PGE 338 Data Analytics and Geostatistics

Lecture 1: Data Analytics

Lecture outline . . .

- Data Analytics Concepts
- Types of Data
- Data Analytics Examples

Introduction

General Concepts

Statistics

Probability

Univariate

Bivariate

Spatial Analysis

Machine Learning

Uncertainty Analysis

Michael Pyrcz, The University of Texas at Austin

Statistics Moment

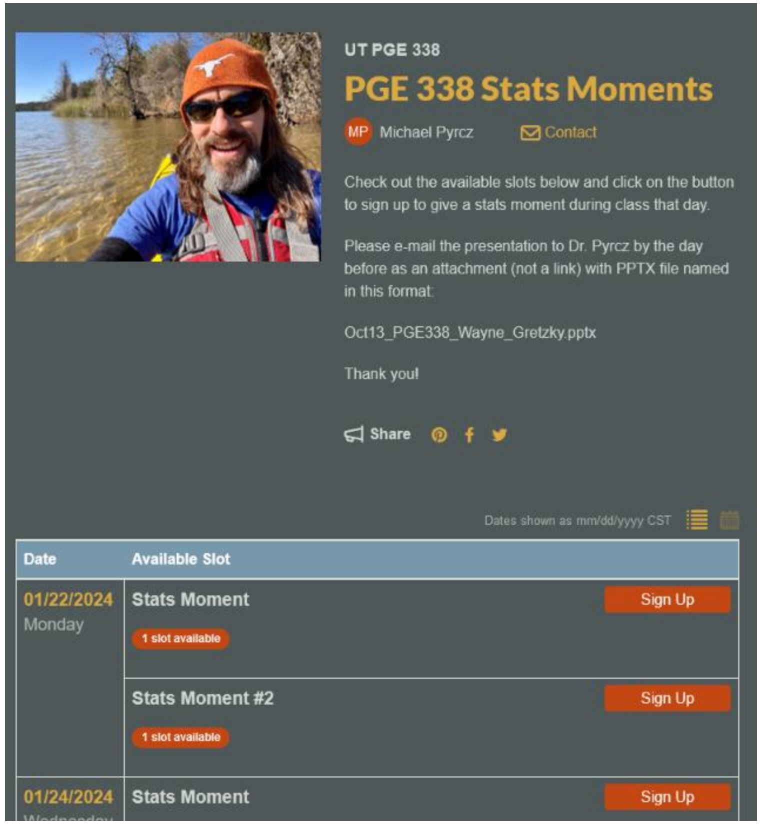
As part of your participation grade everyone will do a statistics moment.

- find an example of statistical impact in industry, society, etc. Sorry no sports related.
- professional / workplace standards, no politics, etc.
- 2 slides maximum and present on it in class, 3 minutes maximum.
- **see the announcement with SignUpGenius site to sign up! grab a spot early!**

Send e-mail with attached .pptx file
by the day before you stats moment.

- please don’t send a link to the cloud
- use filename format: [date]_PGE338_[FirstName]_[LastName].pptx
  e.g., Oct13_PGE337_Wayne_ Gretzky.pptx
- we will do one or more per class

Goal: statistics as a lens to discover new things!



Sign up on SignUpGenius at:
https://www.signupgenius.com/go/508044FAFA72BA3F94-47343741-pge338

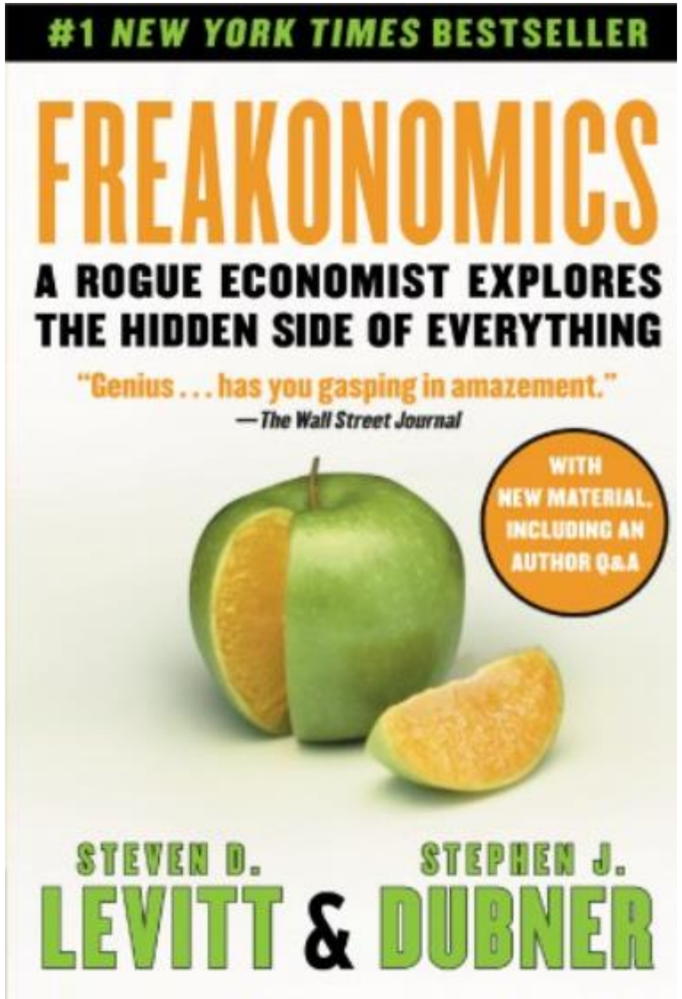
Example of Source Material

Freakonomics

(Levitt and Dubner, 2009)

On the unintended consequences of incentives:

- Sumo wrestlers cheat
  - Wrestlers that met their required number of wins threw matches
- Teachers in Atlanta were cheating
  - Correcting answers on students’ benchmark tests
- Remember this when you:
  - Motivate your asset team, unit or division
  - Working with kids
  - Assess your own actions professionally for bias



Statistics Moment
Seed Ideas

Some Ideas for Statistics Moments, Impact on Society or Industry

- **Interesting Statistics from Oil**
  - e.g., 40% of ocean cargo is oil! Discuss the impact of this!
- **Oil Field HES (Health Environment and Safety)**
  - What are the major risks in our field? Share look backs and lessons-learned!
- **Projects**
  - Size or projects, distribution of new projects world-wide. How big is our enterprise!
- **Energy Use**
  - Allocation geographically and between energy sources. Where?
- **Industry Trends**
  - Changing workforce? New tech? Where are we going?

Statistics Moment
by Chris Cho

Exponential Growth

- Average Paper Thickness is 0.1mm (0.0039 in).
- As Number of folds increase, so the thickness grows exponentially.
- **3 folds**: fingernail
- **10 folds**: Hand width
- **11 folds**: MythBusters record.
- **24 folds**: 1 Kilometer/3280 ft
- **30 folds**: From Earth to space(100km)
- **42 folds**: Moon

Captures from MythBusters Folder Paper Seven Plus Times
https://www.youtube.com/watch?v=65Qzc3_NtGs

Statistics Moment
by Chris Cho

Exponential Growth

- **51 folds**: From the earth to the Sun
- **81 folds**: Your paper will be 127,786 light-years, almost as thick as the Andromeda Galaxy
- **90 folds**: Your paper will be 130.8 million light-years across, bigger than the Virgo Supercluster
- **103 folds**: Reach the outside of the observable Universe, which is estimated at 93 billion light-years in diameters.
- **Current Record is 12 folds:** single piece of toilet paper 4000 ft (1200 m) in length, broke conventional belief of 7 max

Britney Gallivan Guinness World Record holder with 12 folds.

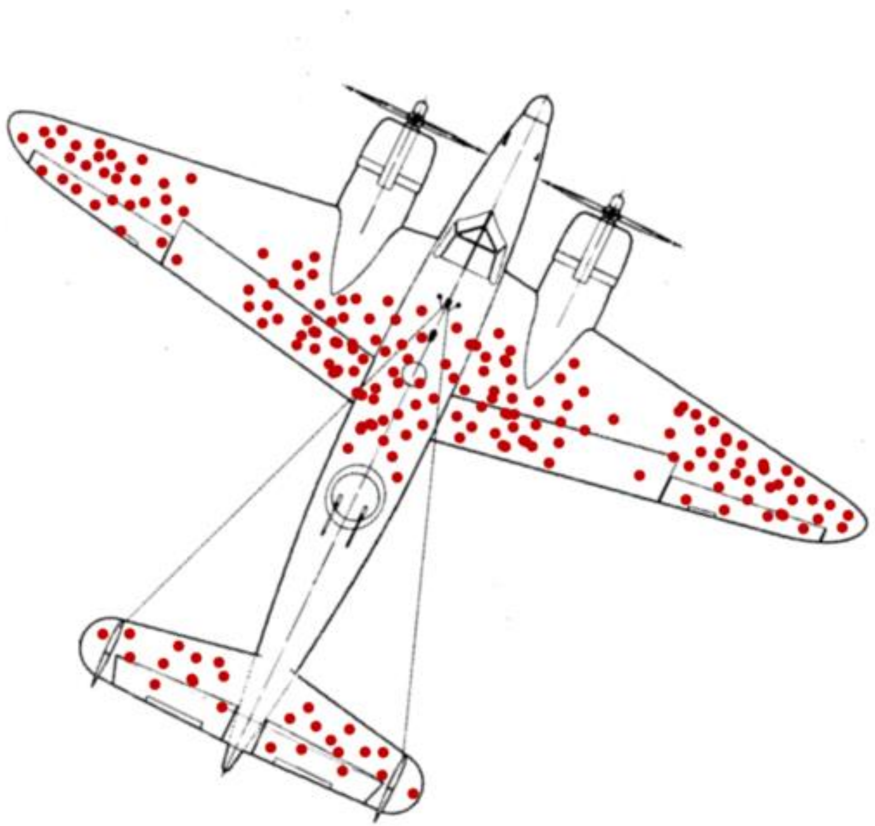
Example Statistics Moment on Survivorship Bias

Michael Pyrcz, The University of Texas at Austin

Survivorship Bias: a form of selection bias resulting from retaining samples that "survived" some previous selection process.

During WWII the Center for Naval Analyses (@CNA_org Twitter) compiled a dataset of bomber damage to assess where reinforcement was needed.

Statistician Abraham Wald recognized this was a case of survivorship bias, he determined:



Hypothetical dataset of aircraft damage for planes that returned to based.

Image from https://en.wikipedia.org/wiki/Survivorship_bias#/media/File:Survivorship-bias.png

Example Statistics Moment on Survivorship Bias

Michael Pyrcz, The University of Texas at Austin

Survivorship Bias: a form of selection bias resulting from retaining samples that “survived” some previous selection process.

During WWII the Center for Naval Analyses (@CNA_org Twitter) compiled a dataset of bomber damage to assess where reinforcement was needed.

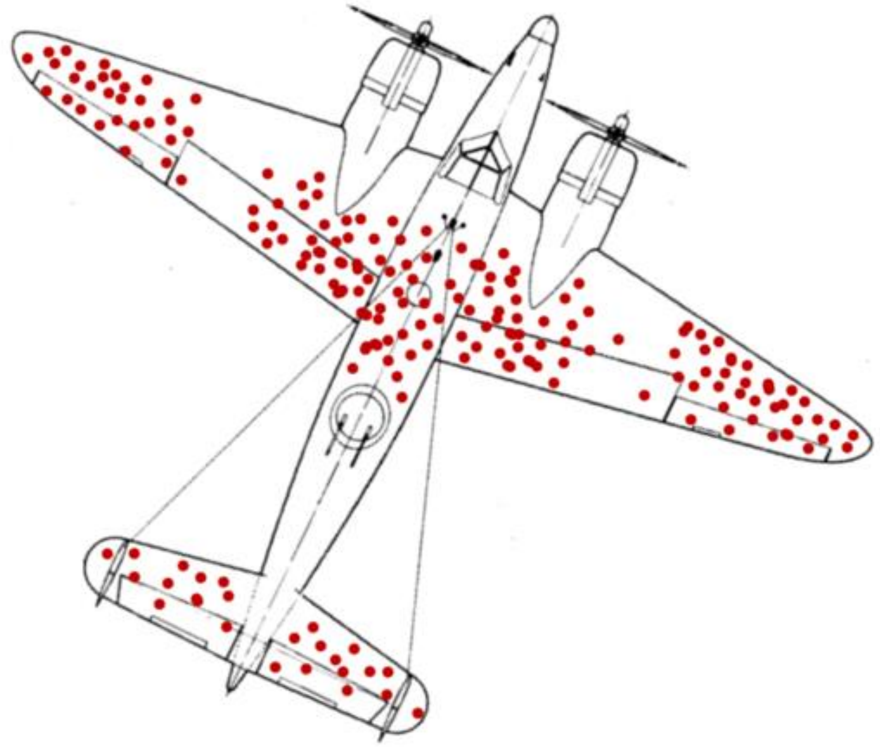
Statistician Abraham Wald recognized this was a case of survivorship bias, he determined:

- **Add armor where there is less damage, those planes don’t make it back to base!**

Impact of Statistics Knowledge: Preselection often occurs with subsurface data sets:

- drill in the best locations
- difficulty in extracting and measuring bad rock
- gather data from successful projects

We can use this statistics knowledge to debias and avoid incorrect conclusions and poor decisions.



Hypothetical dataset of aircraft damage for planes that returned to based.

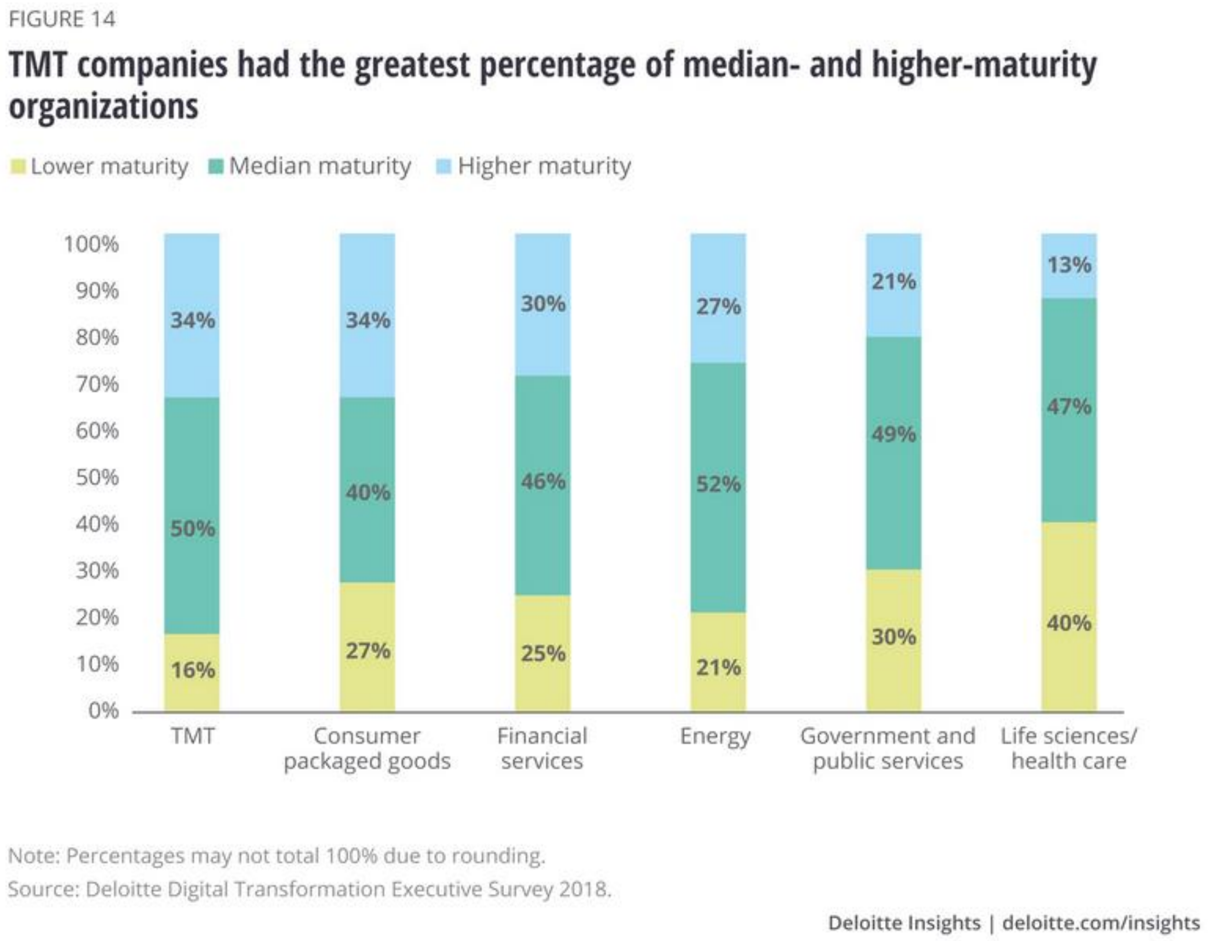
Image from https://en.wikipedia.org/wiki/Survivorship_bias#/media/File:Survivorship-bias.png

Motivation

- We are not alone, digital transformations are underway in all sectors of our economy
- Every energy company that I visit is working on this right now
- We are giving you an edge to succeed in this new environment!

Let’s establish some based concepts, a foundation to build on:

- statistics, data analytics, sampling, data types



Readiness of sectors of our economy for the digital transformation by Deloitte, 2019.

