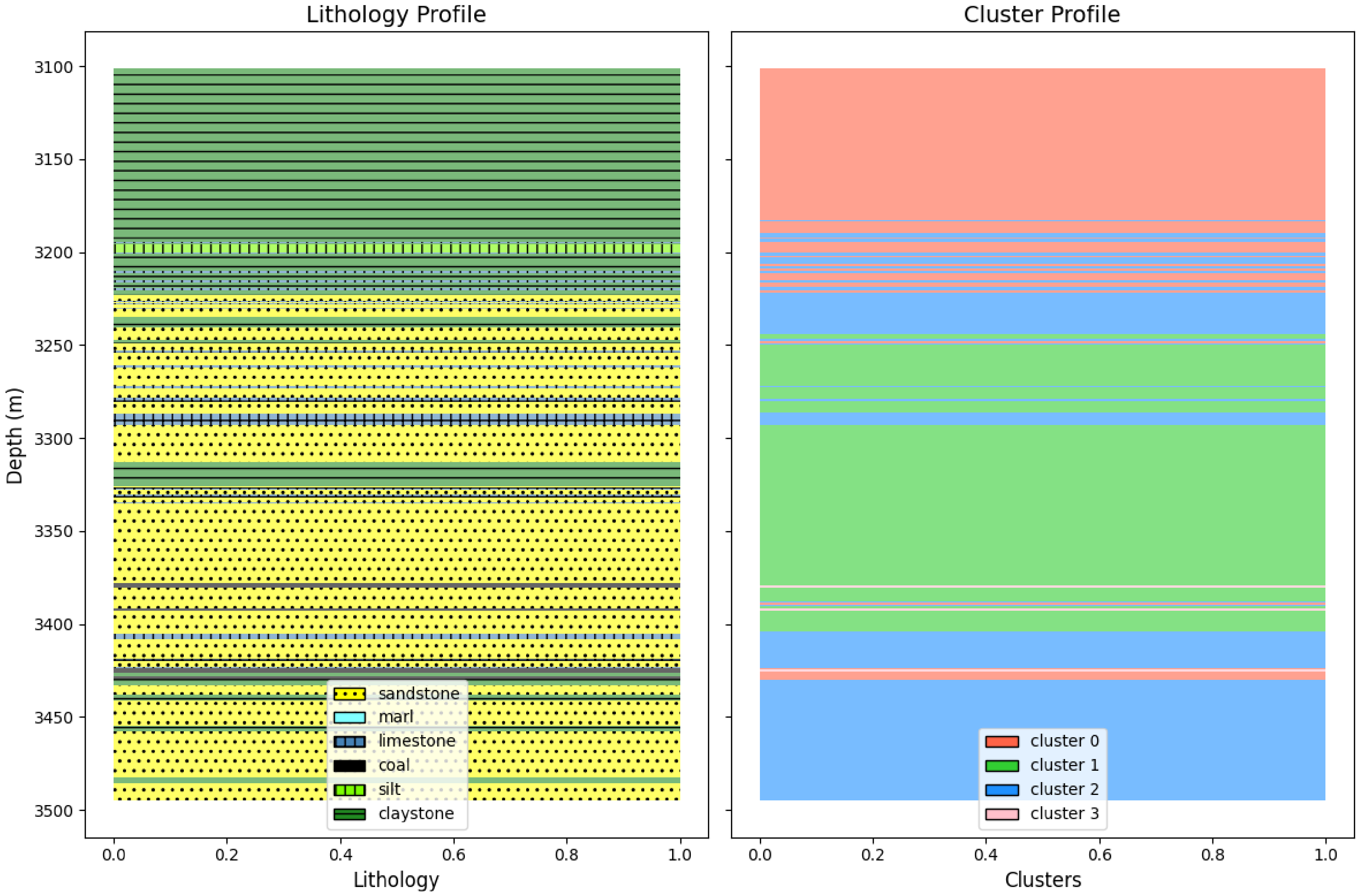
- Repeat Tasks 2.1, 2.2, 2.3 and compare the results for the training and the validation wells.

Select the number of k clusters: 4

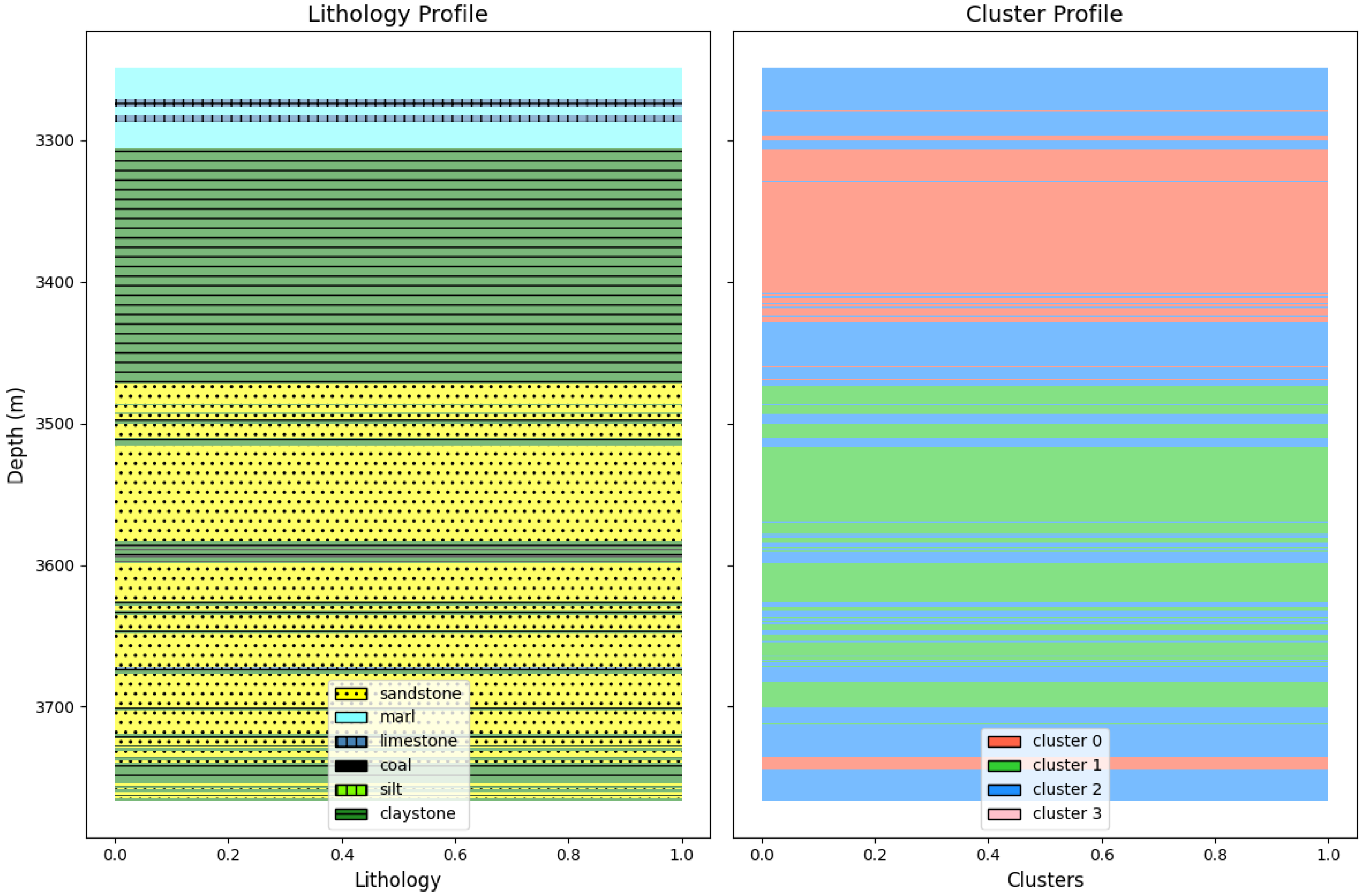
- Inertia plot – elbow method



Training well: 15_9-F-4



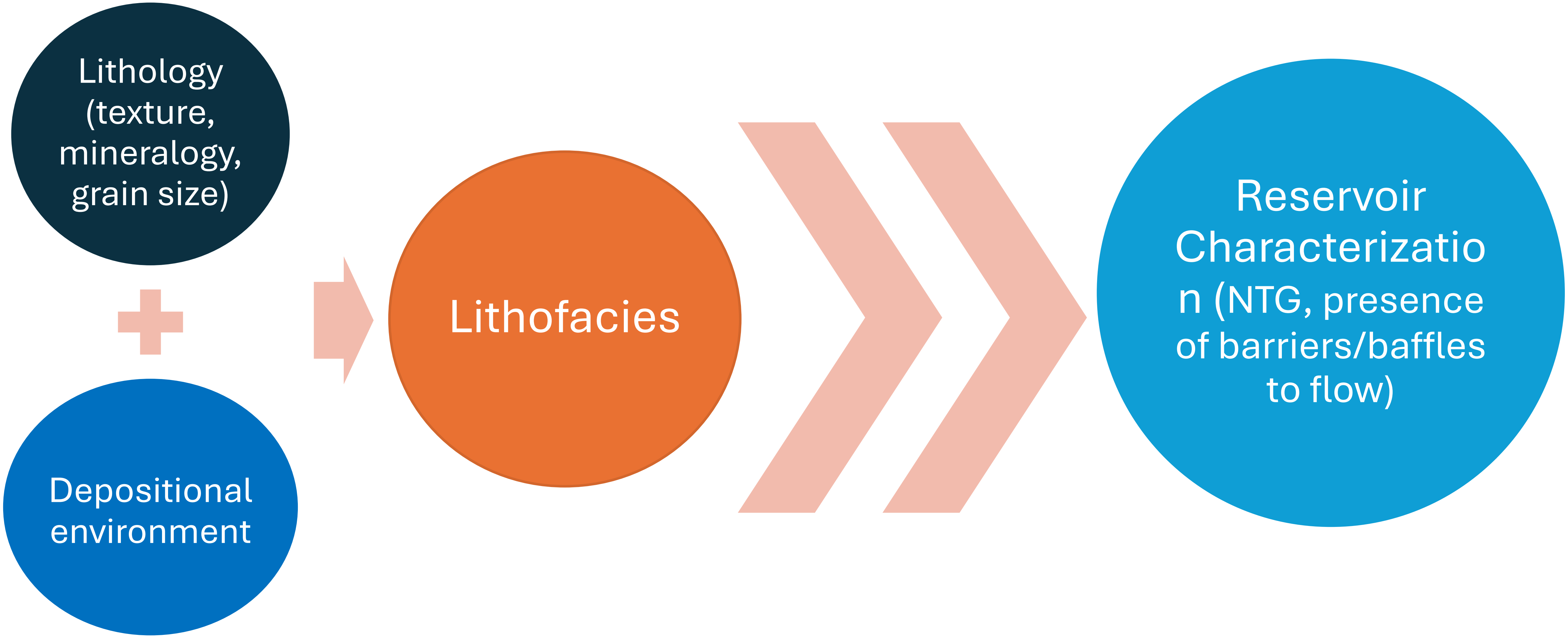
Training well: 15_9-F-5



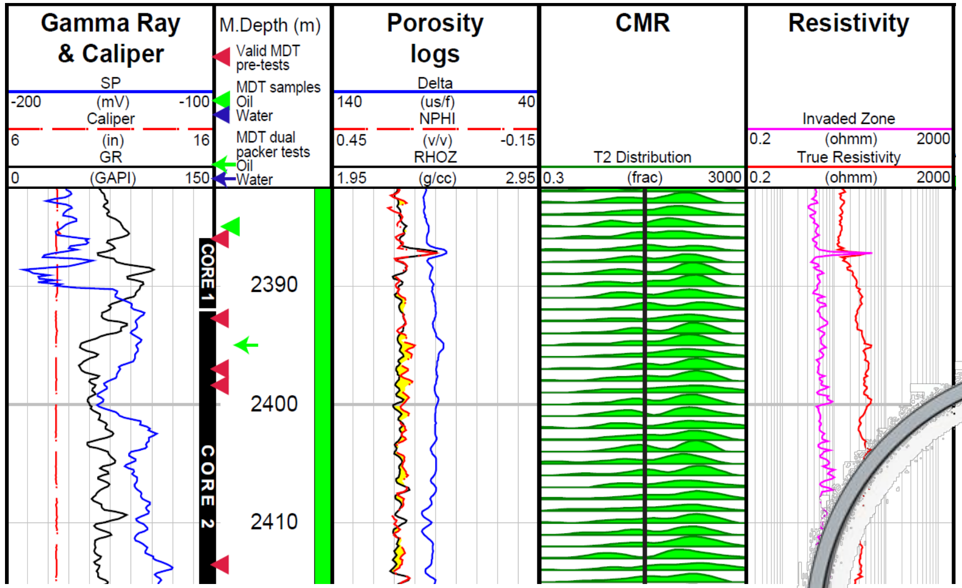
Hands-on AI application to wireline data

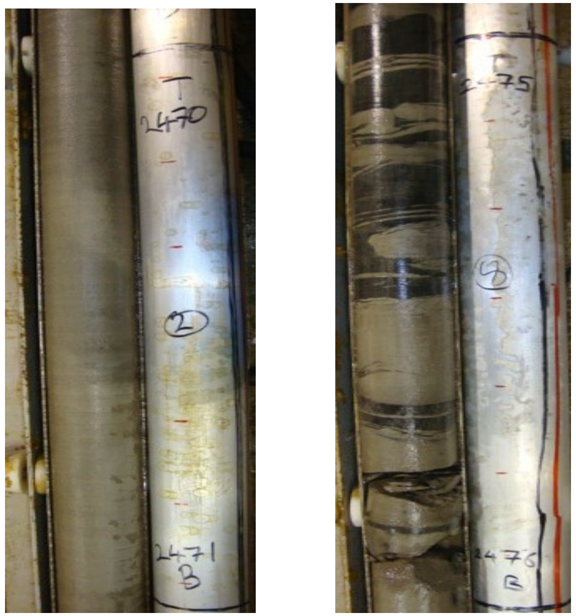
- Volve field
- Unsupervised clustering
- Supervised classification

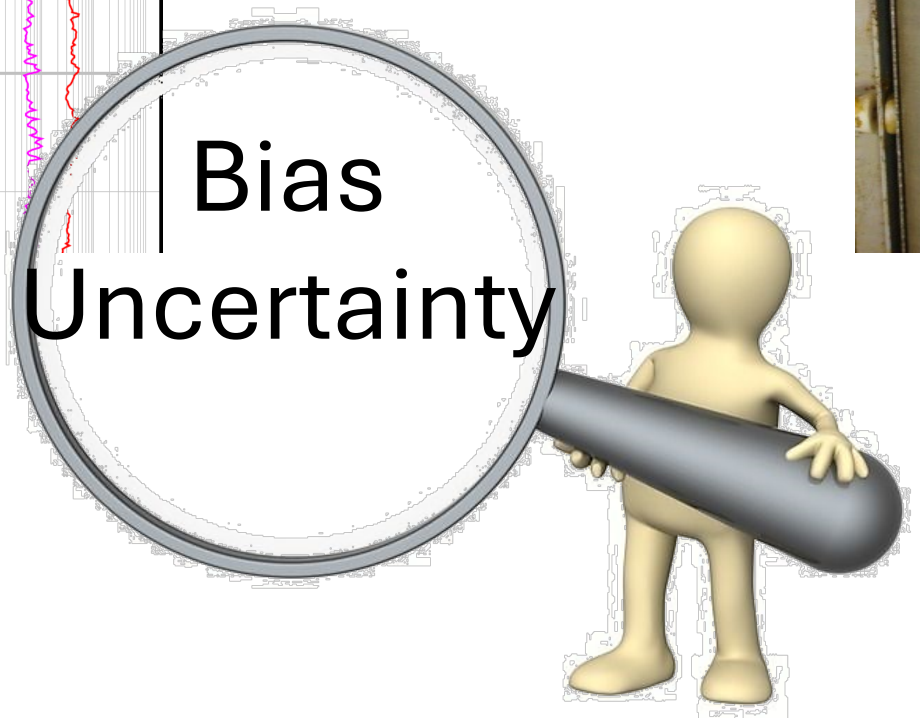
Facies Classification



Facies classification challenge







Petrophysiycal interpretation

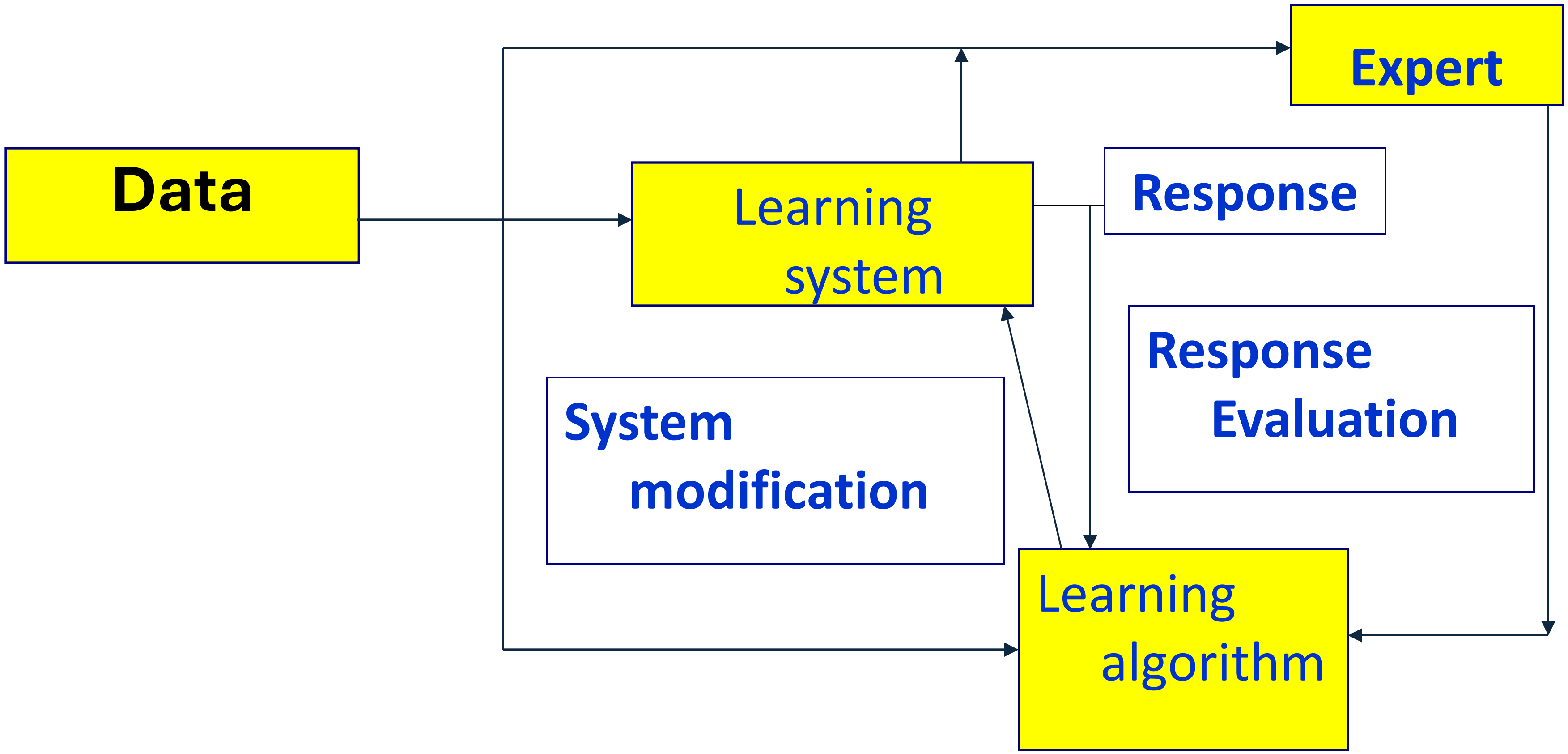
Sedimentological interpretation

Halotel, McCarty, Gardiner, Demyanov, 2019, IAMG conference

Supervised classiciation

- K-Nearest Neighbour
- Random Forest

Supervised Learning



M. Kanevski

K- Nearest Neighbours

- A well known approach in different sciences under different names: Voronoi polygons, Dirichlet cells, Thiessen polygons, method of analogues, area-of-influence,...

- We consider k-NN as a benchmark
  - "vanilla" model
  - kNN is local, nonparametric

K-NN prediction:

- K-NN methods use those *k*-observations in the data set T closest in input space to prediction point *x* to estimate *Y*

$$\hat{Y}(x) = \frac{1}{k} \sum_{x_i \in N_k(x)}^{k} y_i$$

where k is the neighbourhood of *x* defined by the closest points in the set

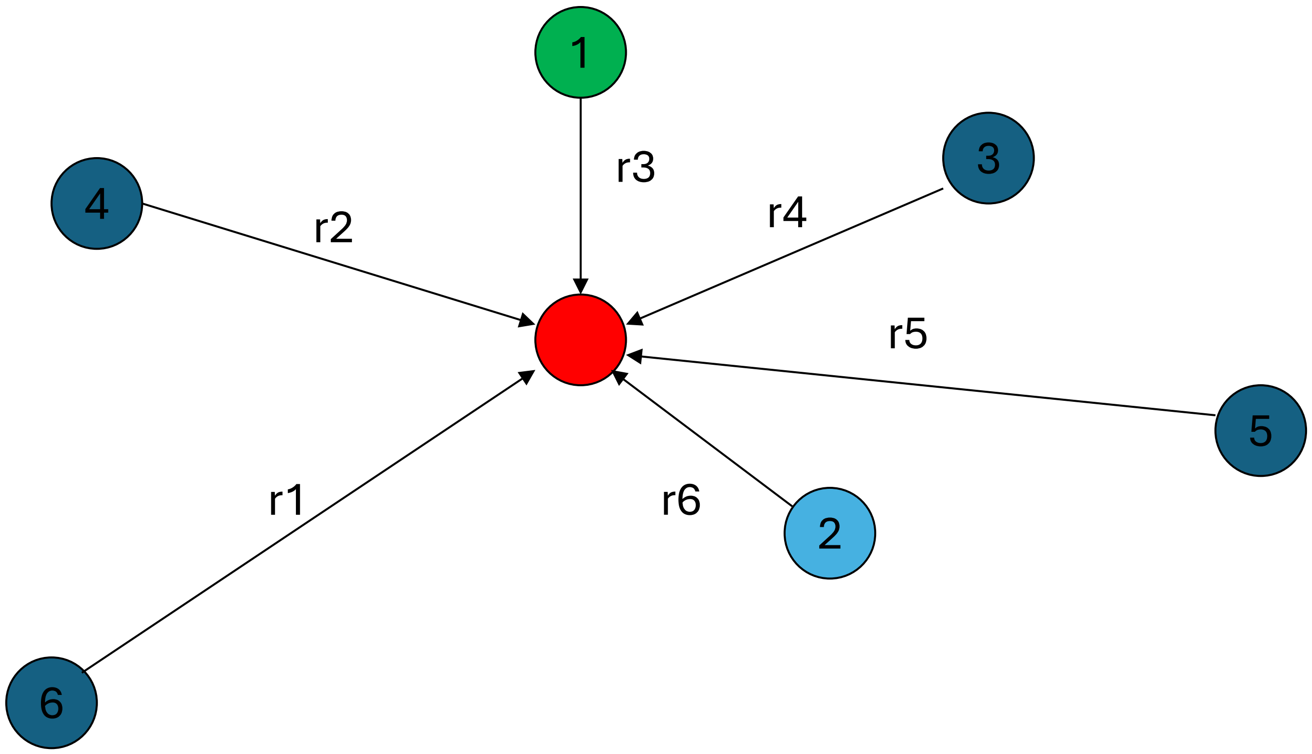
Q: how to select hyper-parameter k?

K-NN Interpolation

$$\hat{Y}(x) = \frac{1}{k} \sum_{x_i \in N_k(x)}^{k} y_i$$

1-NN: $Y(x) = y_1$

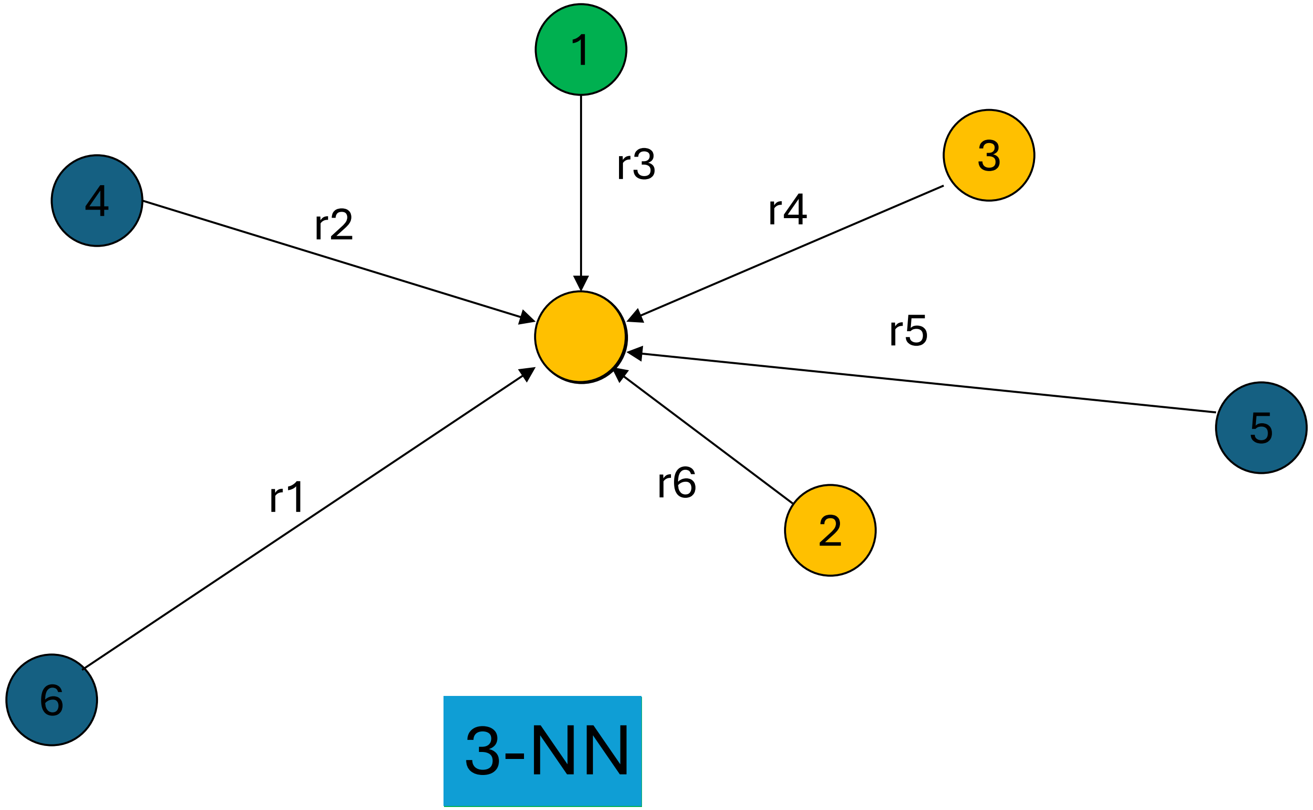
2-NN: $Y(x) = (y_1 + y_2)/2$



k-NN Classifier

- These classifiers are memory-based and do not require any model to be fit!

- Given a query point *x*, we find the *k* training points closest in the distance to *x* and then classify using MAJORITY vote among the k neighbors.



K-NN notes

- Despite of its simplicity, k-NN has been successful in a large number of classification problems, including handwritten digits, satellite image scenes and ECG patterns.

- It is often successful where each class has many possible prototypes, and the decision boundary is very irregular.

- Because it uses only the training point closest to the query point, the bias of the 1-nn estimate is often low, but the variance is high.

KNN
- how to find K?

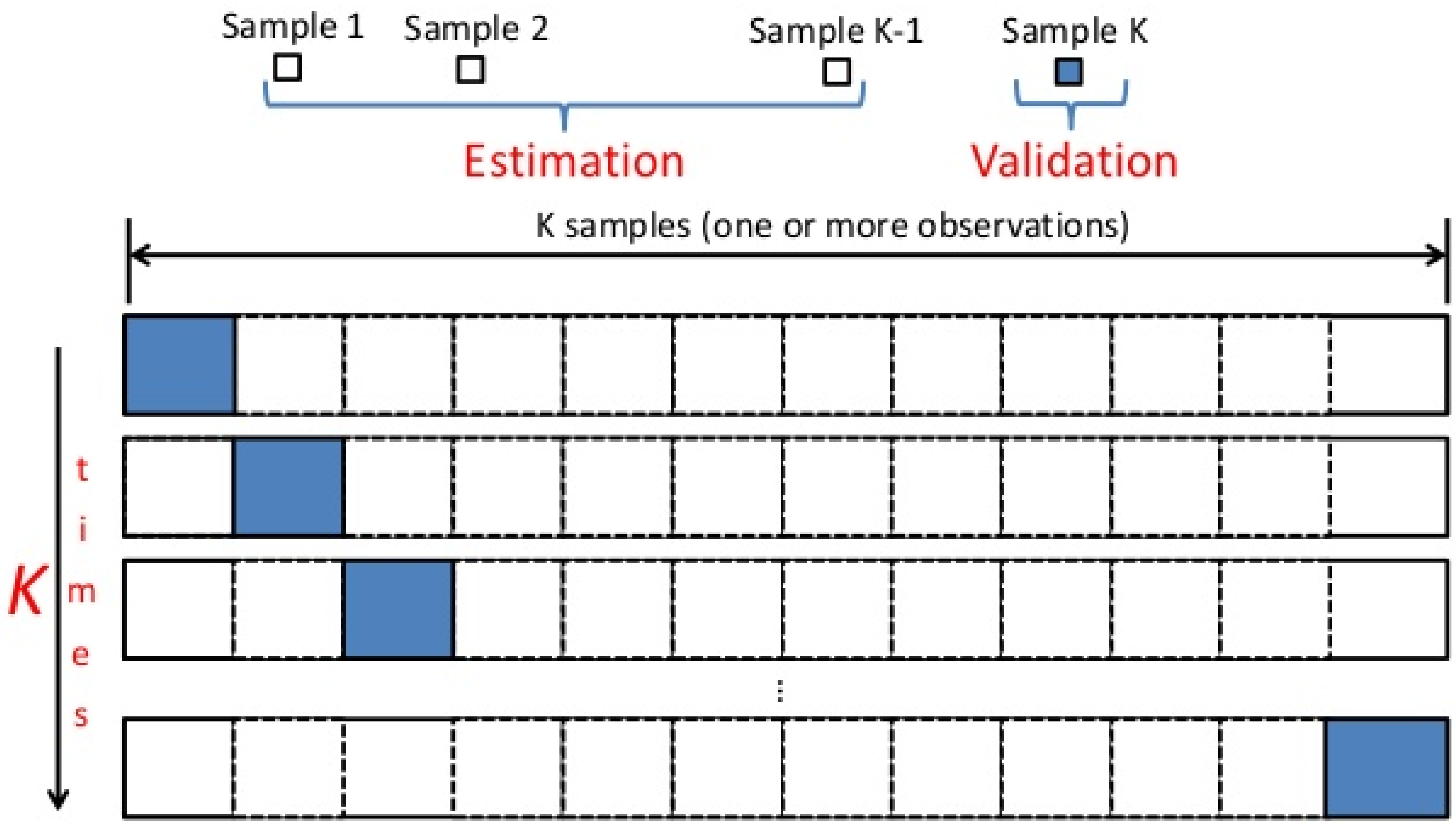
Cross-validation
Classification

Cross-validation

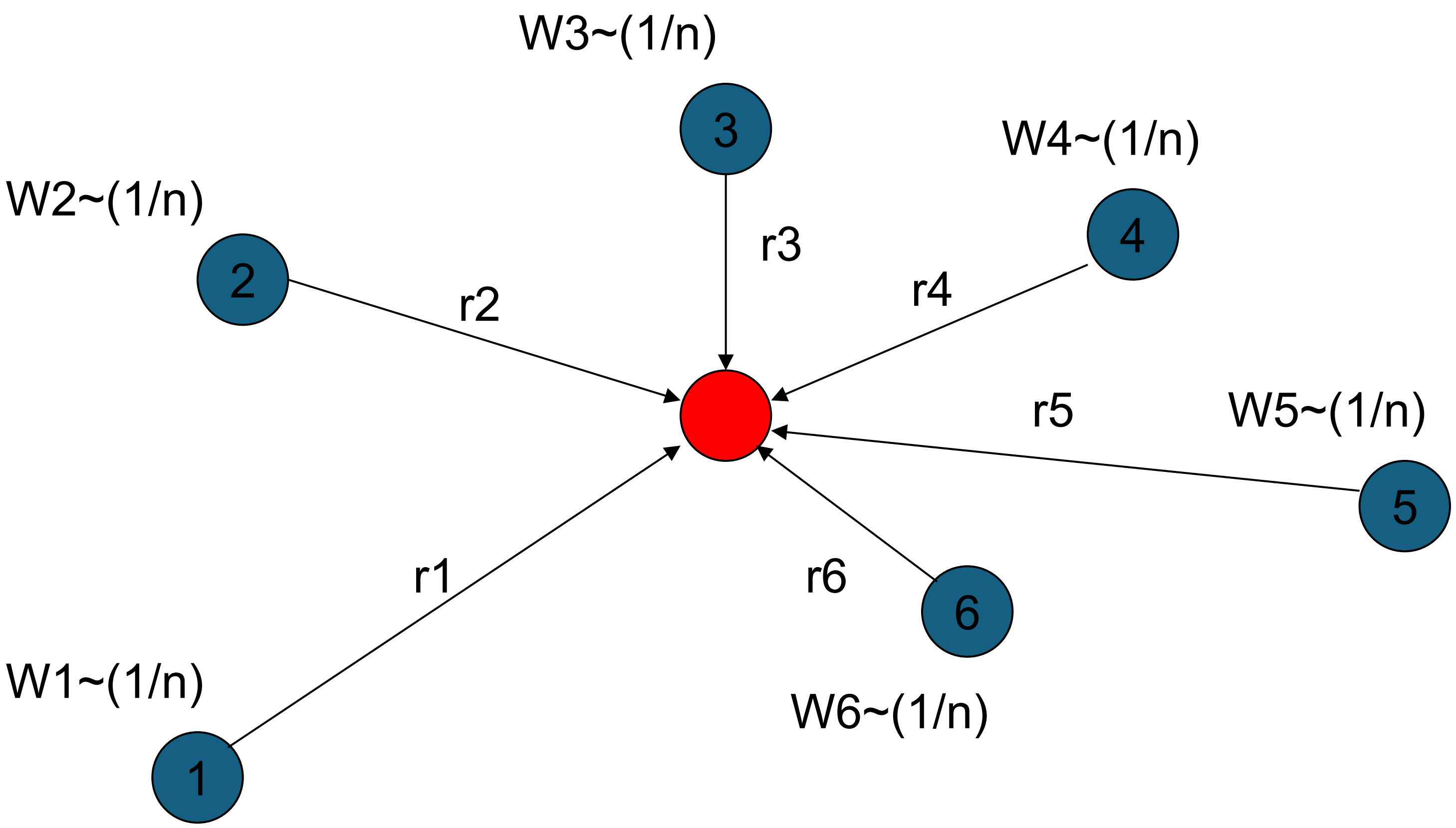
- Predict a value at the data location without using the date value available at this point
- A way to tune the prediction model parameters given the available data
- Compares the predicted value with the data one in one point at a time by removing the data point temporality from the data set.
- Cross-validation – leave-1-out
- Fold cross-validation – leave-fold-out (or jack-knife)

Cross-validation: How it works?

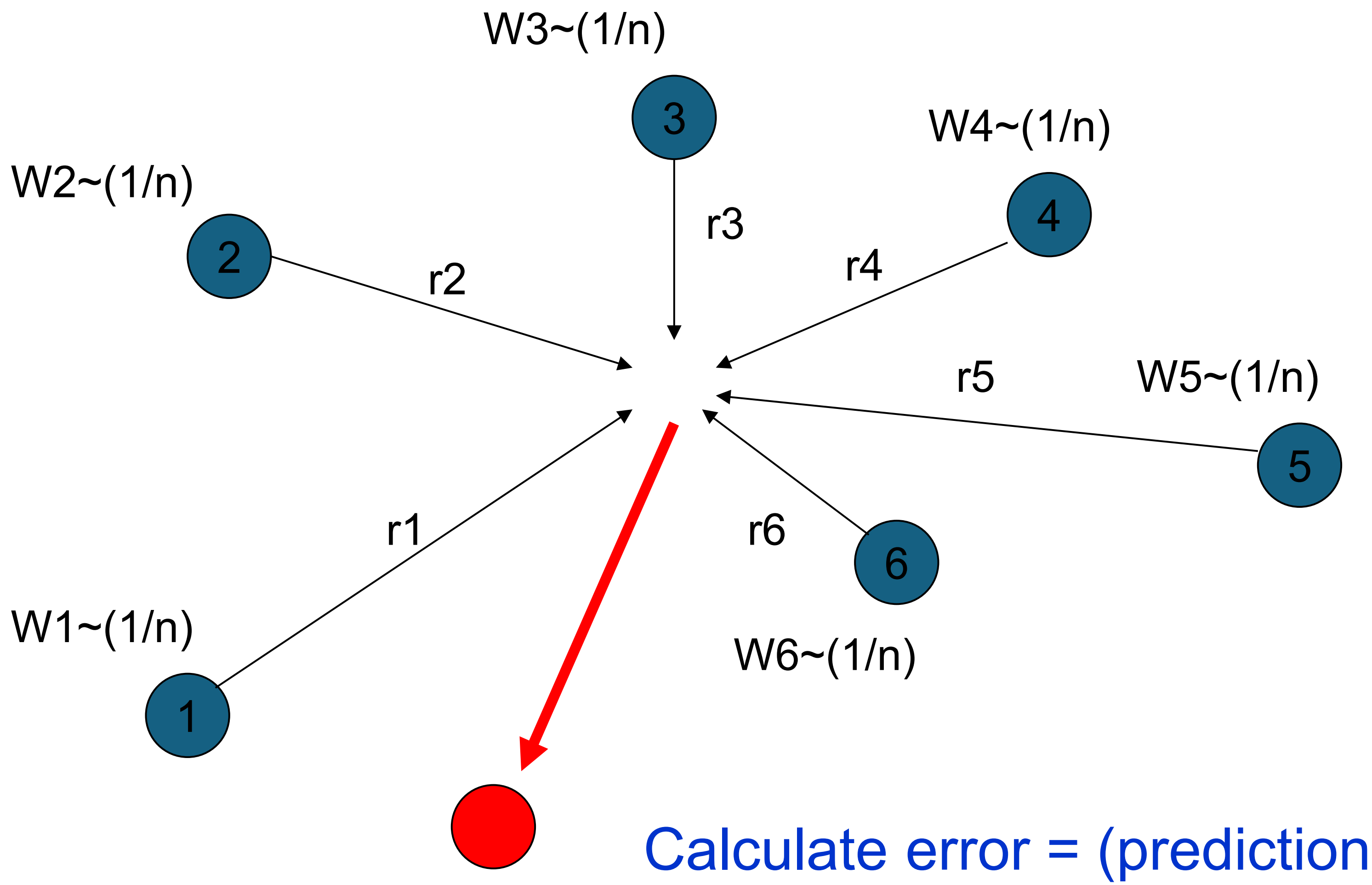
- K-fold cross-validation:



k-NN Interpolation (n=6 ?)

