2-Day Course – Spatial Modeling with Geostatistics

Prof. Michael J. Pyrcz, Ph.D., P.Eng.
Associate Professor

Hildebrand Department of Petroleum & Geosystems Engineering
University of Texas at Austin

Bureau of Economic Geology, Jackson School of Geosciences
University of Texas at Austin

“In two days, what a geoscientists needs to know about geostatistics, and workflows to get you started with applying geostatistics to impact your work.”

Spatial Modeling with Geostatistics
Introduction

Lecture outline . . .

- **Basic Concepts**
- **Motivation**
- **Overview**

Class Objectives

Teach theory and practical methods for geostatistics for geologists and data scientists.

Learning Objectives:

- the benefits and uses of geostatistics,
- the common spatial and uncertainty modeling workflows,
- how to better integrate their domain knowledge into the geostatistical model.

Provide knowledge and resources to start geoscientists building their own workflows.

Initial experience with workflow construction with open source

Everything Geoscientists Need to Know About Geostatistics in 2 Days

Topic	Application to Subsurface Modeling	Example
Probability Theory	Methods to infer marginal, conditional and joint probabilities for subsurface events with frequentist and Bayesian frameworks.	Porosity and permeability relationships Facies trends from seismic
Representative Sampling	Workflows for identifying and mitigating spatial sampling bias and accounting for scale.	Well declustering Seismic debiasing
Spatial Data Analysis	Quantification of spatial continuity Interpretation and modeling of spatial continuity	Facies definitions Trend modeling Characterize heterogeneity
Spatial Estimation	Modeling and utilizing spatial continuity make predictions away from wells and at sub-seismic scale.	Spatial estimation Value of well data Shale block uncertainty
Spatial Simulation	Calculation of model parameter uncertainty, stochastic simulation methods	Predrill assessments Stochastic realizations
Uncertainty Management	Post-processing subsurface realizations and decision making in the presence of uncertainty.	NTG uncertainty from wells Summarizing spatial uncertainty
Machine Learning	Working with many variables, modeling complicated phenomenon.	Dimensional reduction and modeling for unconventionals.

Class Strategy

A combination of lecture, demonstration and hands-on

Lectures

- Provide the fundamental theory with a focus on practice

Demonstrations

- Illustrate the use of geostatistics with open source to solve practical problems
- I will use Python / GSLIB, Excel, and R workflows that are available to students

Hands-on

- Experiential learning with paper-based, Excel, R / R Studio and “gstat” package, and Python / GSLIB

Class Deliverables

Other class deliverables:

- Set of lecture notes
- Set of workflows with synthetic datasets
  - R/Python workflows – code to run in RStudio / Jupyter
  - Markdown docs – code with documentation in HTML
- Demonstrations
  - Excel Spreadsheets
  - GSLIB Workflows with Parameter Files
  - Python GSLIB Wrapped Functions, Markdown documented workflow
- More is available and updated at:
  - GitHub/GeostatsGuy – code, other lecture notes
  - Follow on Twitter @GeostatsGuy

Caveats

I can’t teach you R, or Python nor to be a Geostatistician in 2 days.

But, you will get:

- useful fundamental concepts
- exposure to what you can do
- some experience with the tools
- example workflows that you can build on
- resources to continue of your journey

Introductions

Short Introductions:

Name

Affiliation

Role

Expectations from this Class

What a Geo- or Data-scientist Needs to Know about Geostatistics

Let’s go through an overview of the essential points from this class.

(Geo)statistics
Some Definitions

Statistics is concerned with mathematical methods for collecting, organizing, and interpreting data, as well as drawing conclusions and making reasonable decisions on the basis of such analysis.

Geostatistics is a branch of applied statistics that emphasizes (1) the geological context of the data, (2) the spatial relationship between data, and (3) the different volumetric support and precision of the data.

Why do we work with geostatistics in Petroleum Engineering?

- most of our data has spatiotemporal context, and it's important
  - most of **standard statistical** methods assume independent, identically distributed
- most of our spatial data is at a wide range of scales
  - standard statistical methods do not have methods to account for scale
- flexible uncertainty methods through realizations and scenarios
- well adapted to geological context, communication

(Geo)statistics
Some Definitions

Geostatistics developed from **practice of subsurface estimation** in mining, theory added later.