Chart from https://www2.deloitte.com/insights/us/en/focus/digital-maturity/digital-maturity-pivot-model.html

PGE 338 Data Analytics and Geostatistics

Lecture 1: Data Analytics

Lecture outline . . .

- Data Analytics Concepts

Introduction

General Concepts

Statistics

Probability

Univariate

Bivariate

Spatial Analysis

Machine Learning

Uncertainty Analysis

Michael Pyrcz, The University of Texas at Austin

Statistics, Geostatistics Big Data Analytics

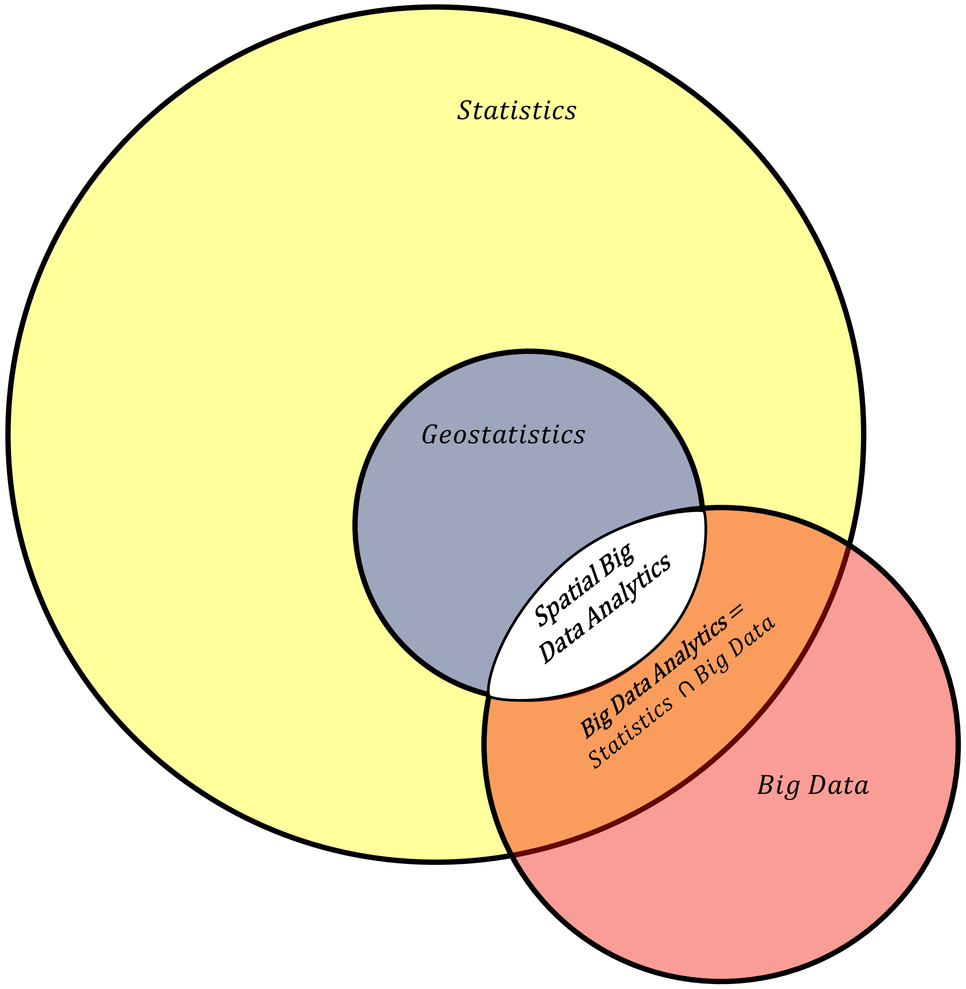
Statistics is collecting, organizing, and interpreting data, as well as drawing conclusions and making decisions.

Geostatistics is a branch of applied statistics that integrates:

1. the spatial (geological) context,
2. the spatial relationships,
3. volumetric support / scale
4. uncertainty.

Data Analytics is the use of statistics with visualization to support decision making.

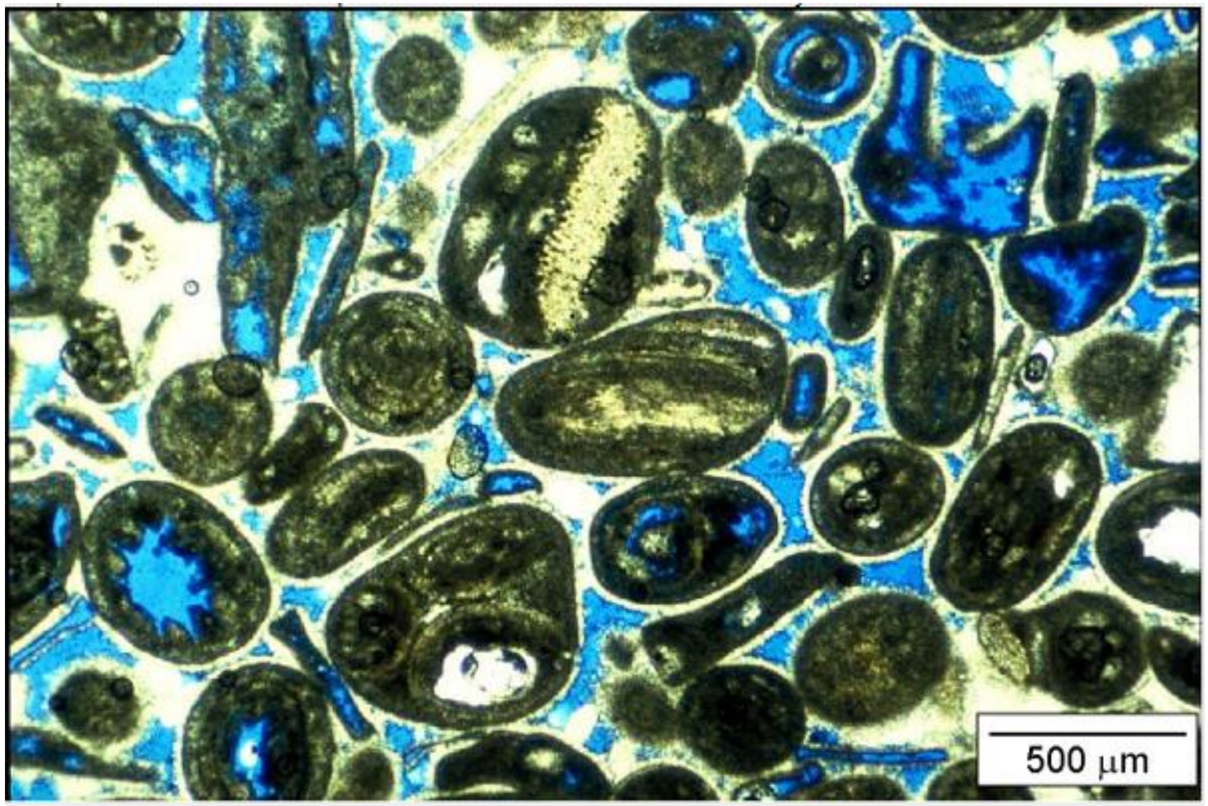
Big Data Analytics is the process of examining large and varied data sets (big data) to discover patterns and make decisions.



Proposed Venn diagram for spatial big data analytics.

Variables / Features

- **Variable / Feature**: any property measured / observed in a study
  - e.g., porosity, permeability, mineral concentrations, saturations, contaminant concentration
  - in data mining / machine learning this is known as a **feature**
  - measure often requires **significant analysis, interpretation,** etc.



Total Porosity
all blue area

Effective Porosity
all connected blue area

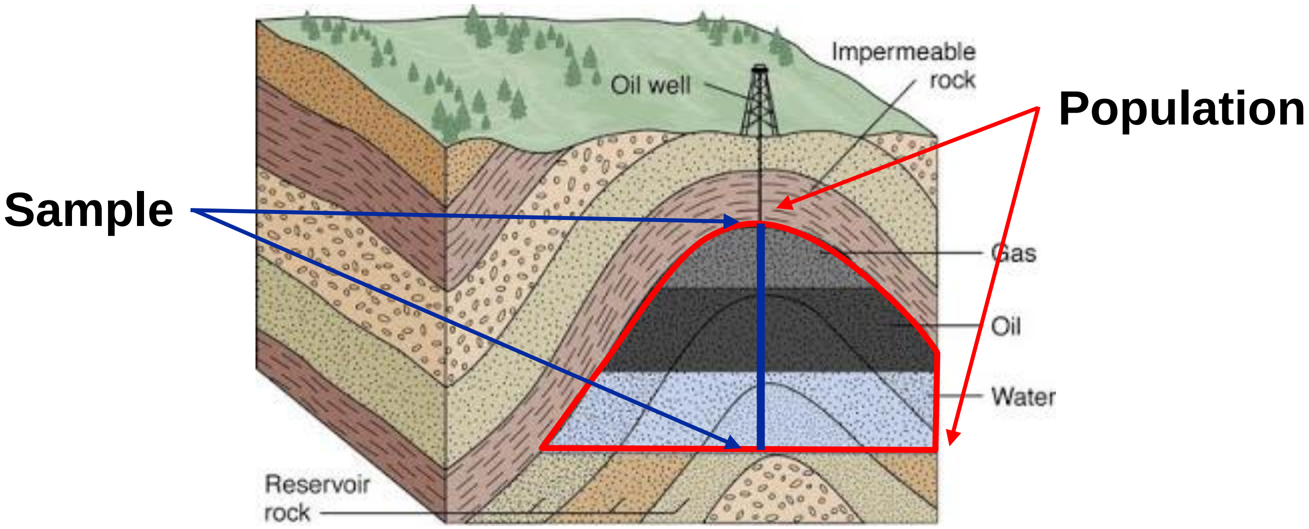
Carbonate thin section from BEG, UT Austin from course by F. Jerry Lucia.

Image from http://www.beg.utexas.edu/lmod/_IOL-CM07/old-4.29.03/cm07-step05.htm

Population and Sample

- **Population**: Exhaustive, finite list of property of interest over area of interest. Generally, the entire population is not accessible.
  - e.g., exhaustive set of porosity at each location within a reservoir

- **Sample**: The set of values, locations that have been measured
  - e.g., porosity data from well-logs within a reservoir



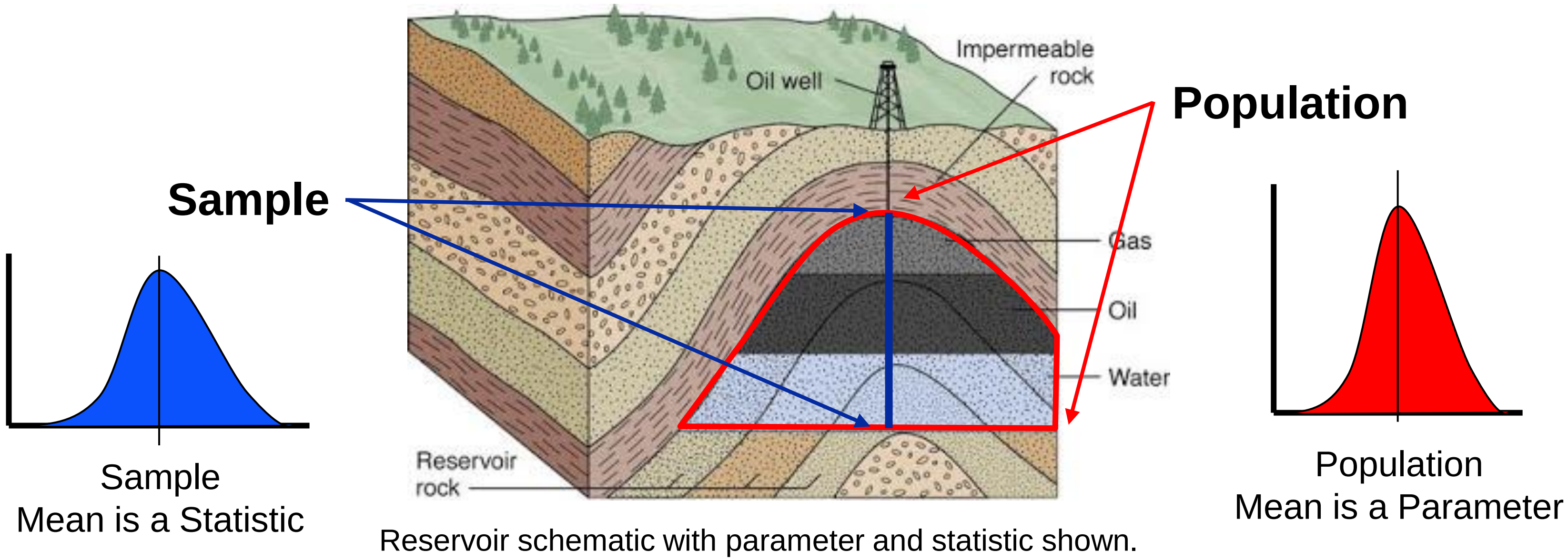
Reservoir schematic with population and sample shown.

Modified image from https://energyeducation.ca/encyclopedia/Oil_and_gas_reservoir

Parameter and Statistic

- **Parameters**: summary measure of a population
  – e.g., population mean, population standard deviation, we rarely have access to this
  – **model parameters** are different, and we will cover this later.

- **Statistics**: summary measure of a sample
  – e.g., sample mean, sample standard deviation, we use statistics as estimates of the parameters



Reservoir schematic with parameter and statistic shown.

Modified Image from https://energyeducation.ca/encyclopedia/Oil_and_gas_reservoir

Predictor and Response Features

Given a Model with our Features

$$Y = f(X_1, \ldots, X_m) + \epsilon$$

e.g., a statistical / machine learning model
note ϵ is a random error term

Predictors / Independent Variables / Features

- input variables, $X_1, \ldots, X_m$

Response(s) / Dependent Variables / Features

- output variable, Y

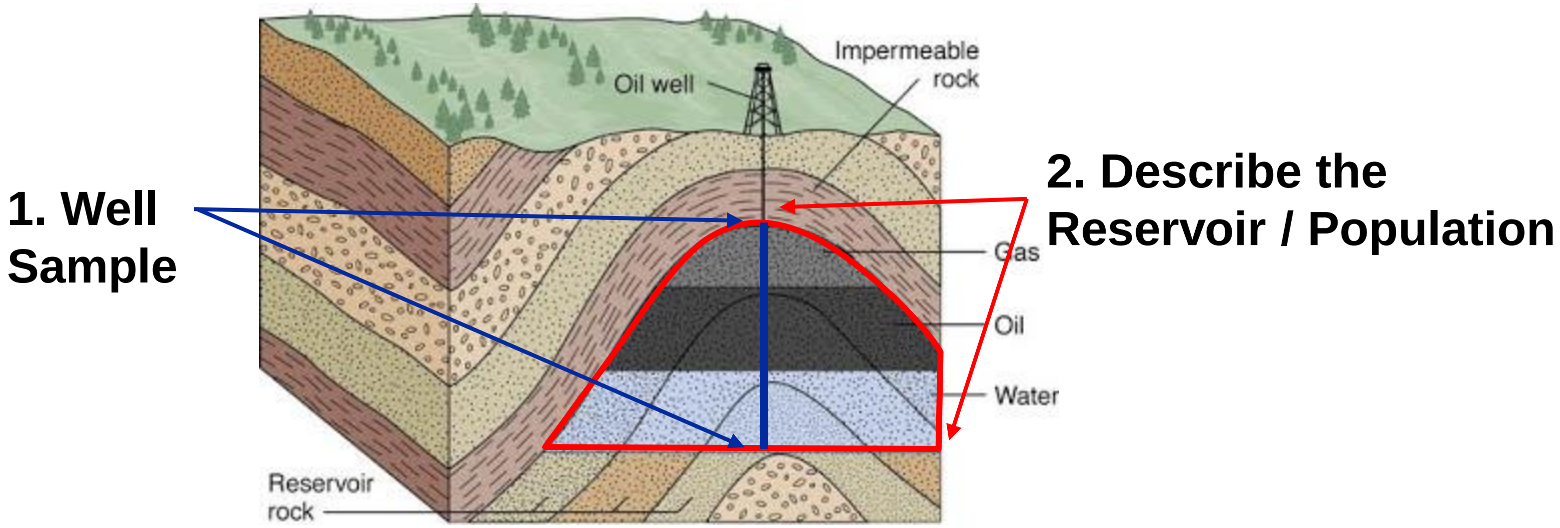
Data Analytics and Machine Learning is All About Building Data-derived Models for 2 Purposes:

1. Inference
2. Prediction

Inference

Inferential Statistics

- Given a random sample from a population, describe the population
- Given the well(s) samples, describe the reservoir



Reservoir schematic with inference problem, given well sample, describe the reservoir, population.

Modified Image from https://energyeducation.ca/encyclopedia/Oil_and_gas_reservoir

Prediction

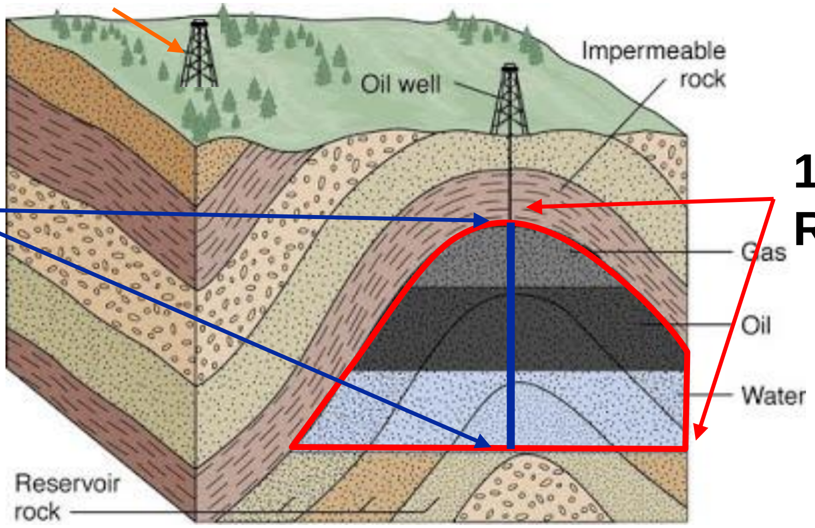
Predictive Statistics

- Predict the samples given assumptions about the population
- Given our model of the reservoir, predict the next well (pre-drill assessment) sample, e.g., porosity, permeability, production rate etc.