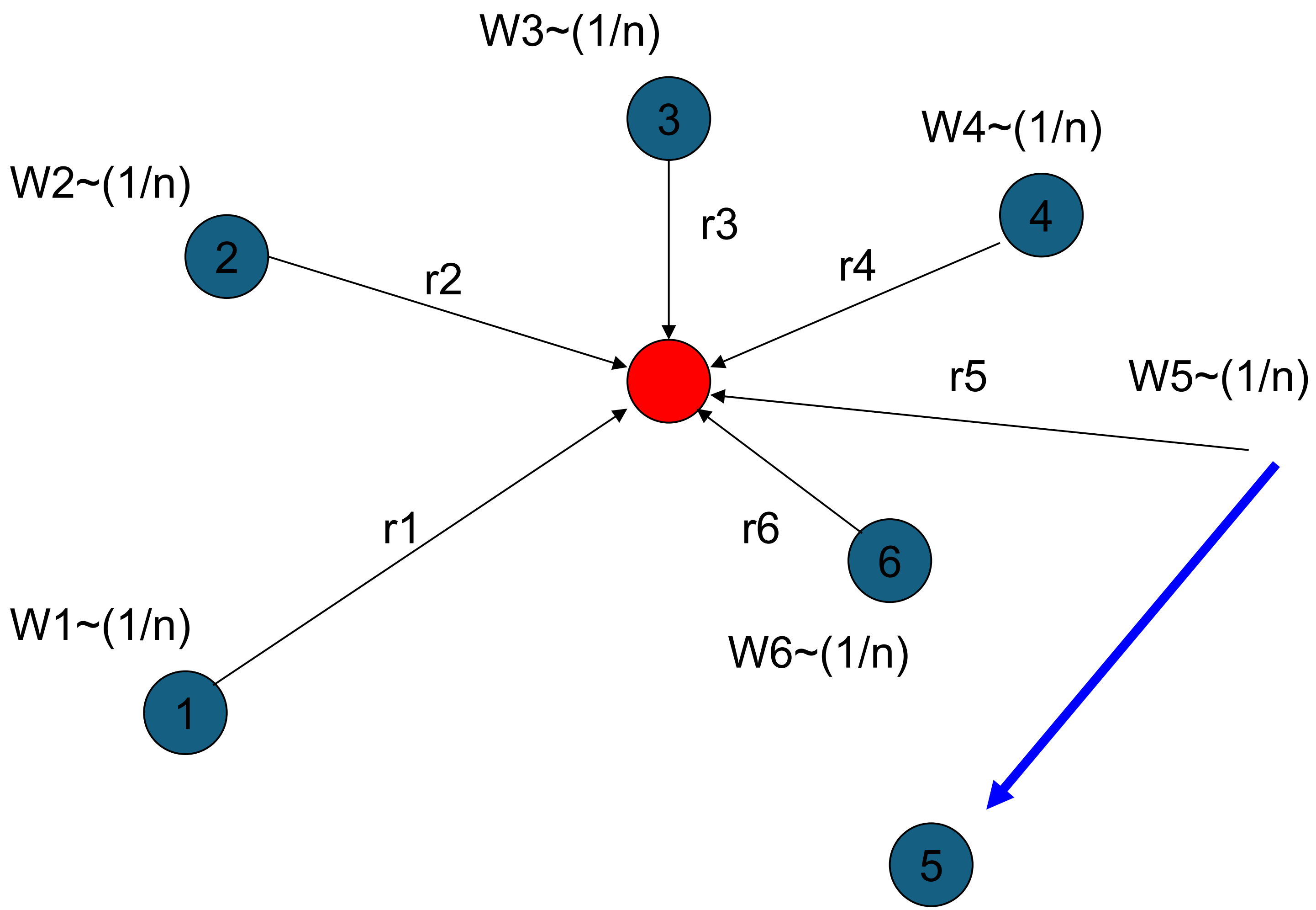


Cross-validation



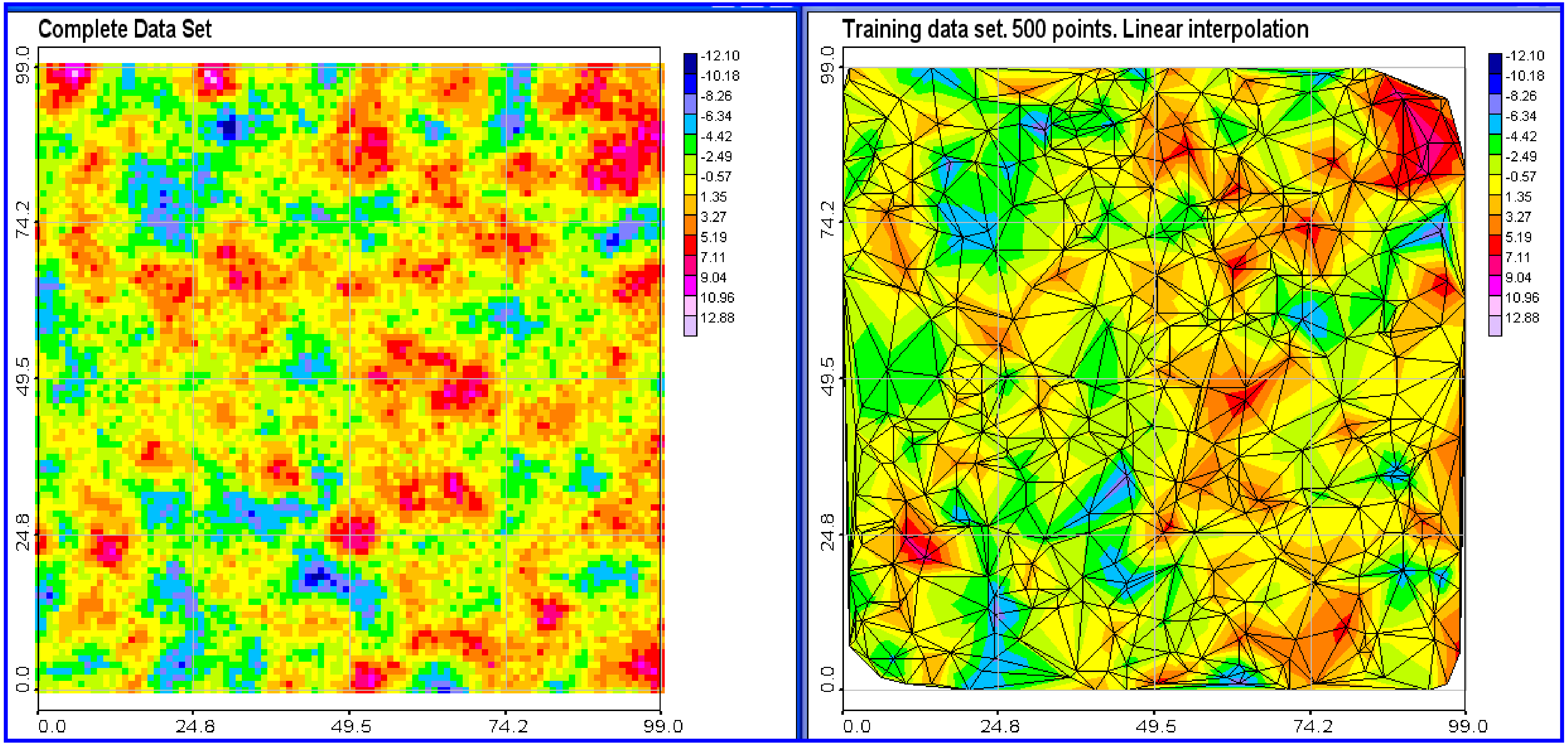
Calculate error = (prediction-data)

Leave-next-one-out, etc



Calculate error = (prediction-data)

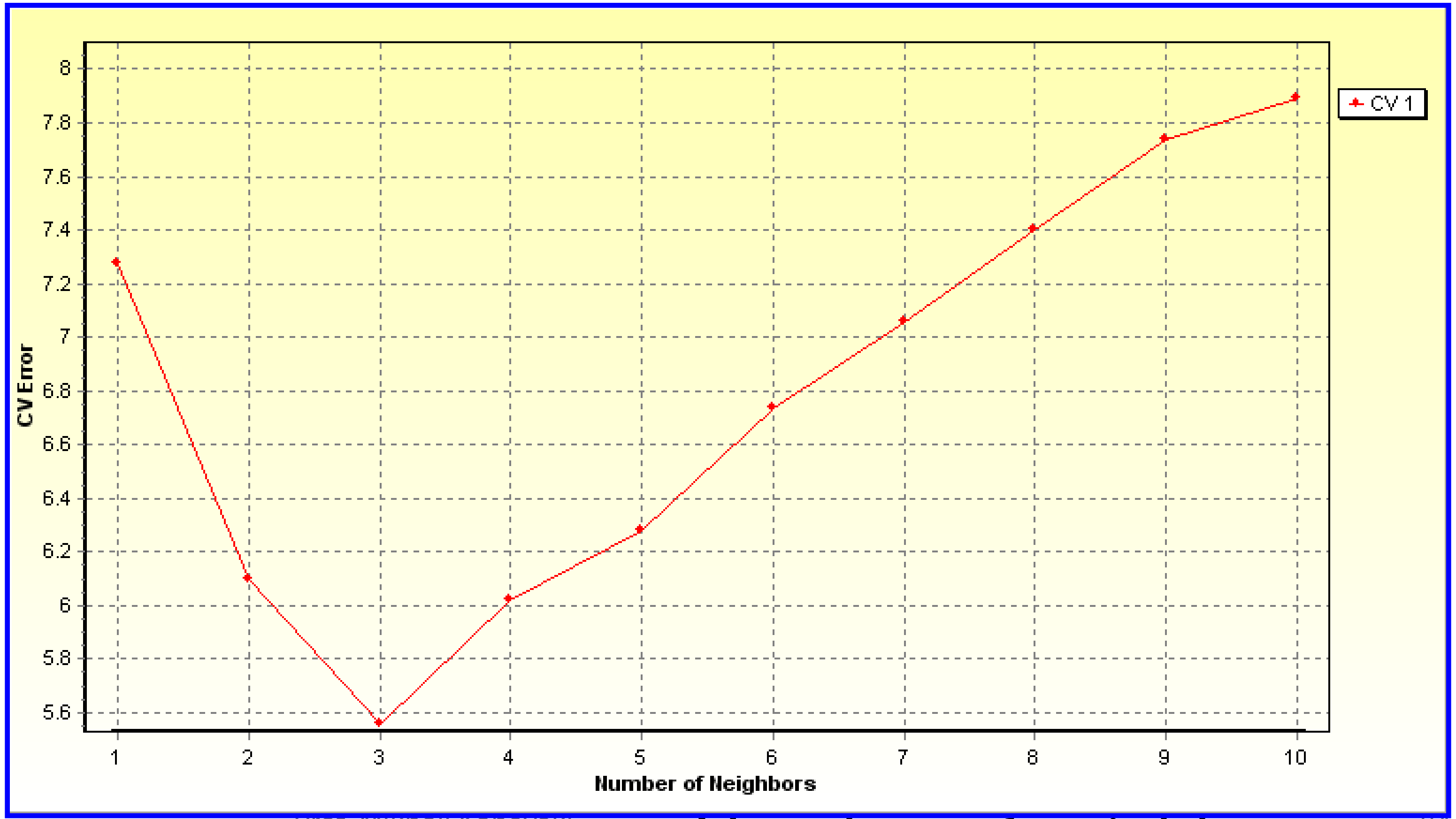
Prediction with a spatially correlate pattern



K-NN Cross-validation error curve

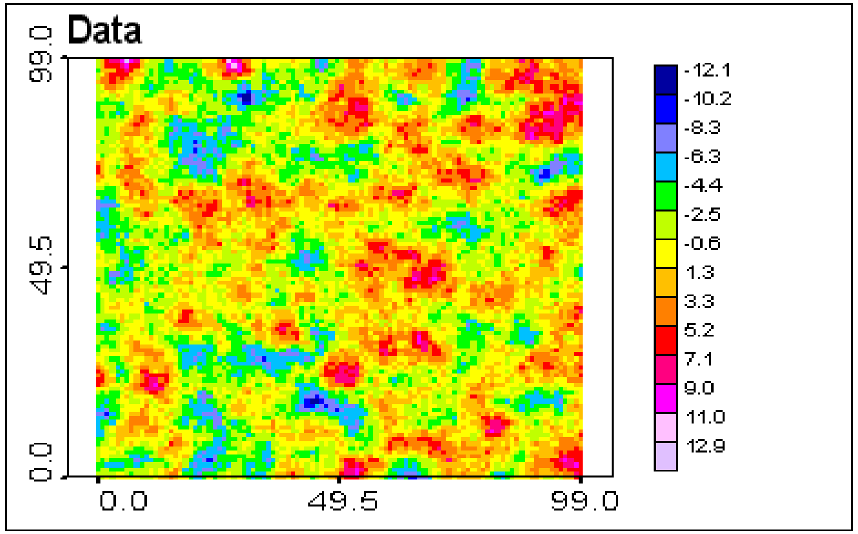
- How many K neighbours give the best prediction?

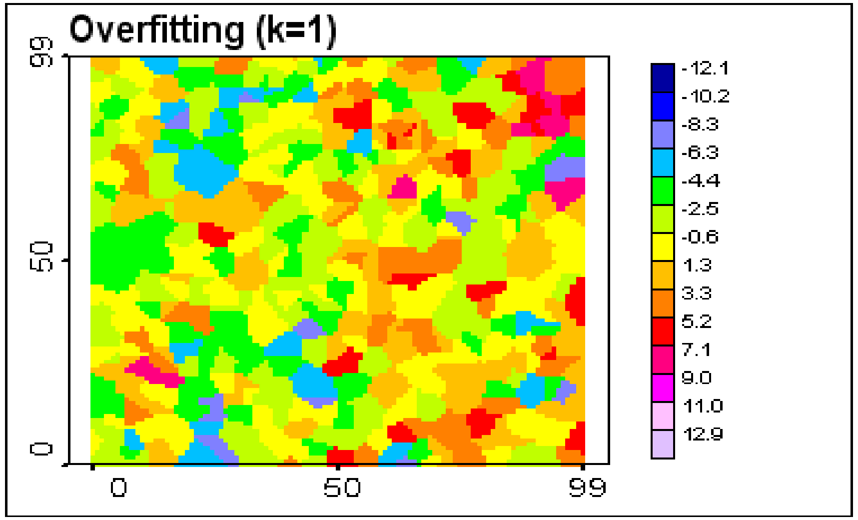
Cross-validation Error

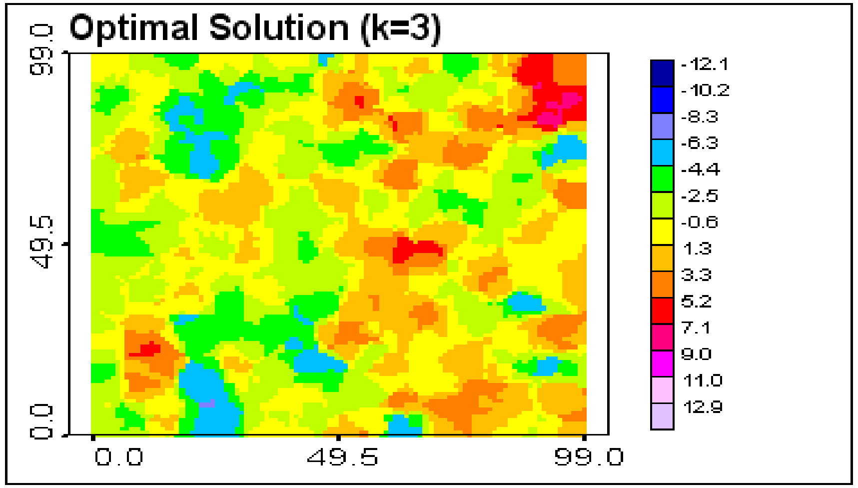


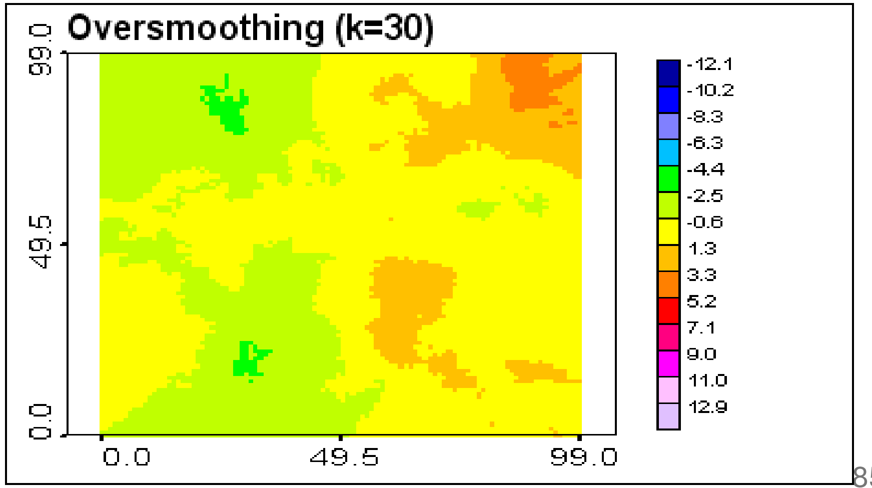
Number of neighbours

K-nn predictions





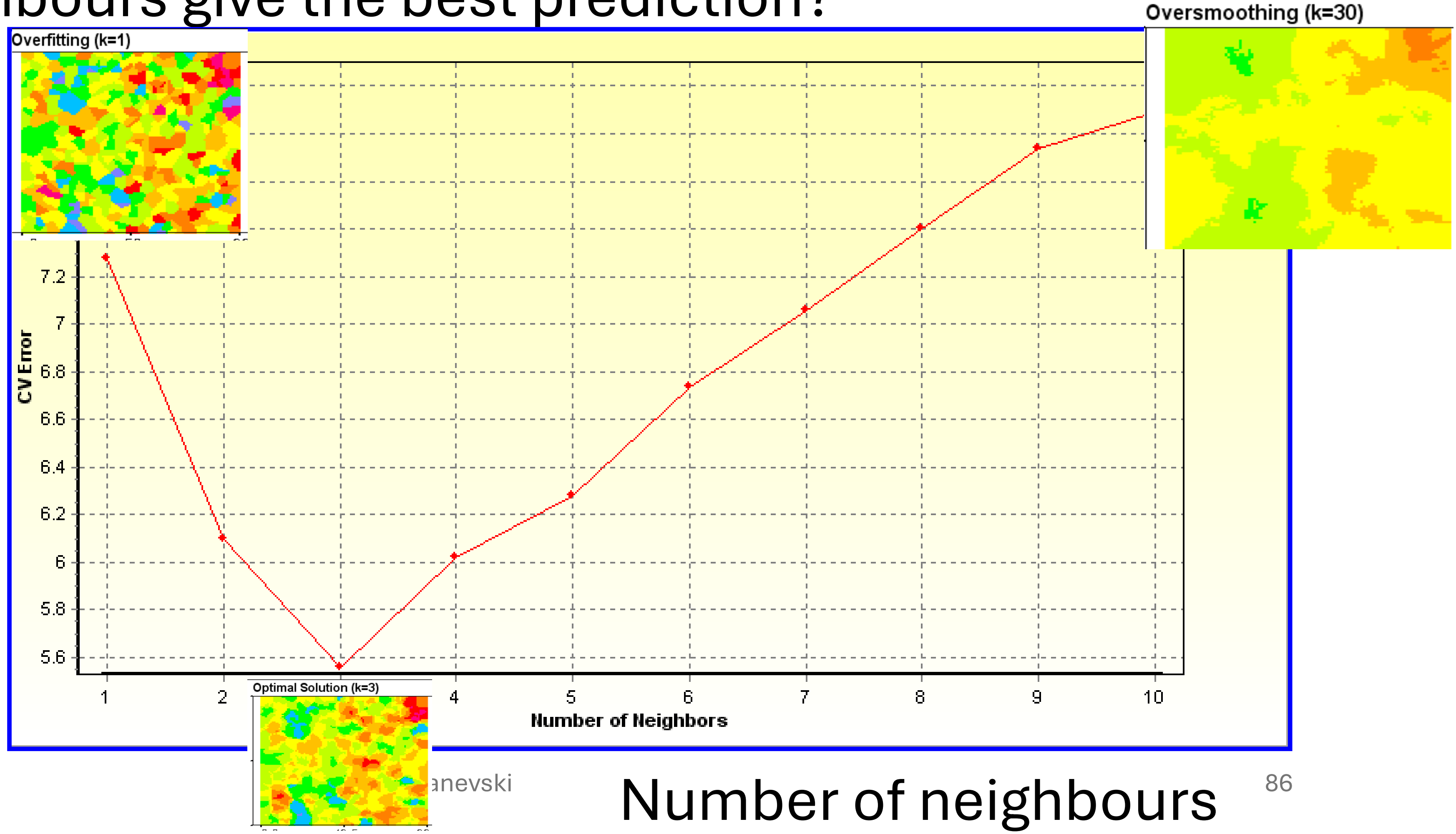


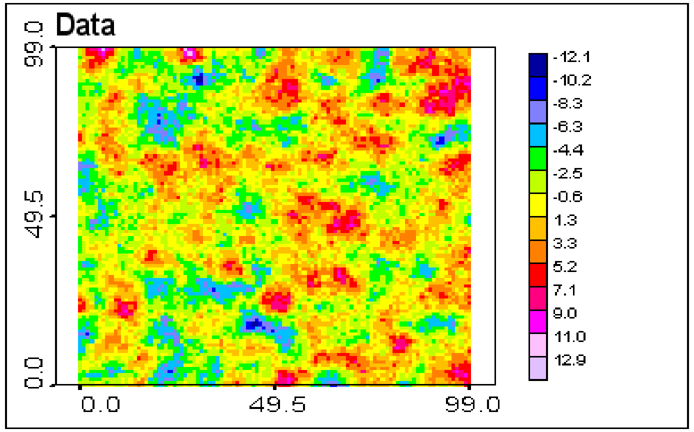


K-NN Cross-validation error curve

- How many K neighbours give the best prediction?

CV Error

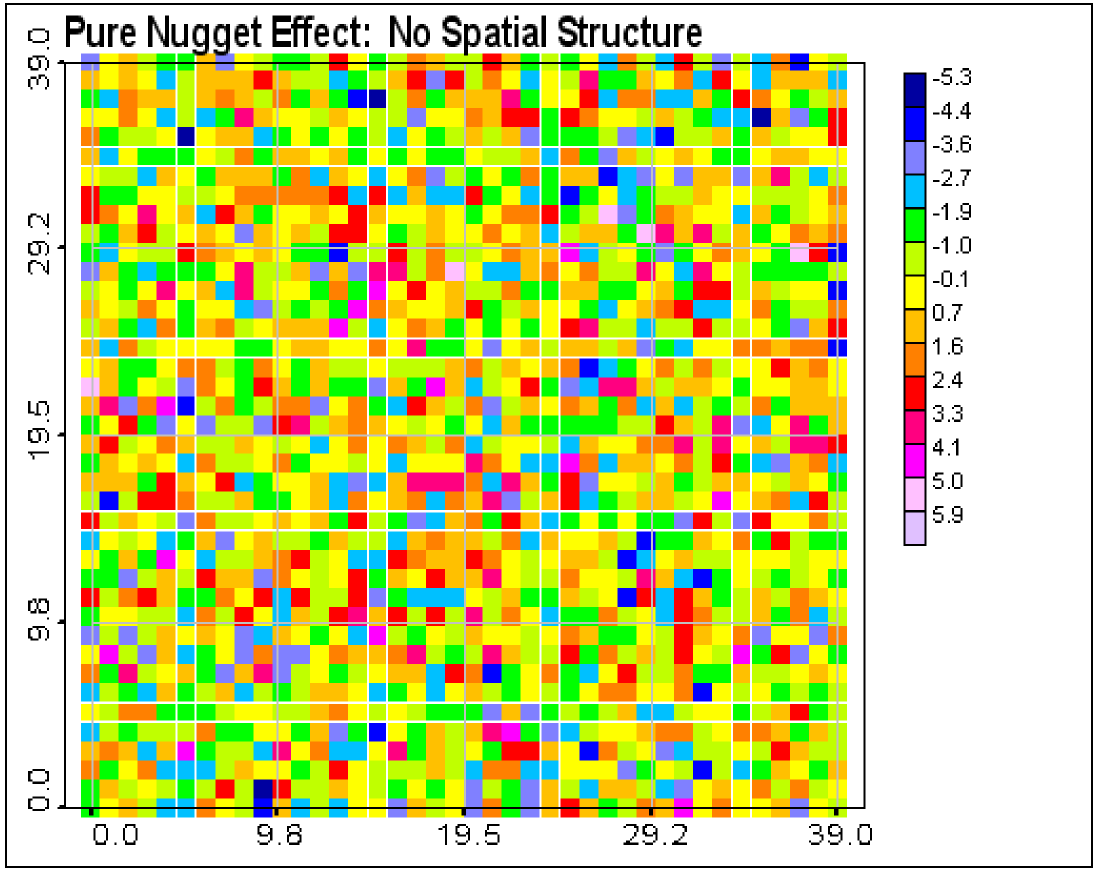




Number of neighbours

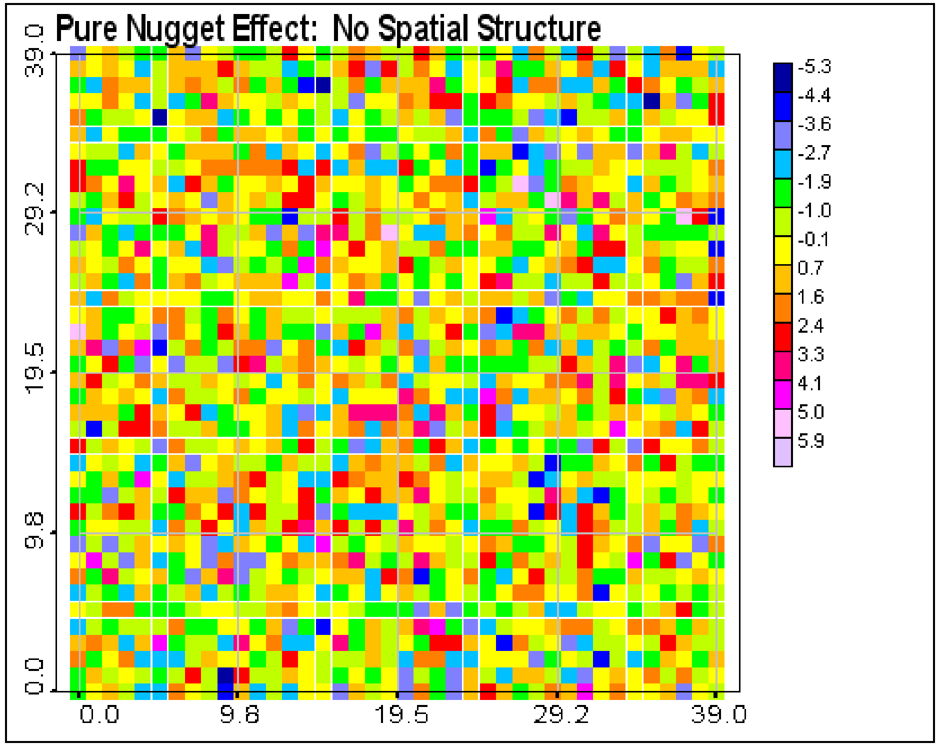
What is there is no spatial correlation - random pattern?

- Take a random value
- Take the mean value
- Does the spatial estimate depend on *x&y*?

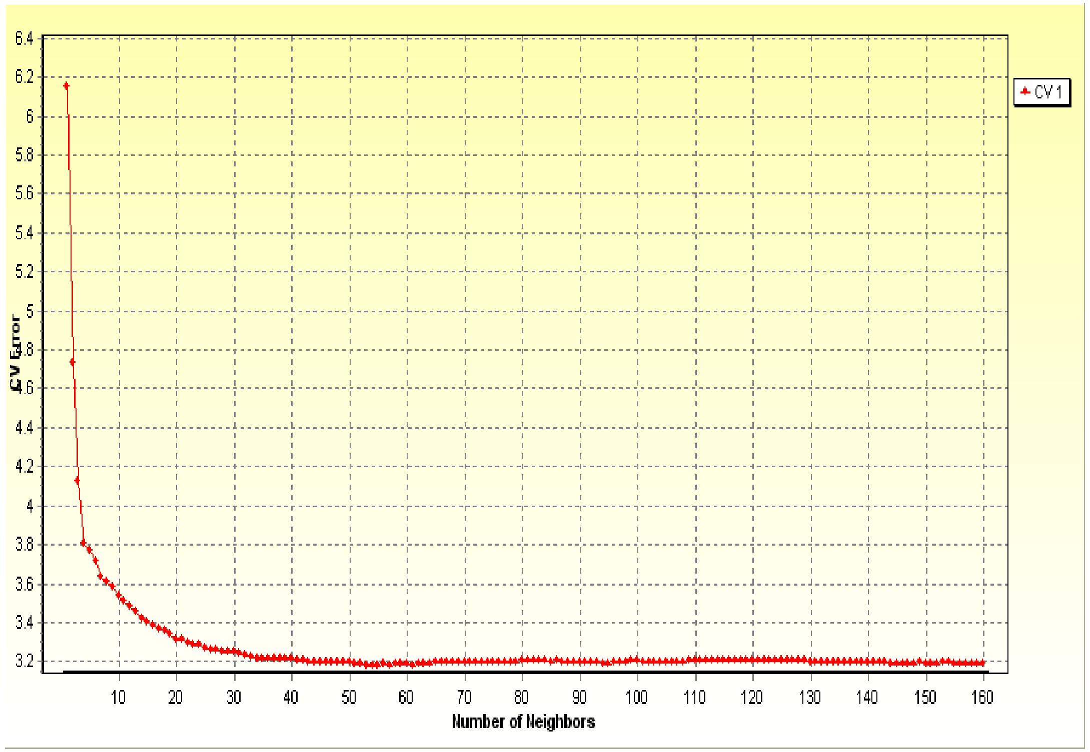


K-NN prediction of a random noise

- k-nn Cross-validation error curve has no minimum!



Error

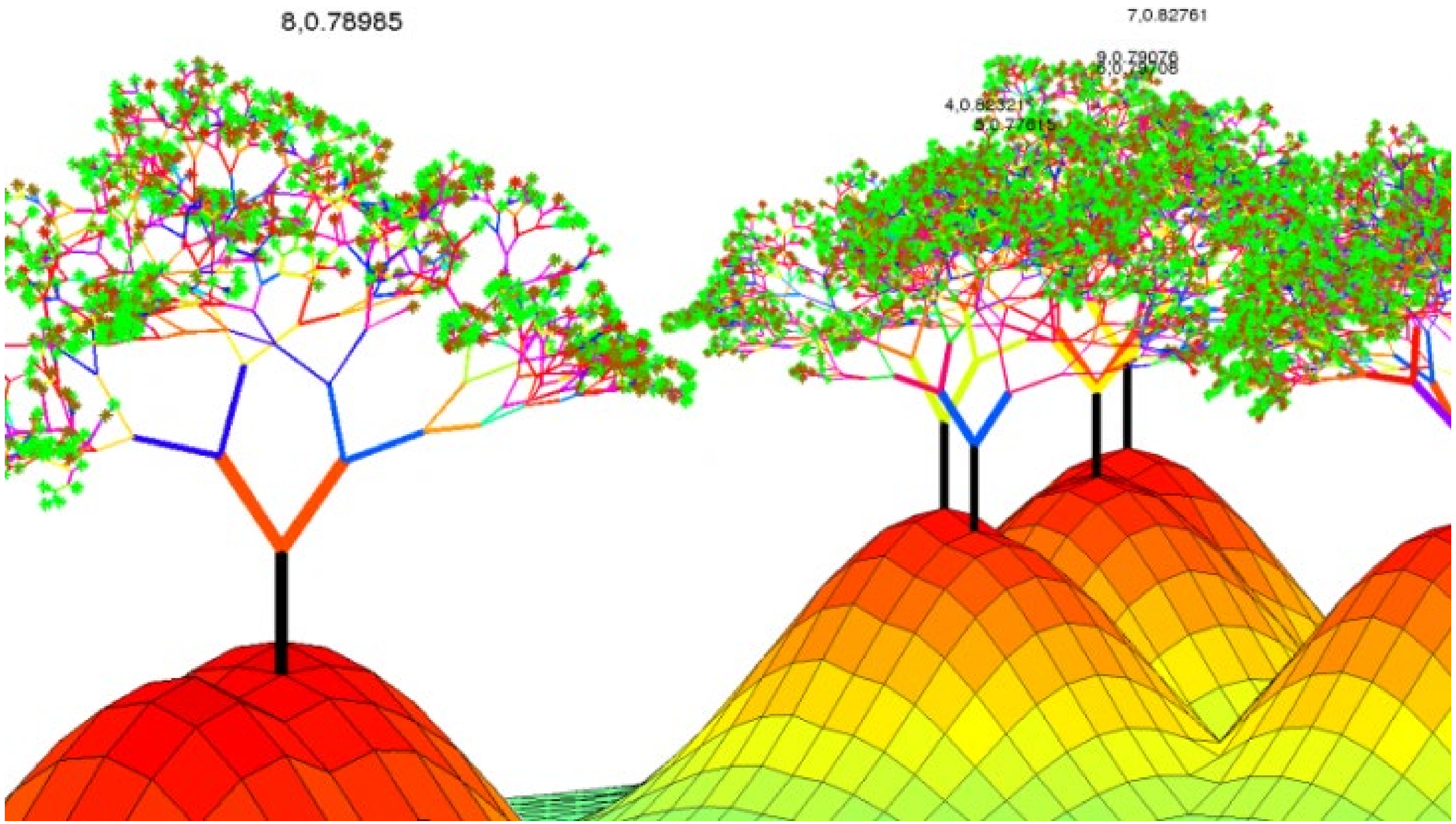


- Spatial x&y coodinates are not relevant input variable predictors

Summary

- Learning patterns from data – signal or noise?
- Avoid overfitting – match perfectly training data but fail to predict new data
- Complexity vs predictive power – bias variance dilemma
- Training / testing / validation sub-sets
- Cross-validation for parameter tuning
- Relevant vs irrelevant features

Random Forest



Classifier: Random Forest

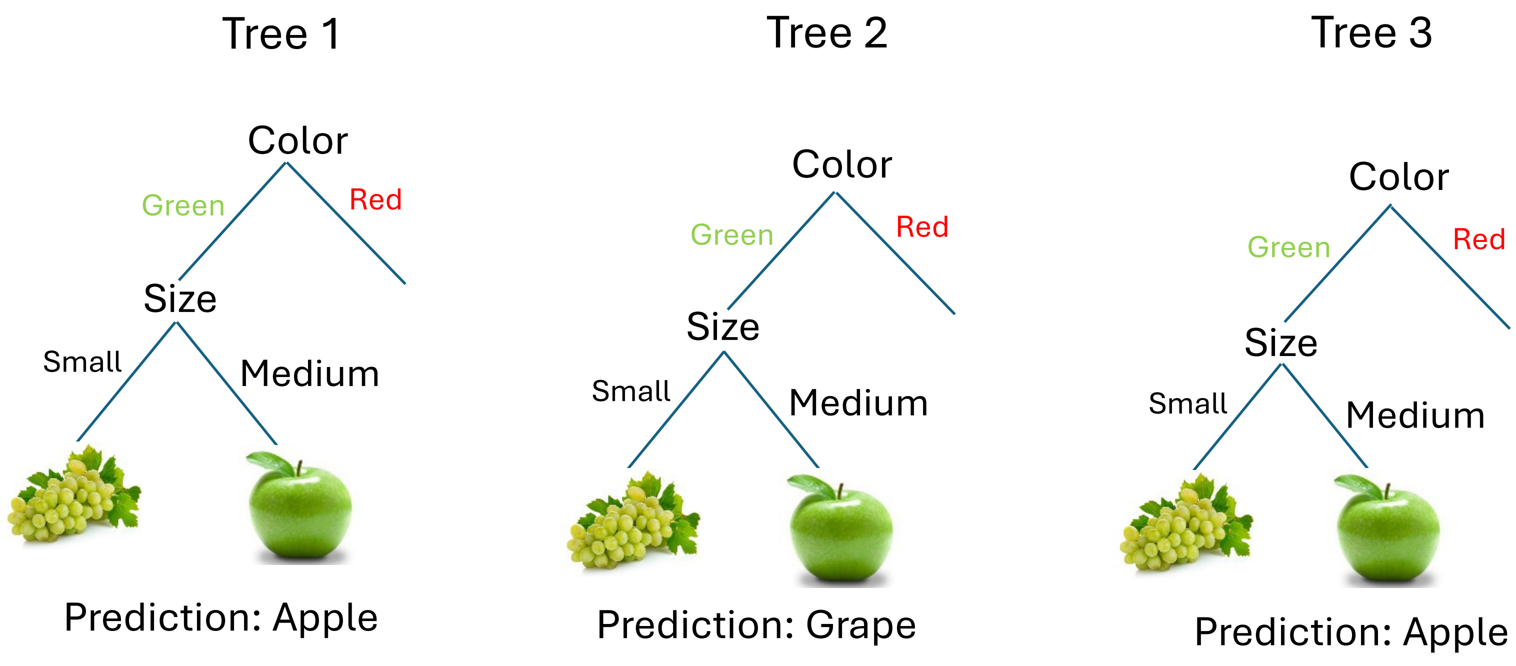
What is a random forest?

An ensemble of decision trees which predict an instance by the most popular class

Where does the random component come from?

Trees constructed on

- Bagging
- Random feature selection

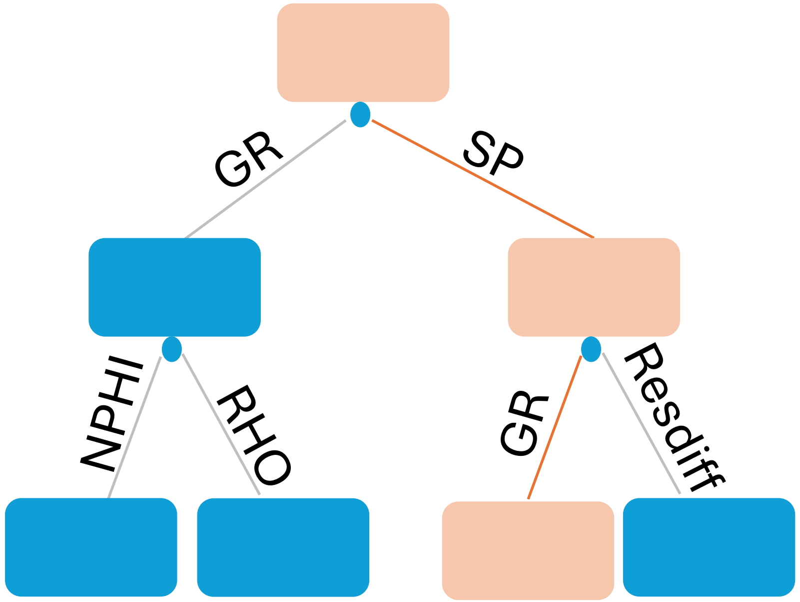


Majority vote: Apple

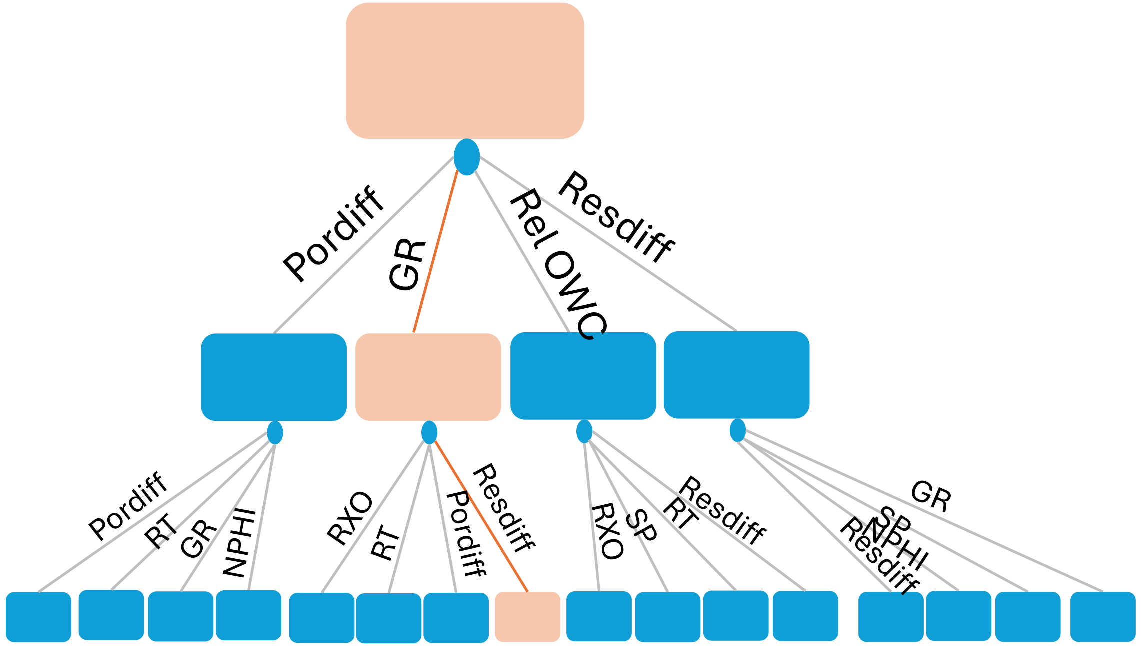
Growing a tree and a forest: RF parameters

Number of features to consider when looking for the best split + number of trees

2 features (out of 9)

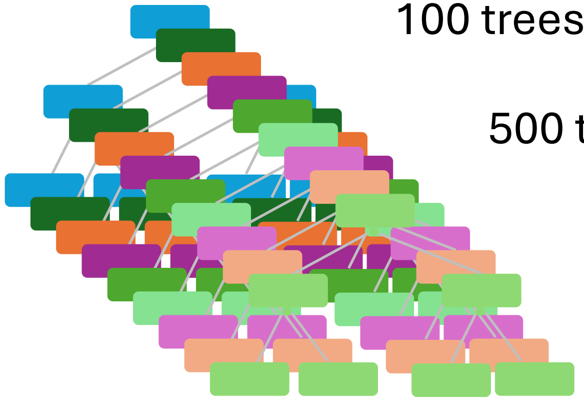


4 features (out of 9)?



Random forest parameters

How many trees in the forest?



100 trees?

500 trees?

900 trees?

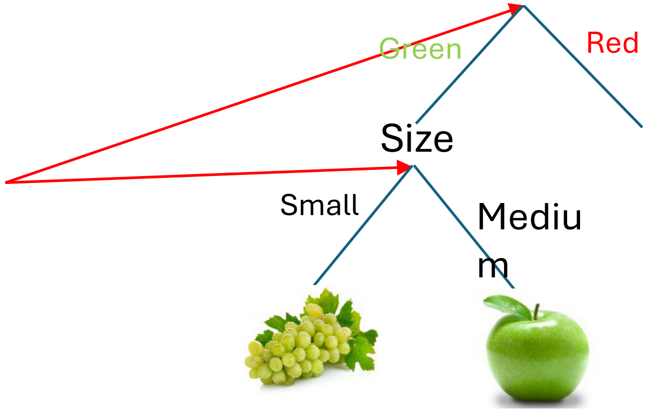
Combination of parameters which yields the lowest prediction error

How many splits at a node?

Each node is split with a set number of features

Tree 1

Nod
e



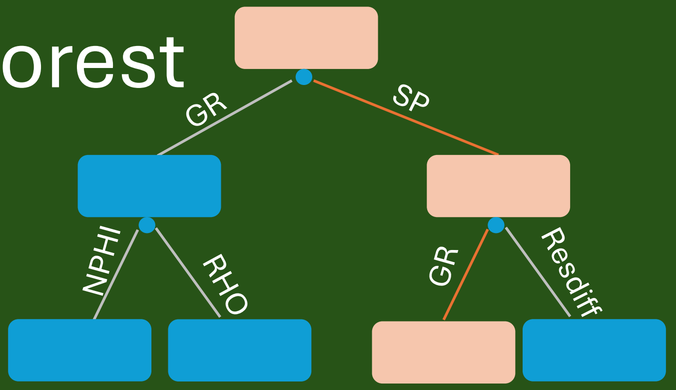
Classifier Selection

What are the right inputs?

INPUT: wireline logs

NPHI DT GR PEF RHOB RT BVW VSH KLOGH

Random Forest (RF)

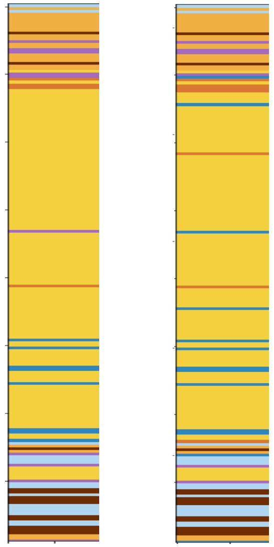


Leo Breiman, 2001

Classification

OUTPUT

Predicted True



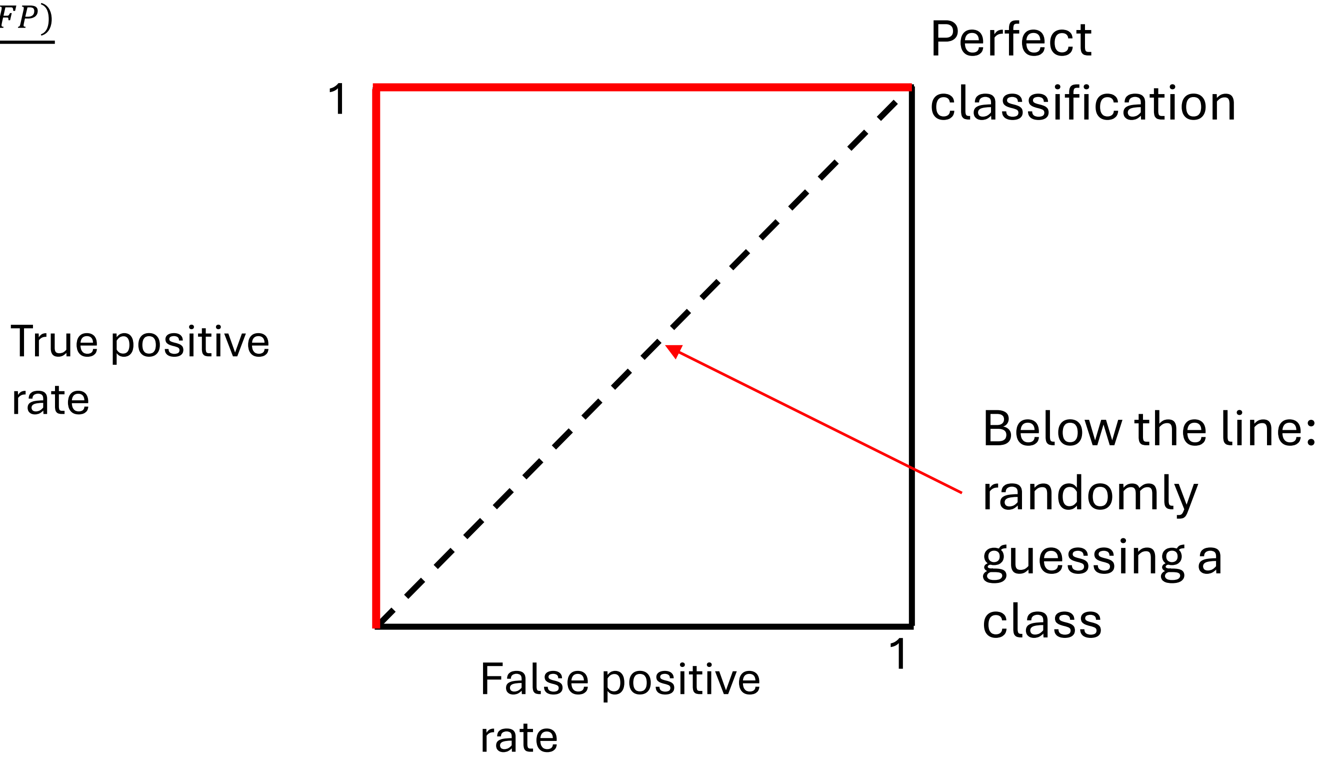
Classification quality

True positive rate = $\frac{Positives\ correctly\ classified\ (TP)}{Total\ positives\ (TP+FP)}$

False positive rate = $\frac{Negatives\ incorrectly\ classified\ (FP)}{Total\ negatives\ (FN+TN)}$

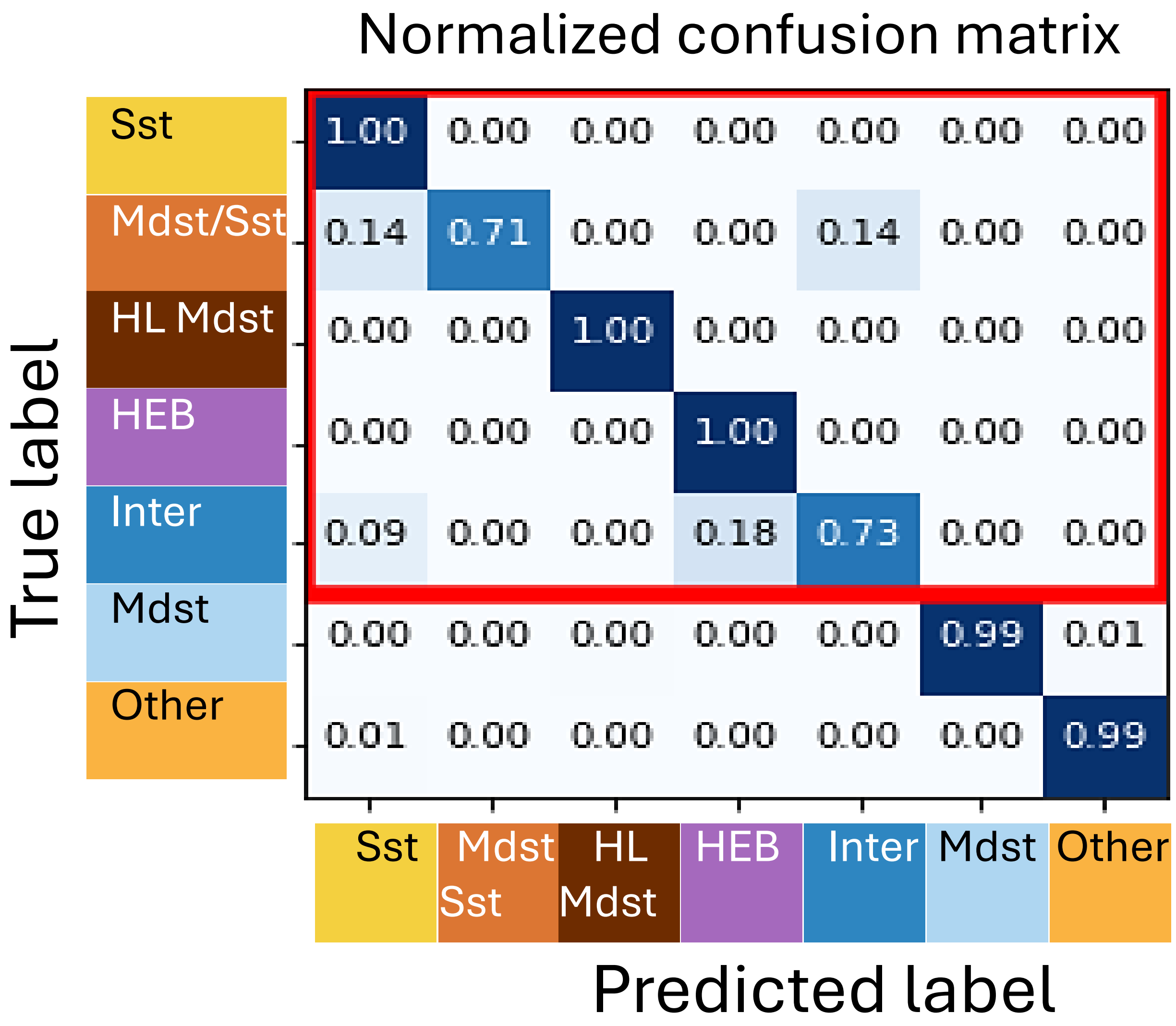
True labels

		A	B
Predicted labels	A	True positive (TP)	False Positive (FP)
	B	False Negative (FN)	True Negative (TN)

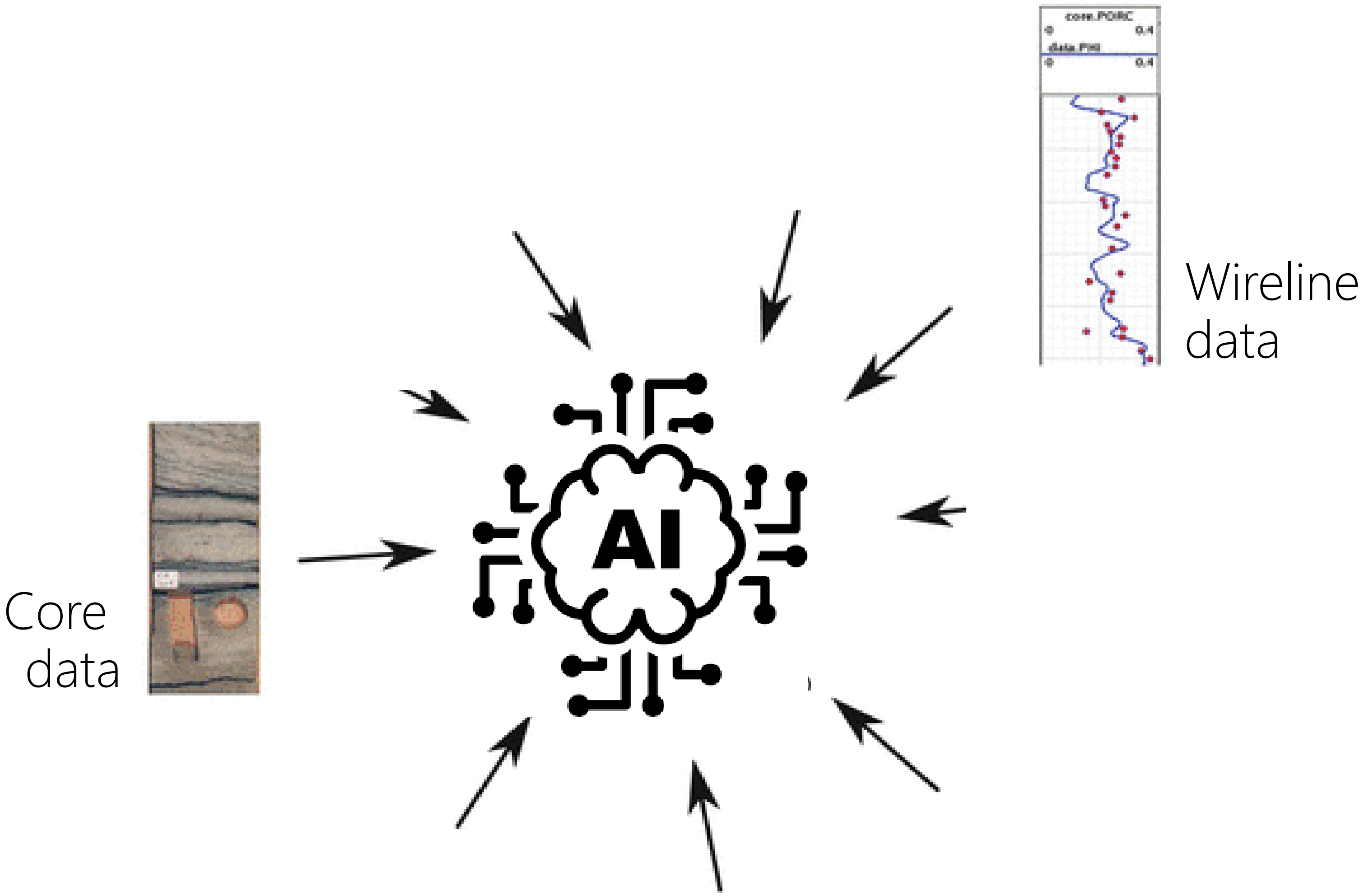


How to evaluate the TEST performance?

Misinterpreted facies
- interbedded Mdst/Sst
and Intermediate



Data integration in reservoir prediction modelling

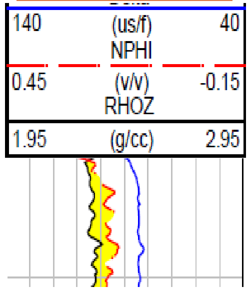


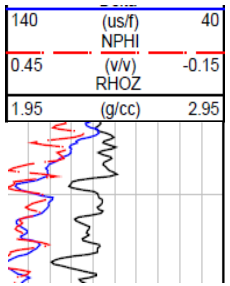
Modified from Rivenæs, J.C., Sørhaug, P., Knarud, R. (2015). Introduction to Reservoir Modelling. In: Bjørlykke, K Petroleum Geoscience. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-34132-8_22

What are the right input features?

INPUT: wireline logs + engineered features

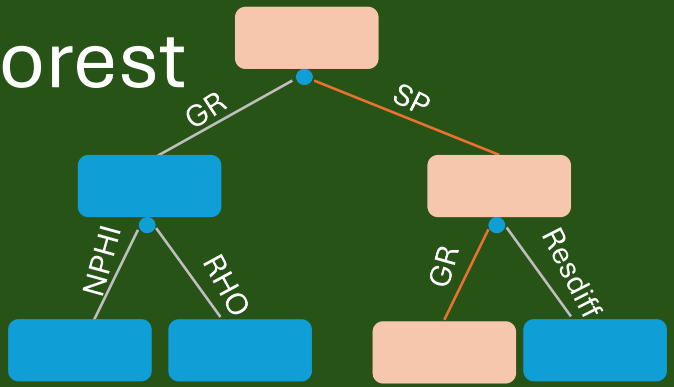
GR	NPHI	RHO	Pordiff	RT	RXO	Resdiff	SP	OWC





Random Forest (RF)

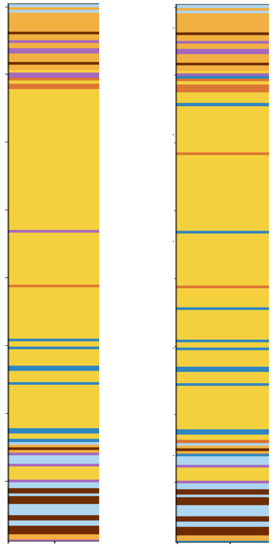
Leo Breiman, 2001



Classification

OUTPUT

Predicted True



Value of Geologically Derived Features in Machine Learning Facies Classification

Julie Halotel,
Vasily Demyanov &
Andy Gardiner

Mathematical Geosciences volume 52, pages 5–29 (2020)

Math Geosci (2020) 52:5–29
https://doi.org/10.1007/s11004-019-09838-0

Check for updates

SPECIAL ISSUE

Value of Geologically Derived Features in Machine Learning Facies Classification

Julie Halotel[1,2] · Vasily Demyanov[1] · Andy Gardiner[1]

Received: 12 November 2018 / Accepted: 1 November 2019 / Published online: 16 November 2019

Abstract The aim of this work is to demonstrate how geologically interpretative features can improve machine learning facies classification with uncertainty assessment. Manual interpretation of lithofacies from wireline log data is traditionally performed by an expert, can be subject to biases, and is substantially laborious and time consuming for very large datasets. Characterizing the interpretational uncertainty in facies classification is quite difficult, but it can be very important for reservoir development decisions. Thus, automation of the facies classification process using machine learning is a potentially intuitive and efficient way to facilitate facies interpretation based on large-volume data. It can also enable more adequate quantification of the uncertainty in facies classification by ensuring that possible alternative lithological scenarios are not overlooked. An improvement of the performance of purely data-driven classifiers by integrating geological features and expert knowledge as additional inputs is proposed herein, with the aim of equipping the classifier with more geological insight and gaining interpretability by making it more explanatory. Support vector machine and random forest classifiers are compared to demonstrate the superiority of the latter. This study contrasts facies classification using only conventional wireline log inputs and using additional geological features. In the first experiment, geological rule-based constraints were implemented as an additional derived and constructed input. These account for key geological features that a petrophysics or geological expert would attribute to typical and identifiable wireline log responses. In the second experiment, geological interpretative features (i.e., grain size, pore size, and argillaceous content) were used as additional independent inputs to enhance the prediction accuracy and

✉ Julie Halotel
julie.halotel@gmail.com

1 Heriot-Watt University School of Energy Geoscience Infrastructure and Society, Edinburgh, UK

2 Present Address: Equinor, Oslo, Norway

Facies Classification

Hands-on

Supervised Classification. Notebook 3

https://colab.research.google.com/drive/1-9Az597WtJgnU_5o0FHbbc41u3TtABs0?usp=sharing

Exercise 3

Task 3.1: KNN classification:

- Select training/validation/testing wells.
- Plot lithology distributions for the three sub-sets and compare.
- Select input log variables and screen their stats between training/validation/test sets.
- Optimise **K** parameter with validation accuracy curve – select max accuracy.
- Compare test predictions with optimal/suboptimal k
- Save plots in Word/PowerPoint (Copy image/Paste)

Task 3.2: Random forest classification

- Train RF classifier
- Make predictions on training/validation/test wells and compare with KNN perditions (Task 3.1).
- Select a difference combination input log variables and compare the outcomes.

Task 3.3: Select different wells for training/validation/test

- Repeat Tasks 3.1, 3.2, compare the outcomes.

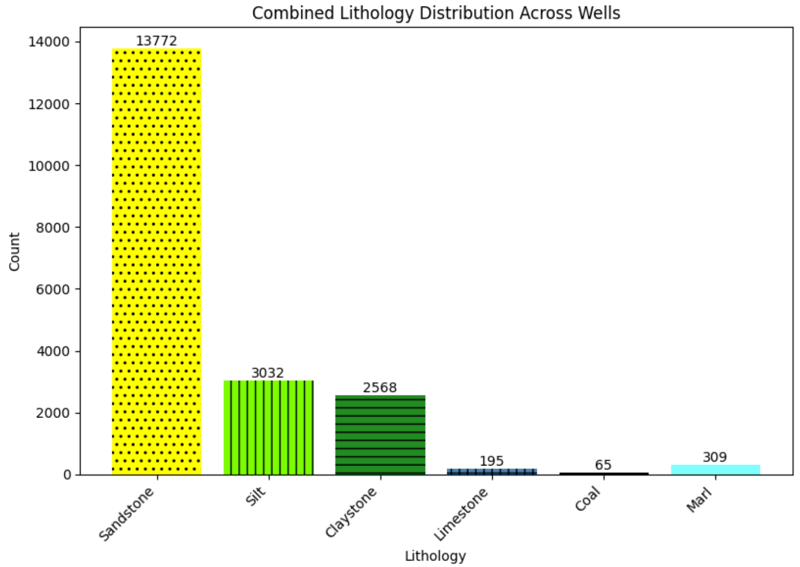
Data sub-set Purpose

- The **training** set is used to fit the models
- The **validation** set is used to estimate prediction error for model selection (tuning hyperparameters)
- The **test** set is used for assessment of the generalization error of the final chosen model

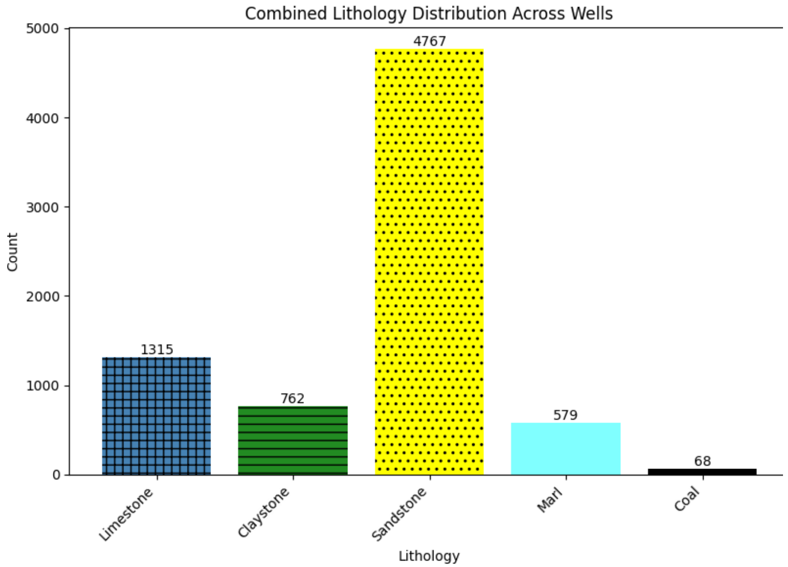
The Elements of Statistical Learning- Hastie, Tibshirani & Friedman 2009