2. Pre-Drill Prediction for Proposed Well

Well Sample

1. Given a Model of Reservoir / Population



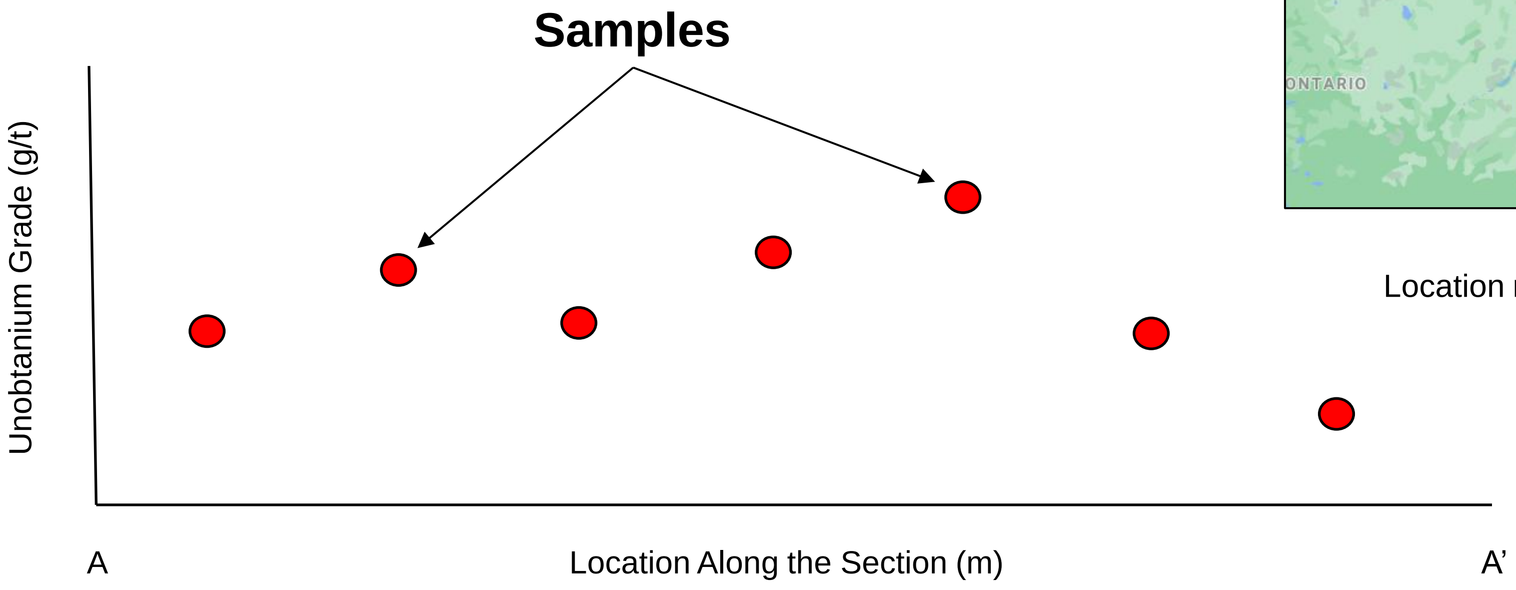
Reservoir schematic with prediction problem, given a model of the reservoir population, make a prediction of the next sample.

Modified Image from https://energyeducation.ca/encyclopedia/Oil_and_gas_reservoir

Deterministic and Statistical Modeling

Let’s build a model, here’s some unobtainium data.

Why do we need a model?



Location map with section $A - A'$.

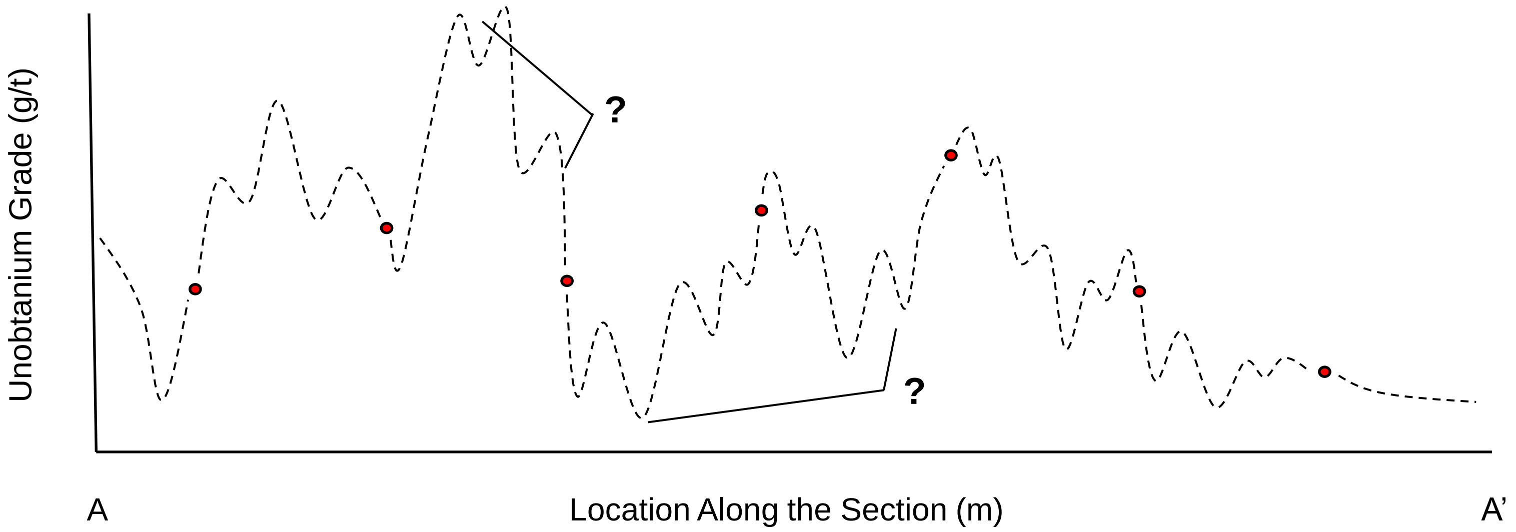
Image is a screen shot from Google maps.

Deterministic and Statistical Modeling

Let’s build a model, here’s some unobtainium data.

Why do we need a model?

We need to understand the population to make decisions, but we only sampled 1/ trillionth.
Where is the unobtanium concentrated, how can we extract it economically? etc.



Sample → Inference → Model of Population → Prediction → Decision

Deterministic and Statistical Modeling

Deterministic Model: assumes the system or process that is completely predictable

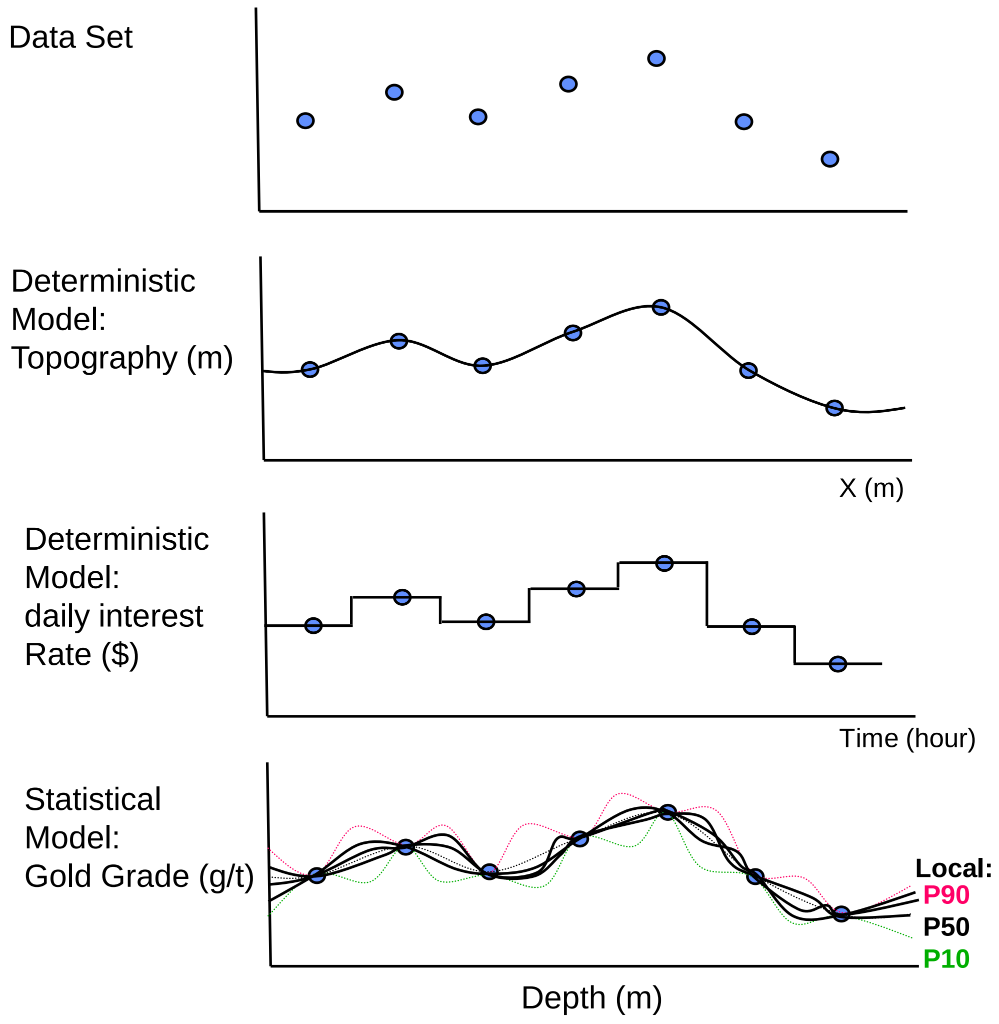
- engineering and geoscience physics and expert judgement

Stochastic Model: system or process that is uncertain, multiple models constrained by statistics

- data-driven, machine learning

Hybrid Model: system or process that includes a combination of both deterministic and stochastic modeling

- most geostatistical models are hybrid models
- e.g., additive deterministic trend models and stochastic residual

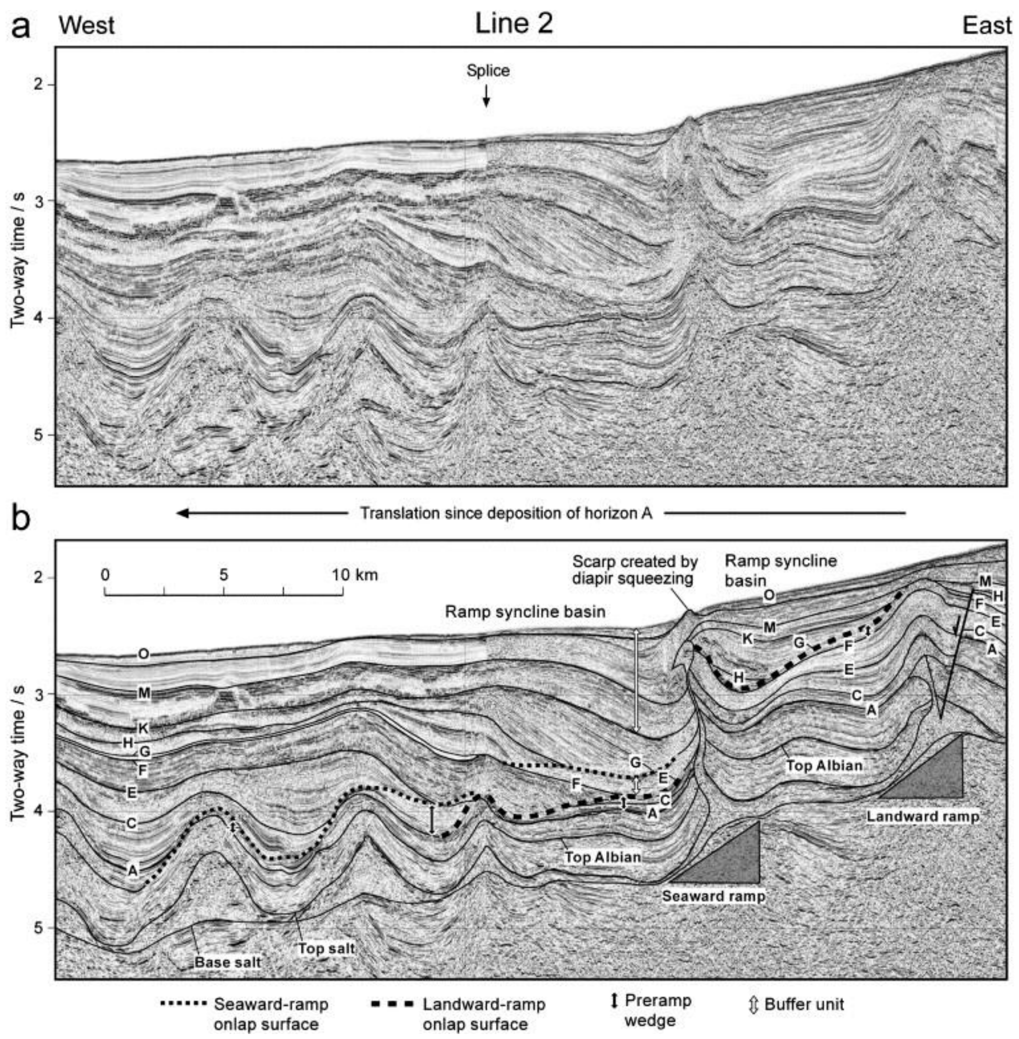


Data and three possible models.

Deterministic Modeling

Deterministic Modeling:

- Examples:
  - ‘by-hand’ interpretation of 3D seismic
  - physics-based flow modeling
- Advantages:
  - integration of physics and expert knowledge
  - integration of various information sources
- Disadvantages:
  - often quite time consuming
  - often no assessment of uncertainty, focus on building one model



Kwanza Basin, Angola stratigraphic model resulting from translation and salt diapirs.

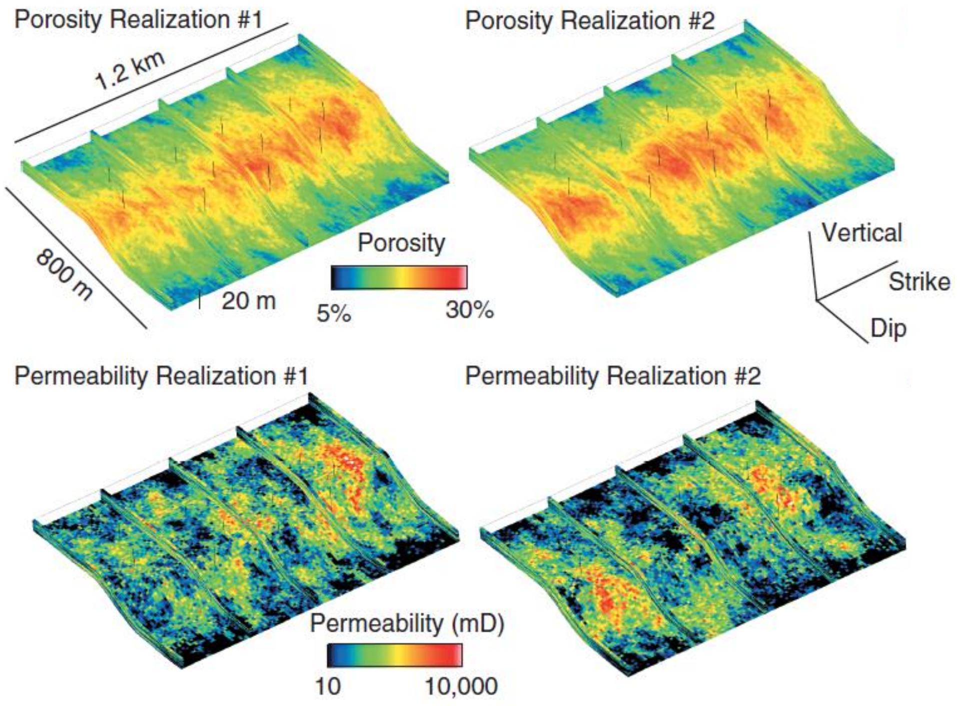
Statistical Modeling

Stochastic Modeling:

- Examples
  - Data-drive, geostatistics, machine learning

- Advantages:
  - speed
  - uncertainty assessment
  - report significance, confidence / prediction intervals
  - honor many types of data
  - data-driven approaches

- Disadvantages:
  - limited physics used
  - statistical model assumptions / simplification



Geostatistical models of reservoir porosity (above) and permeability (below) over a reservoir.

Image from Pyrcz and Deutsch (2014), ‘Geostatistical Reservoir Modeling’.

Data-driven Models

Estimation:

- is process of obtaining the single best value of a subsurface feature at an unsampled location.
- local accuracy takes precedence over global spatial variability.
- too 'smooth', not appropriate for forecasting
- this is most of predictive machine learning!

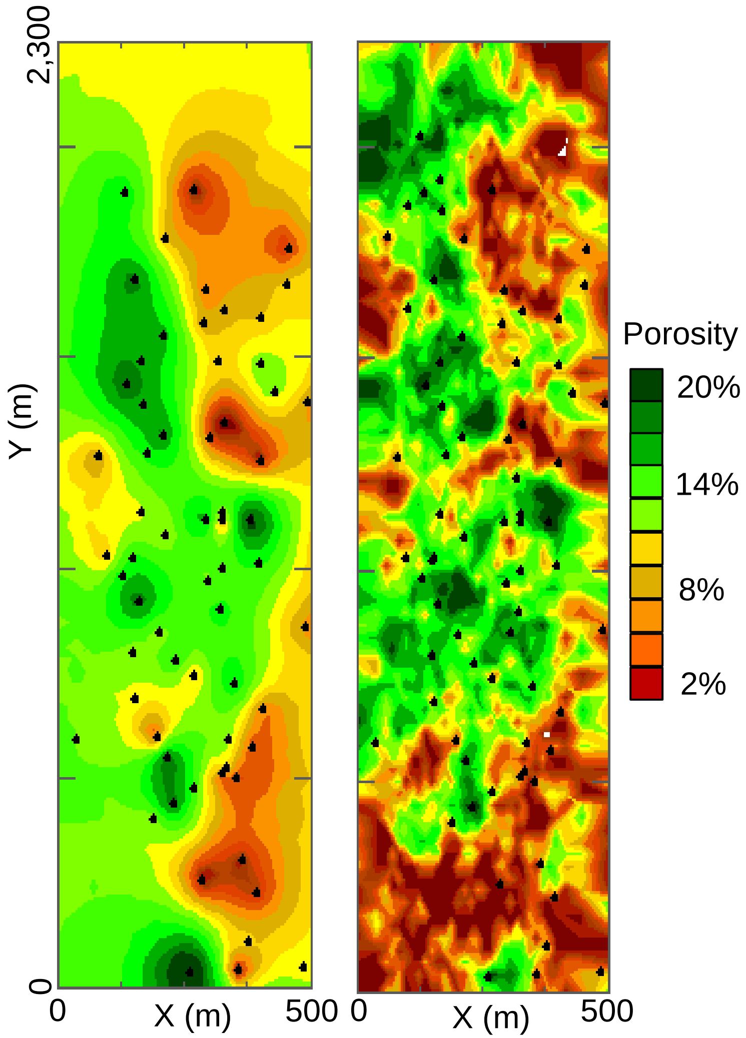
Simulation:

- is process of obtaining one or more good values of a reservoir property at an unsampled location.
- global accuracy, matches the global statistics
- simulation methods tend to produce more realistic feature spatial, univariate distributions.

Why would we prefer simulation?

- we need to reproduce the distributions of features of interest, the extreme values matter
- we need realistic models for flow simulation

Note: We cover simulation with Geostatistics.

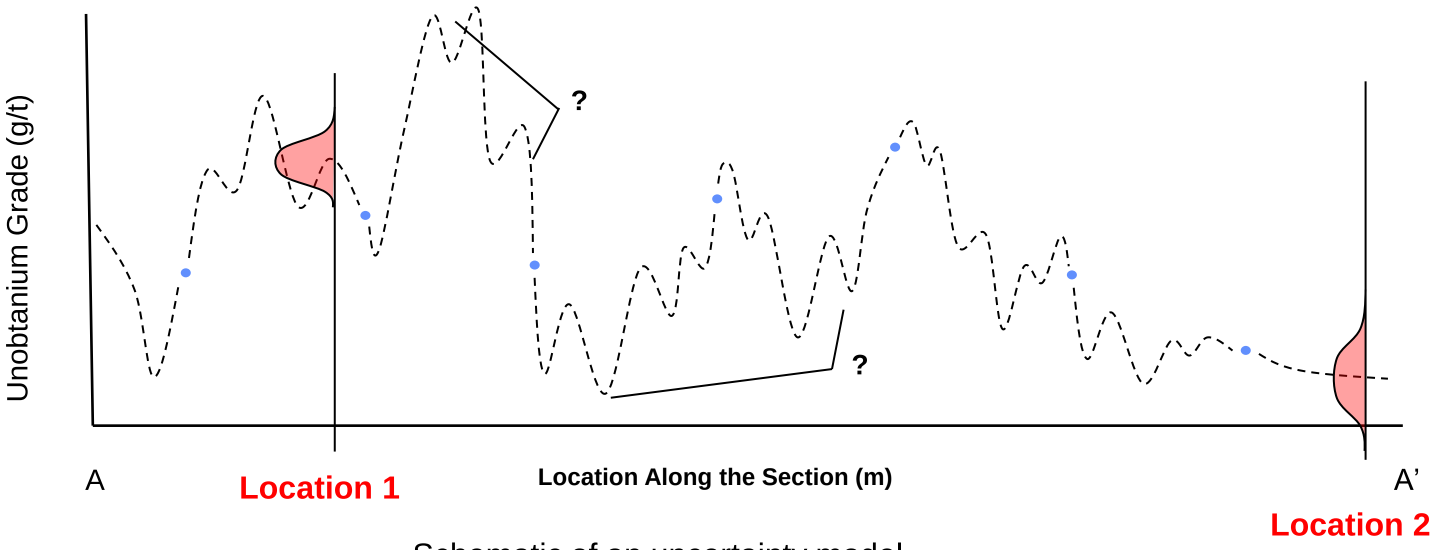


Estimated map (left) and simulated map (right).

Uncertainty Modeling

Uncertainty Modeling:

- the range of possible values for a feature at a location
- limitation in our precision of our measures or models
- uncertainty is a model, there is no objective uncertainty
- uncertainty is caused by our ignorance
- sparse sampling, measurement error and bias, and heterogeneity increase uncertainty



Schematic of an uncertainty model.

All models are wrong, but some are useful', George Box (1976)

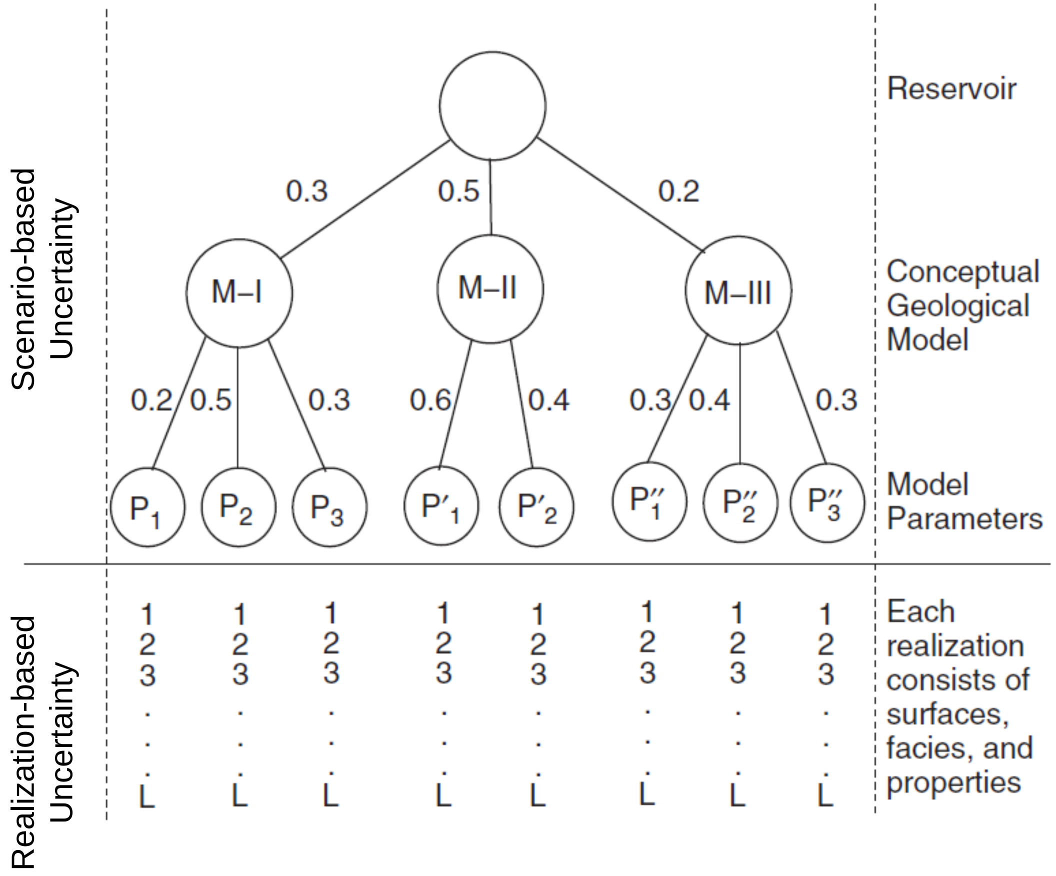
Uncertainty Modeling

Uncertainty Modeling:

Generate multiple simulated models that jointly represent the ‘**space of uncertainty’**.

- **Scenarios**
  - change the input parameters or other modeling choices
  - model parameters and choices uncertainty
  - e.g., use multiple possible average porosity means, low, mid and high over the models

- **Realizations**
  - hold input parameters constant and change random number seed
  - spatial uncertainty
  - e.g., hold the porosity mean constant and observe changes in porosity away from the wells over multiple realizations



Uncertainty modeling framework, scenarios and realizations.

Image from Pyrcz and Deutsch (2014), ‘Geostatistical Reservoir Modeling’.

Reservoir Modeling

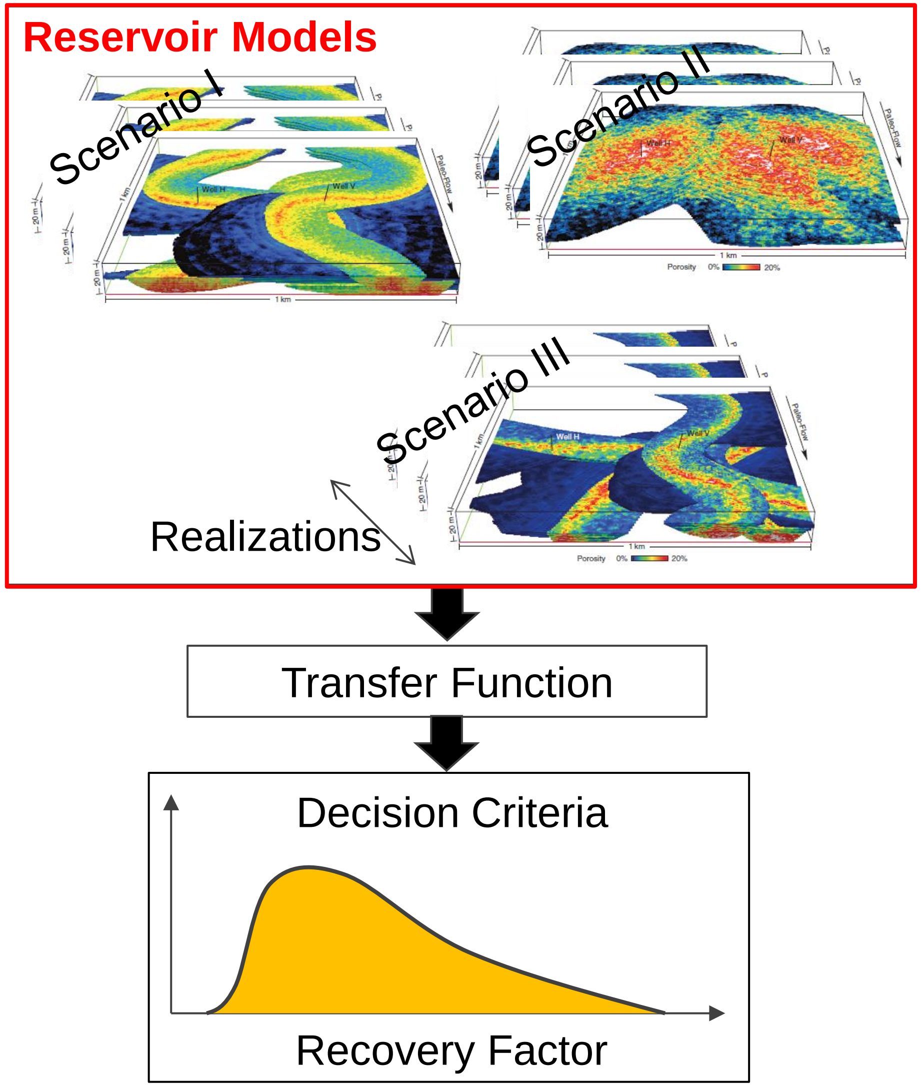
Reservoir Modeling

1. Integrate all available information to build multiple scenarios and realizations to sample the uncertainty space

2. Apply all the models to the transfer function to map to a decision criteria

3. Assemble the distribution of the decision criteria

4. Make the optimum decision accounting for this uncertainty model



The standard reservoir modeling workflow.

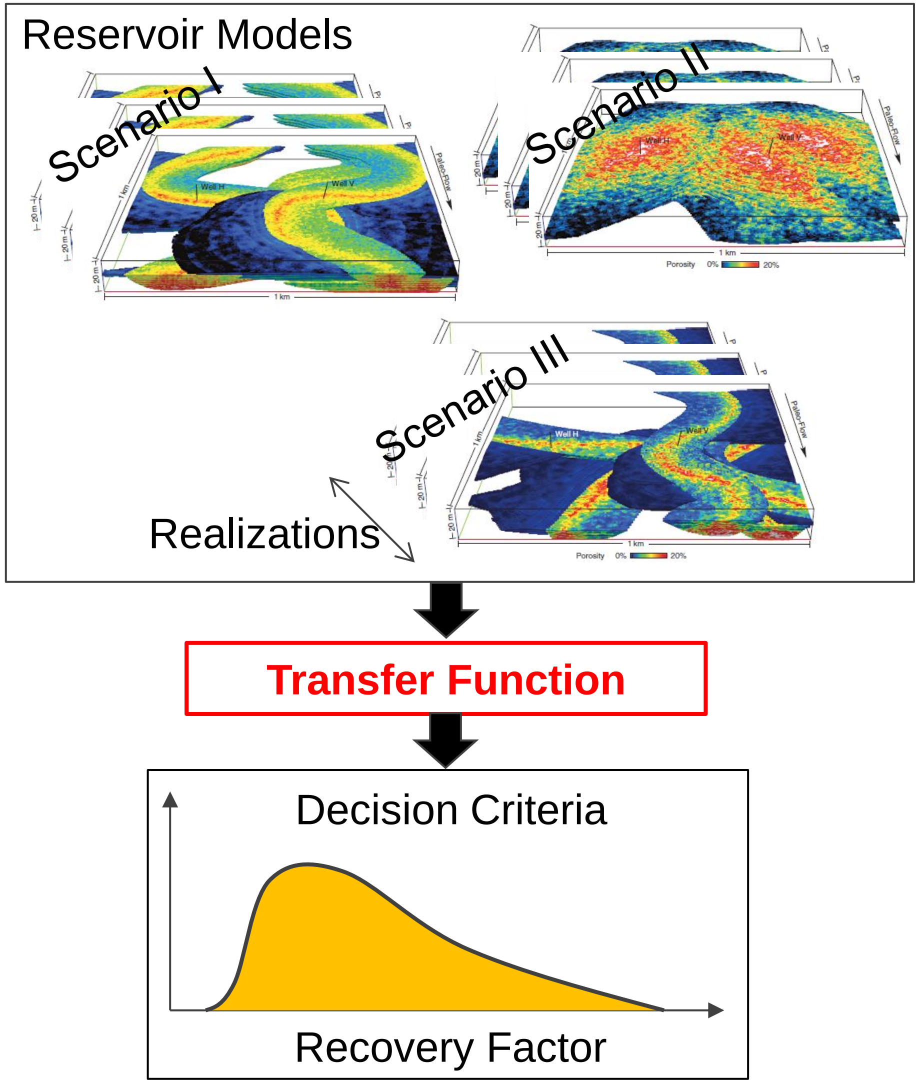
Transfer Function

Transfer Function

Calculation applied to the model (deterministic or statistical) to calculate a decision criterion

Examples:

- transport and bioattenuation of a soil contaminant
- volumetric calculation for oil-in-place
- heterogeneity metric for estimation of recovery factor
- flow simulation for production forecast
- mine plan and rock homogenization for mineral resources



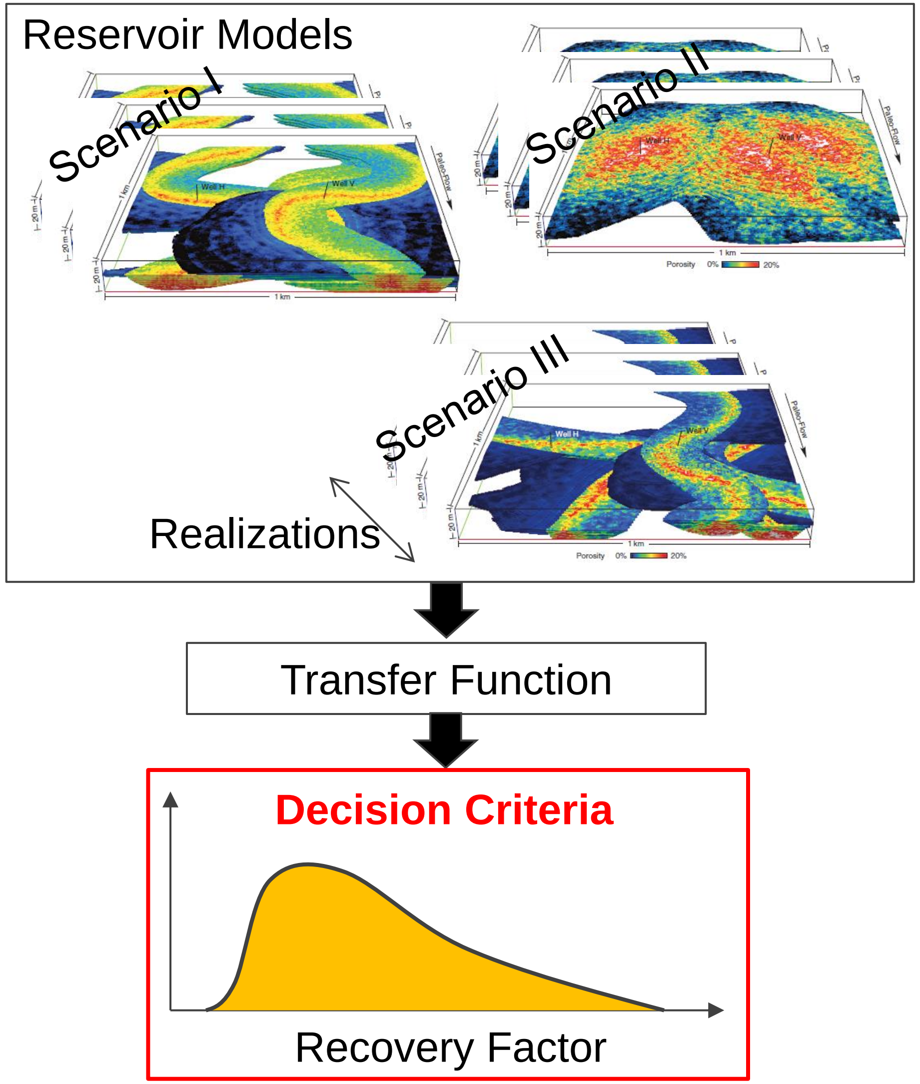
The standard reservoir modeling workflow.