Lithology distributions

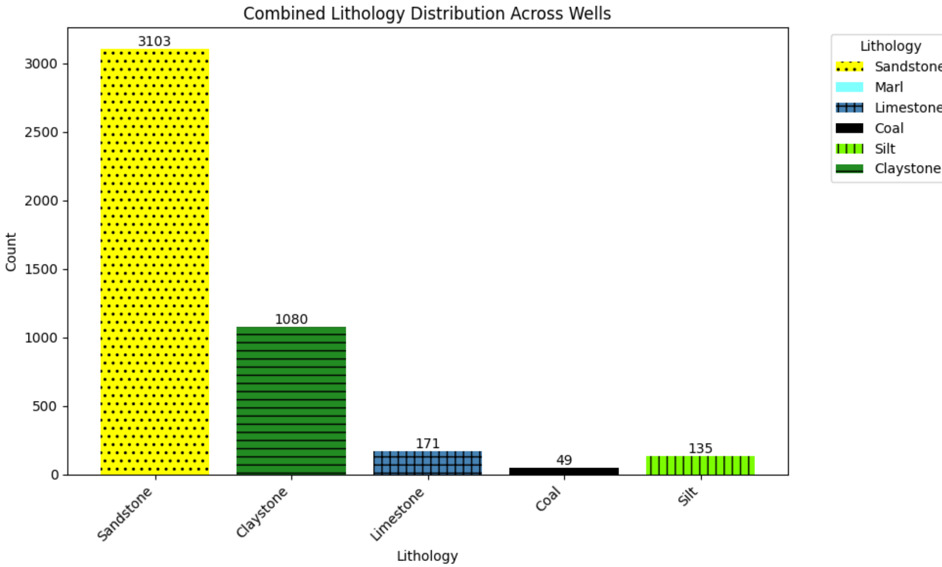
Training wells



Validation wells

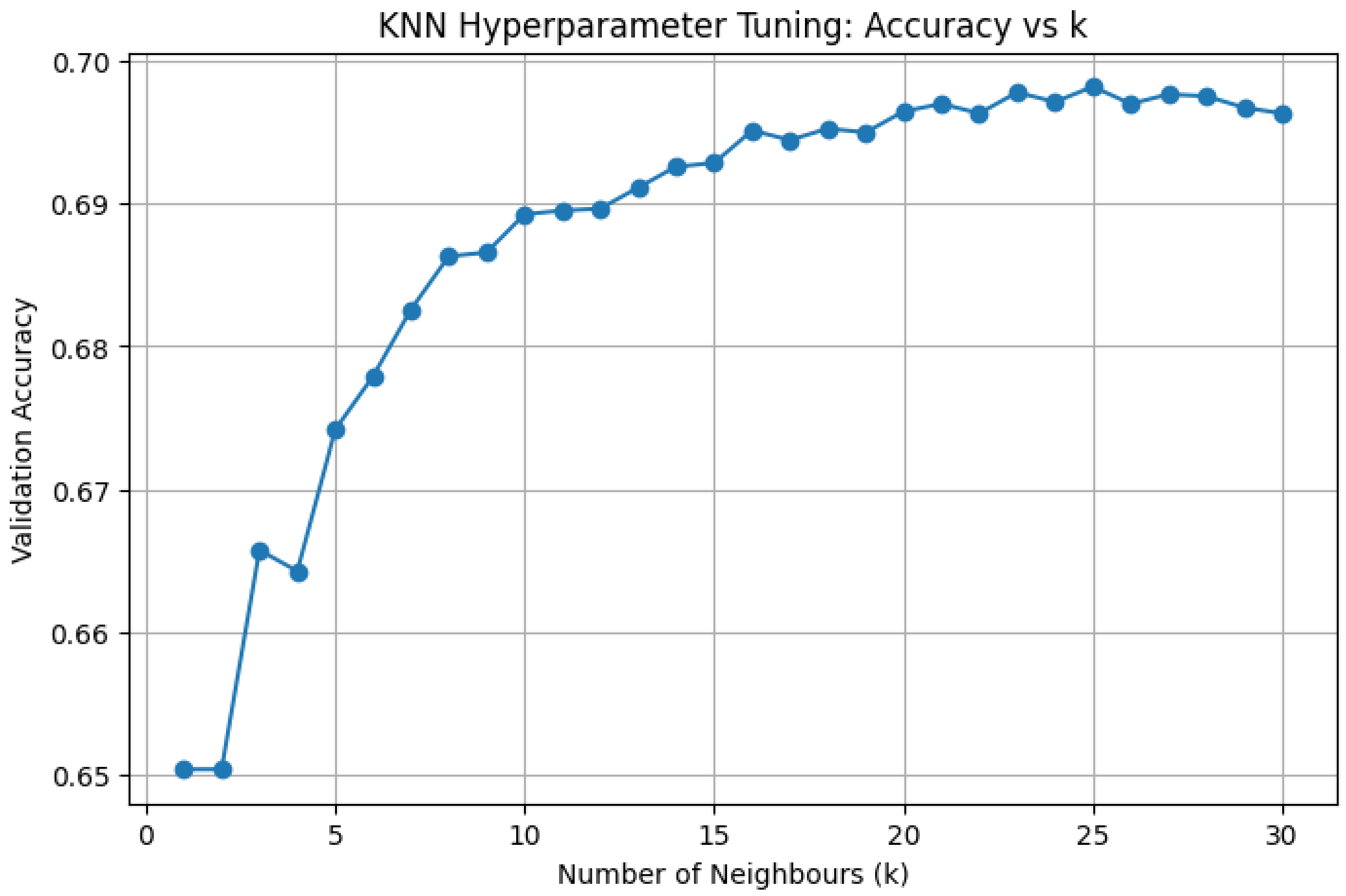


Test wells



Validation accuracy curve

- Max validation accuracy
- k = 25

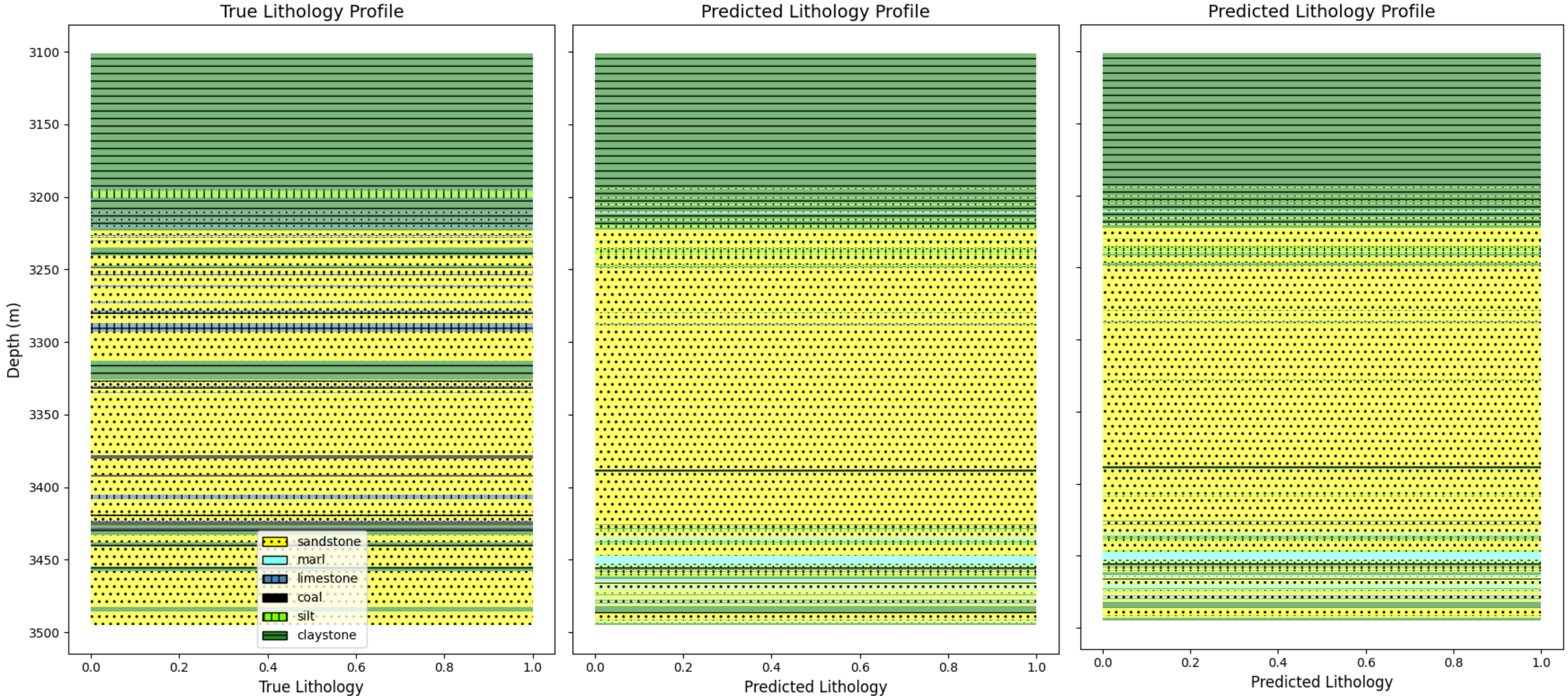


Validation well lithology classification

- Reference

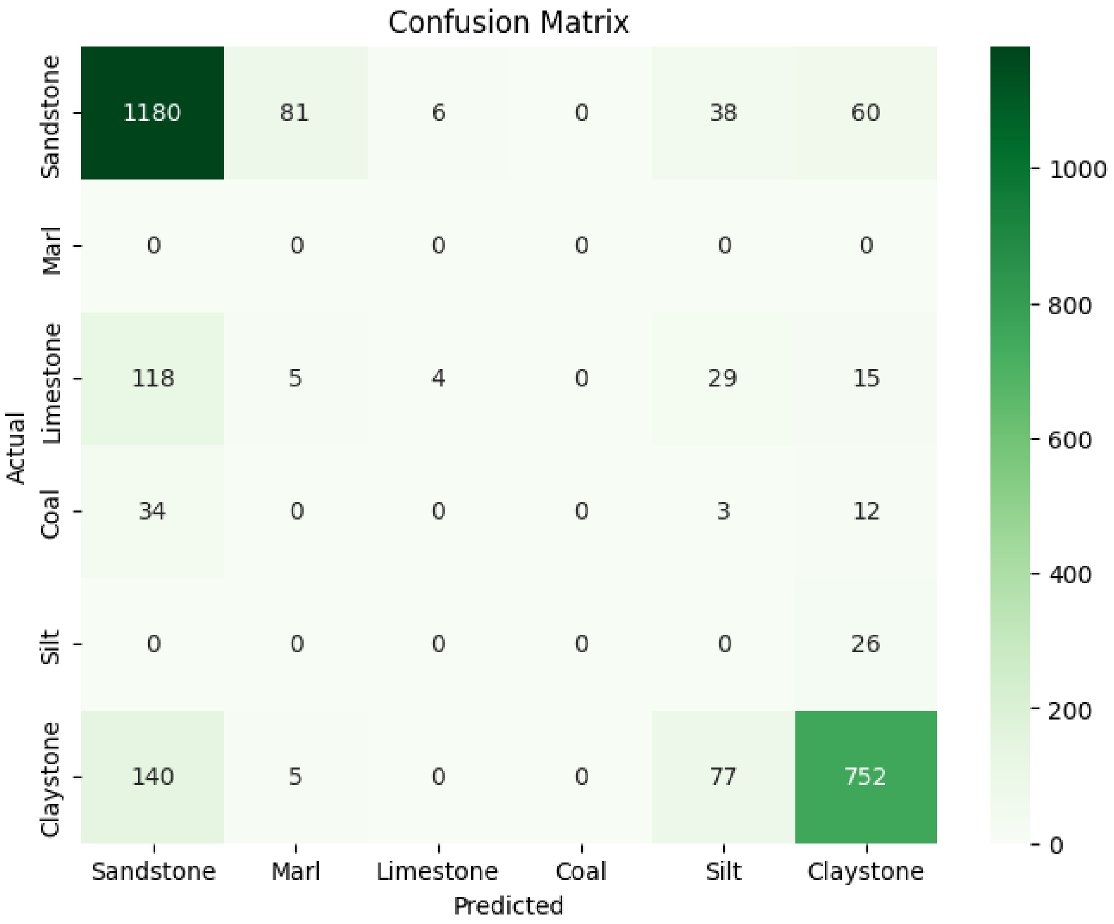
k = 25, error= 0.749

k = 10, error= 0.740

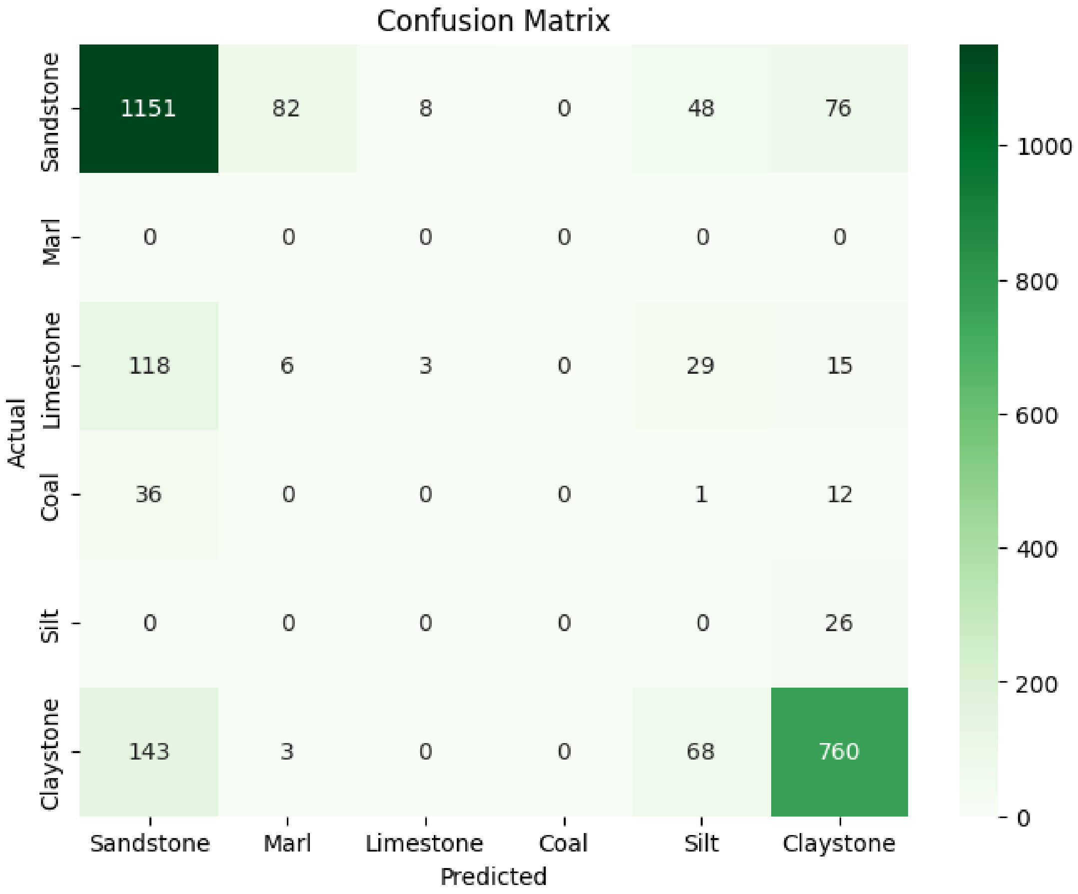


Confusion matrix: KNN

- k = 25

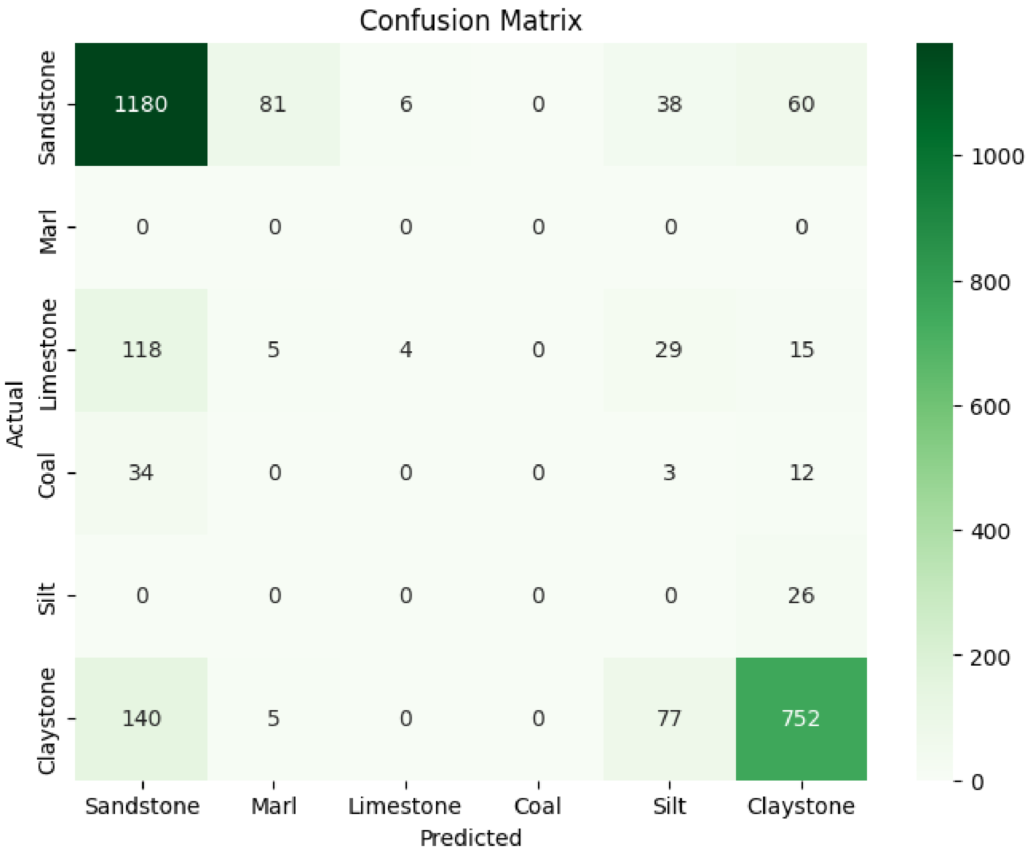


- k = 10

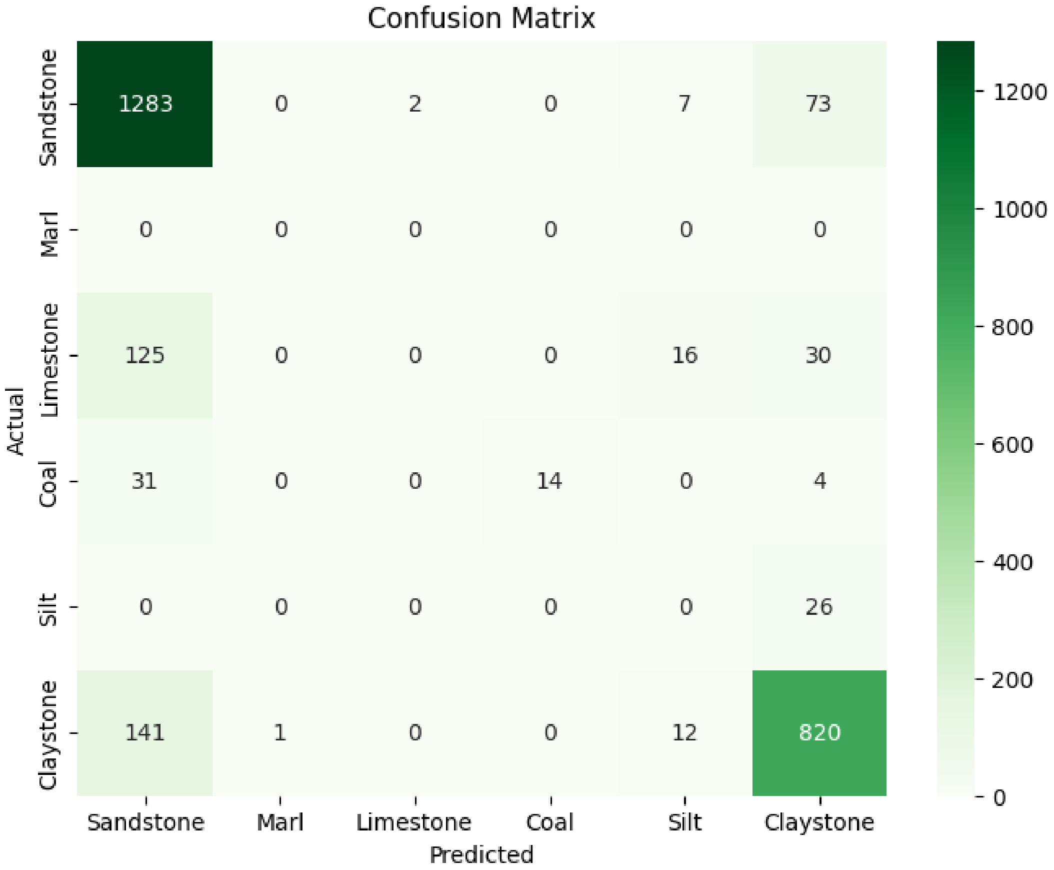


Confusion matrix: KNN vs RF

- KNN



- Random Forest

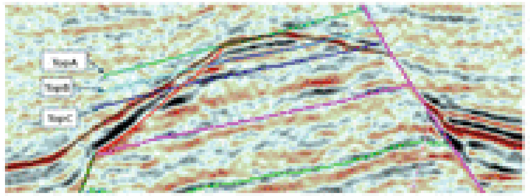


Try on your own data

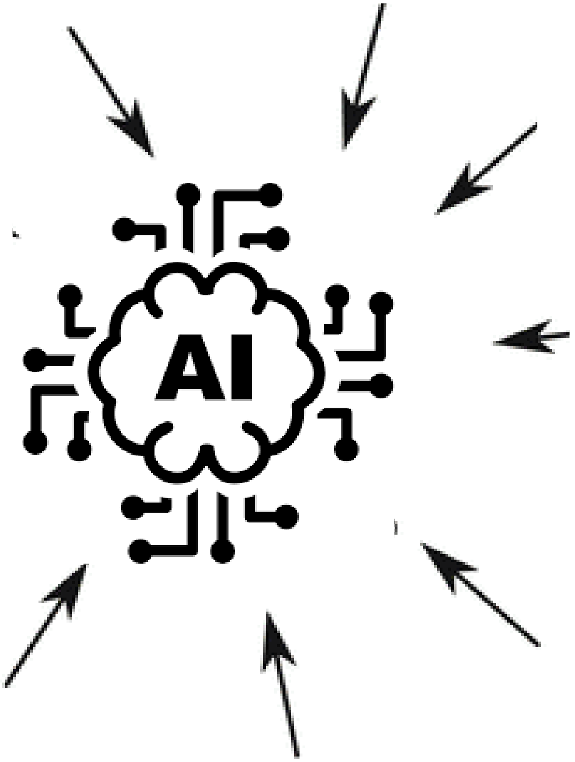
Unsupervised Seismic Interpretation

- Interpretational uncertainty
- Seismic pattern recognition
- Unsupervised seismic segmentation
- Object detection
- Real field application

Data integration in reservoir prediction modelling



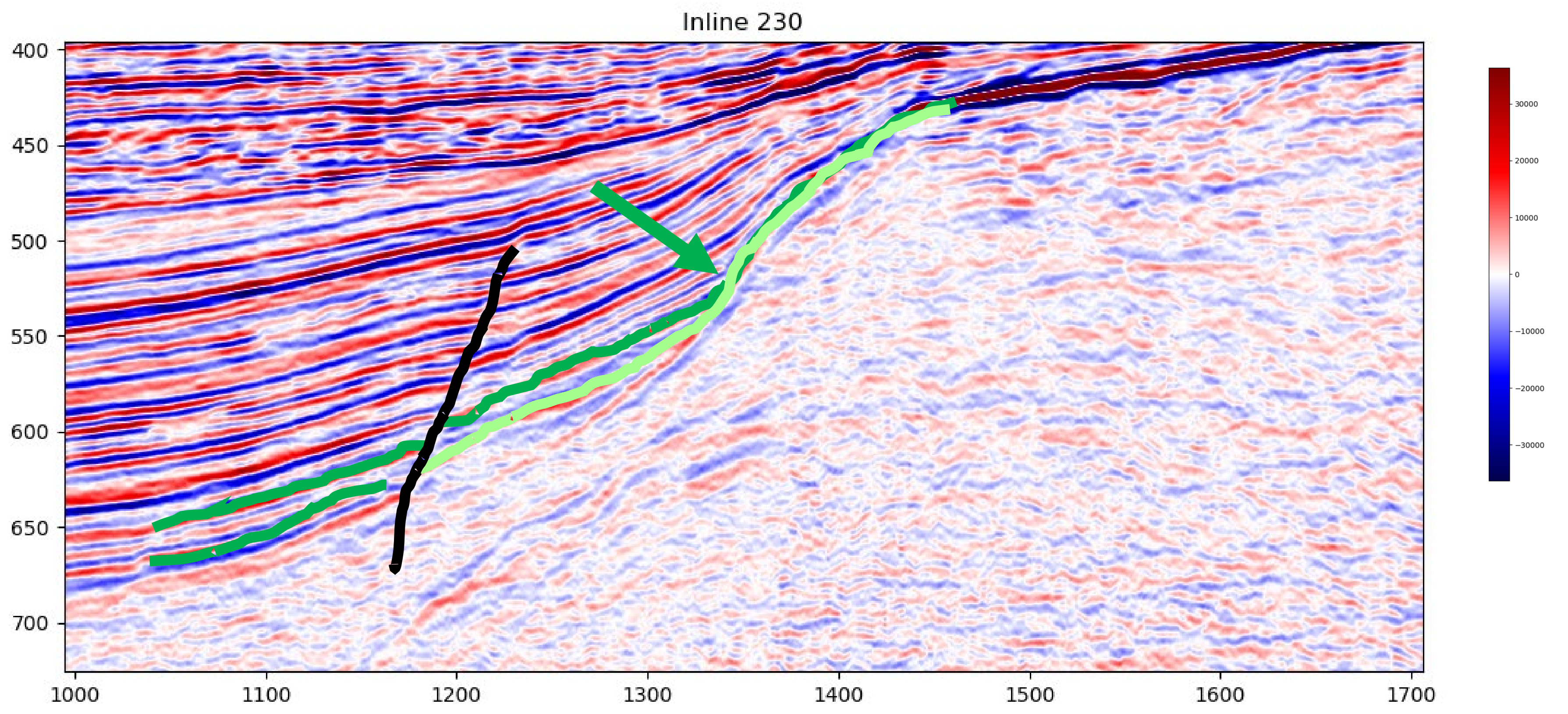
Seismic



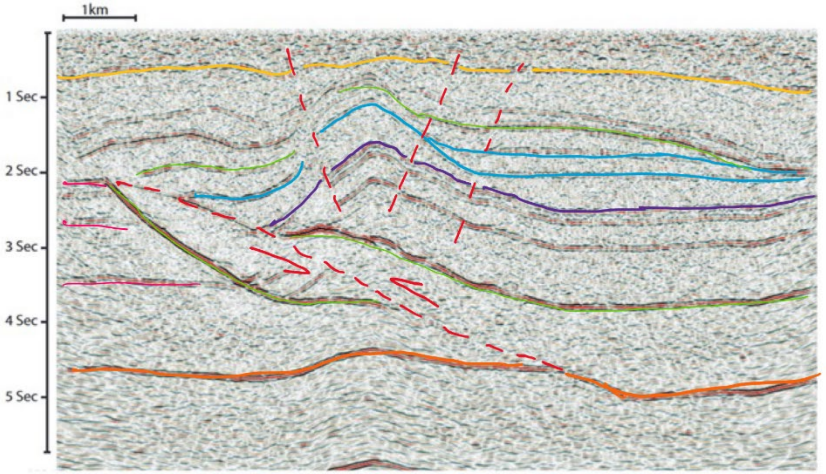
Modified from Rivenæs, J.C., Sørhaug, P., Knarud, R. (2015). Introduction to Reservoir Modelling. In: Bjørlykke, K Petroleum Geoscience. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-34132-8_22

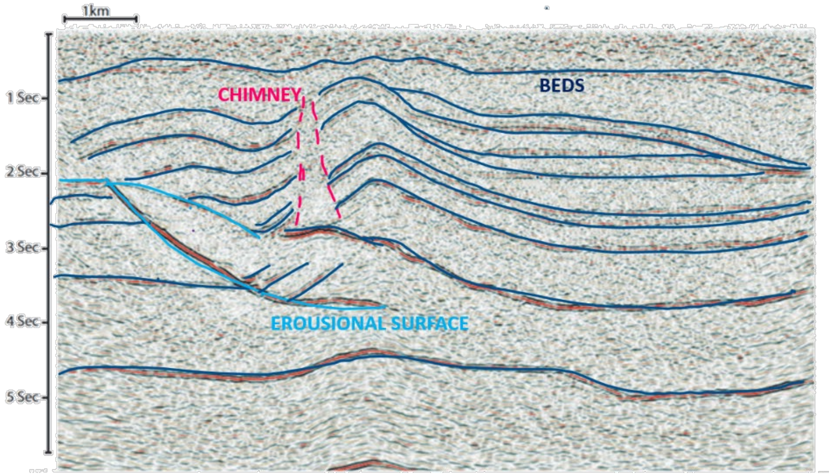
Why is it a challenge?

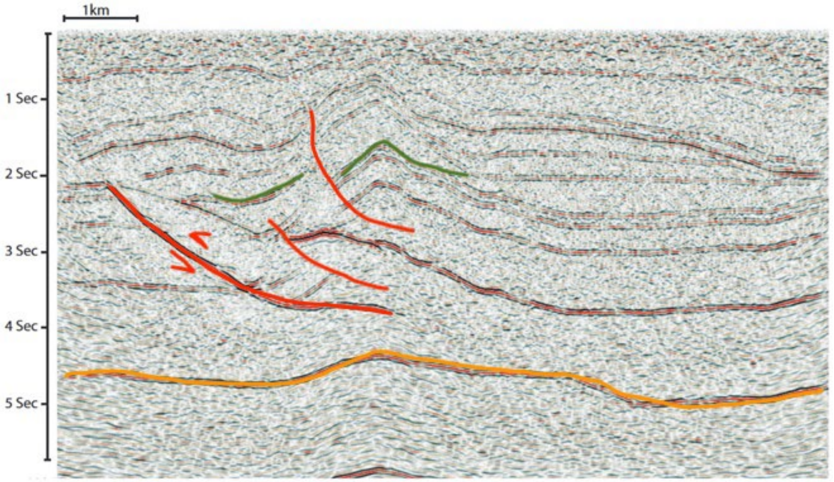
- Detecting fans?
- Starts? Ends? Path?
- Fans overlapping?
- Deal with faults

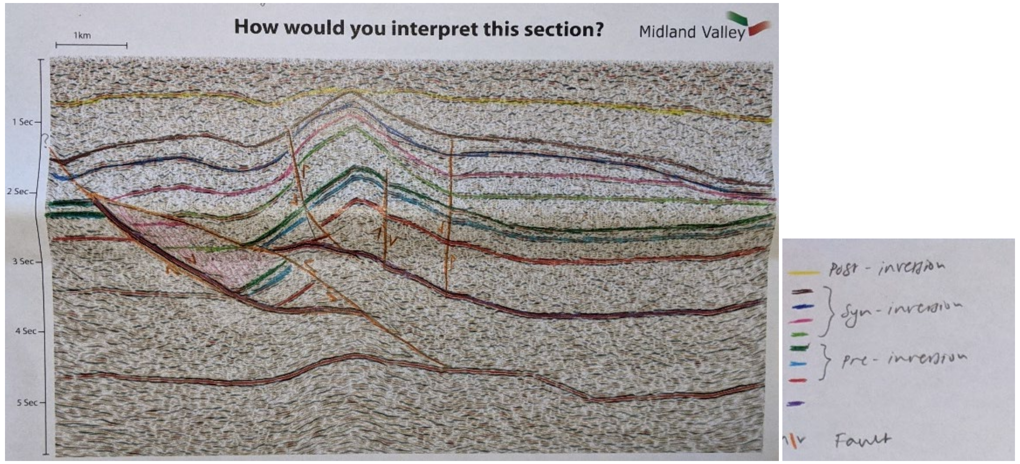


Interpretational Uncertainty



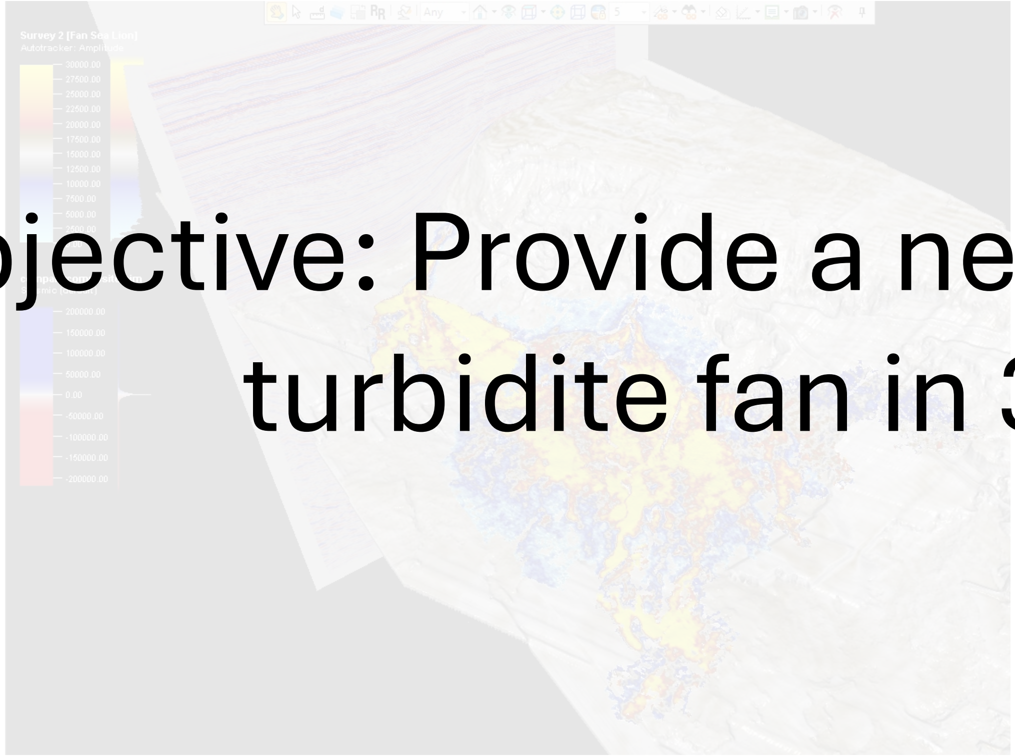




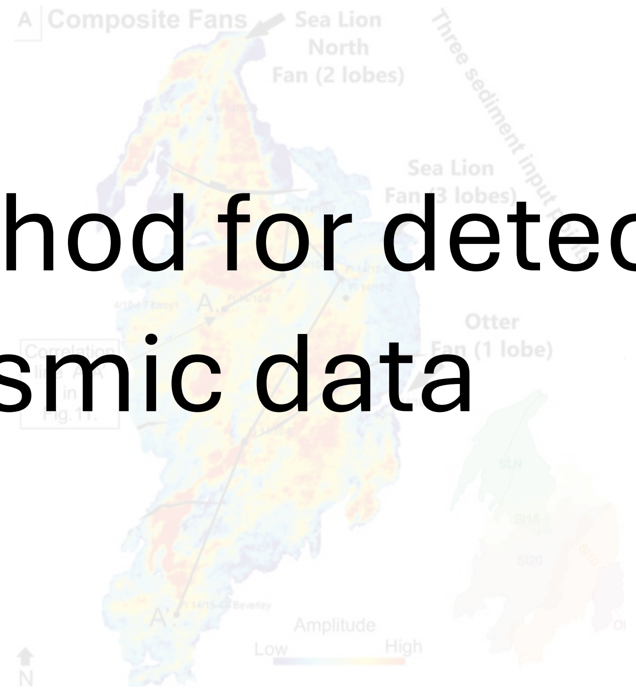


Challenges of fan detection

Fan interpretation in the North Falkland Basin (Sea Lion field)



[Petrel, Schlumberger]



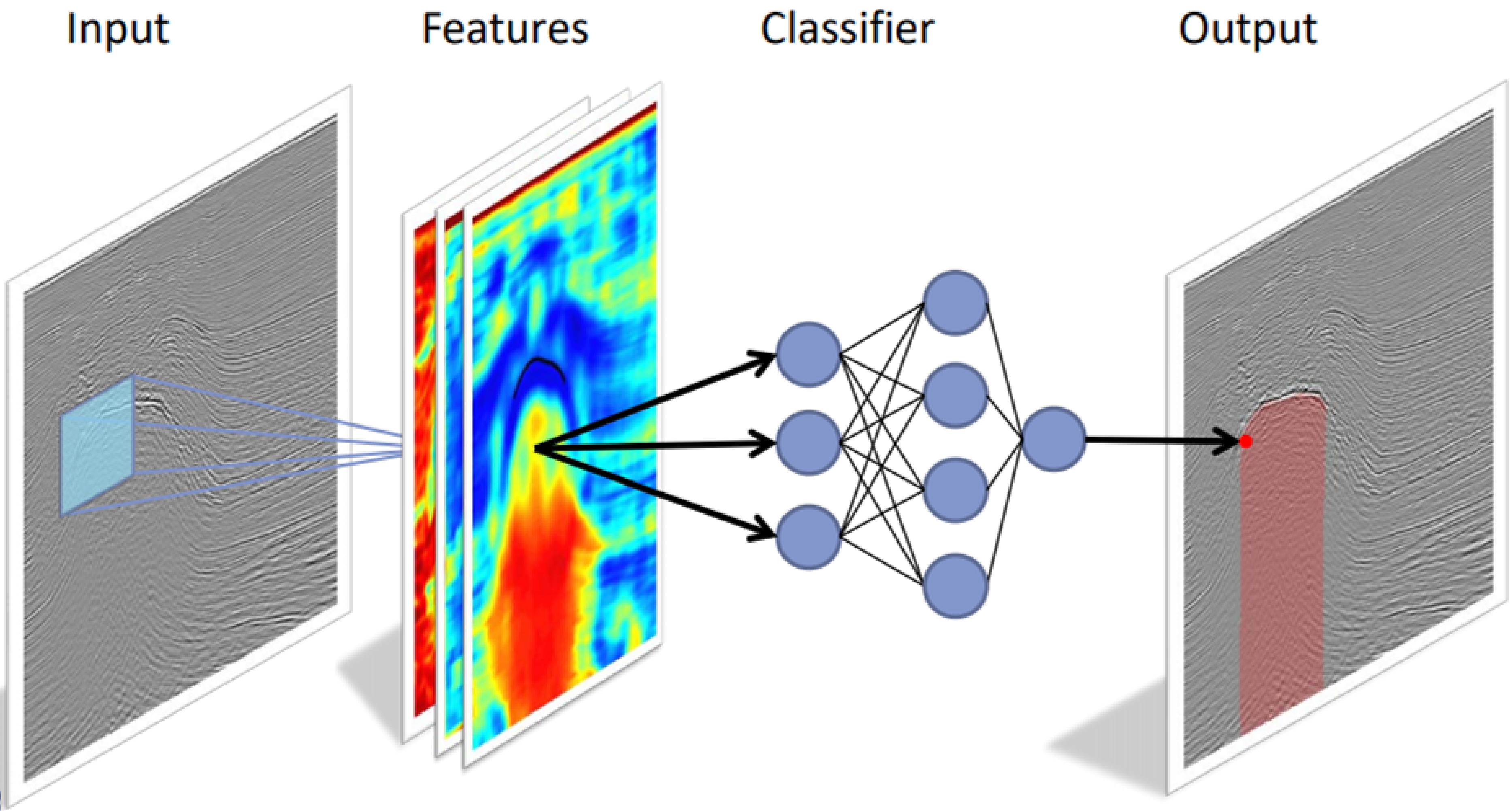
[Dodd et al., 2019, Sedimentology]

Objective: Provide a new method for detecting turbidite fan in 3D seismic data

Motivations for a data-driven approach

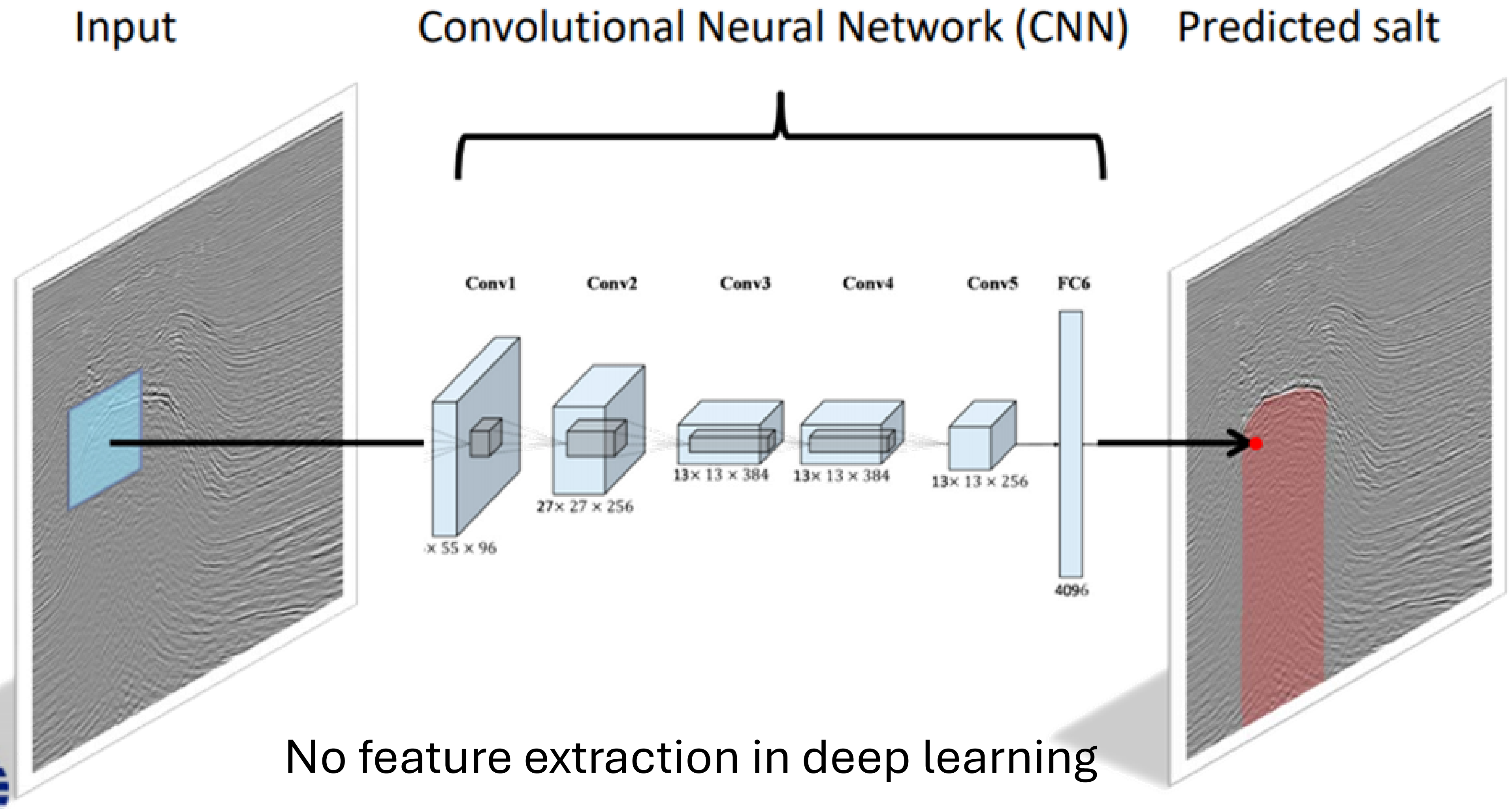
- Embedding all information at once (3D information, seismic attributes, different sources information...)
- Unbiased by colours, zooming...
- Repeatable process
- Multiple realizations to quantify uncertainties