Transfer Function

Decision Criteria

A feature to supports decision making and should be as close to dollars as possible.

Examples:
- contaminant recovery rate
- oil-in-place
- Lorenz coefficient
- recovery factor or production rate
- recovered mineral grade and tonnage



The standard reservoir modeling workflow.

Safe-Stats
Using Statistics on the Job

Hadley Wickham, Chief Scientist at RStudio, known for development of open-source statistical packages for R to make statistics accessible and fun (http://hadley.nz/).

Read Hadley Wickham's paper:

Teaching Safe-Stats, Not Statistical Abstinence
(https://nhorton.people.amherst.edu/mererenovation/17_Wickham.PDF)

- **Teaching:** We need to rethink statistics curriculum – we risk becoming irrelevant!
- **Currently:** Stats tends to be taught as avoid, unless you are an "statistician" or with one
  - Otherwise, you will cause great harm
  - But there are not enough professional statisticians
  - Rather than stigmatize amateur, new tools should be safer to use
- **My Goal:** Teaching practical workflows, with user-friendly tools for wider use of best-practice, robust statistics!

Hadley Wickham photograph from

Reference Only – Not on Test

Image from https://en.wikipedia.org/wiki/Hadley_Wickham

Cognitive Biases

In any modeling there will be choices. We must understand and mitigate our own biases.

Example of Cognitive Biases:

1. **Anchoring Bias**: too much emphasis on first piece of information. Studies have shown that first piece of information could be completely irrelevant!
2. **Availability Heuristic**: overestimate importance of information available to them. “My grandpa smoked 3 packs a day and lived to 100”.
3. **Bandwagon Effect**: probability increases with the number of people holding the belief.
4. **Blind-spot Effect**: fail to see your own cognitive biases.
5. **Choice-supportive Bias**: probability increases after a commitment, decision is made.
6. **Clustering Illusion**: seeing patterns in random events.
7. **Confirmation Bias**: only consider new information that supports current model.
8. **Conservatism Bias**: favor old data to newly collected data.
9. **Recency Bias**: favor the most recently collected data.
10. **Survivorship Bias**: focus on success cases only.

Robust use of statistics / data analytics protects use from bias.

Reference Only – Not on Test

PGE 338 Data Analytics and Geostatistics

Lecture 1: Data Analytics

Lecture outline . . .

- Types of Data

Introduction

General Concepts

Statistics

Probability

Univariate

Bivariate

Spatial Analysis

Machine Learning

Uncertainty Analysis

Data Dimensionality

1D data recorded in a sequence of distance or time

- Well log, gamma ray (shale indicator), neutron density (porosity indictor)
- Injection and production data

2D sampling and models

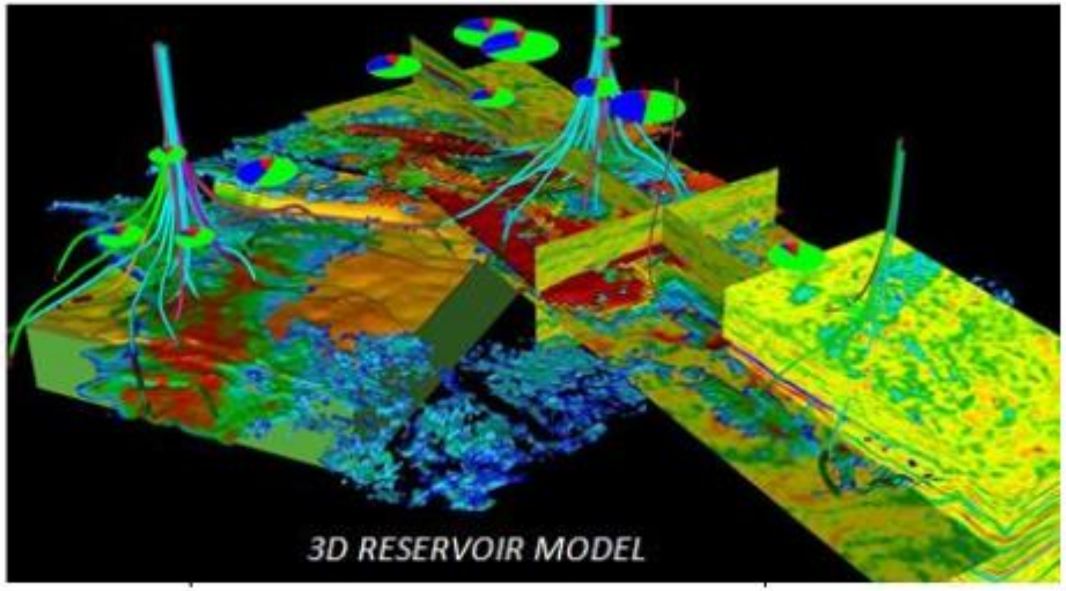
- Geologic maps, seismic line, spatial analysis of thin sections, …

3D sampling and models

- 3D seismic volumes, sets of correlated well logs, reservoir modeling, flow analysis

4D sampling and models

- 4D seismic data with baseline and multiple 3D monitor seismic surveys



3D reservoir model with various data sources

This is ‘**Data Variety**’, a criteria for big data, more later!

Image from http://www.oil-gasportal.com/reservoir-management/integrated-reservoir-modeling.

Types of Data

Measurement Types:

- **Categorical / Nominal (Classes)**
  - Example: Grains in sandstones can belong to categories including quartz, feldspar, without natural ordering
- **Categorical / Ordinal:** categories and the ordering of the categories are important
  - Example: Geologic age, hardness
- **Continuous / Interval:** the intervals between numbers are equal
  - Example: Celsius scale of temperature (arbitrary zero)
- **Continuous / Ratio**: numerical value truly indicate the quantity being measured
  - Example: Kelvin scale of temperature, porosity, permeability, saturation
- **Discrete**: used for any type of binned / grouped data (continuous or categorical)

Categorical Data

↓

- Nominal scale
- Ordinal scale

Continuous Data

↓

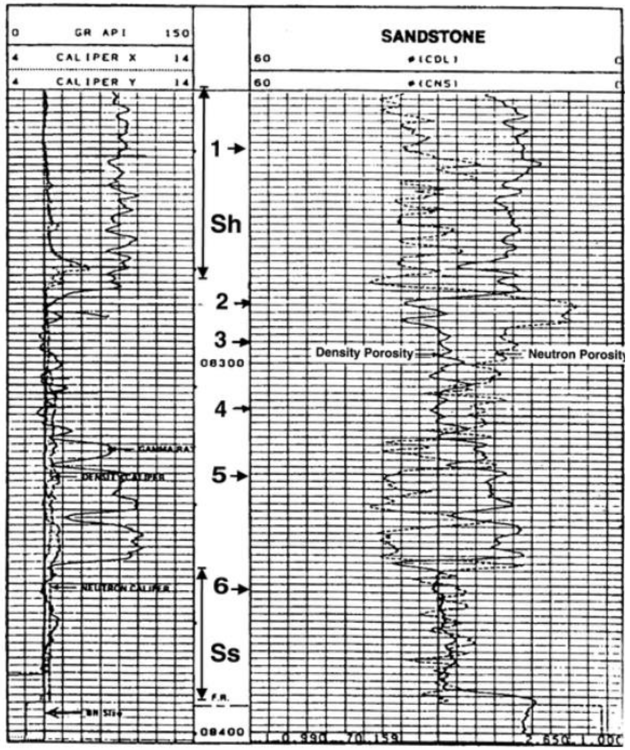
- Interval scale
- Ratio scale

Types of Data

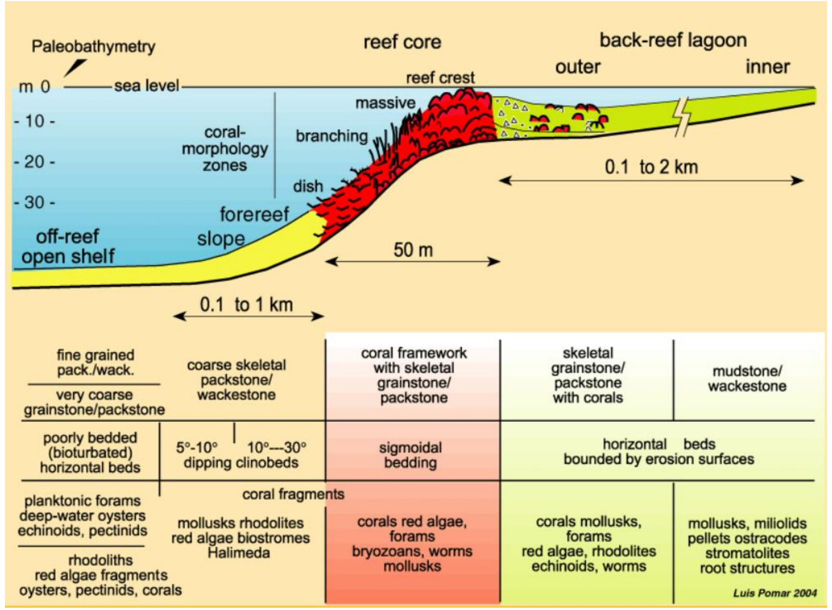
Types of Data:

- **Quantitative Data** – information about quantities that can be written in numbers.
  - Example: age, porosity, saturation

- **Qualitative Data** – information about quantities that you cannot directly measure, require interpretation of measurement
  - Example: rock types, facies



Quantitative data, density and neutron log for measuring porosity from a sandstone unit.



Qualitative data, interpretation of lagoon carbonate depositional system, Pomar (2004).

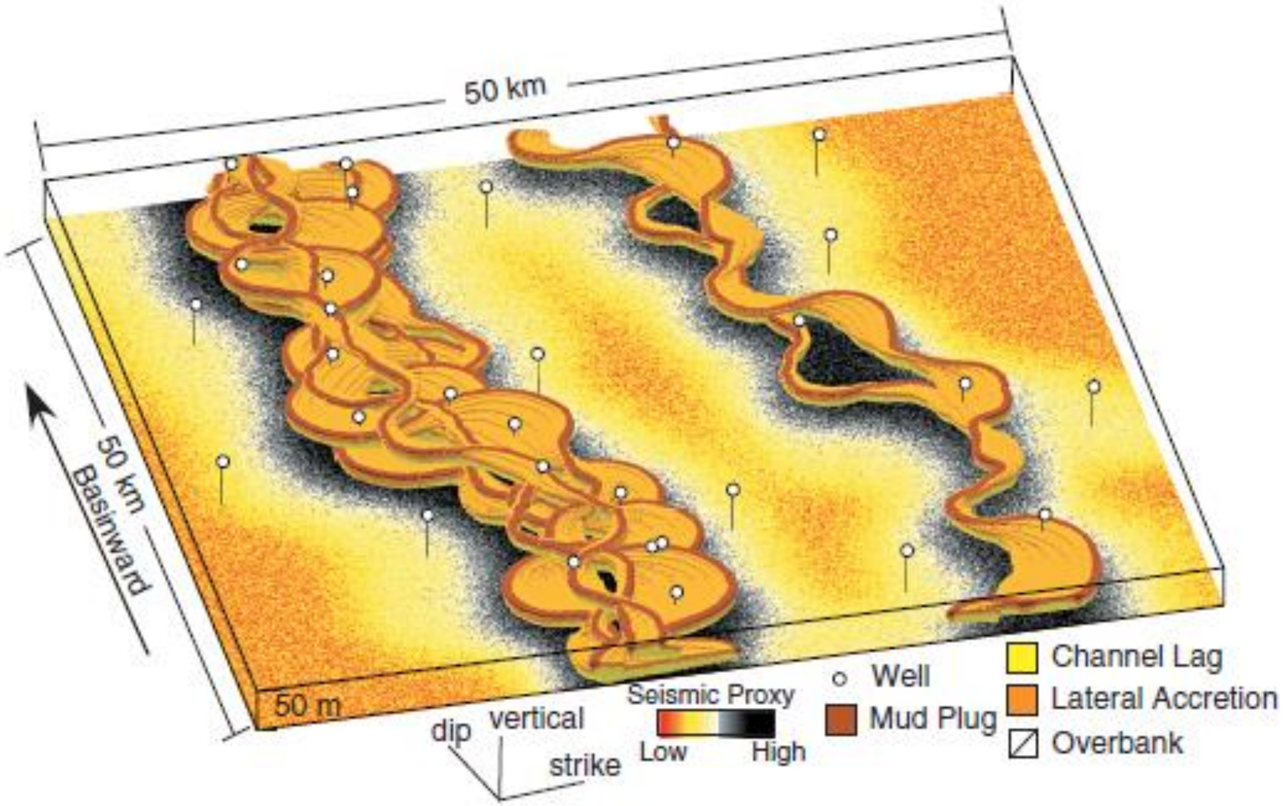
Images from (left) http://wiki.aapg.org/Density-neutron_log_porosity and (right) http://www.sepmstrata.org/page.aspx?&pageid=54&3.

Types of Data

Types of Data:

- **Hard Data** – data that has a high degree of certainty. Usually based on a direct measurement.
  - Example: well core- and log-based porosity, lithofacies

- **Soft Data** – data that provides indirect measures of the property of interest with a significant degree of uncertainty
  - Example: probability density function for local porosity calibrated from acoustic impedance



Fluvial reservoir model, channel sand bodies, hard well data (○) and soft seismic data (seismic proxy map).

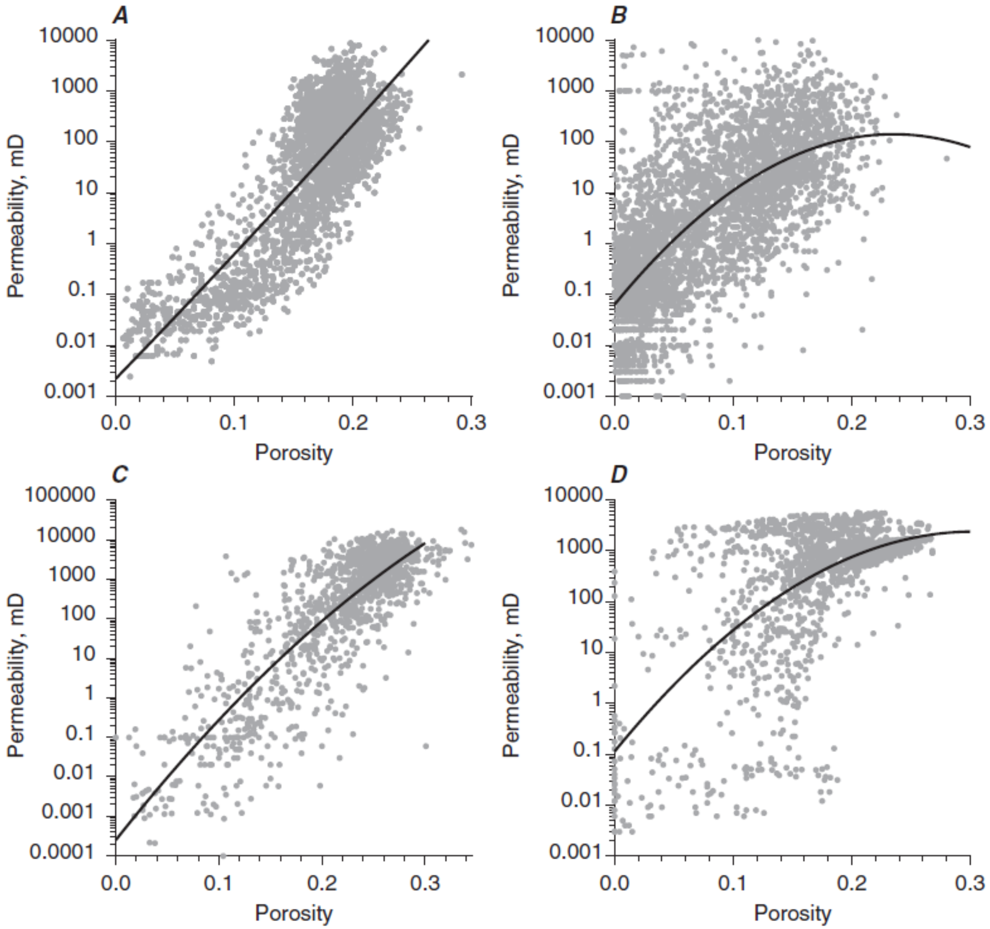
Image from Pyrcz and Deutsch, 2014, Geostatistical Reservoir Modeling.

Types of Data

Types of Data:

- **Primary Data** – the variable of interest. The target for building a model.
  - Example: porosity measures from cores and logs used to build a full 3D porosity model.

- **Secondary Data** – another variable / feature that provides information about the primary data through a relationship / calibration.
  - Example: acoustic impedance to support modeling porosity and porosity to support modeling permeability.



Examples scatterplots illustrating permeability (primary data) relationships with porosity (secondary data).

Image from Pyrcz and Deutsch, 2014, Geostatistical Reservoir Modeling.

Describing Data

Types of Data

- The following discussion is a very cursory treatment.
- Multiple classes would be required to cover each
- We just explain what they are and summarize their coverage, scale and information type

Coverage

- What proportion of the reservoir / population has this data available typically?
- e.g., a couple of meters around wells, everywhere etc.

Scale / Support Size

- What is the scale of the individual data measures?
- e.g., pore scale, cm^3 scale, m^3 scale, reservoir unit scale etc.

Information Type

- What does the data tell us about the subsurface?
- e.g., grain size, fluid type, layering etc.

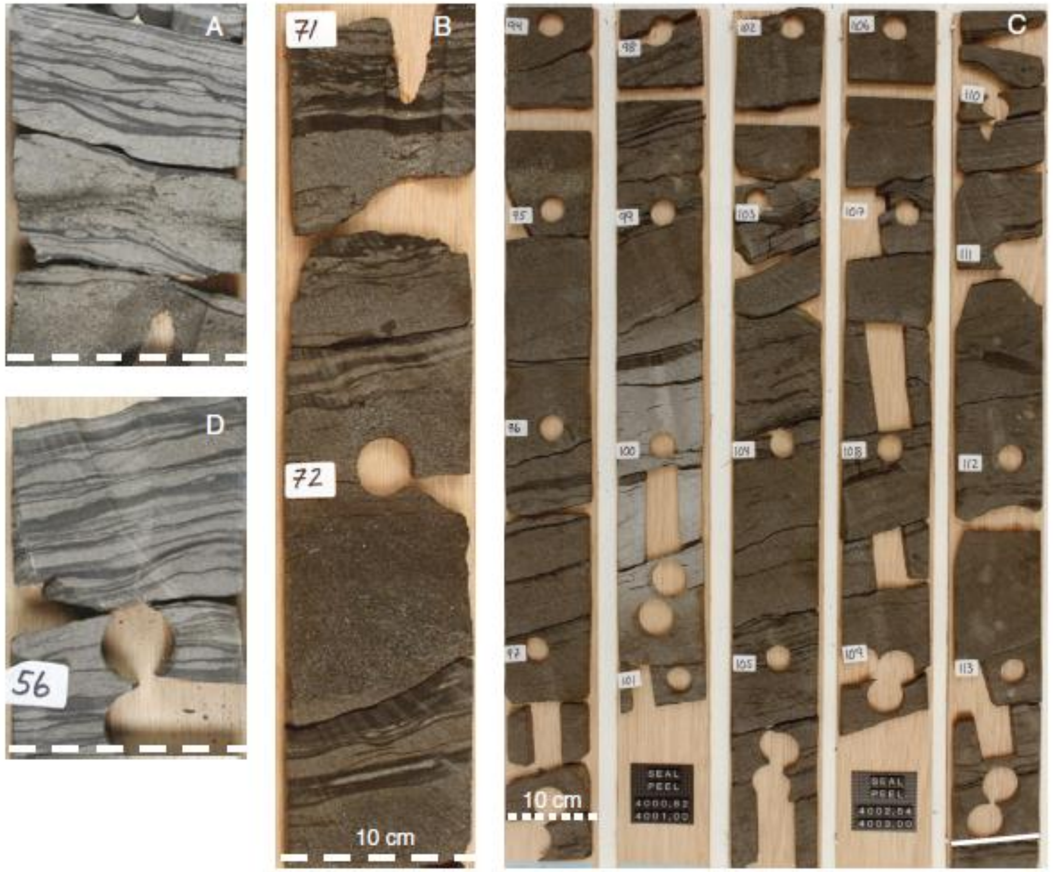
Types of Data

Core Data

- Expensive / Time Consuming to Collect
  - infrequent / incomplete coverage of well
  - at select locations

Petrology, Stratigraphy

- Excellent for quantitative measures such as grain size and porosity
- Interpretations are critical to support the entire reservoir concept / framework for prediction
- Integration of facies, porosity important calibration for all well logs



Sectioned core photographs of the Cook Formation, a shallow marine sandstone reservoir from the North Sea, fluvial / deltaic depositional setting with general progradation upward, Folkestad et al. (2012).

Reference Only – Not on Test

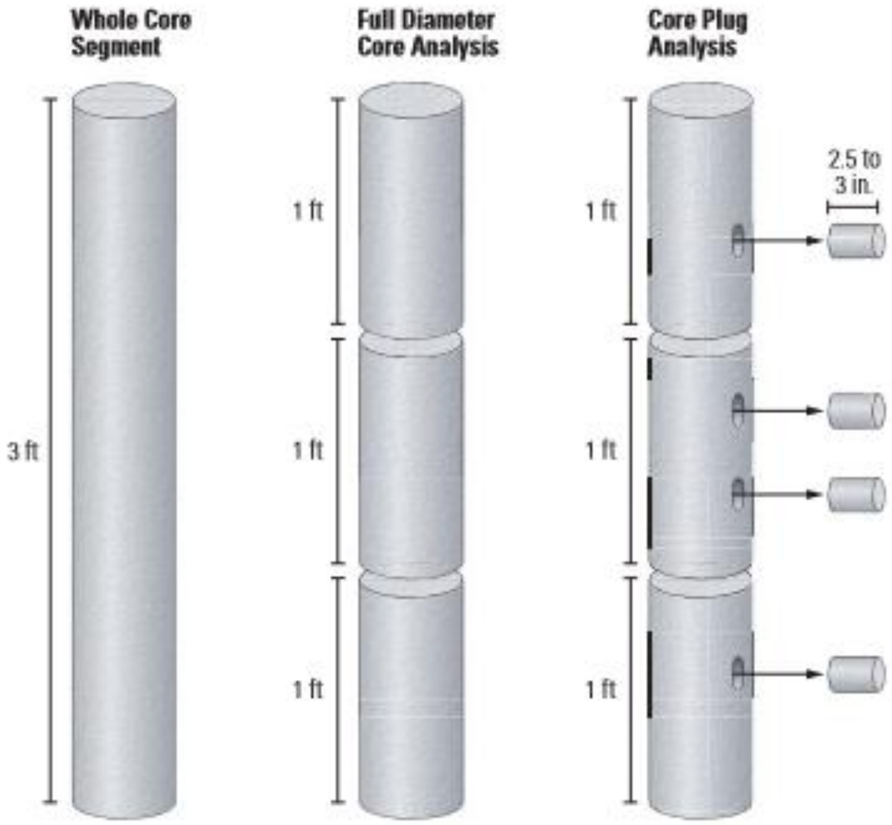
Image from Pyrcz and Deutsch, 2014, Geostatistical Reservoir Modeling.

Types of Data

More on Core Data

Routine Core Analysis
- Porosity, permeability and saturation
- Core gamma logging for calibration to well logs
- Core tomography (CT) scans to assess pore structure

Special Core Analysis
- Electrical measurements for calibration of spontaneous potential (SP) and nuclear magnetic resonance (NMR) well logs.
- Mercury injection for pore throat distribution
- Relative permeability for multiphase flow character



Schematic of whole core, full diameter and core plug extracted for core analysis.

Reference Only – Not on Test

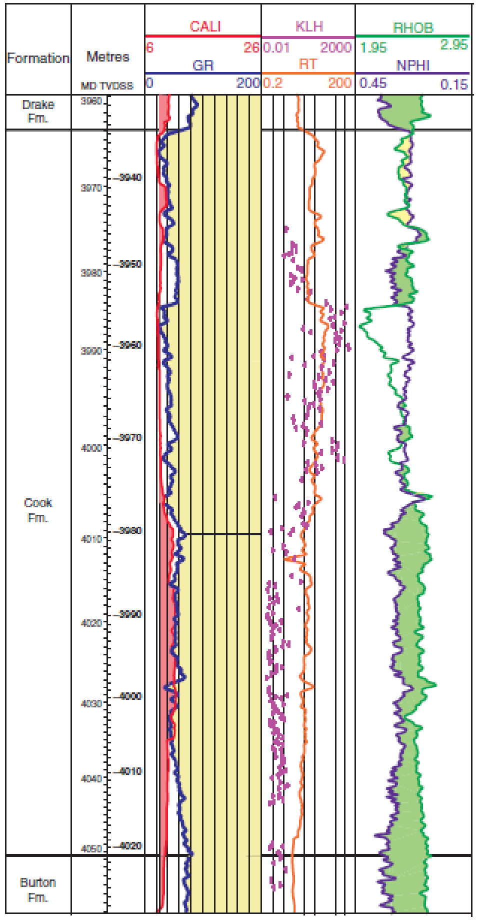
Image from http://www.slb.com/~/media/Files/resources/oilfield_review/ors13/sum13/02_core_truth.pdf

Types of Data

Well Log Data

- Common / Wide Coverage / Suite of Logs

- Examples:
  - Multiple indirect measures of near bore
    - Resistivity and spontaneous (SP)
      - Bed boundaries, fraction of shale
      - Fluids
    - Gamma ray
      - Gamma ray counter to detect organic rich shale
    - Nuclear magnetic resonance
      - Use for medical imaging
      - Respond to presence of hydrogen protons
      - Quantity and type of fluids



Well logs from the Cook Formation, a shallow marine sandstone reservoir from the North Sea, fluvial / deltaic depositional setting with general progradation upward, Folkestad et al. (2012).

Reference Only – Not on Test

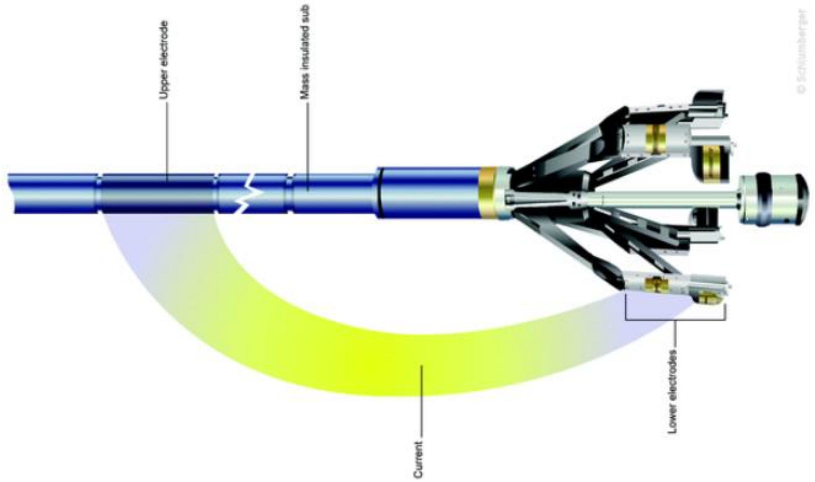
Image from Pyrcz and Deutsch, 2014, Geostatistical Reservoir Modeling.

Types of Data

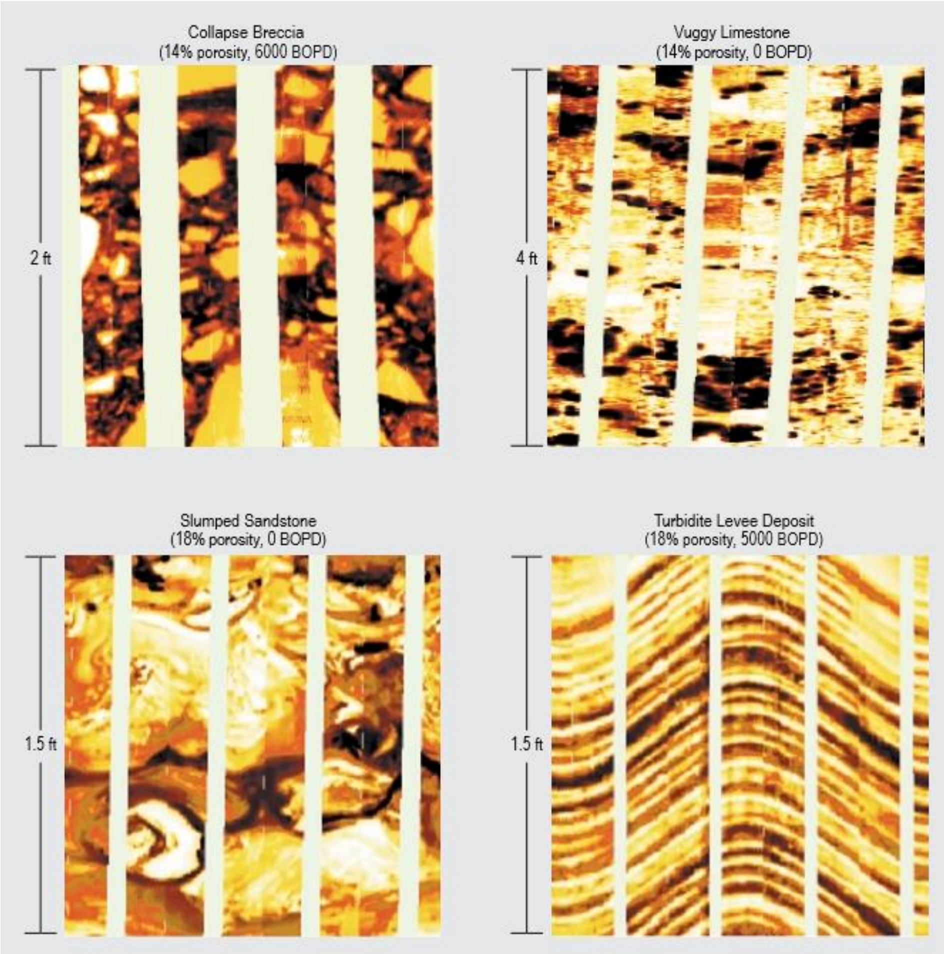
Well Log Data – Image Logs

Centimeter-scale microresistivity images of bore hole walls, e.g., Fullbore formation Microimager (FMI) with:

- 80% bore hole coverage
- 0.2 inch (0.5 cm) resolution vertical and horizontal
- 30 inch (79 cm) depth of investigation
- observe lithology change, bed dips and sedimentary structures.



Fullbore formation microimager logging tool.



FMI Image Log examples with geological interpretation.