Machine learning for seismic classification



Anders U. Waldeland, 2018

Deep learning for seismic classification

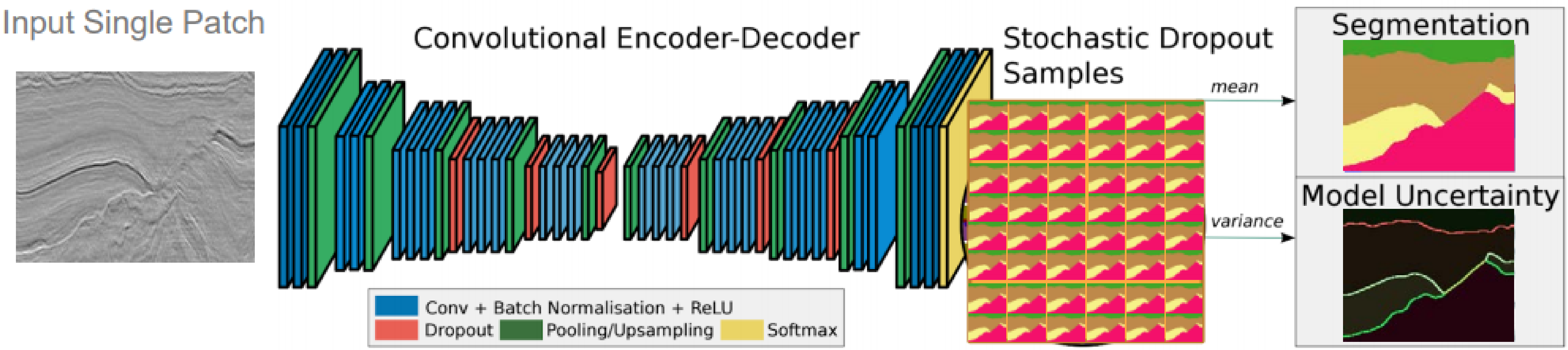


No feature extraction in deep learning

Anders U. Waldeland, 2018

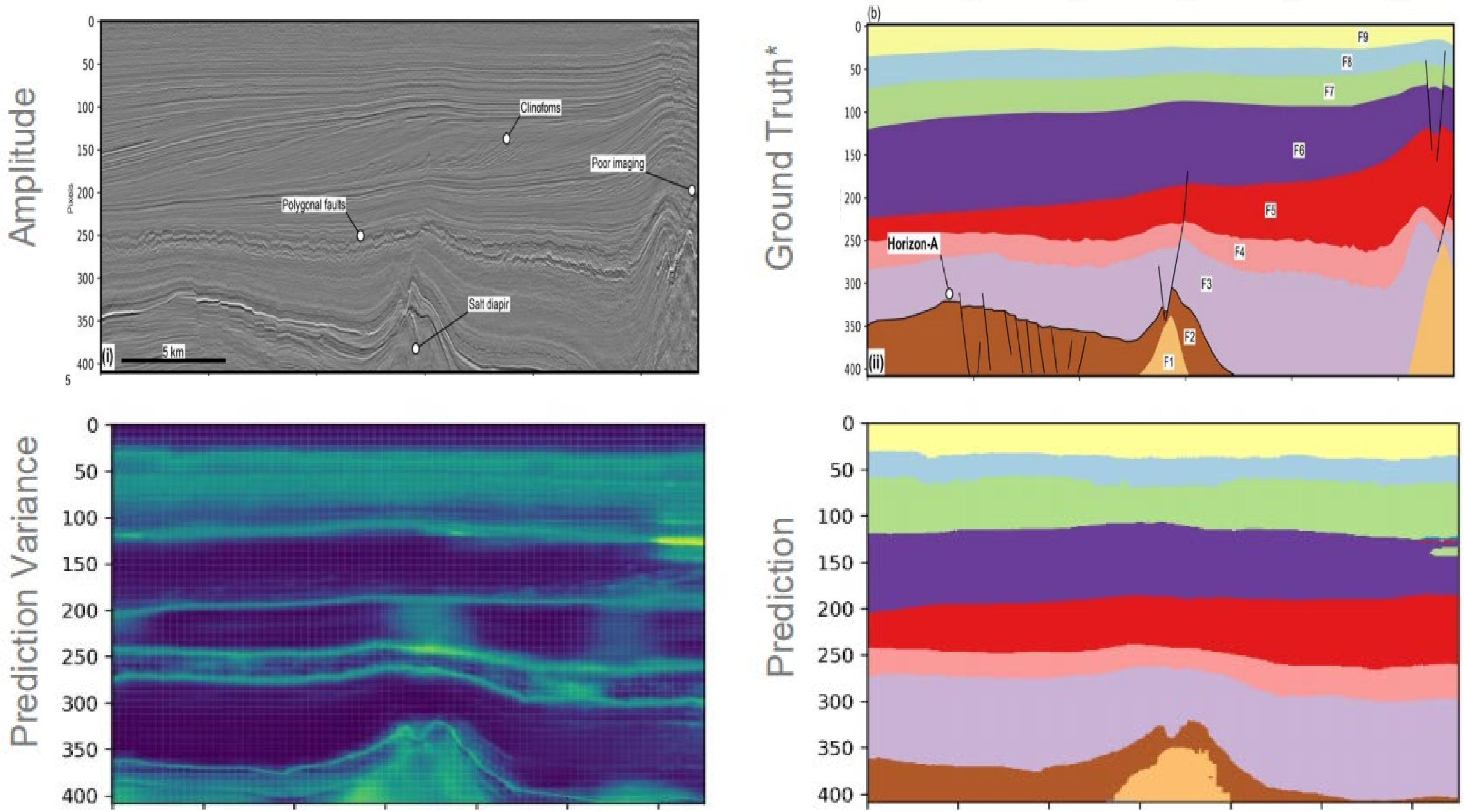
CNN for stratigraphic interpretation

Model Architecture – Bayesian ConvNet: Segnet



Lukas Mosser, 2018, PhD Imperial College

CNN vs manual interpretation

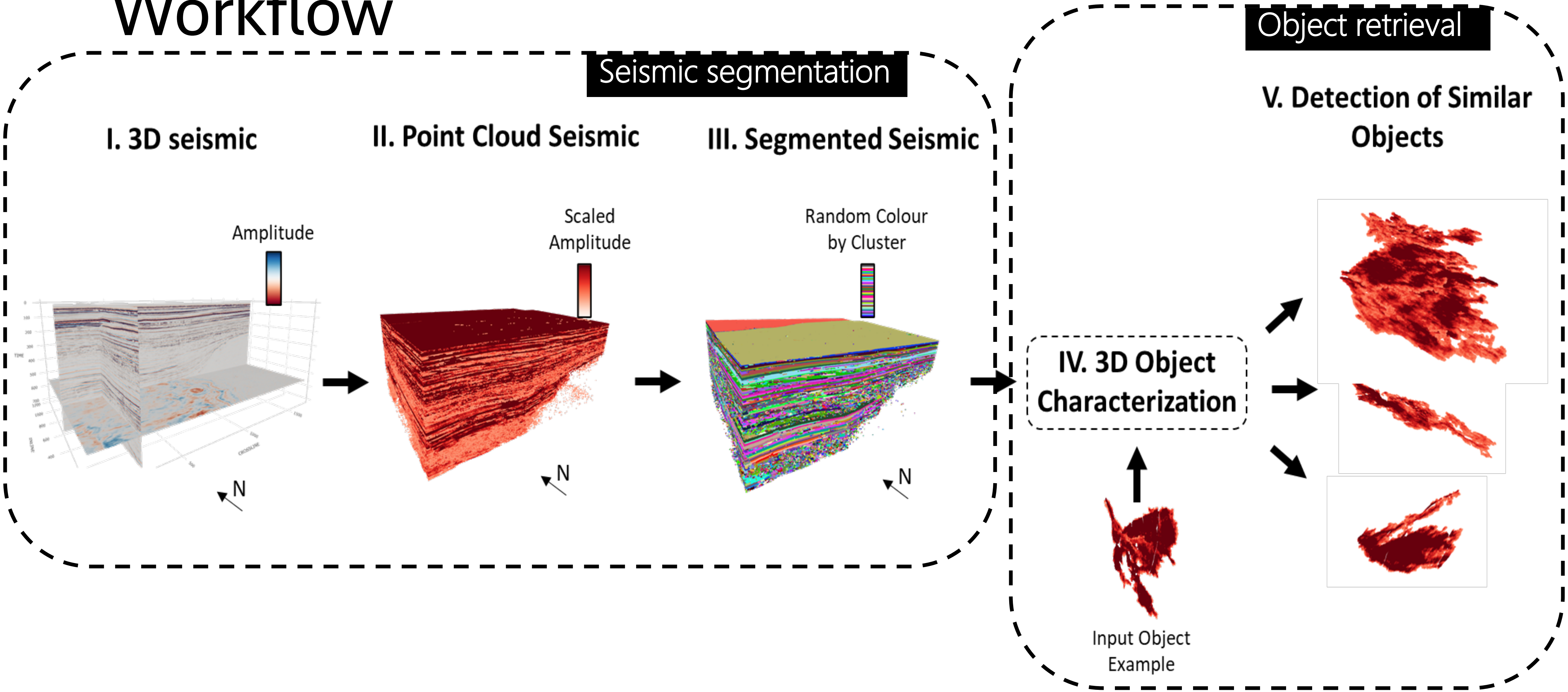


Lukas Mosser, 2018, PhD Imperial College

Unsupervised Seismic Interpretation

Quentin Corlay
Fast Detection of Geobodies in 3D Seismic with Unsupervised Segmentation
PhD thesis, 2023

Workflow



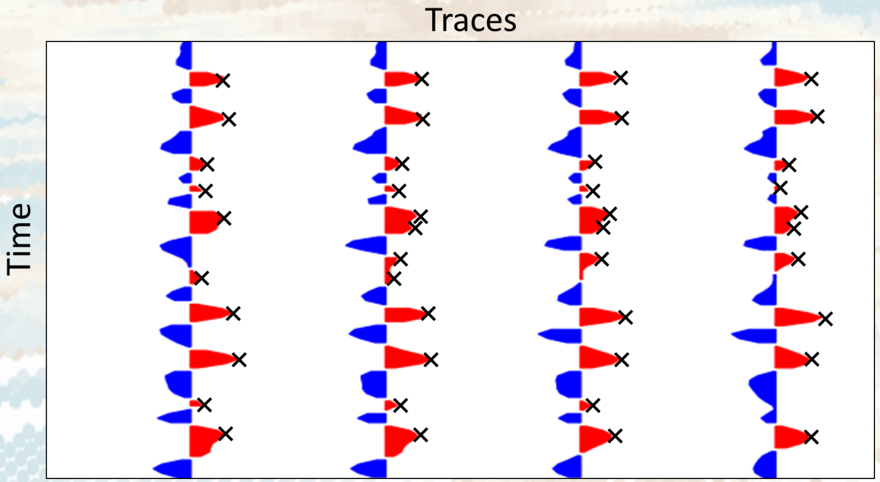
Seismic segmentation
I. 3D seismic
II. Point Cloud Seismic
III. Segmented Seismic
Amplitude
Scaled Amplitude
Random Colour by Cluster
N
N
N
Object retrieval
V. Detection of Similar Objects
IV. 3D Object Characterization
Input Object Example

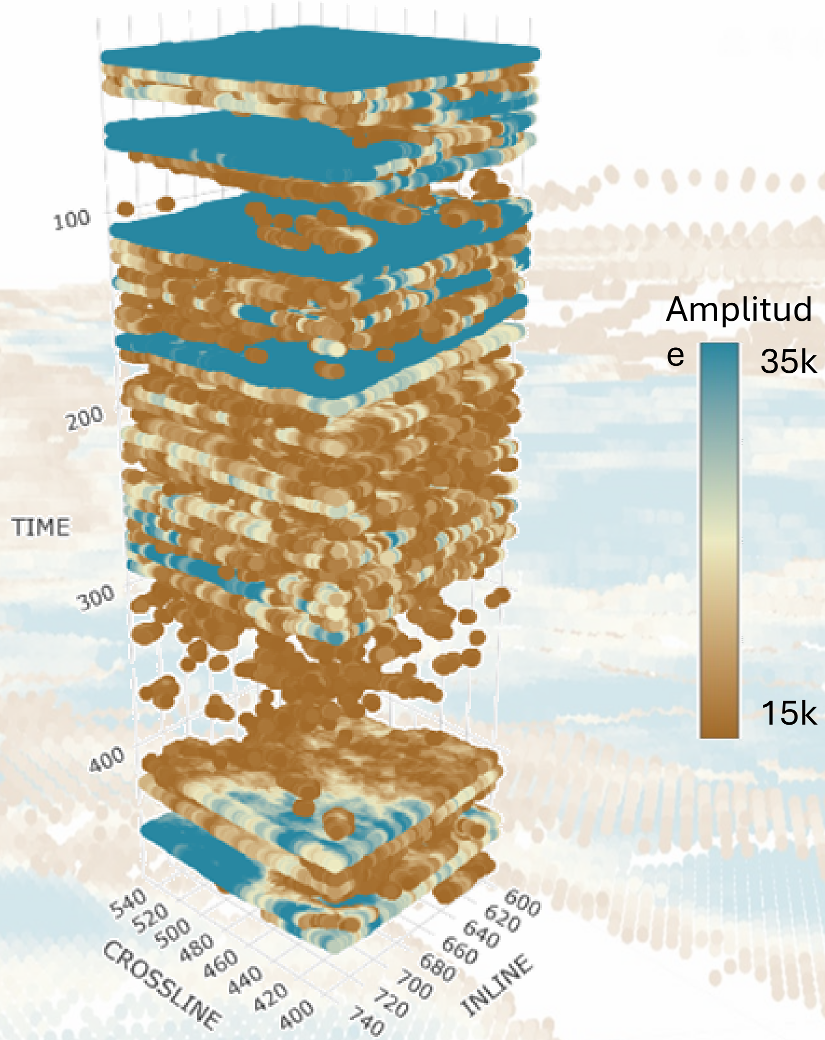
Point Cloud Seismic

Sparse representation of 3D seismic

Point selection:

- local extrema extraction in the trace direction
- seismic attributes filters





Spatial Segmentation - DBSCAN

Density-Based Spatial Clustering with Application of Noise

Unsupervised machine learning clustering algorithm based on spatial density

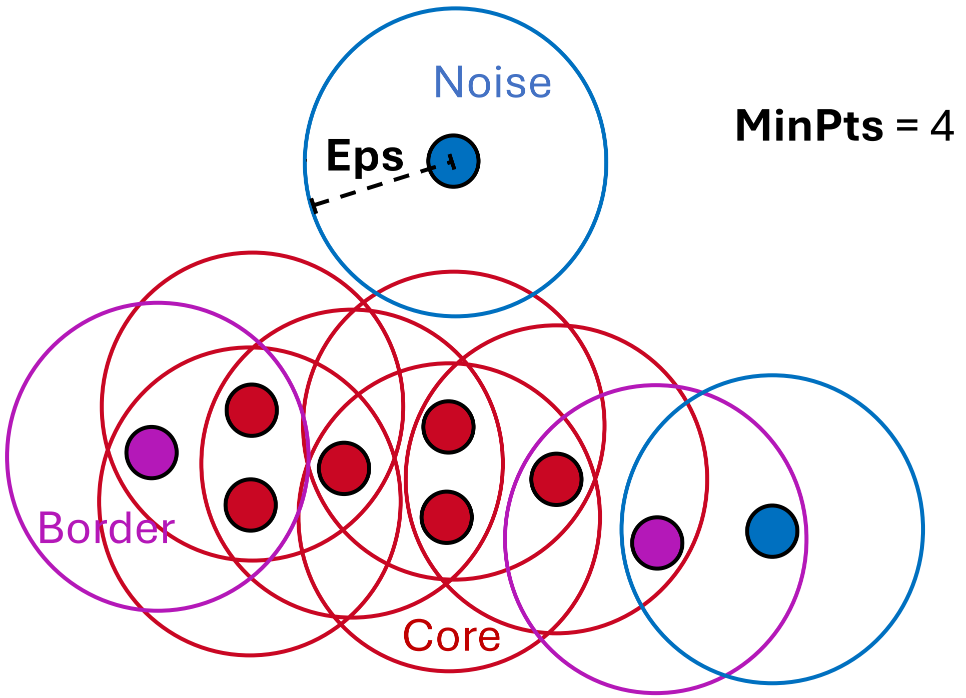
Epsilon-Neighbourhood:

$$N_{Eps}(p) = \{q \in D \mid dist(p,q) \leq Eps\}$$

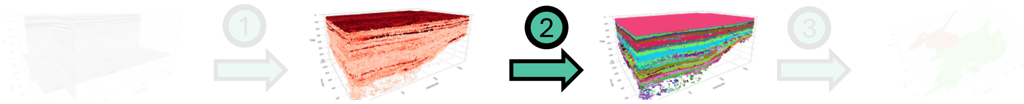
MinPts:
Minimum number of point in an Eps-Neighbourhood for a point to be considered as a Core point

A cluster regroups the ensemble of connected core points and their Eps-Neighbourhood.

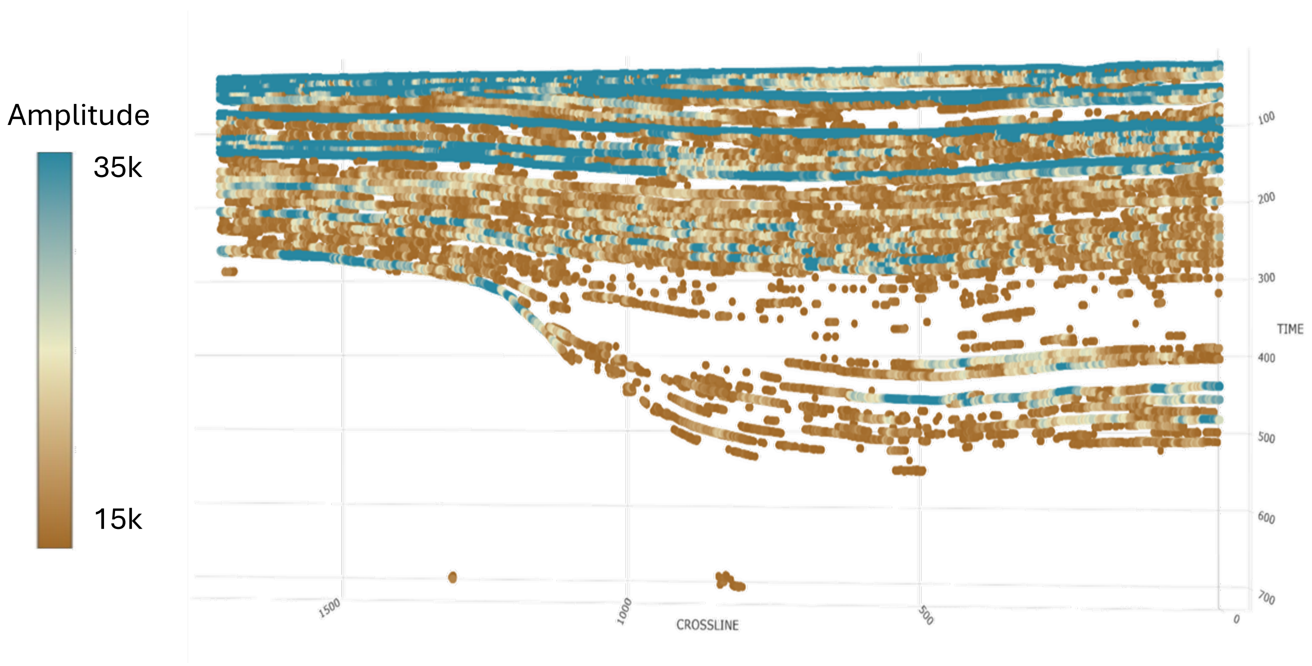
The points not included in a cluster are considered as Noise

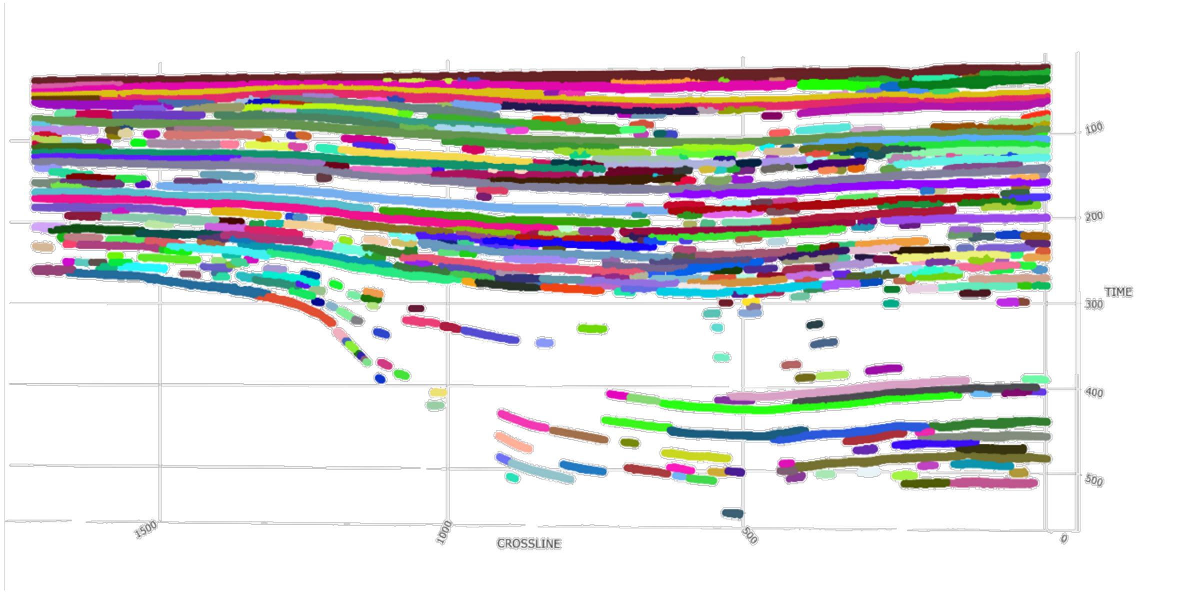


[Ester et al., 1996]

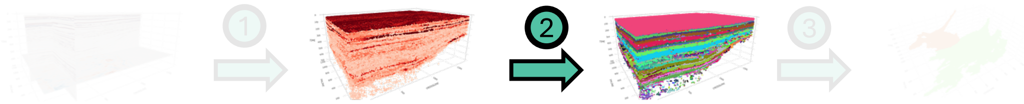


Segmented Point Cloud Seismic



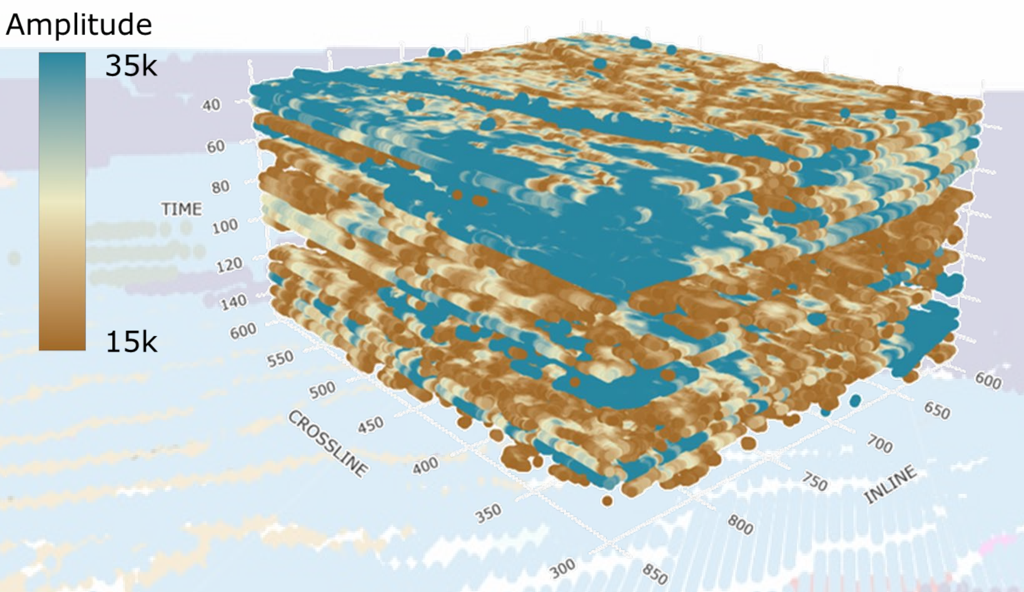


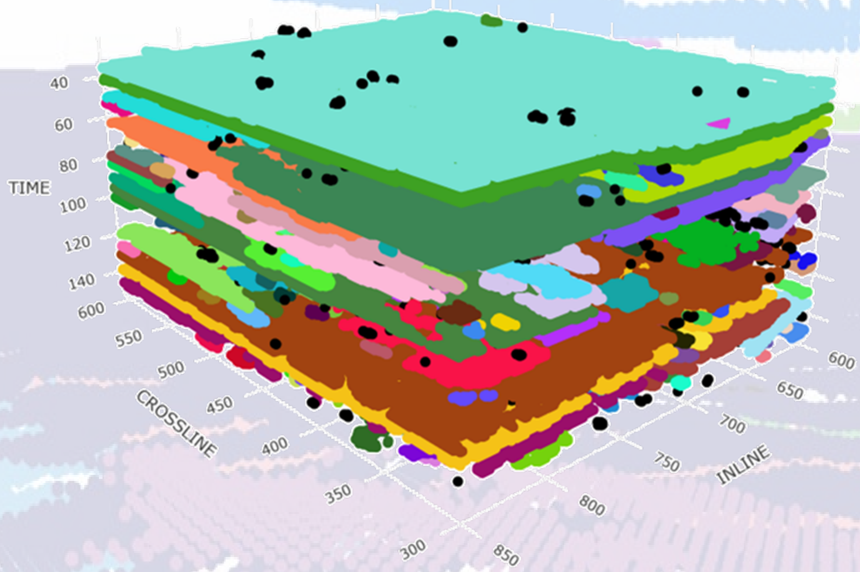
Every cluster is associated with a random color
Noise has been removed from the display



Segmented Point Cloud Seismic

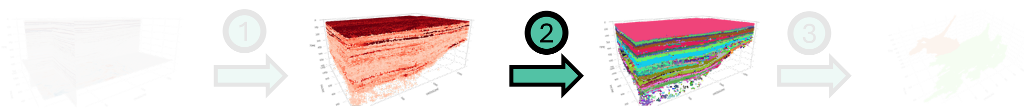
Small 3D area: seismic cube (250 inlines × 300 crosslines × 100 timeslices)





Clusters associated with random color
Noise appears in black

393,299 points are extracted and segmented – 9s on a standard workstation

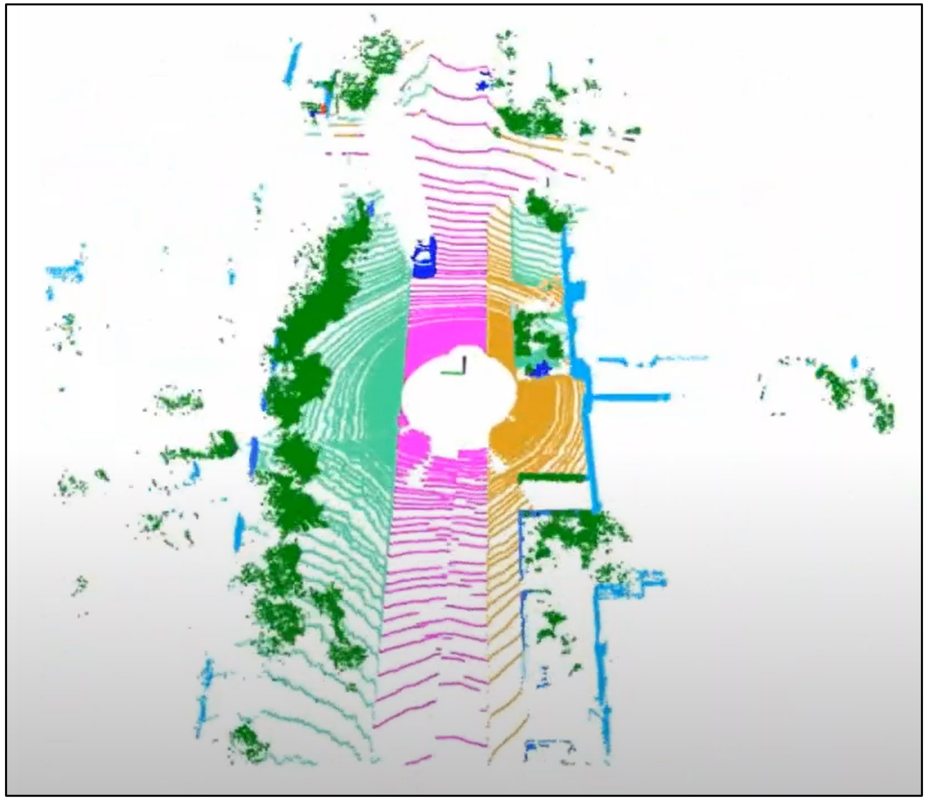


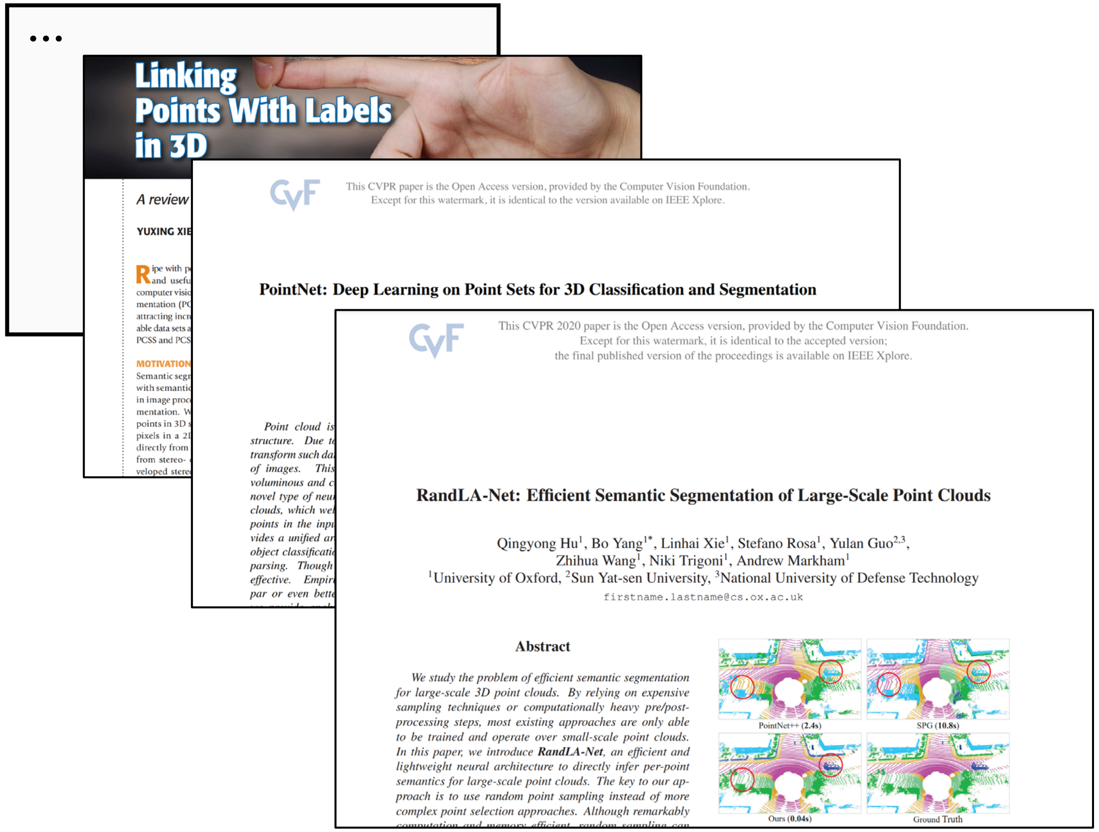
Unsupervised Seismic Interpretation

- Interpretational uncertainty
- Seismic pattern recognition
- Unsupervised seismic segmentation
- Object detection
- Real field application

Discussion - Point Cloud Semantic Segmentation with Deep Learning

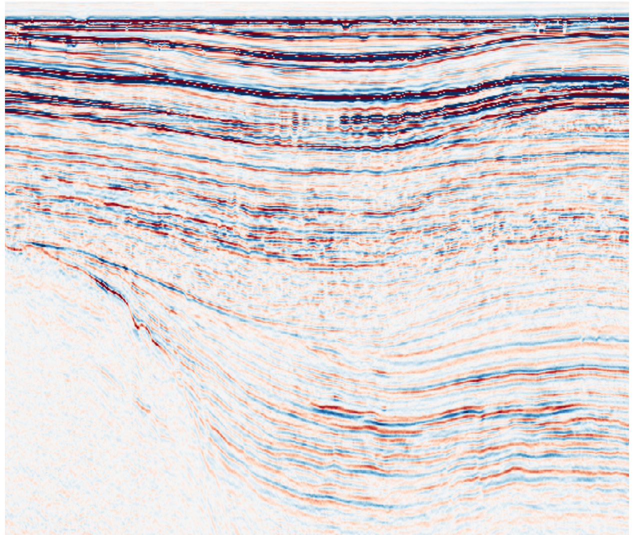
A all literature is available on point cloud semantic segmentation

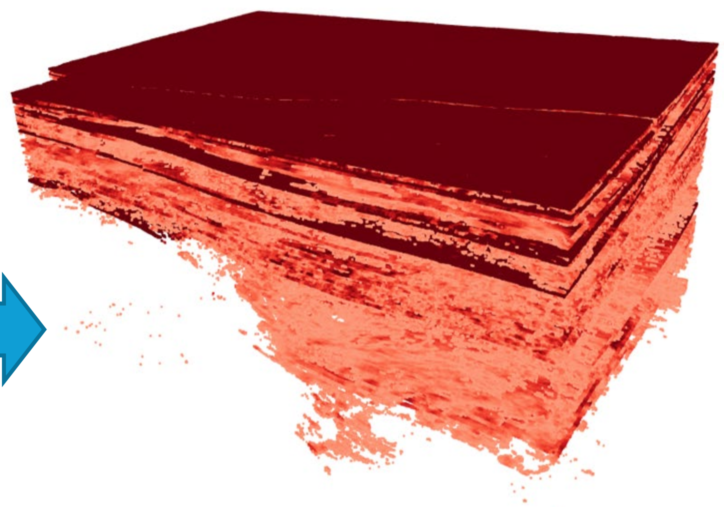


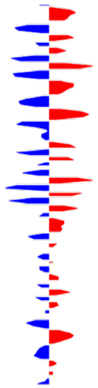


[Hu et al., 2020, *IEEE/CVF Conference on Computer Vision and Pattern Recognition*]

Step 1 - Seismic point cloud



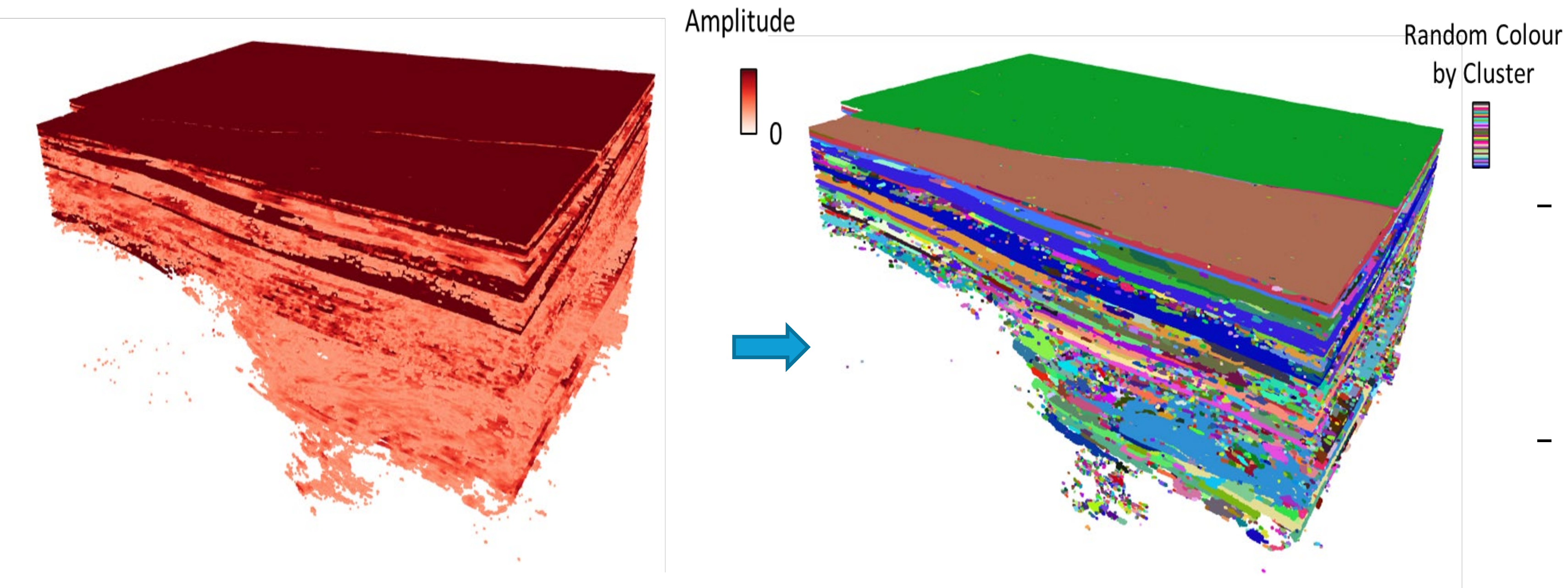




- Extrema extraction in the trace direction
- Filtering points on coherence
- Filtering points on amplitude

Advantages:
- Creates space between objects in the seismic
- Reduces the volume of data

Step 2 - Seismic segmentation



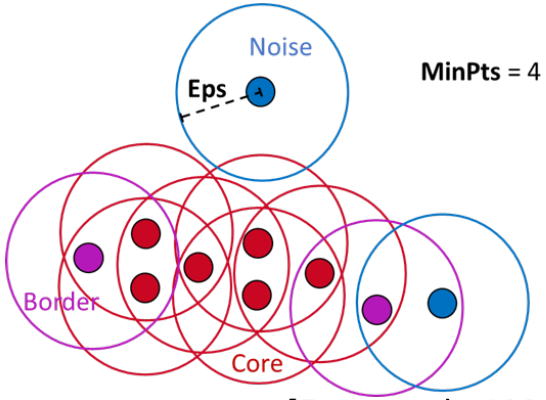
- Density-Based Spatial Clustering with Application of Noise (DBSCAN)

- Clusters based on density connectivity between the points

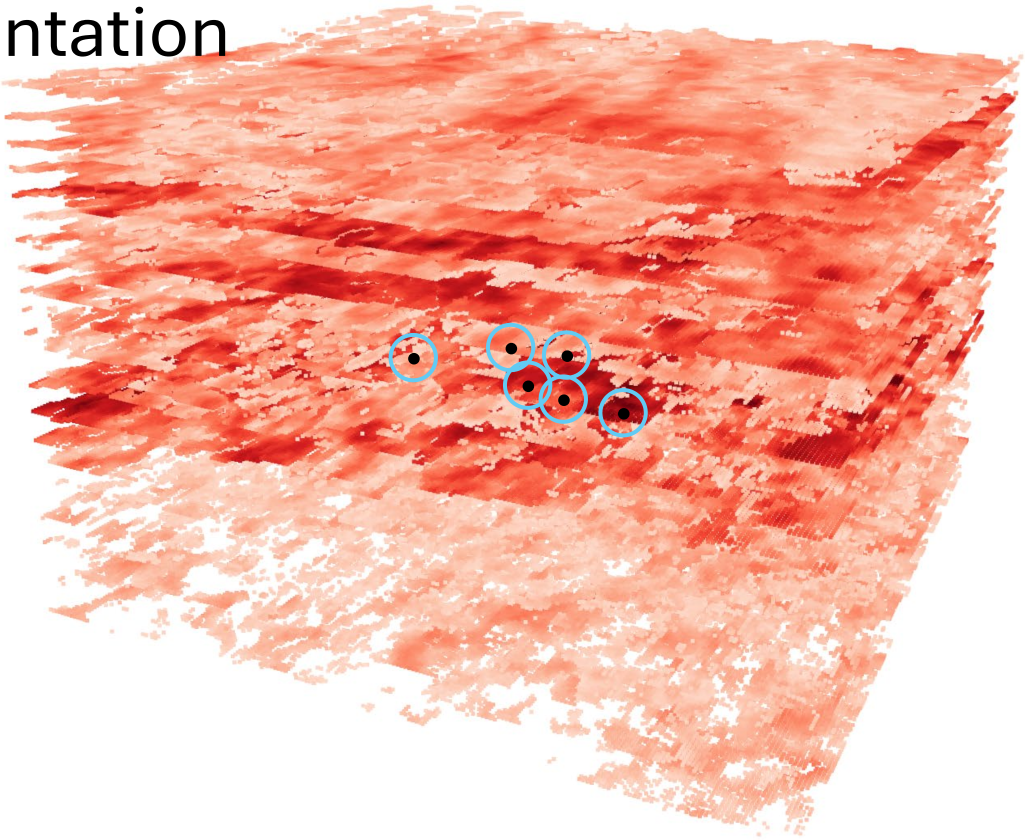
Step 2 - Seismic segmentation

- Spatial segmentation

DBSCAN



[Ester et al., 1996]