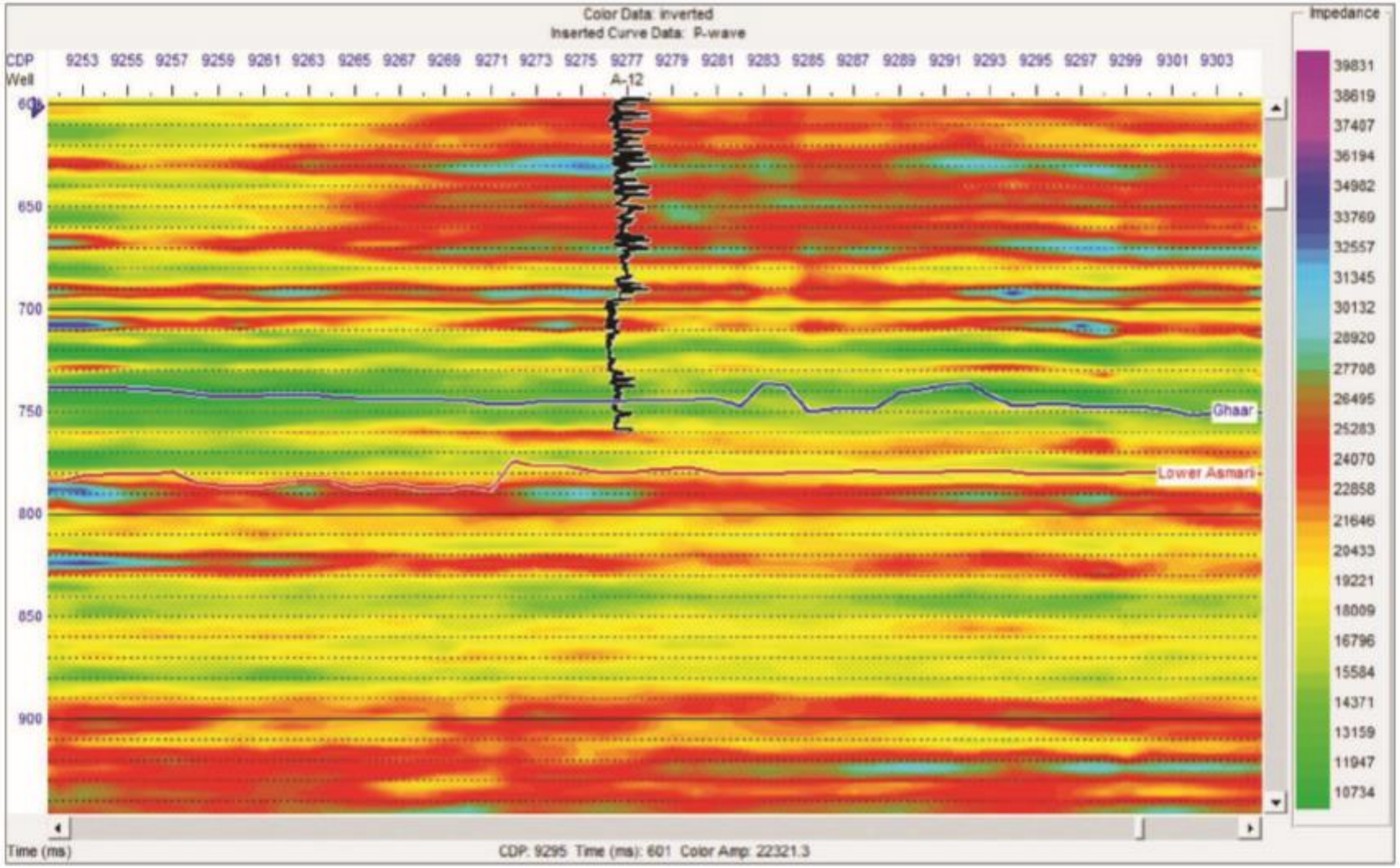
Reference Only – Not on Test

Images from (left) http://petrowiki.org/File:Vol5_Page_0403_Image_0001.png and (right) http://www.slb.com/~/media/Files/ evaluation/brochures/wireline_open_hole/geology/fmi_br.ashx

Types of Data

Seismic Data

- Seismic reflections (amplitude) data inverted to rock properties, e.g., acoustic impedance, consistent with well sonic logs
- Provides framework, soft information on reservoir properties, e.g., porosity and facies.



Acoustic-impedance section from model-based inversion on the seismic section tied to a well. The black well-log curve is the sonic log, Jafari et al. (2017).

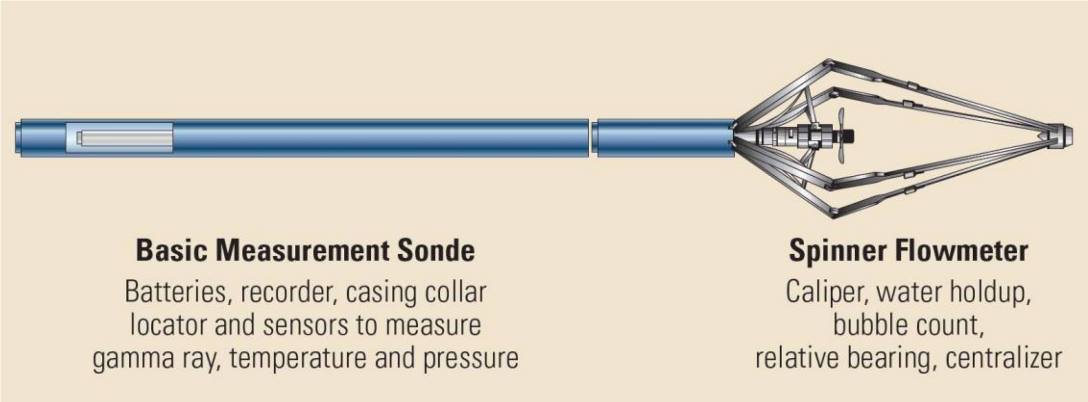
Reference Only – Not on Test

Image from https://library.seg.org/doi/pdf/10.1190/tle36060487.1

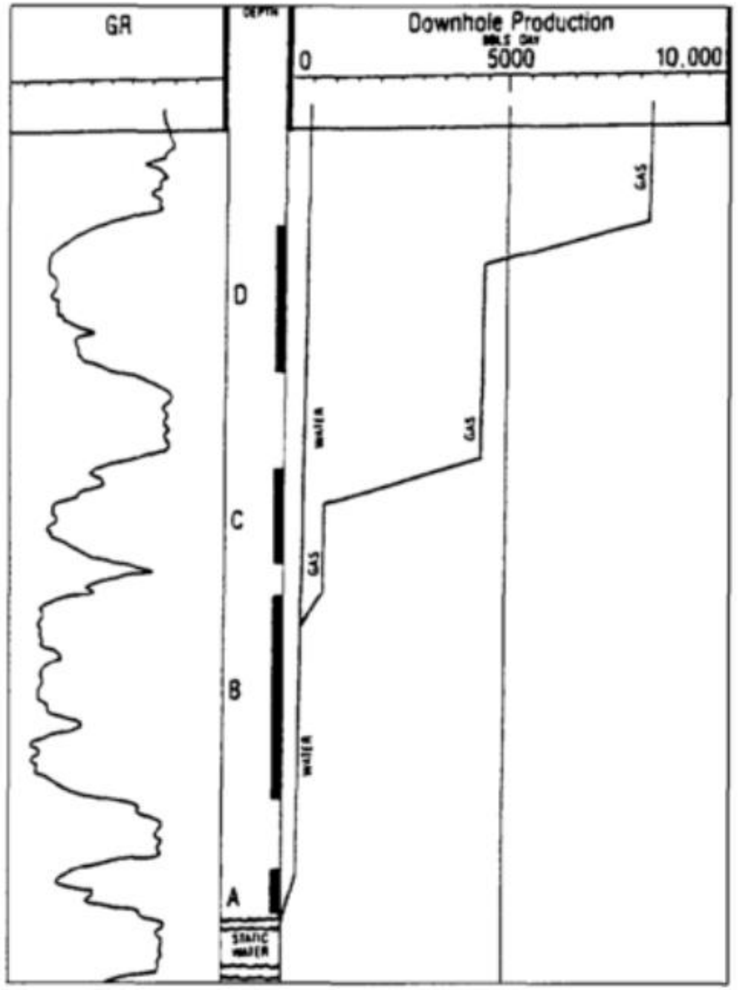
Types of Data

Production Data

- Bottom hole pressure, fluid production (rates, types, temperatures etc.)
- Production may be comingled over multiple producing intervals, unless production logging tool (PLT) results are available
- **Most important ground truth** to be matched with a reservoir model.



Example production logging toolstring.



Production log from a production logging tool (PLT).

Images from (left) https://www.slb.com/resource-library/oilfield-review/defining-series/defining-production-logging and (right) http://wiki.aapg.org/Production_logging.

Reference Only – Not on Test

Data Summary

Should know the top part of this table, red box is for reference only.

Type	Resolution	Coverage	Information Type
Core	$\simeq \infty$	In Well Bore	Lithology, pore and sedimentary structures
Well Log	10 cm	Near Bore	Facies, porosity, minerology
Image Log	5 mm	Near Bore	Sedimentary structures, faults
Seismic	10 m	Exhaustive	Framework, trends, facies, porosity
Production	10–100 m	Drainage Radius	Volumes, connectivity, permeability
Analog			
Mature Fields	10–100 m	≤ Complete	Validation, prior for all
Outcrop *Geomorphology*	$\simeq \infty$ $\simeq \infty$	none none	Concepts, input statistics Concepts
Shallow Seismic	≥ Element	none	Concepts, input statistics
Experimental Stratigraphy	$\simeq \infty$	none	Concepts
Numerical Process	≥ Complex	none	Concepts

A general summary of data types, resolution, coverage and information type.

In Red Box Not on Test / Reference Only

PGE 338 Data Analytics and Geostatistics

Lecture 1: Data Analytics

Lecture outline . . .

- Data Analytics Examples

Introduction

General Concepts

Statistics

Probability

Univariate

Bivariate

Spatial Analysis

Machine Learning

Uncertainty Analysis

Data Analytics Examples

Variogram and Trend-based Uncertainty Away from a Single Well

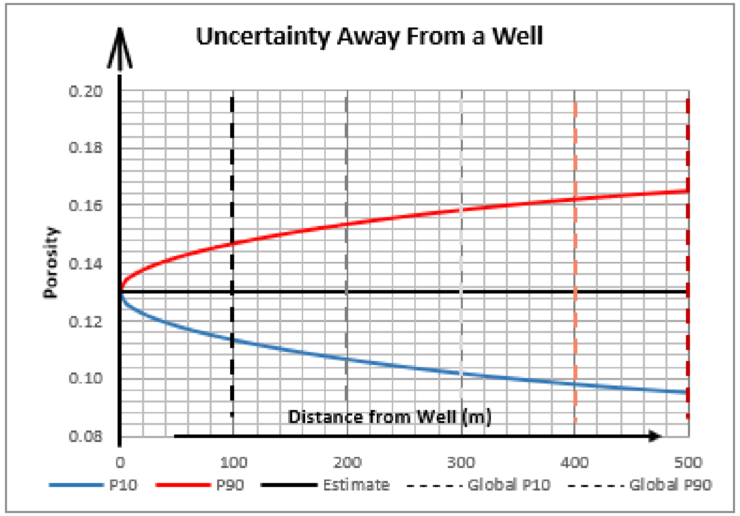
Michael Pyrcz, Geostatistics at Petroleum and Geosystems Engineering, University of Texas at Austin (mpyrcz@austin.utexas.edu)

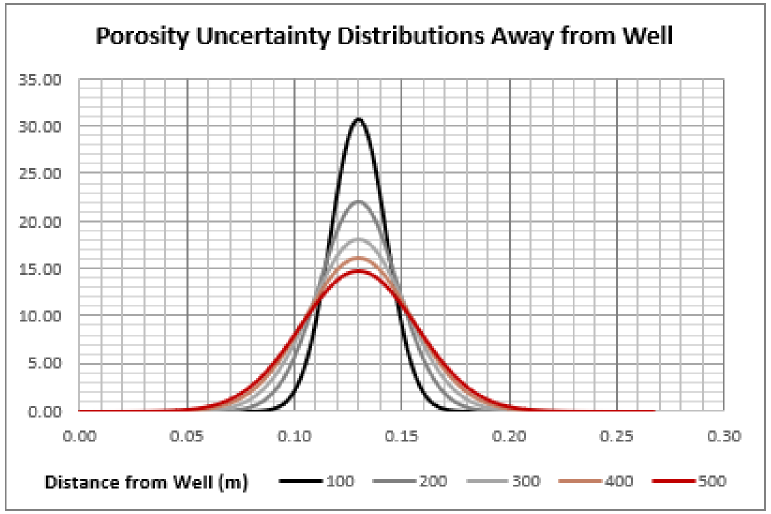
Instructions: set the (1) well porosity value, (2) global porosity variance, (3) trend slope away from the well, and (4) variogram parameterized by the relative nugget effect and sp

	Spatial Model			
Porosity at Well	Well Value	13%		
Total Variance	Global Std./Var.	3%	0.0009	
Change away from well	Trend ΔPor / m	0%	%/m	
Proportion Random	Nugget	0%		
Proportion Correlated	Spherical	100%		
Distance of Correlation	Range	800	m	

Distance	0	5	10	15	20	25	30	35	40	45	50
Estimate	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13
Rel. Var.	0%	1%	2%	3%	4%	5%	6%	7%	7%	8%	9%
St. Dev.	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
P10	0.13	0.13	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12
P90	0.13	0.13	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14
GlobalP10	-0.09	-0.09	-0.09	-0.09	-0.09	-0.09	-0.09	-0.09	-0.09	-0.09	-0.09
GlobalP90	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35

Distance	55	60	65	70	75	80	85	90	95	100
Estimate	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13
Rel. Var.	10%	11%	12%	13%	14%	15%	16%	17%	18%	19%
St. Dev.	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
P10	0.12	0.12	0.12	0.12	0.12	0.12	0.11	0.11	0.11	0.11
P90	0.14	0.14	0.14	0.14	0.14	0.14	0.15	0.15	0.15	0.15
GlobalP10	-0.09	-0.09	-0.09	-0.09	-0.09	-0.09	-0.09	-0.09	-0.09	-0.09
GlobalP90	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35





Excel-based demonstration of spatial uncertainty that we will use later in the spatial estimation unit (Pyrcz et al., 2016).

Reference Only – Not on Test

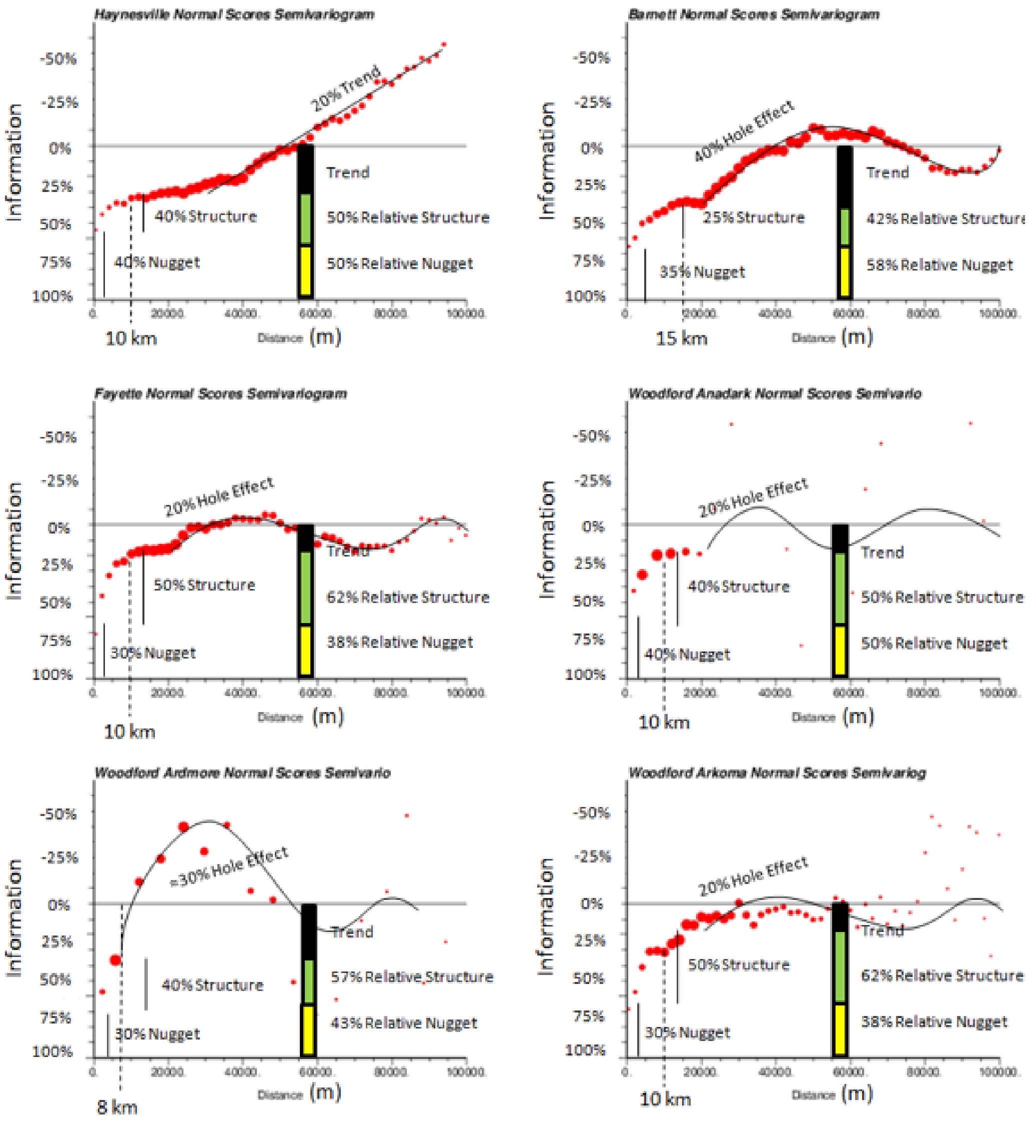
Data Analytics Examples

Quantification / Comparison

Abstraction to a small set of parameters allows us to detect features, learn new insights

We learned:

- the within pad well production variability
- ability to predict offset well production



Quantification of spatial continuity of shale gas production rates (Pyrcz et al., 2016).