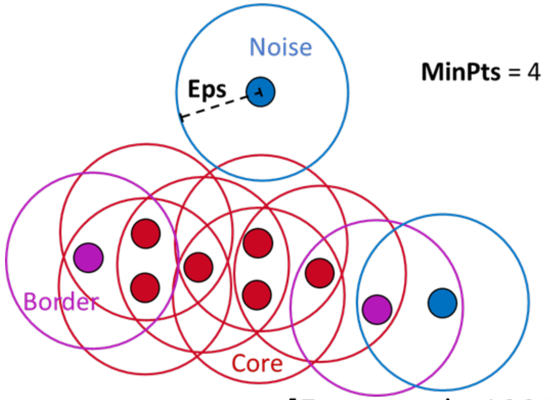


DBSCAN parameters
Eps: Epsilon
MinPts: Minimum number of points

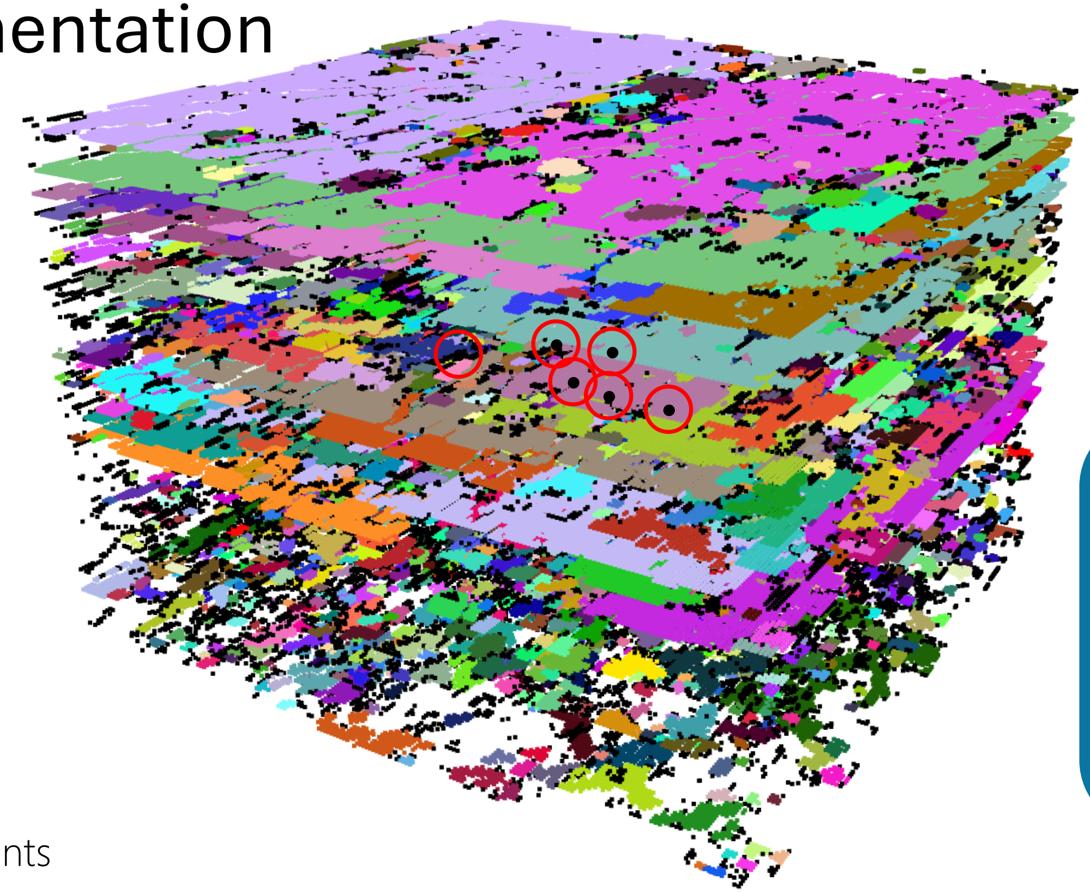
Step 2 - Seismic segmentation

- Spatial segmentation

DBSCAN



[Ester et al., 1996]



DBSCAN parameters
Eps: Epsilon
MinPts: Minimum number of points

Advantages:
- Segments are not size or shape dependent
- Robust to noise
- Extremely fast

Seismic segmentation

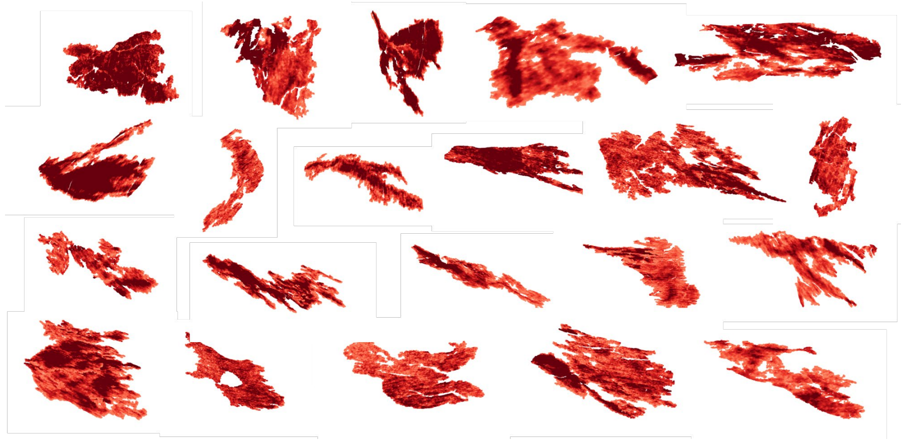
[Corlay et al., 2020]

Unsupervised Seismic Interpretation

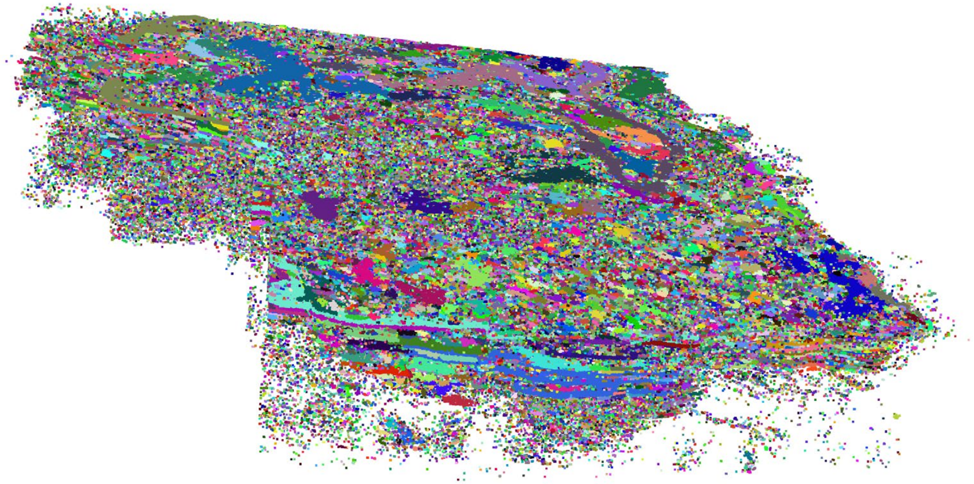
- Interpretational uncertainty
- Seismic pattern recognition
- Unsupervised seismic segmentation
- Object detection
- Real field application

Problems – No labelled datasets

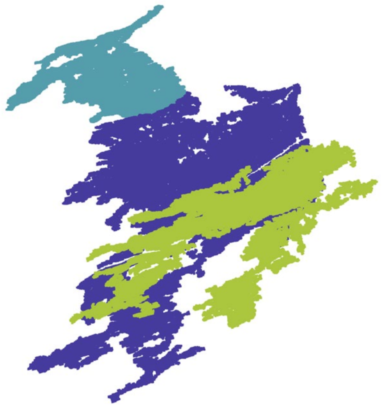
2. Diversity of objects in the seismic



1. No labelled datasets

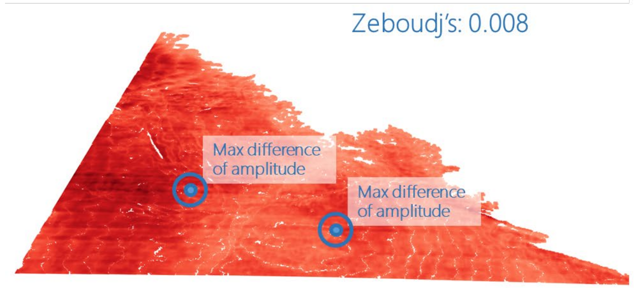


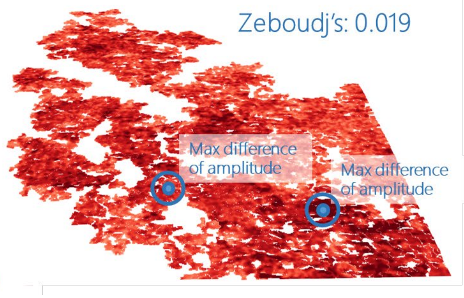
3. Diversity of shapes of turbidite fans



Segment Characterisation with Feature Engineering

1. Amplitude spatial distribution





Zeboudj's Internal disparity (CI): $CI(R_j) = \frac{1}{S_j} \sum_{s \in R_j} max\{C(s,t), t \in W(s) \cap R_j\}$

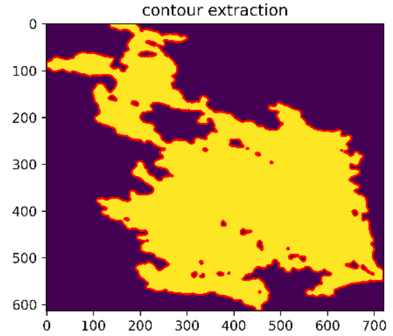
$$C(s,t) = \frac{|Amp(s) - Amp(t)|}{Amp_{max}}$$

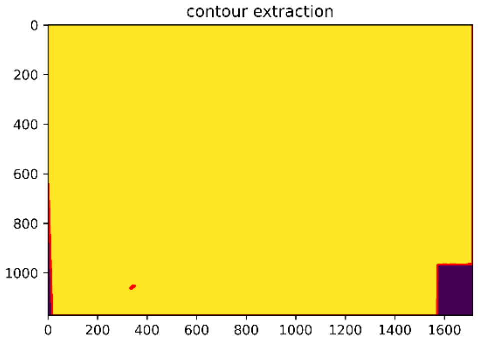
[Zéboudj, R., 1988, Doctoral thesis]

2. Contour ratio

Surface (S)

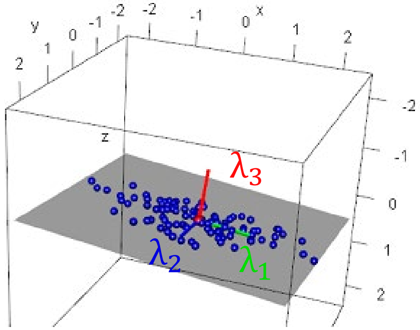
Contours





$$\frac{d_{contours}}{S}$$

3. Aspect ratio



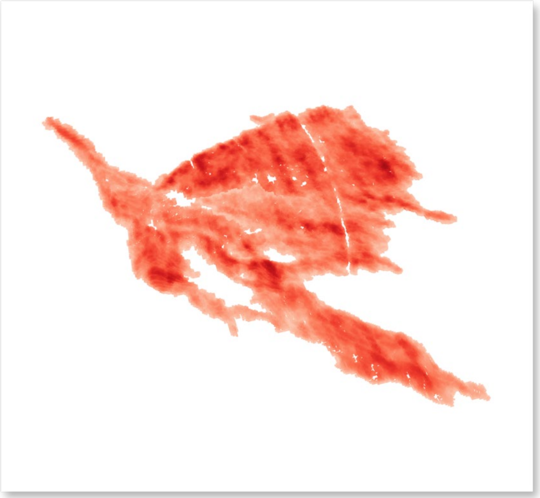
$$\frac{\lambda_1 - \lambda_2}{\lambda_1}$$

$$\frac{\lambda_3}{\lambda_1}$$

Table of extracted features

segmentID	nPoints	Amplitude Mean	Semblance Mean	Zeboudj Distance	Outline Ratio	linearity	slope	orientation	rZHigh	rZLow
21851	-0.085	0.433	0.351	-0.916	0.836	0.076	-0.078	0.911	-0.836	-0.679
27634	-0.244	0.598	-0.590	-0.060	0.920	0.077	-0.078	1.167	-0.306	-0.445
105043	-0.255	1.007	0.109	-0.016	-0.319	0.074	-0.078	-0.789	-0.693	0.148

Approach 1: binary classification



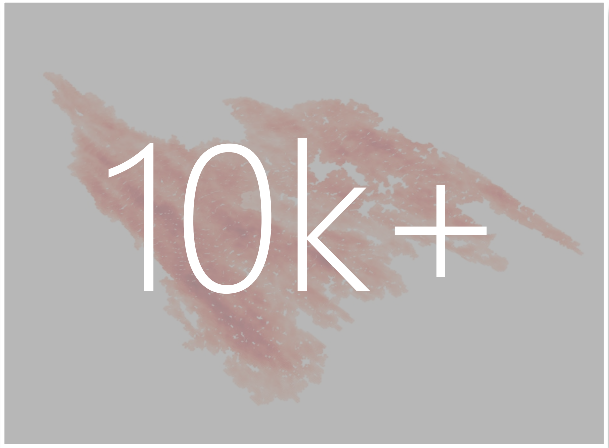
A turbidite fan

Not a turbidite fan

Object detection

Approach 1: binary classification

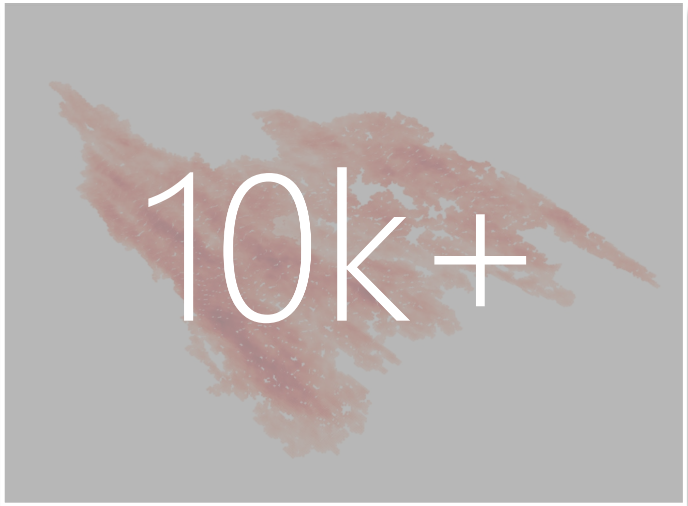
A turbidite fan
3 examples



Not a turbidite fan
10k+ examples

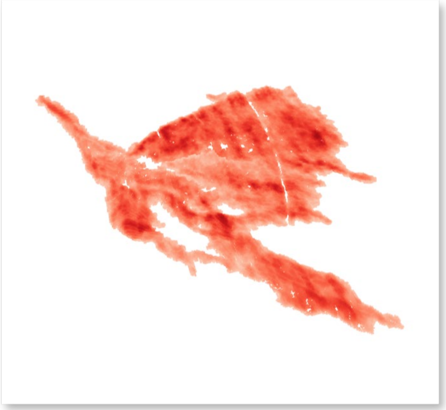
Problems:
- 3 fan examples / 10k+ segments
- False positives?

Approach 2: closest objects retrieval



Database of segments (10k+)

Find the n-closest segments



Example of object to find

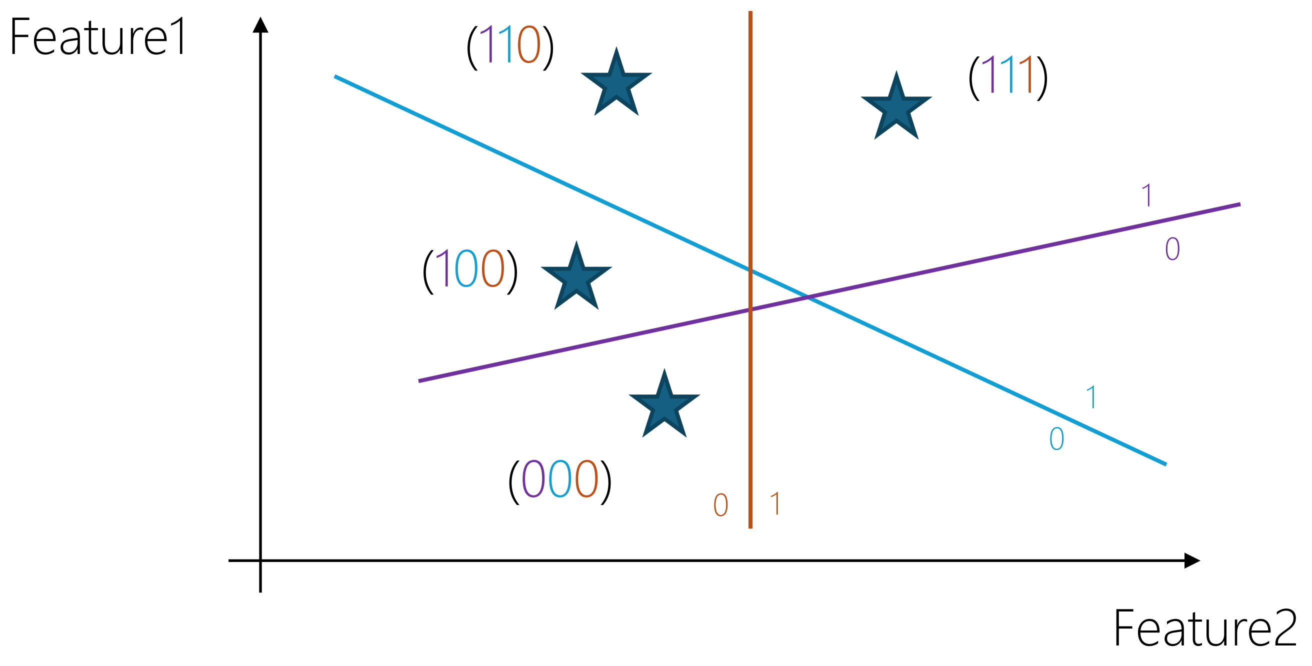
Object detection

Approach 2: closest objects retrieval

Binary Hashing for Approximate Nearest Neighbour

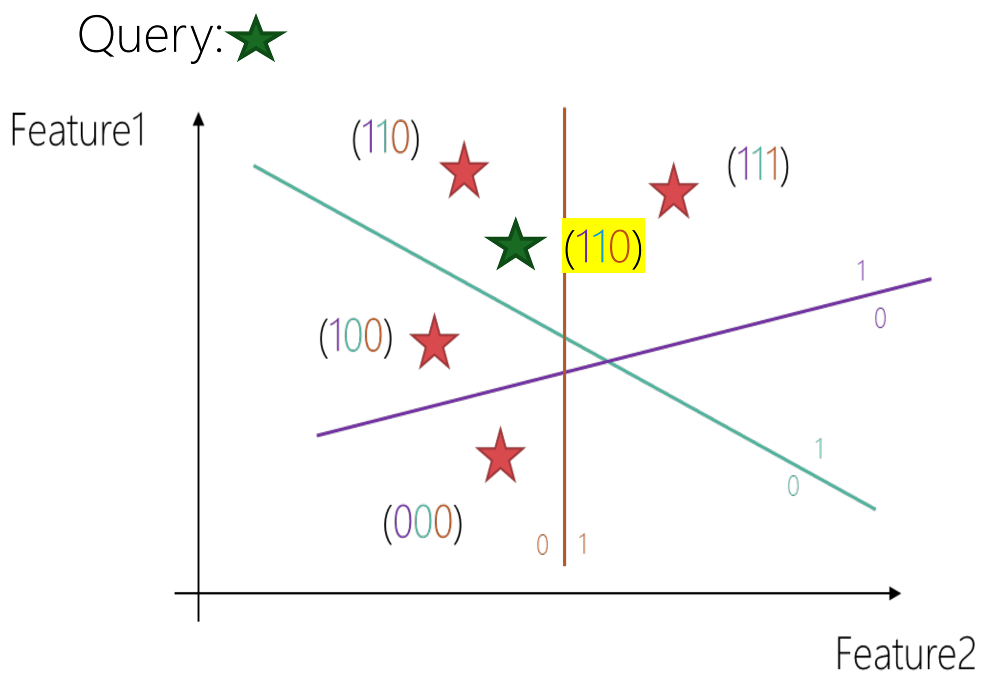
[Lin et al., 2015, IEEE]

Step 1 Binary code creation by feature hashing



Step 2 Object retrieval

Query:

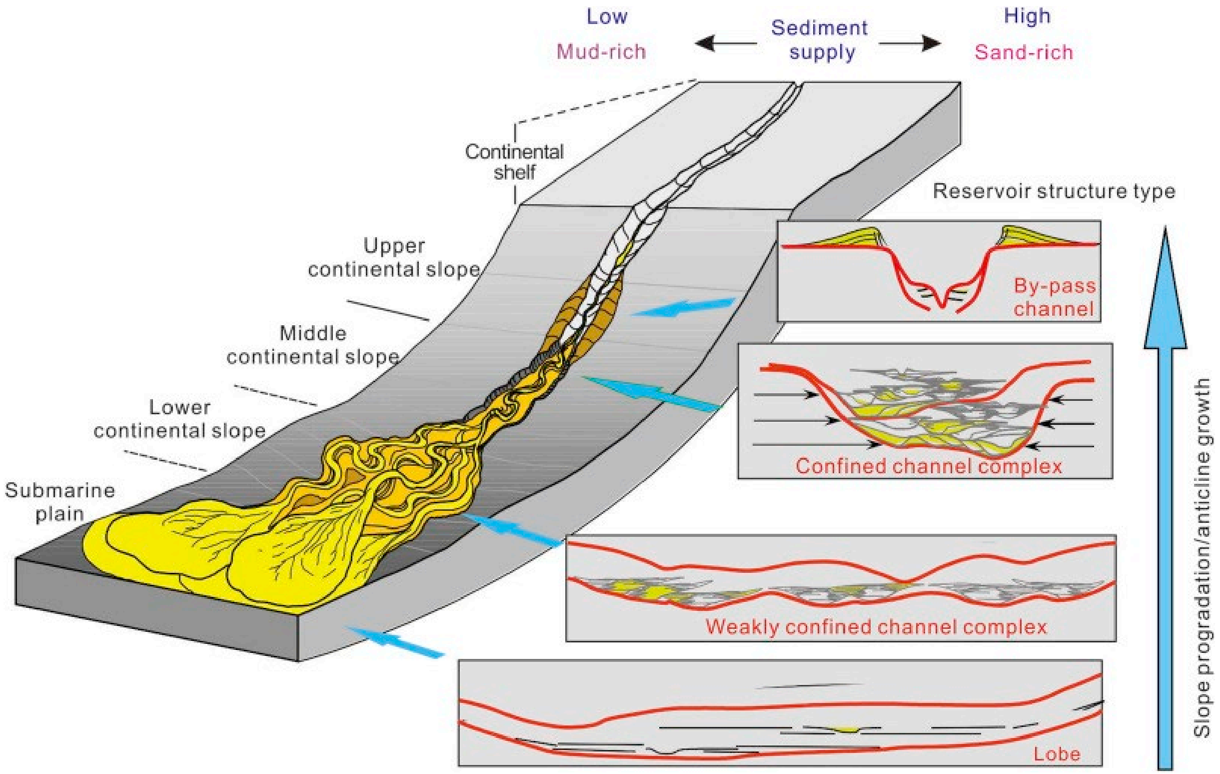


Unsupervised Seismic Interpretation

- Interpretational uncertainty
- Seismic pattern recognition
- Unsupervised seismic segmentation
- Object detection
- Real field application

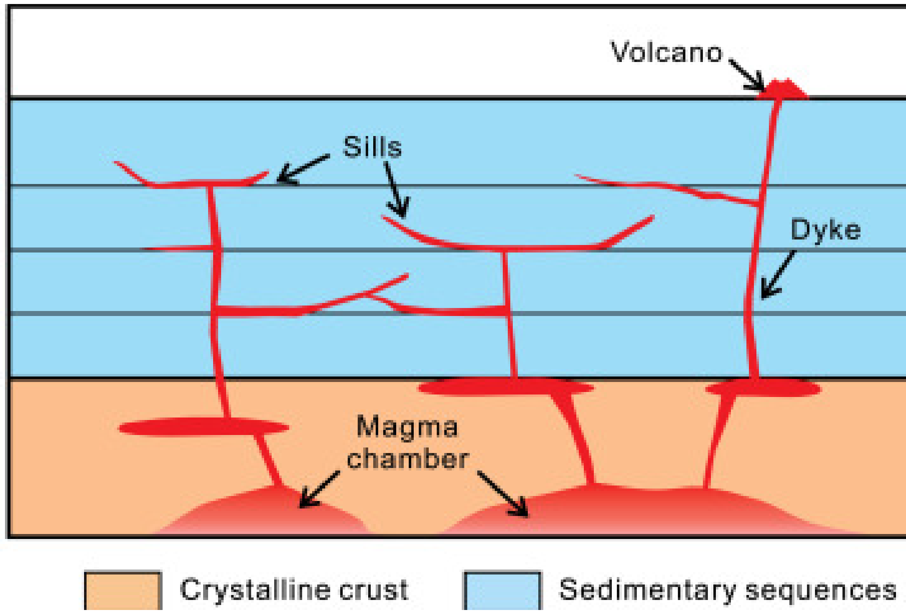
Geobody detection – 2 type of objects

Turbidite fan



[Huang, Y., 2018, Petroleum Research]

Sill intrusions

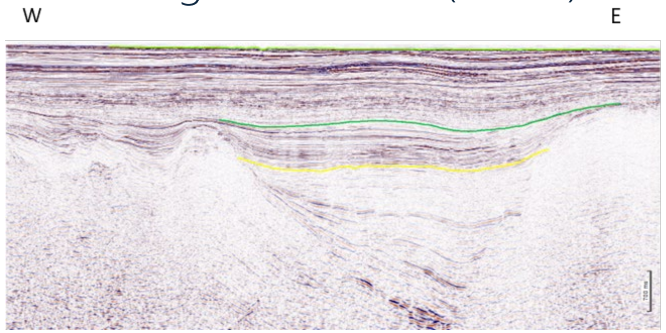


[Yao Z. et al., 2020, Marine and Petroleum Geology]

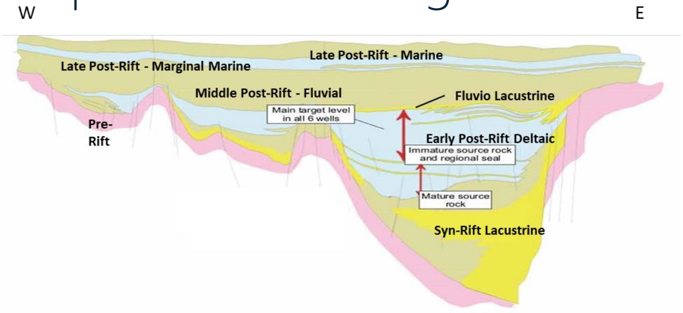
North Falkland Basin

North Falkland Basin

A modern high resolution (12.5m) 3D seismic

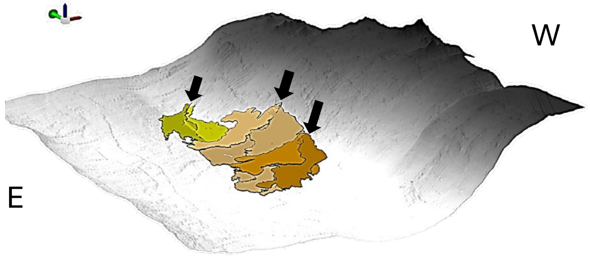


A deep lacustrine half graben



[Bamford, 2010, GeoExPro]

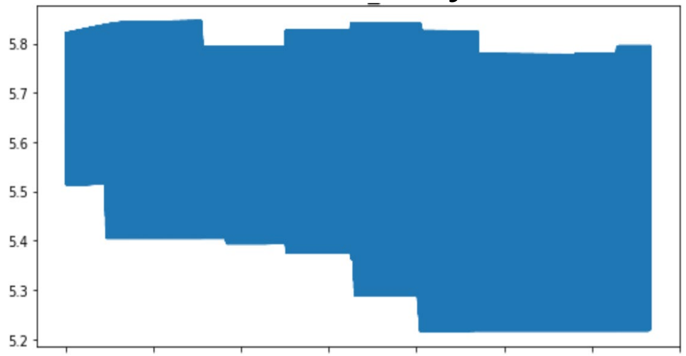
Tubidite fans

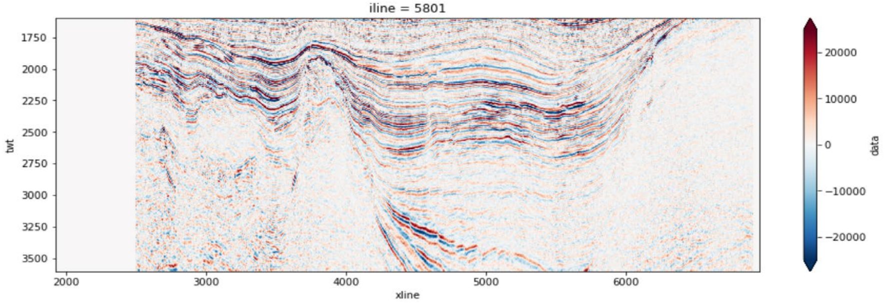


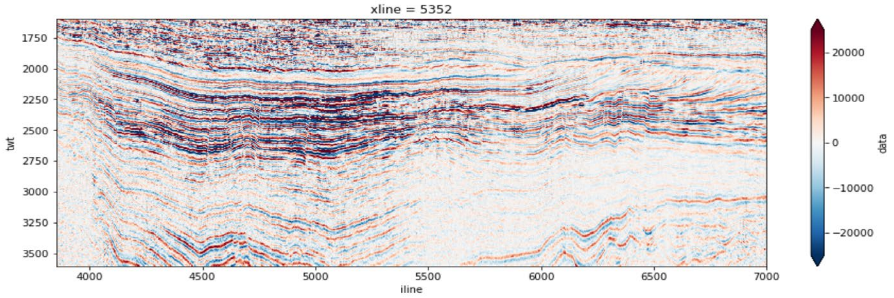
[Presentation Dodd & McCarthy, 2017]

Very fast, automatic & unsupervised full seismic segmentation

3D Geometry

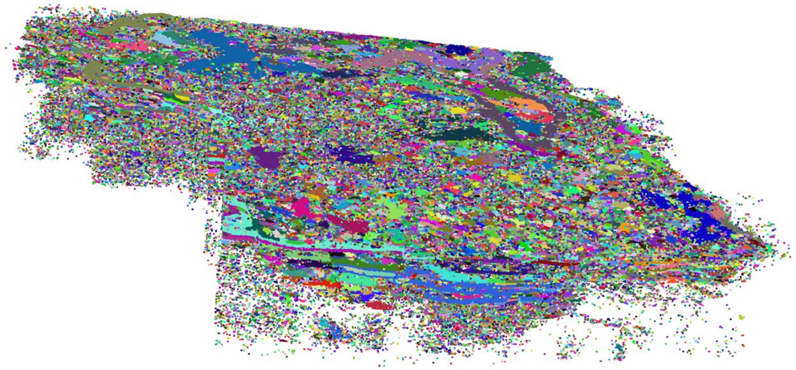






Sizeable seismic (79GB)

Results



574634 segments

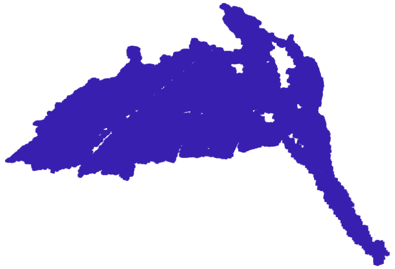
step	time
Extract extrema maxima/minima	13min 12s
Normalize and filter amplitude	9s
Compute and filter semblance	8min 41s
Spatial segmentation (DBSCAN)	9min 8s

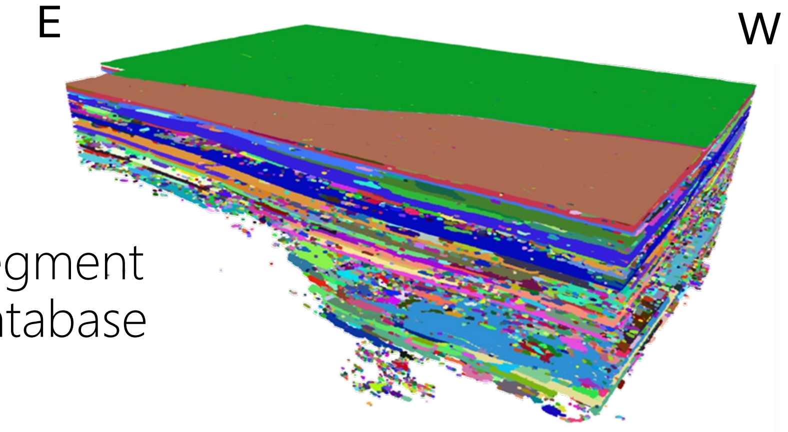
Point cloud creation ~25min

Turbidite fan detection

Inputs:

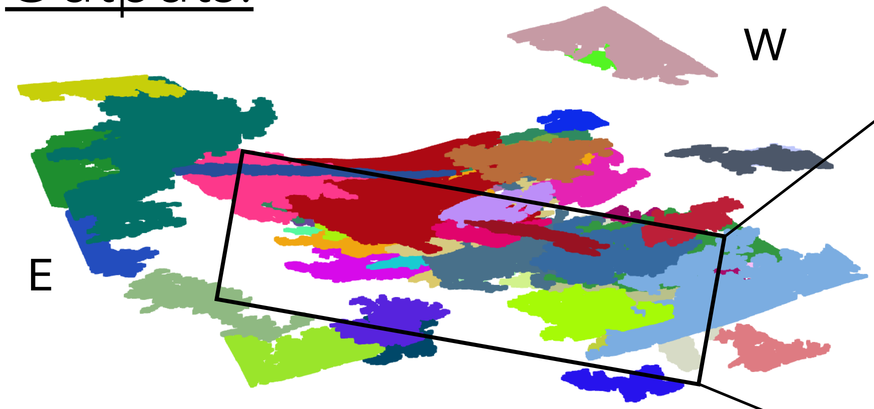
Segment to retrieve



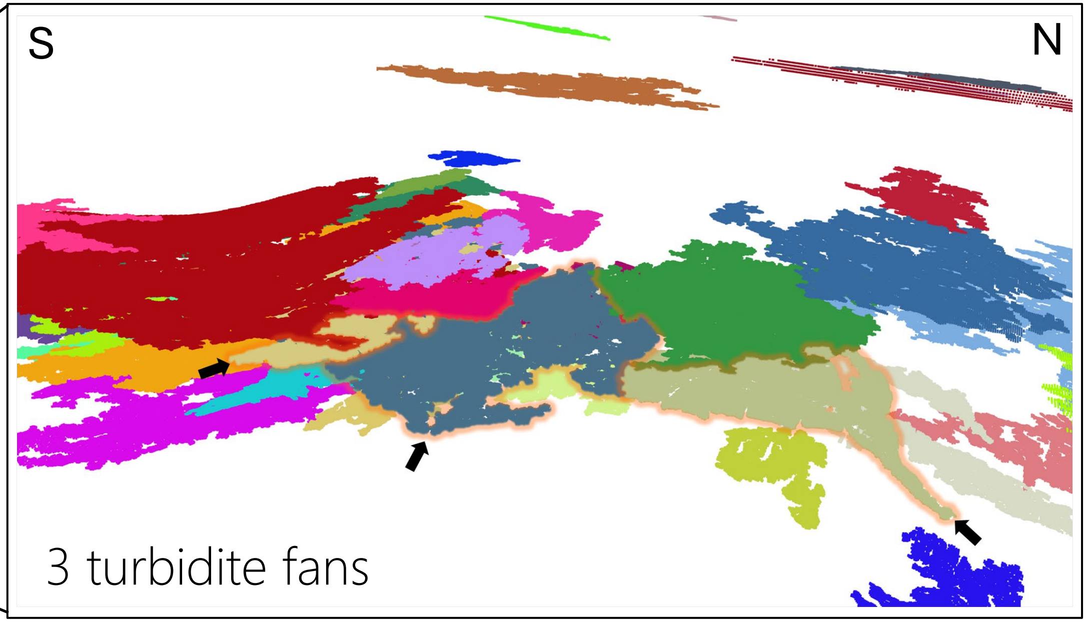


Segment database

Outputs:

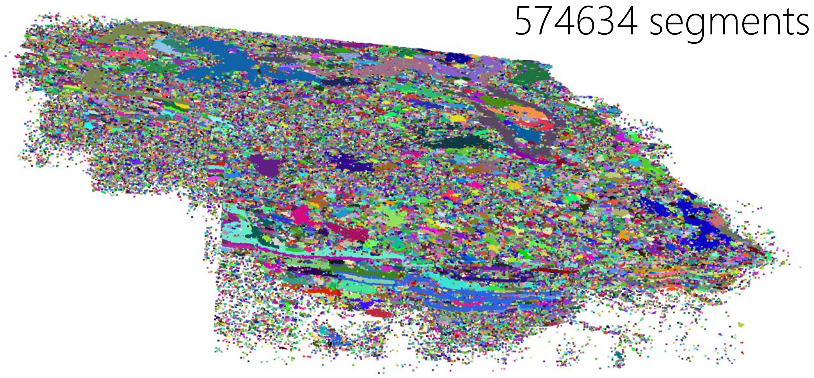


50-closest segments in the database



NFB – Turbidite fan detection

Full seismic segmentation

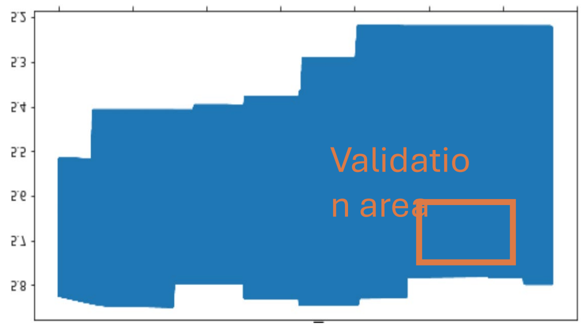


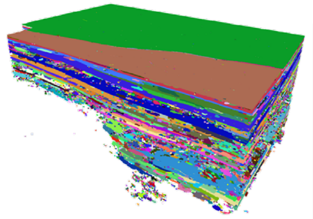
Sizeable seismic (30GB) (3150 iL, 5032 xL, 501 ts)

step	time
Point Cloud extraction	25min
Spatial segmentation (DBSCAN)	9min

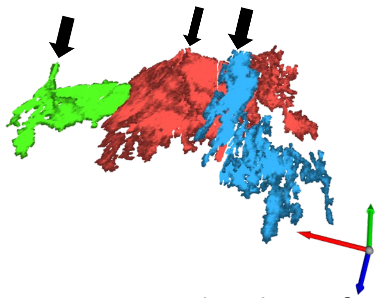
Extremely fast

Validation methodology





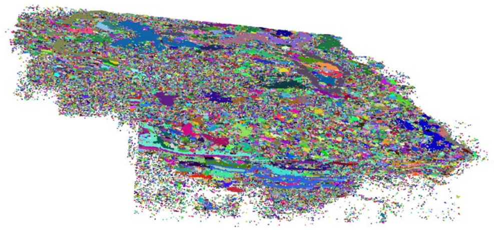
Validation dataset

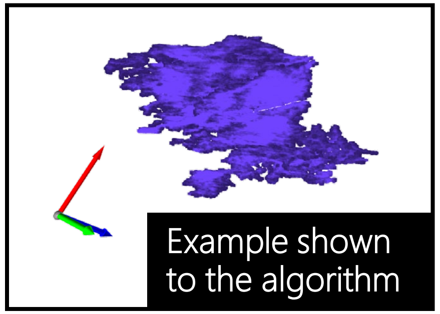


3 known turbidite fans

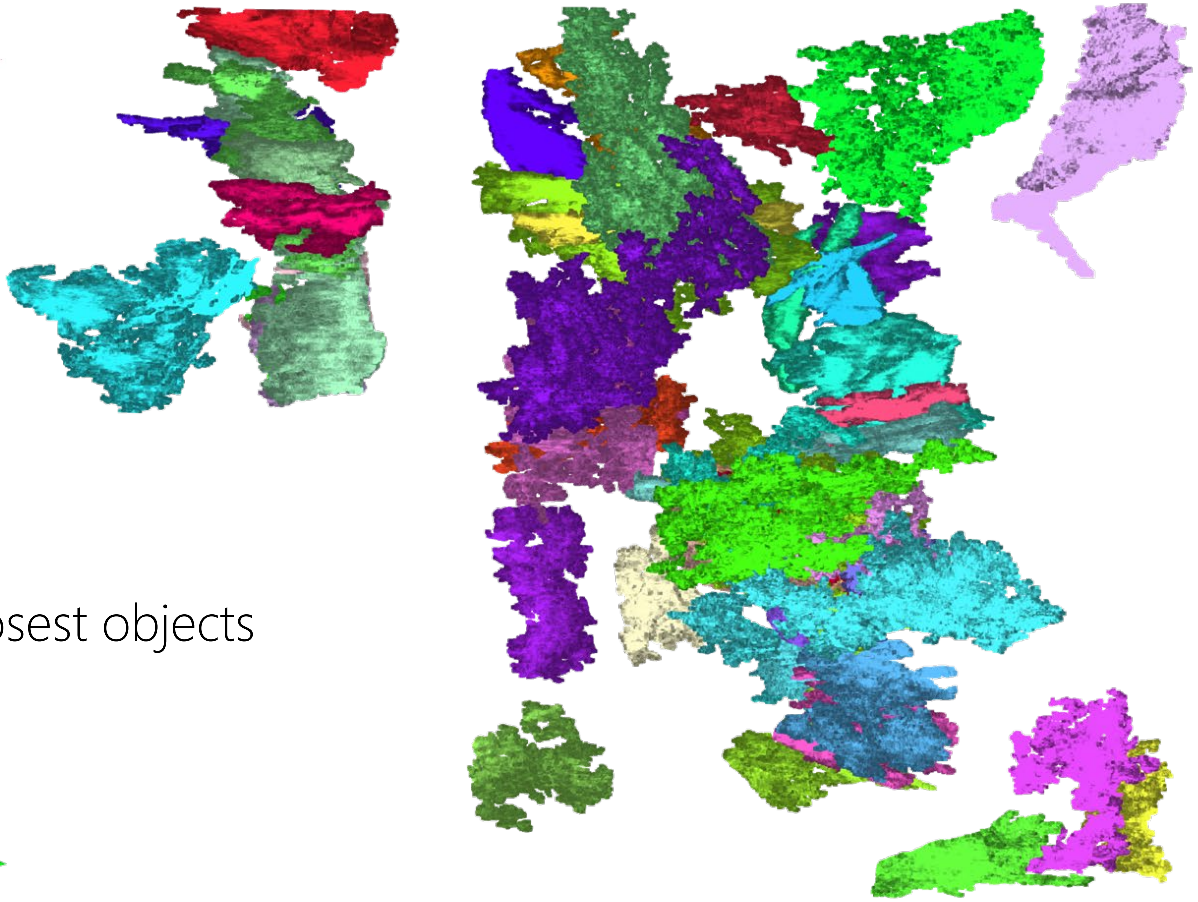
Retrieve unknown turbidites?

NFB – Turbidite fan detection

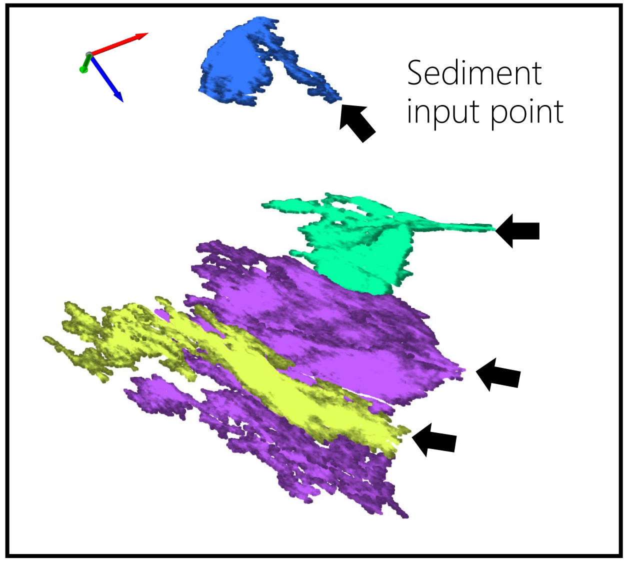




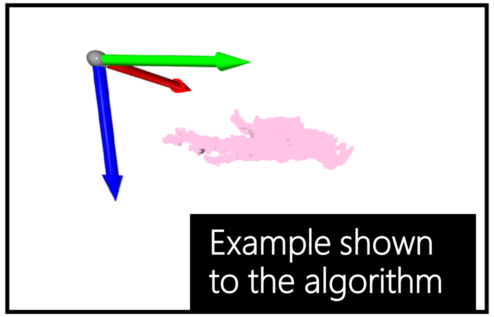
Results turbidite fan detection



50 closest objects

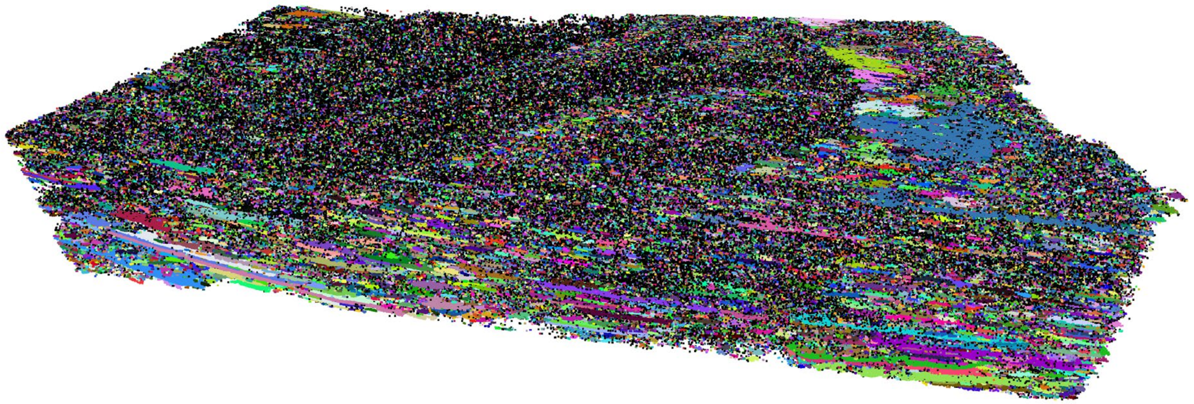


Fisa – Sills intrusion detection



Full seismic segmentation

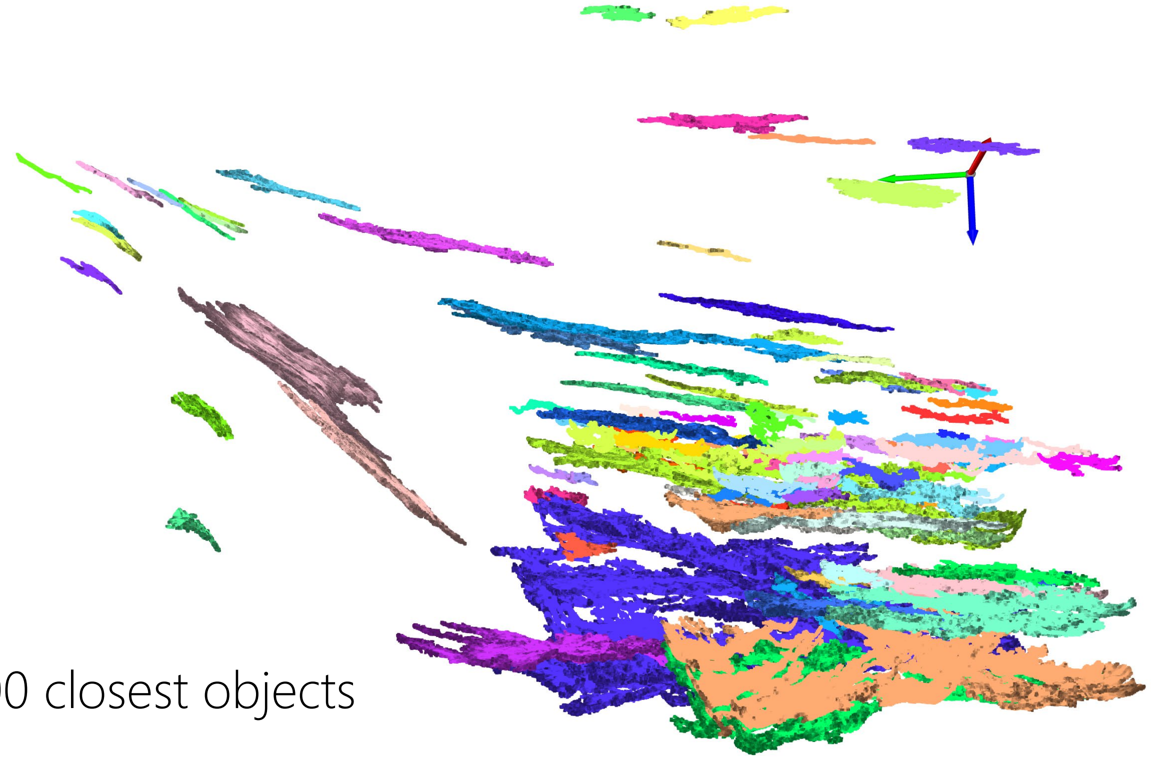
1,594,216 segments



Sizeable seismic (56GB) (2616 iL, 5692 xL, 1001 ts)

step	time
Point Cloud extraction	1h11min
Spatial segmentation (DBSCAN)	18min

Results sills detection



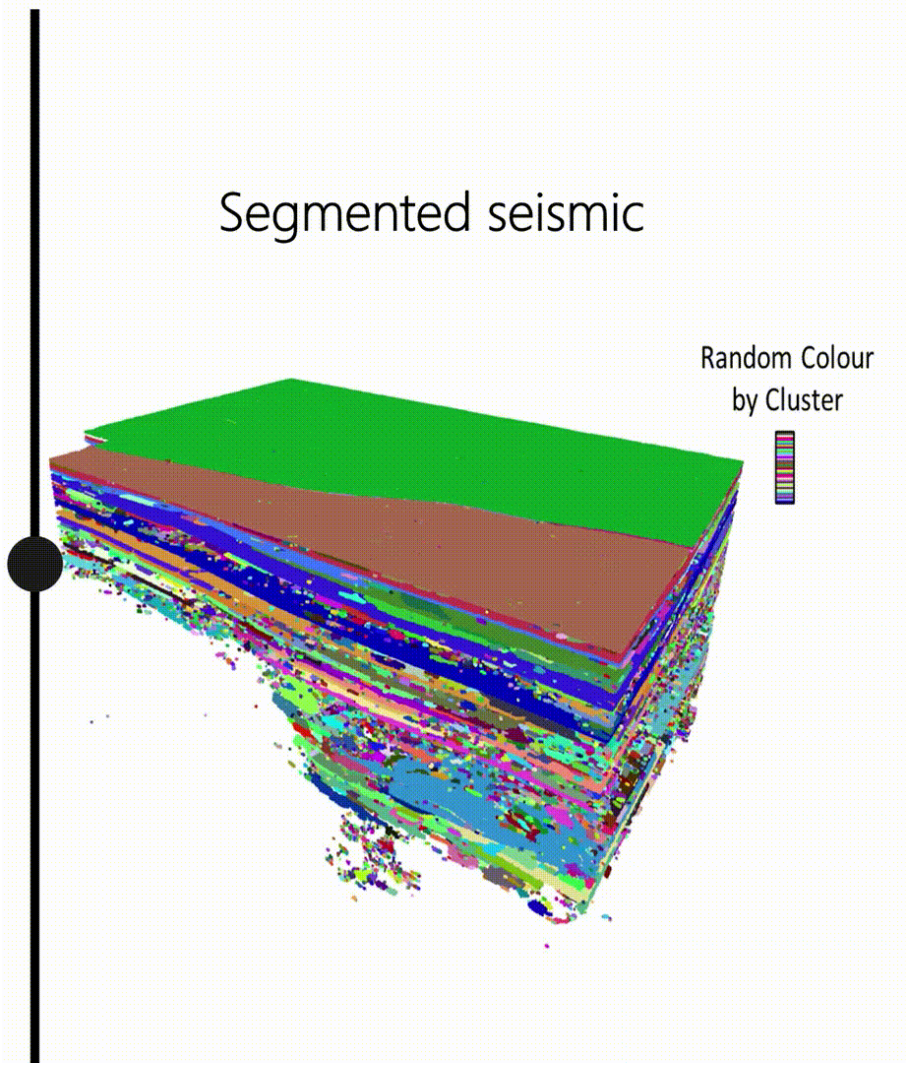
100 closest objects

Conclusion

- Fully data-driven method for detecting turbidite fans
- Allow to screen through a small number of possible targets looking for a particular geobody
- Adaptive to detection of other types of geobodies

pySeismic Python code available:

[GitHub repository](https://github.com/GeoDataScienceUQ/pyseismic/)
https://github.com/GeoDataScienceUQ/pyseismic/

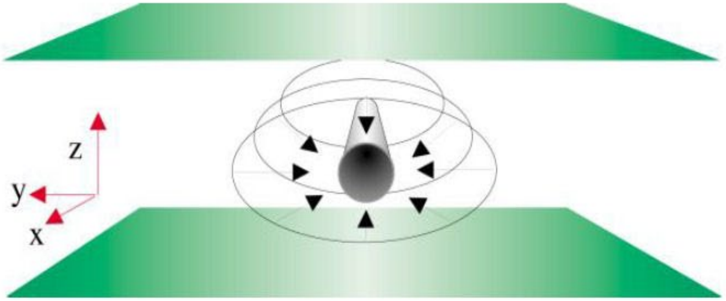


Pressure Transient Analysis (PTA)

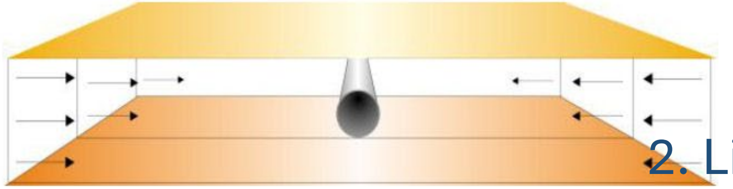
- PTA basics: flow regimes, k, skin
- Transient analysis illustration
- Permeability estimation from PTA
- Practical exercise

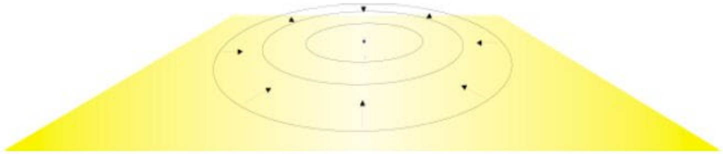
Pressure Transient Analysis (PTA): flow regimes

1. Radial (cross-section) flow, similar as to vertical well



2. Linear flow between top and bottom





3. Radial (reservoir) flow, similar as to vertical well

Top view (vertical well)

1. Radial flow (reservoir)

2. Linear (boundary dominated)

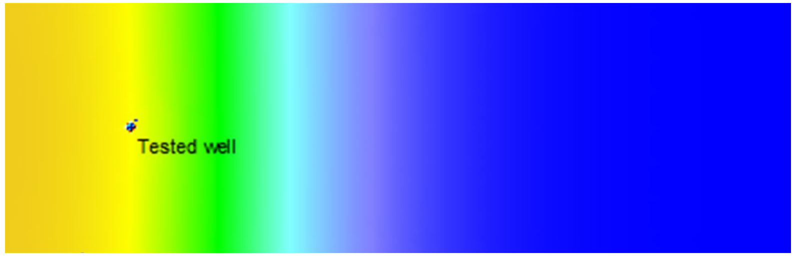
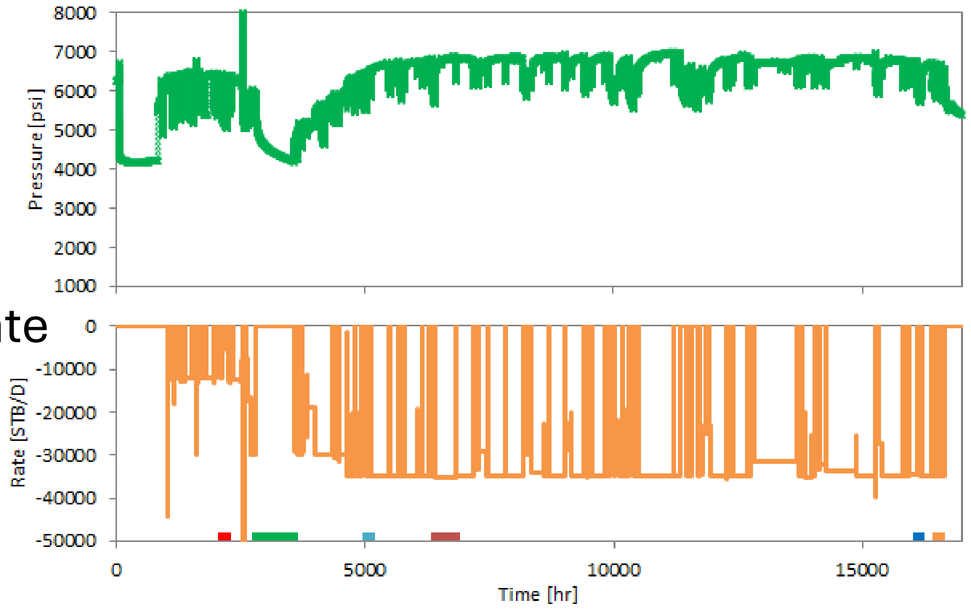


Illustration from [Dynamic Data Analysis / Kappa Eng., 2011]

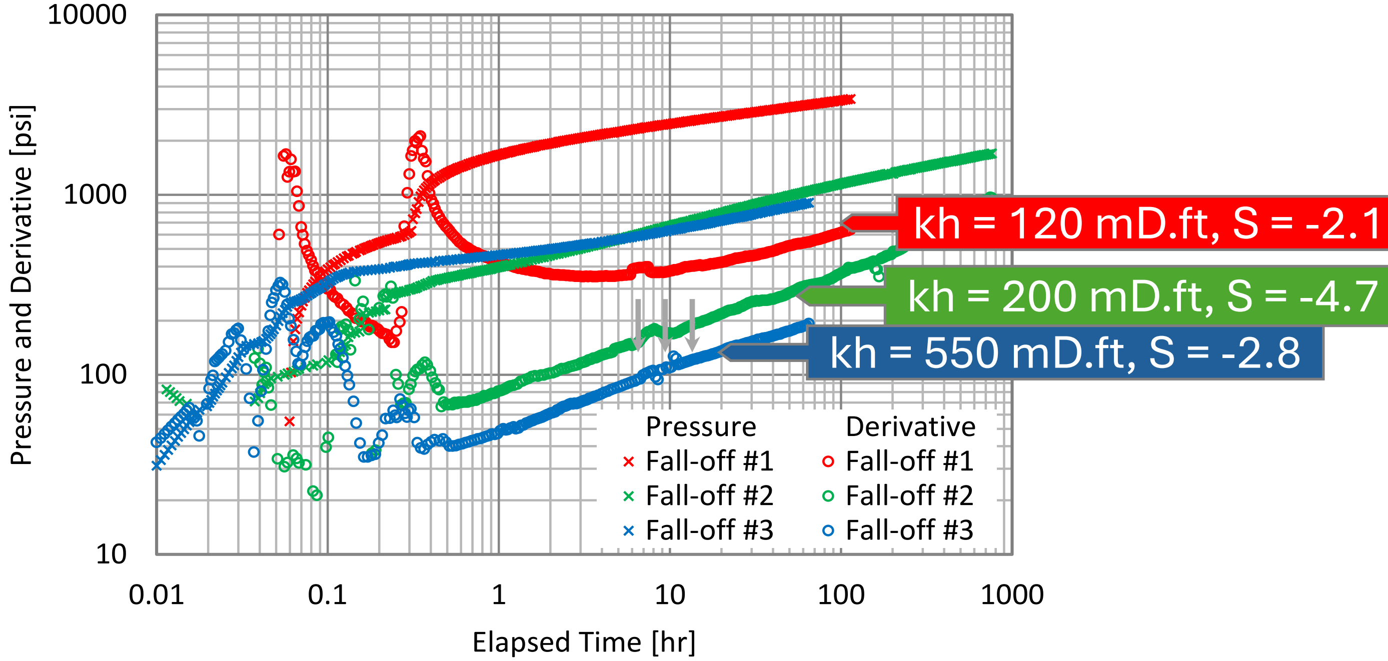
Pressure transients

Pressure (BHP):

Rate

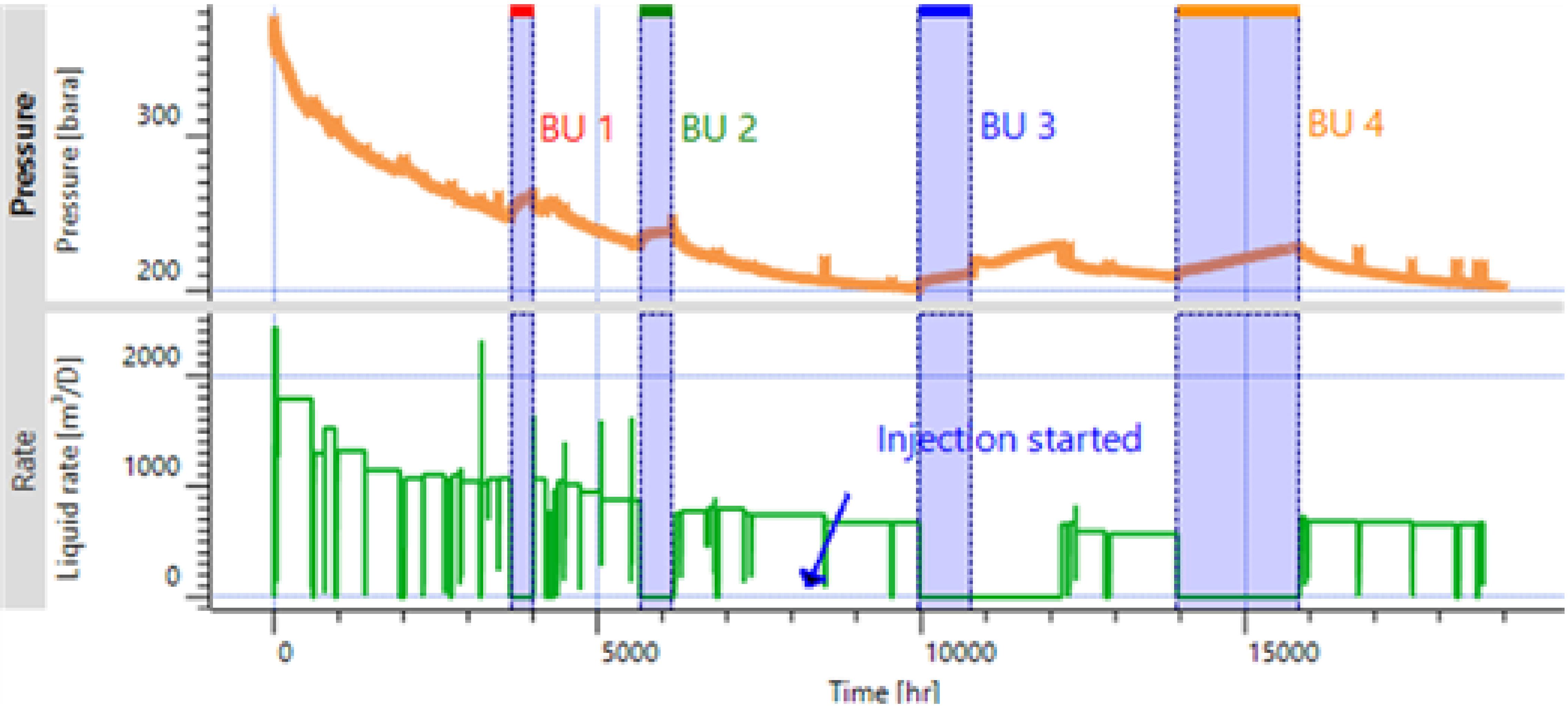


History of a horizontal multi-frac water injector with variety of fall-off and injection periods

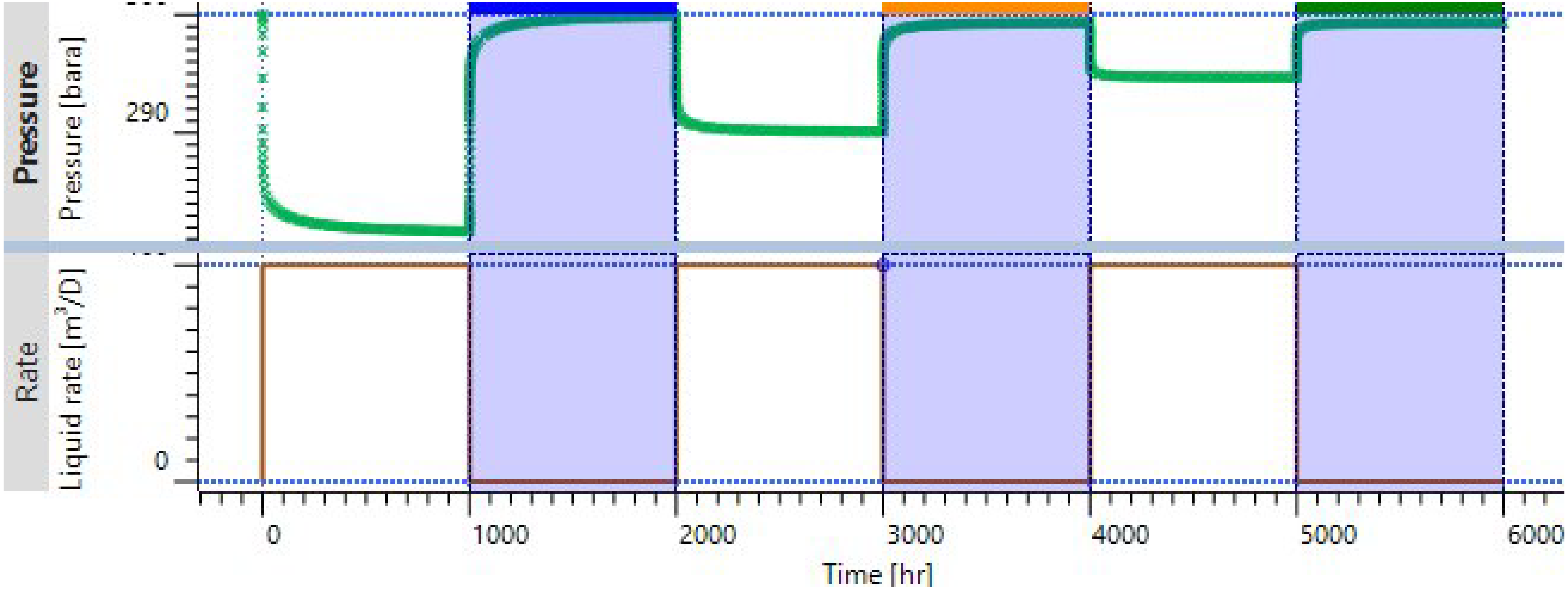


Shifting fall-off pressure derivatives down indicates kh growth with time / pressure build-up

Pressure and rate measurements



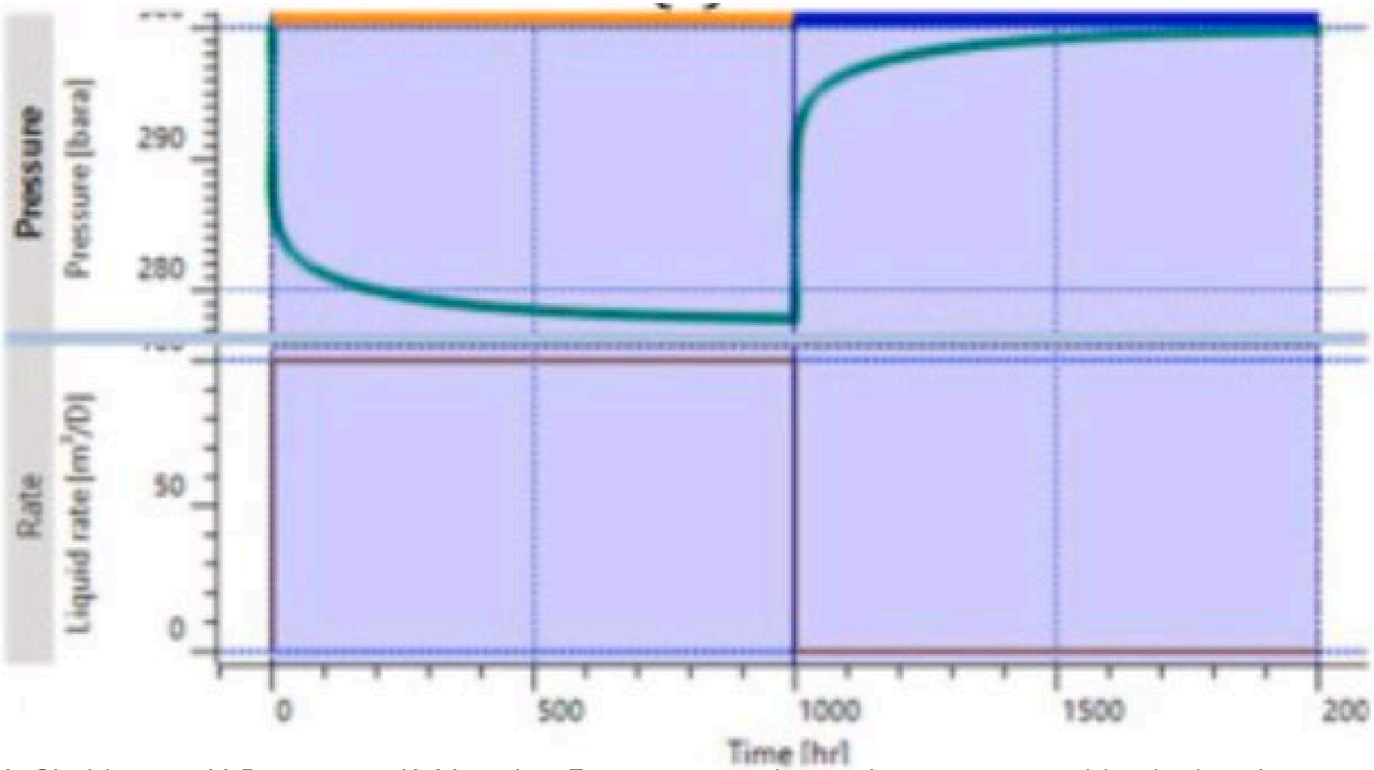
Well pressure transients: flow/shut-in



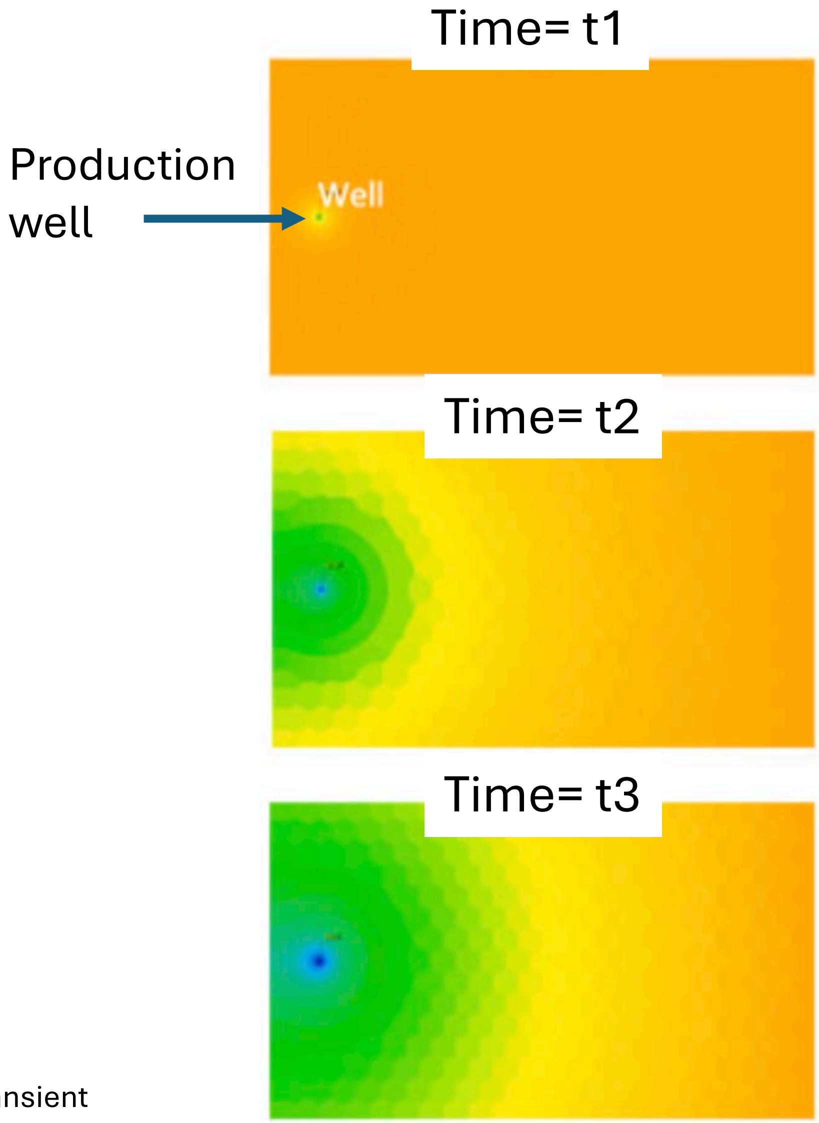
Pressure transient analysis (PTA)

Production well:

flow in interval shut-in interval



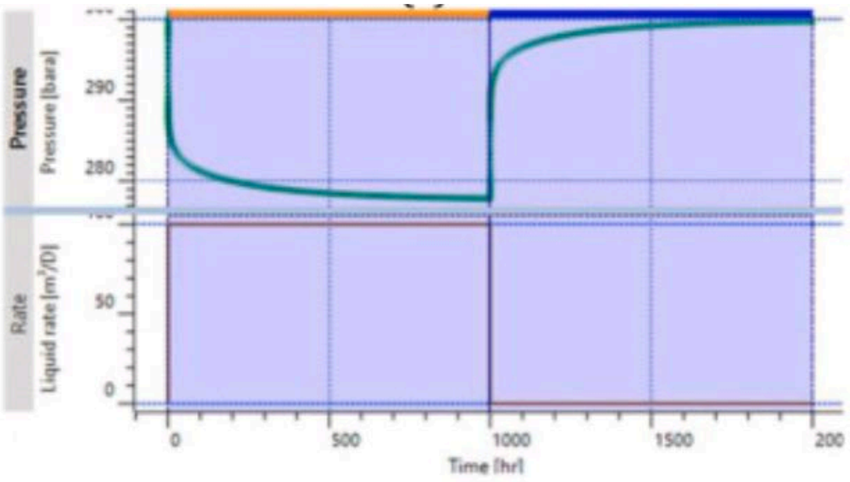
Pressure field change:

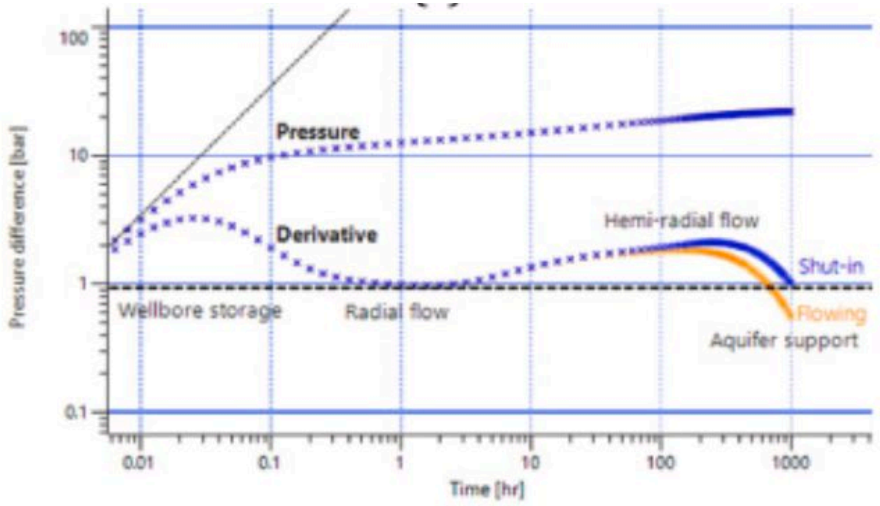


V. Starikov, A. Shchipanov, V. Demyanov, K. Muradov, Feature extraction and pattern recognition in time-lapse pressure transient responses, Geoenergy Science and Engineering, 2024, https://doi.org/10.1016/j.geoen.2024.213160

Pressure transient analysis (PTA)

PTA helps to estimate well and reservoir parameters (skin and permeability) by analysing Pressure vs Time response to a well flow rate change.