Reference Only – Not on Test

Data Analytics Examples

Model of Uncertainty

sparse sampling

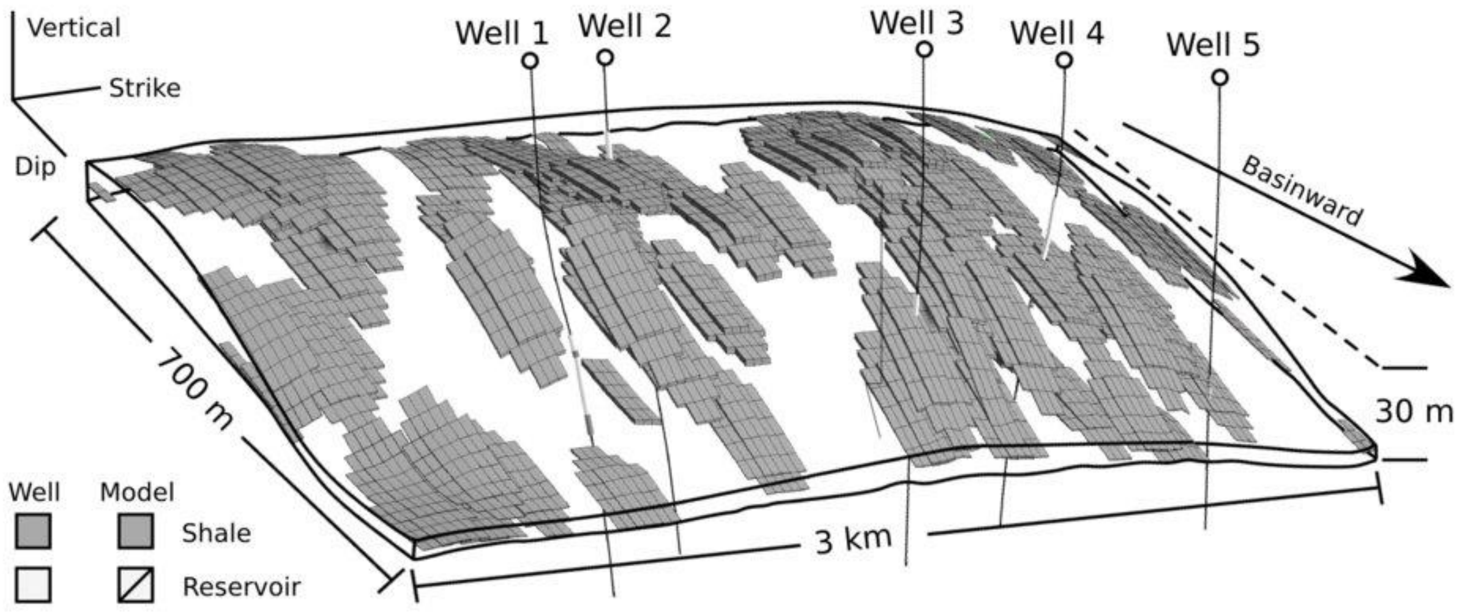
\+ heterogeneity

uncertainty

if we had enough data and understood the phenomenon perfectly there is no uncertainty, no need for a statistical stochastic.

We learned:

- the impact of shale drapes on production for a deltaic reservoir



Can't know exactly where the shales are from 5 wells and given the shale discontinuity. (Pyrcz and Deutsch, 2014)

Reference Only – Not on Test

Image from Pyrcz and Deutsch, 2014, Geostatistical Reservoir Modeling.

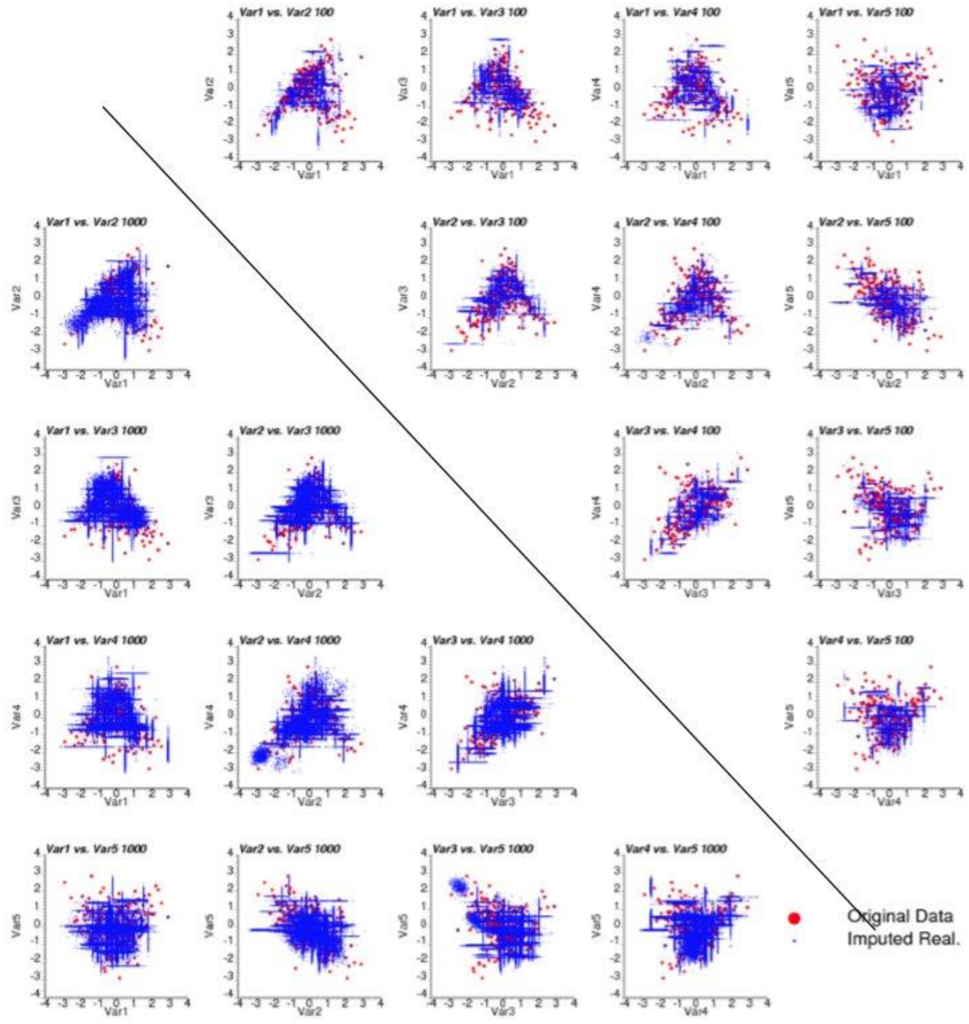
Data Analytics Examples

Too Big / Too Complicated Massive Multivariate

due to the curse of dimensionality, we cannot get enough samples to characterize the system, need to used a statistical multivariate model

We learned:

- workflows to project the data to a lower dimensionality that we can model practically



Multivariate modeling accounting for complicated relationships (Barnett and Deutsch, 2012)

Reference Only – Not on Test

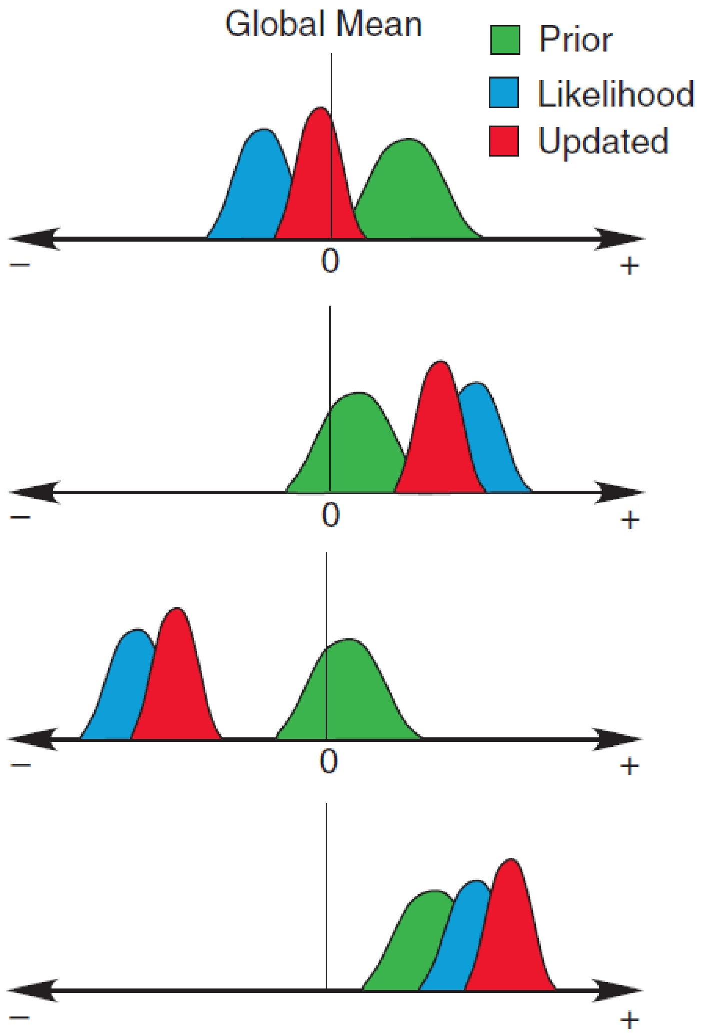
Data Analytics Examples

Combining / Updating with New Information

Bayesian Updating

- need statistical models to integrate multiple information sources

We Learned:
- the value of new information
- the impact of new information on the uncertainty distribution



Bayesian updating under the assumption of Gaussian (Pyrcz and Deutsch, 2012).

Reference Only – Not on Test

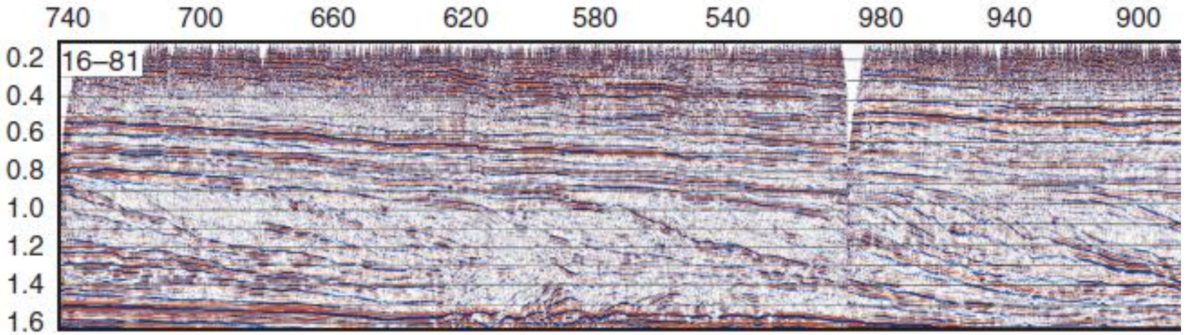
Image from Pyrcz and Deutsch, 2014, Geostatistical Reservoir Modeling.

Data Analytics Examples

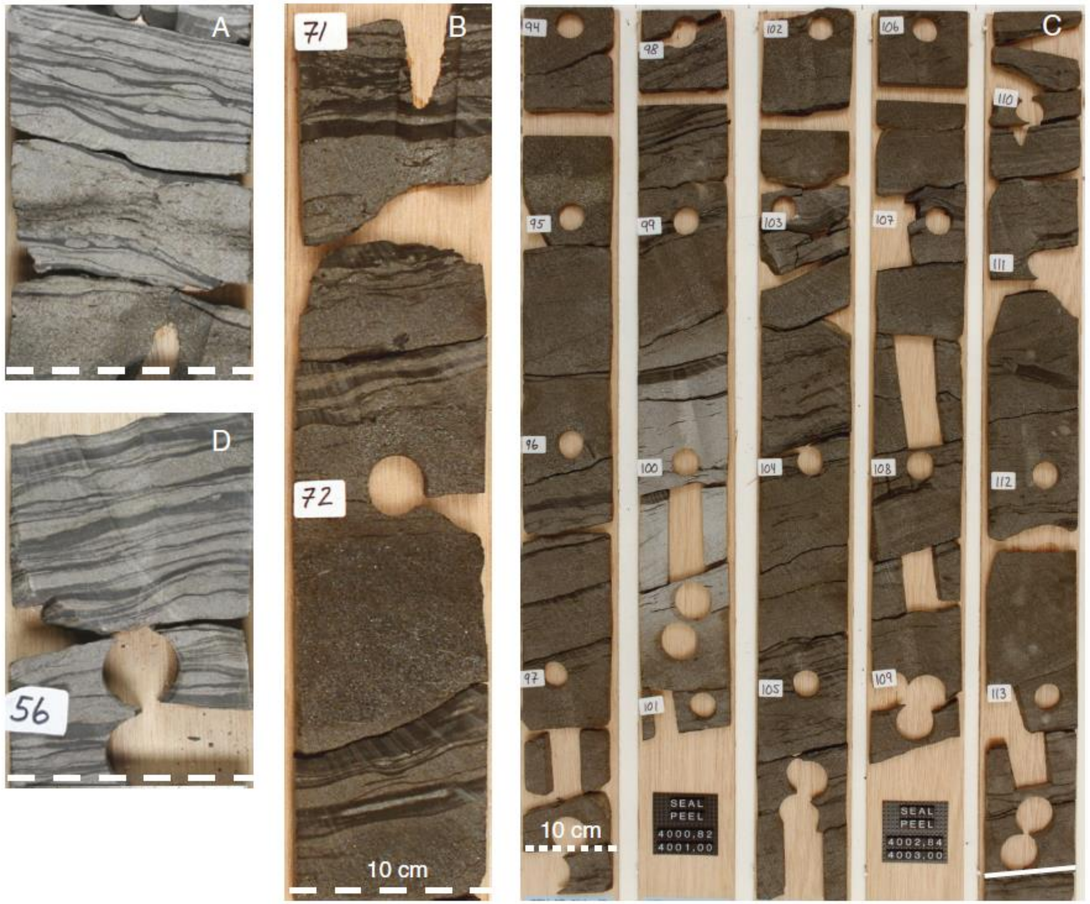
Accounting for Scale

Pores to Production

statistical models for change of support size



A Seismic Line for the Torok Formation Clinoforms of the Lower Cretaceous in the National Petroleum Reserve of Alaska, USA Just South of the Harrison Bay on the Coast of the Beaufort Sea. Seismic lines are shot by the USGS and are available in public domain. Figure provided in high resolution by Professor Chris Kendall, available at the Society for Sedimentological Research Stratigraphy Website.



Sectioned core photographs of the Cook Formation, a shallow marine sandstone reservoir from the North Sea, Folkestad et al. (2012).

Reference Only – Not on Test

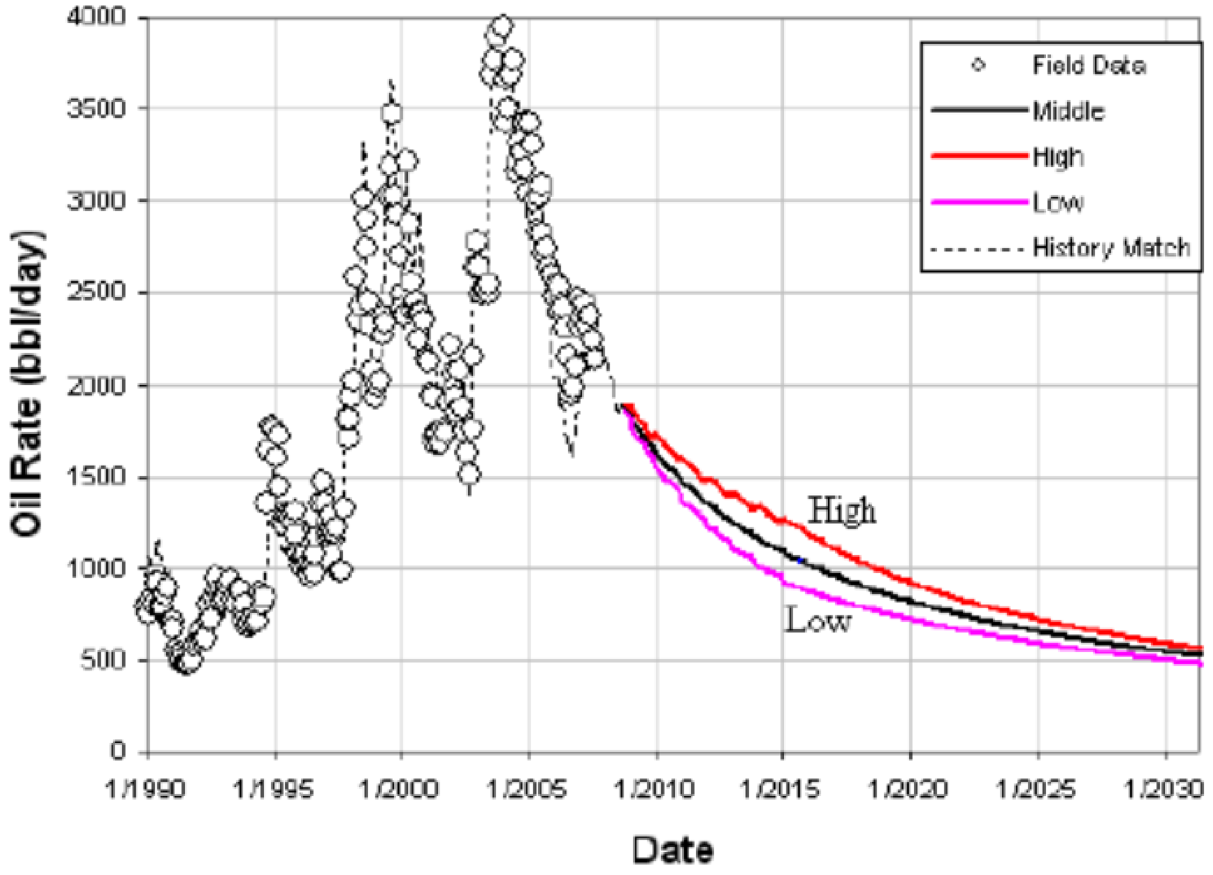
Images from Pyrcz and Deutsch, 2014, Geostatistical Reservoir Modeling.

Data Analytics Examples

Forecasting / Decision Making

Decision Support

use all available information to build forecast uncertainty models to optimize very expensive project decisions



Reservoir forecasting with uncertainty (Yang, 2009).

Reference Only – Not on Test

PGE 338 Data Analytics and Geostatistics

Lecture 1: Data Analytics

Lecture outline . . .

- Data Analytics Concepts
- Sampling
- Types of Data
- Data Analytics Examples

Michael Pyrcz, The University of Texas at Austin