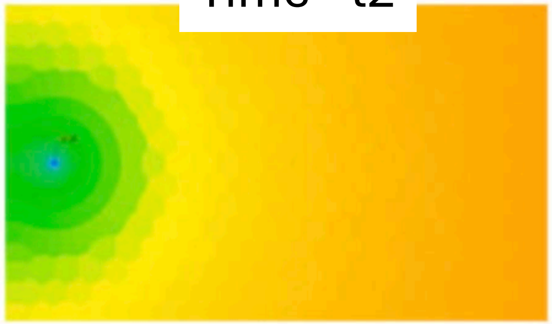




Pressure field change:

Time= t1

Time= t2



Time= t3

V. Starikov, A. Shchipanov, V. Demyanov, K. Muradov, Feature extraction and pattern recognition in time-lapse pressure transient responses, Geoenergy Science and Engineering, 2024, https://doi.org/10.1016/j.geoen.2024.213160

Permeability estimation: Slope analysis

Pressure follows linear relationship with log(t) during Radial Flow regime.

Once Radial Flow Identified:

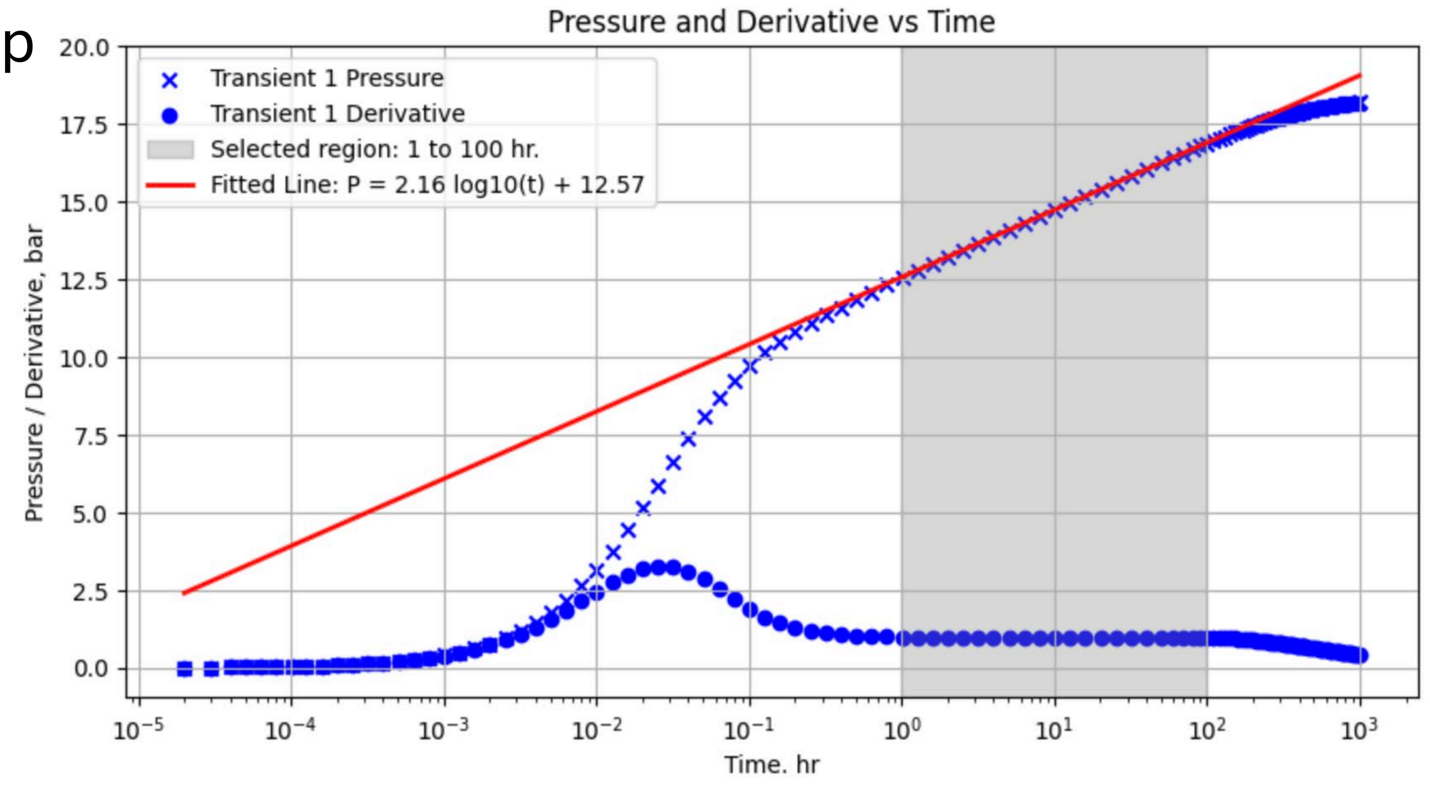
start = 1hr, end = 100hr

Permeability can be estimated from the slope *a* of the:

$$P \sim a*\log(t)+b$$

Model permeability k = 100 mD

Estimated permeability k = 99.39 mD



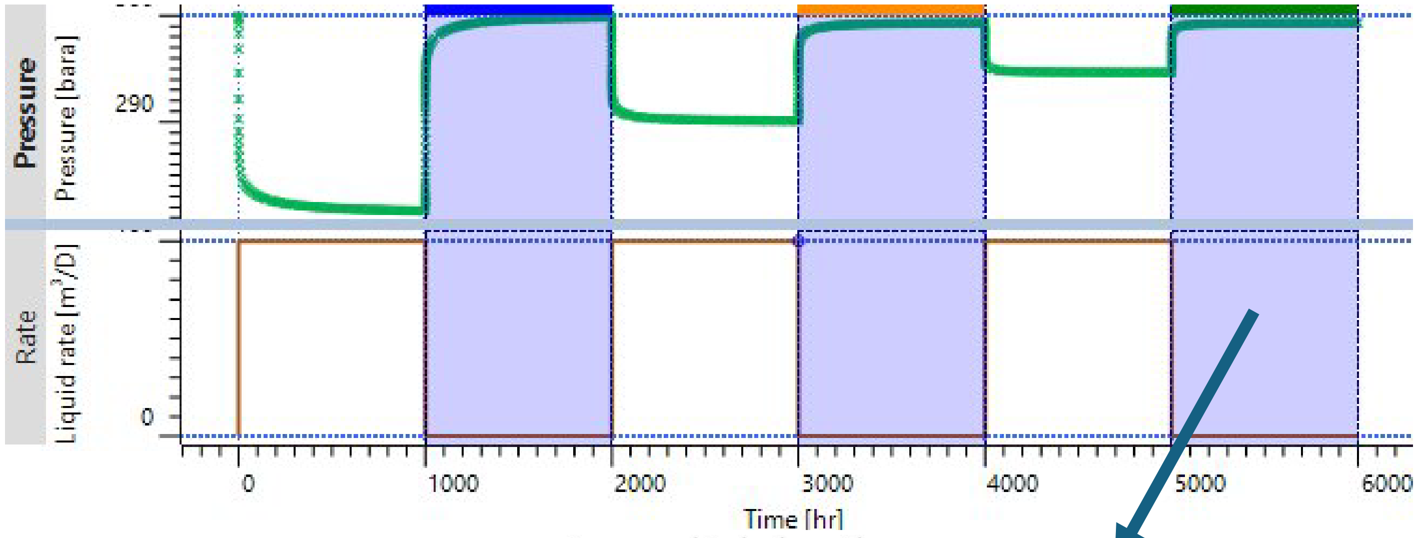
Pressure Transient Analysis (PTA)

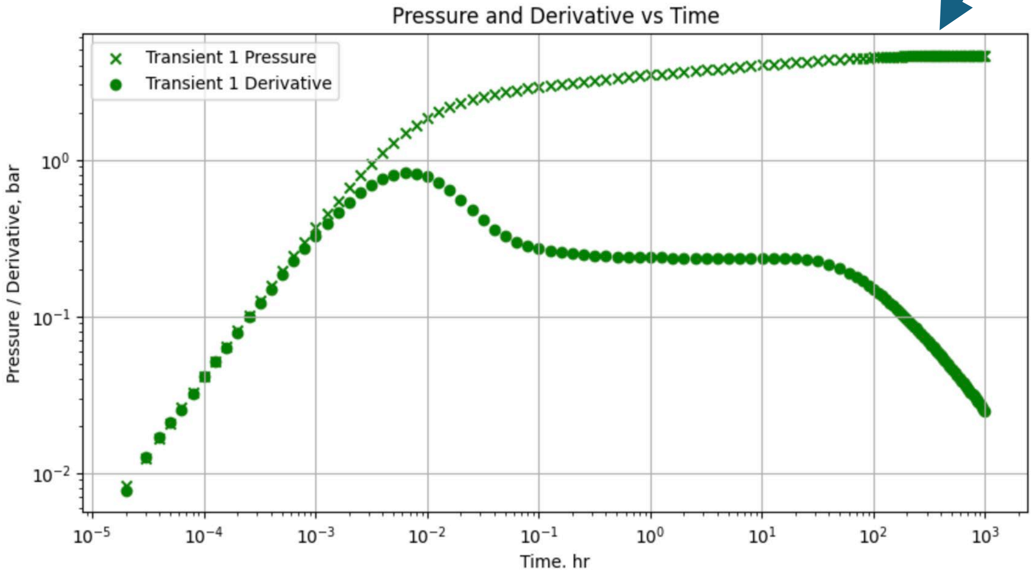
Hands-on

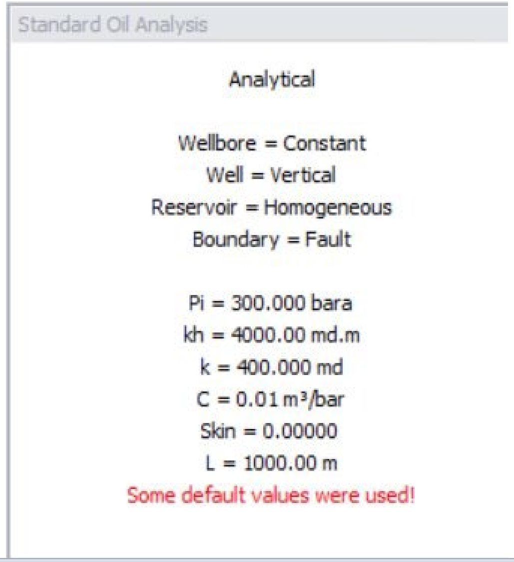
Unsupervised Clustering. Notebook 4

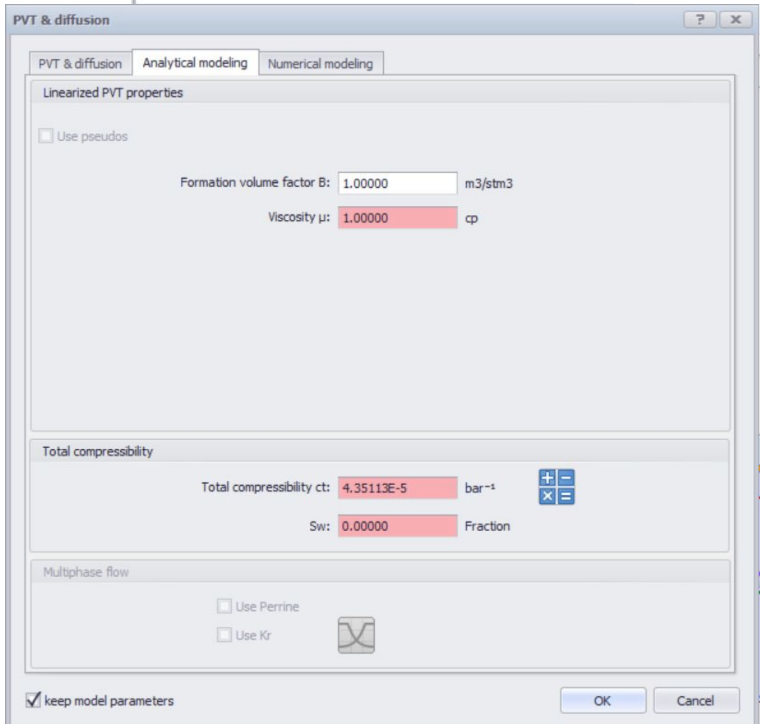
https://colab.research.google.com/drive/1TEFEJ7wAeNW9ocSHJsidmeOvxS_pq4KP?usp=sharing

Pressure and derivative





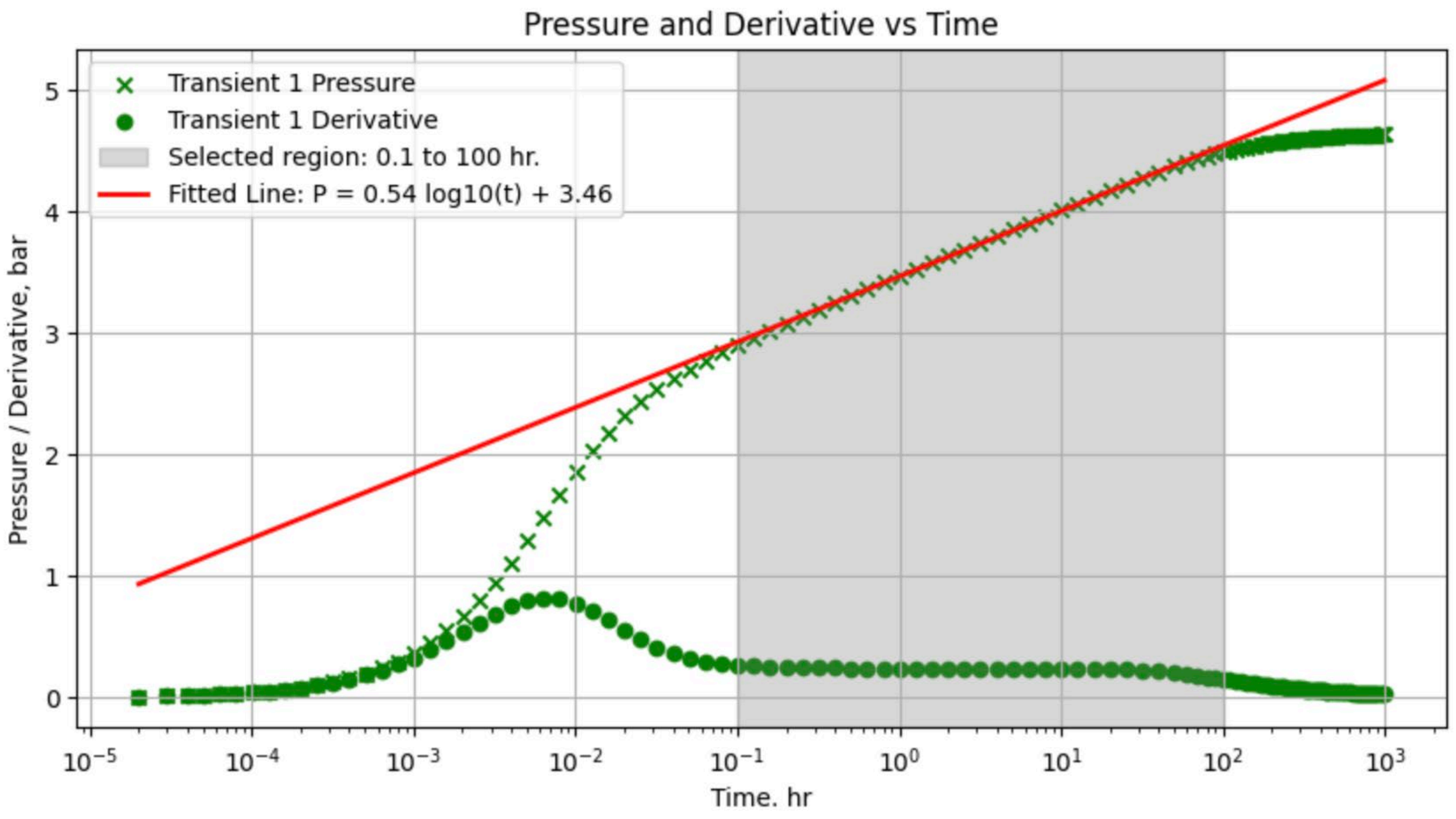




Exercise 4.1

Assumed Radial Flow period:
Start = 0.1 hr
End = 100 hr
Estimated permeability - ?

Actual model permeability:
K = 400mD



Exercise 4

Task 4.1: Manual PTA

- Select a radial flow period (flat dP derivative) and calculate permeability.
- Select a different radial flow period and compare how it impacts permeability calculation

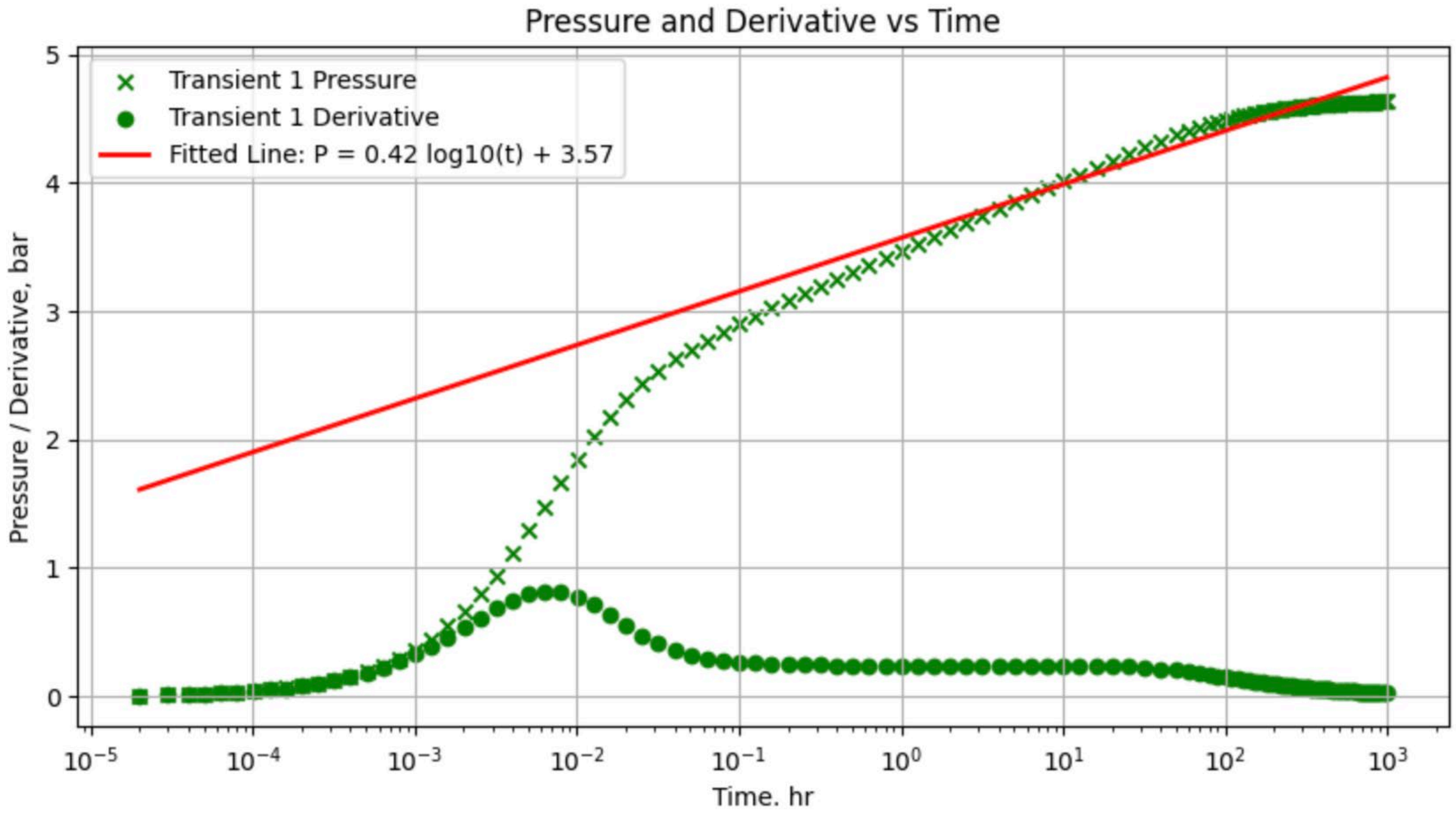
Task 4.2: Automated flow pattern classification with pta-learn

- Automatically identify radial flow period and calculate permeability, compare with the one in calculated in Task 4.1.
- Apply automatic PTA classification for a different well.

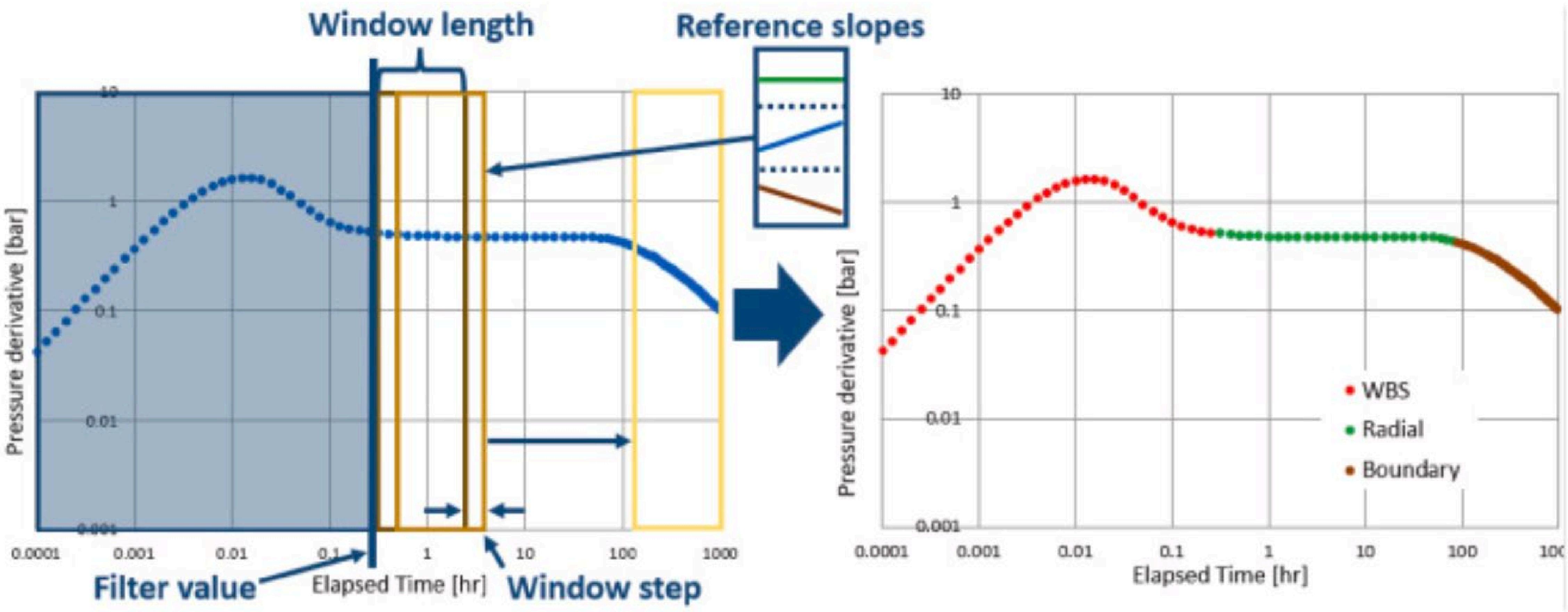
Exercise 4.1

How does region selection affect permeability calculation?
Start = ?
End = ?
Estimated permeability - ?

Actual model permeability:
K = 400mD



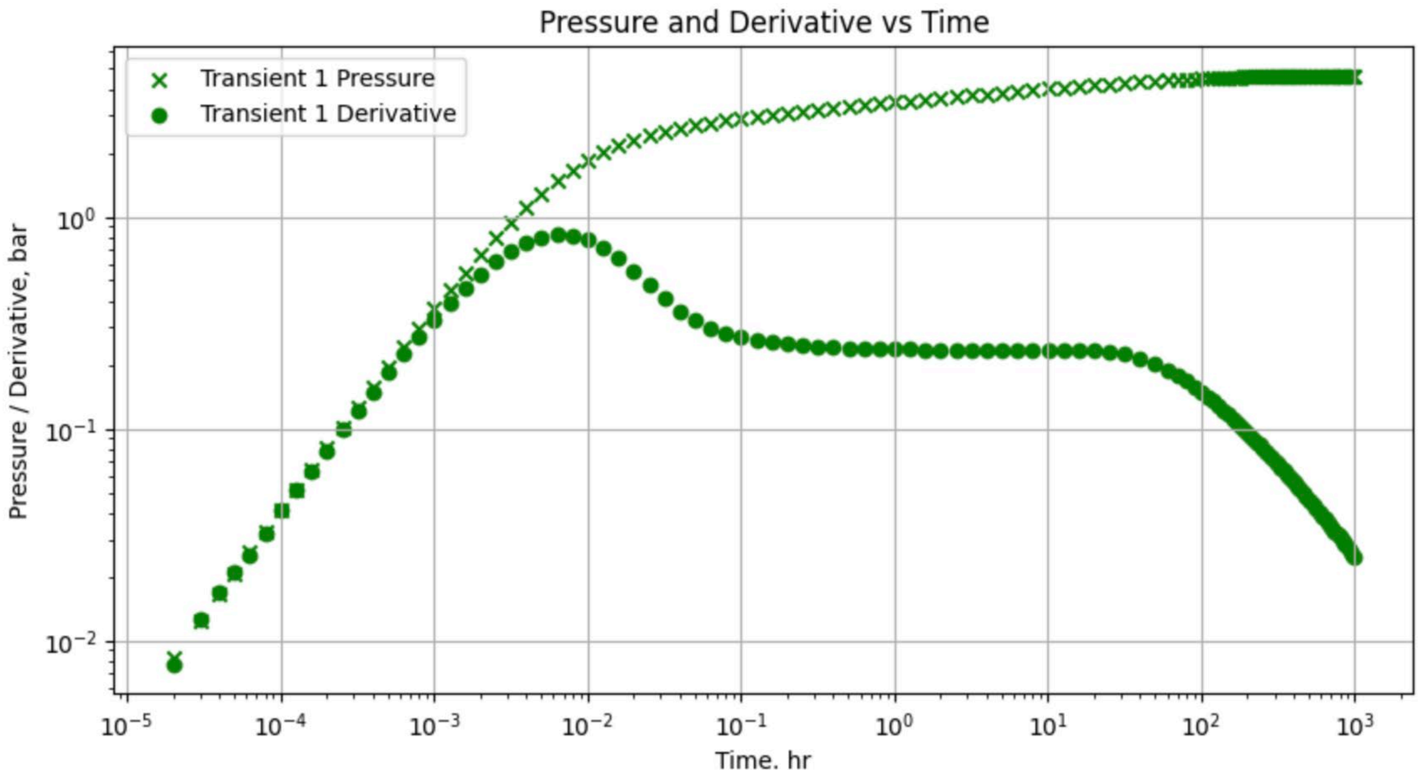
pta-learn: Automated feature extraction



V. Starikov, A. Shchipanov, V. Demyanov, K. Muradov, Feature extraction and pattern recognition in time-lapse pressure transient responses, Geoenergy Science and Engineering, 2024, https://doi.org/10.1016/j.geoen.2024.213160

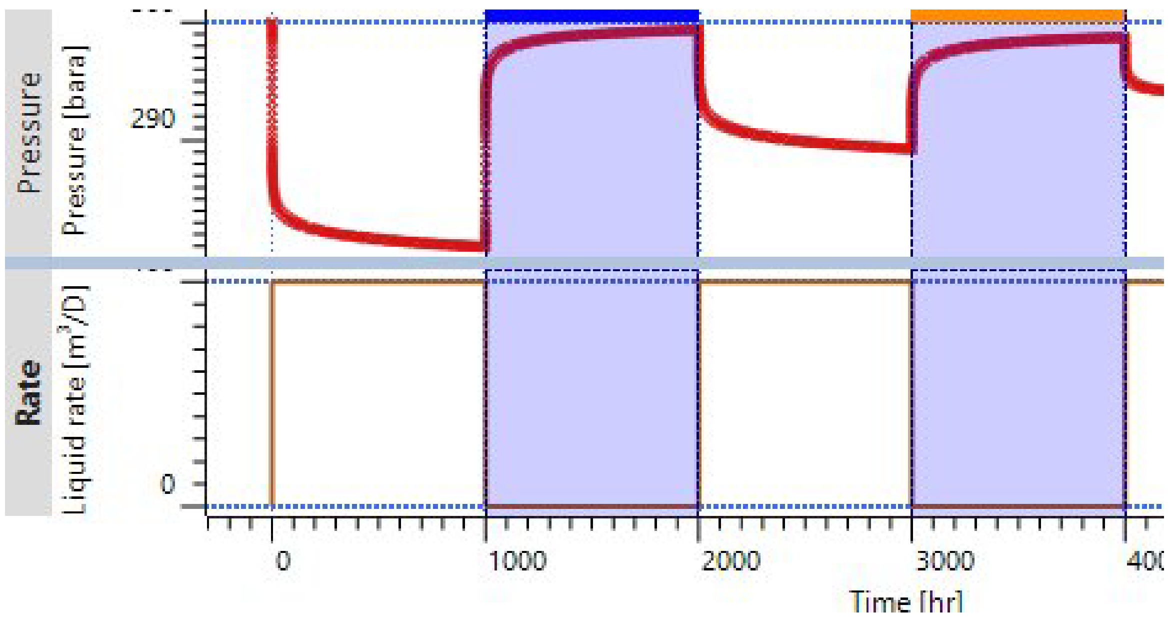
Exercise 4.2

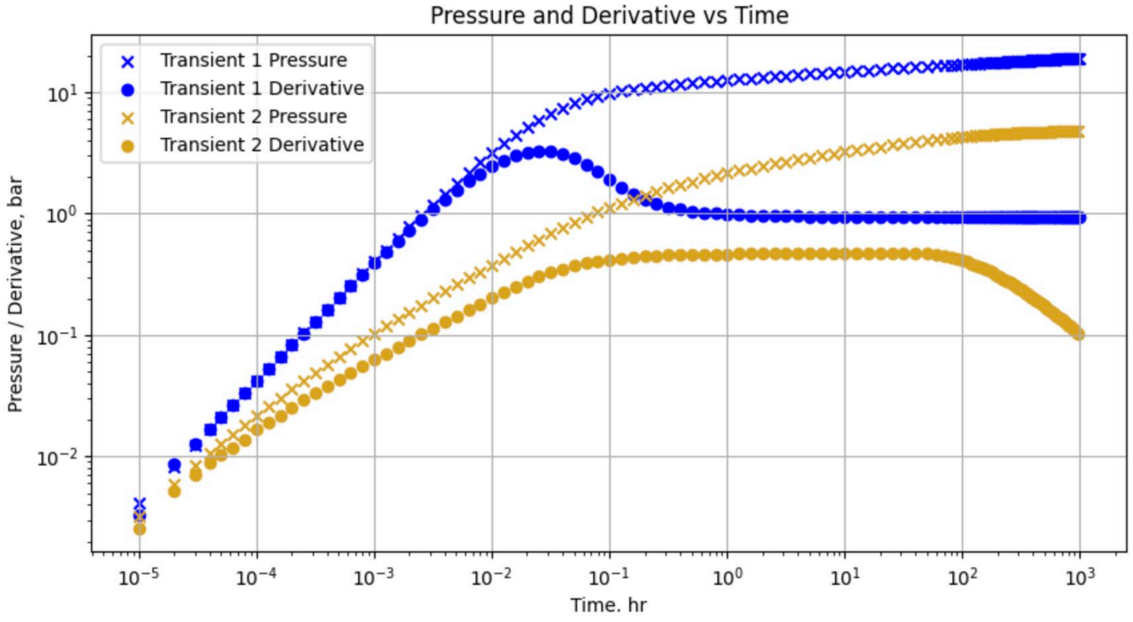
1. Extract Flow regime features
2. Identify radial flow regime
3. Calculate permeability



Exercise 4.2

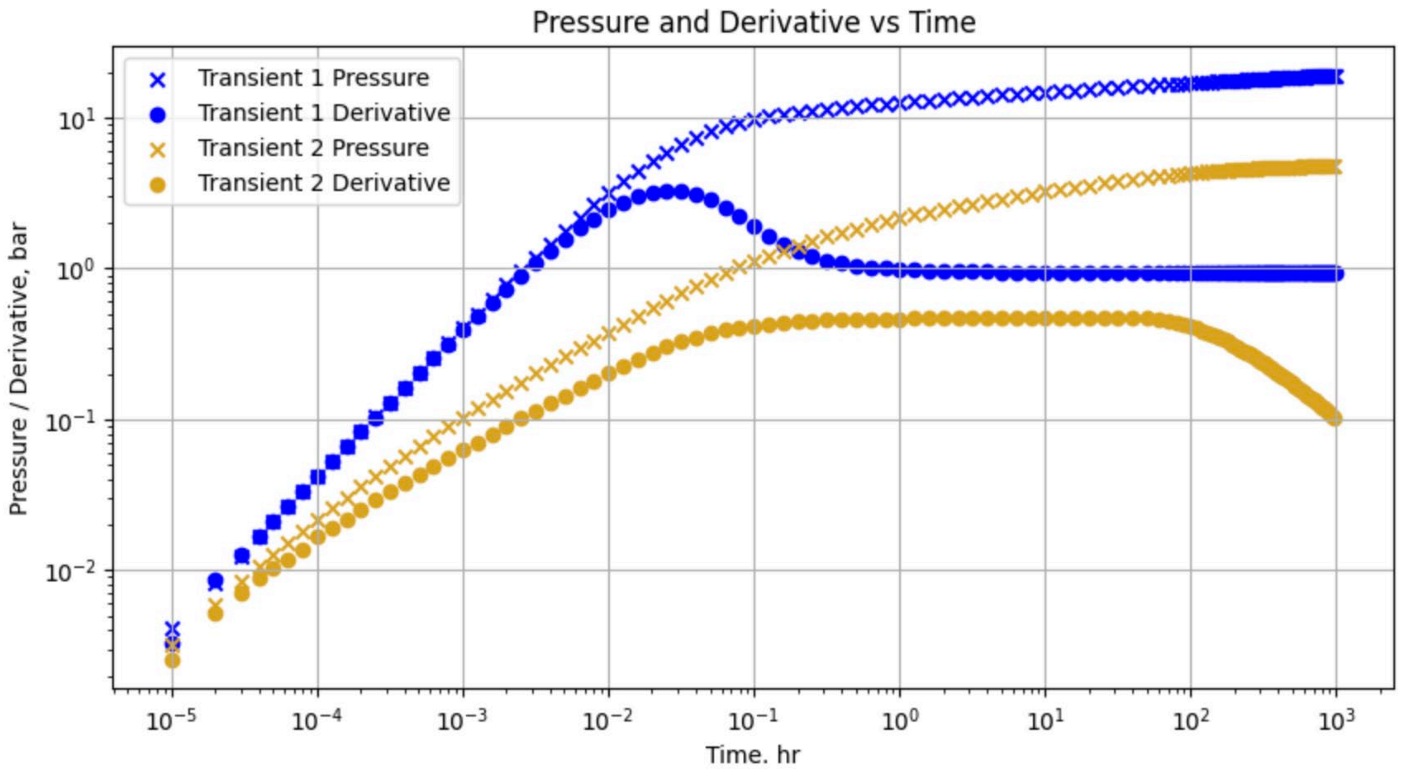
Given well history, use extract pressure transient responses.





Exercise 4.2

1. Extract Flow regime features
2. Identify radial flow regime
3. Calculate permeability



Try on your own data

Thank you for attending

Professor Vasily Demyanov, Farah Rabie, Vitalii Starikov

geodatascience.hw.ac.uk

References

- Halotel, J., Demyanov, V. & Gardiner, A. (2020) *Value of Geologically Derived Features in Machine Learning Facies Classification*. Mathematical Geosciences, 52, 5–29.
- Konoshonkin, D., Shishaev, G., Matveev, I., Volkova, A., Rukavishnikov, V., Demyanov, V. & Belozerov, B. (2020) *Machine Learning Clustering of Reservoir Heterogeneity with Petrophysical and Production Data*, SPE Europec Featured at 82nd EAGE Conference and Exhibition, SPE-200614-MS
- V. Demyanov, Q. Corlay, A. Nathanail, C. Sun4and D. Arnold (2025*) AI-based Reservoir Modelling Workflows - an Illustrative Overview*, European Association of Geoscientists & Engineers, Fifth EAGE Digitalization Conference & Exhibition, Mar 2025, Volume 2025, p.1 – 5, DOI: https://doi.org/10.3997/2214-4609.202539041
- Q. Corlay, V. Demyanov, D. McCarthy and D. Arnold (2020) *Turbidite Fan Interpretation in 3D Seismic Data by Point Cloud Segmentation Using Machine Learning*, EAGE 2020 Annual Conference & Exhibition Online pp 1 – 5.
- Q. Corlay, *Detection of Geobodies in 3D Seismic using Unsupervised Machine Learning*, PhD thesis, Heriot-Watt University, 2023
- V. Starikov, A. Shchipanov, V. Demyanov, K. Muradov (2024) *Feature extraction and pattern recognition in time-lapse pressure transient responses*, Geoenergy Science and Engineering,
- A. Shchipanov, L. Kollbotn, G. Namazova (2023) *Evaluation of Well Interference from Time-lapse PTA,* Journal of Petroleum Exploration Production Technology, **13**, 1591–1609