Concept	Geological Expression	Geostatistical Expression
Major changes in relationships between reservoir bodies	Architectural complexes and complex sets	Regions—separate units and model with unique methods and input statistics
Changes in reservoir properties within reservoir bodies	Basinward and landward stepping Fining/Coarsening up	Nonstationary mean
Stacking patterns of reservoir bodies	Organization, disorganization, compartmentalization, compensation	Attraction, repulsion, minimum and maximum spacing distributions, interaction rules
Major direction of continuity	Paleo-flow direction	Major direction of continuity, locally variable azimuth model
Relationship between vertical and horizontal continuity	Walther's Law	Geometric and zonal anisotropy
Distinct reservoir property groups	Lithofacies, depositional facies, and architectural elements	Reservoir categories, stationary regions
Heterogeneity	Architecture	Spatial continuity model geometric parameters, training image patterns

Common concepts, it all translates.

Stationarity
Substituting time for space

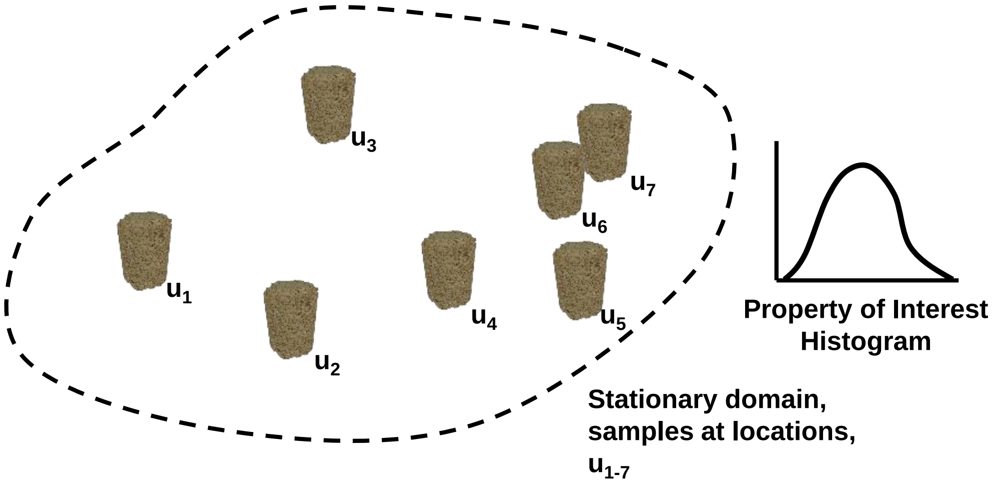
Any statistic requires replicates, repeated sampling (e.g. air or water samples from a monitoring station). In our geospatial problems repeated samples are not available at a location in the subsurface.

Extracted Sample at Location, u_1 **Hole Remains at Location, u_1**

Instead of time, **we must pool samples over space** to calculate our statistics. This decision to pool is the decision of stationarity. It is the decision that the subset of the subsurface is all the “same stuff”.

Stationarity
Substituting time for space

The decision of the stationary domain for sampling is an expert choice. Without it we are stuck in the "hole" and cannot calculate any statistics nor say anything about the behavior of the subsurface between the sample data.

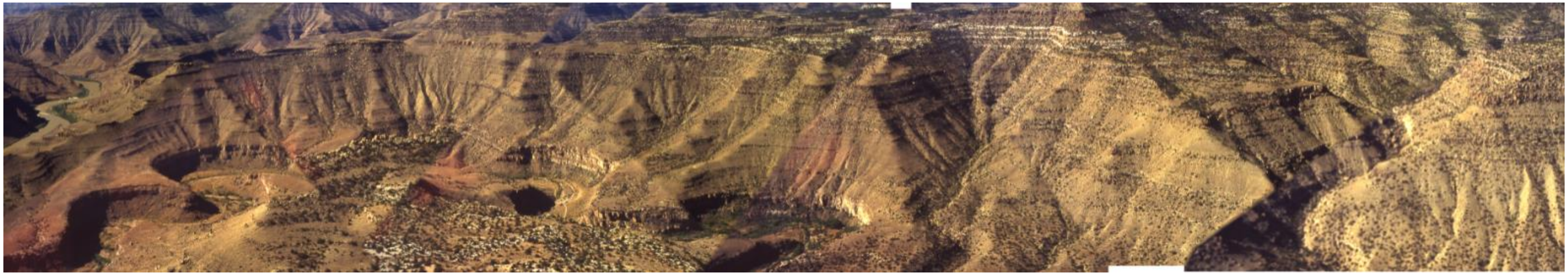


Import License: choice to pool specific samples to evaluate a statistic.

Export License: choice of where in the subsurface this statistic is applicable.

Stationarity
Definition 1: Geologic

Geological Definition: The rock over the stationary domain is sourced, deposited, preserved, and postdepositionally altered in a similar manner, the domain is map-able and may be used for local prediction or as information for analogous locations within the subsurface; therefore, it is useful to pool information over this expert mapped volume of the subsurface.



Stationarity
Definition 2: Statistical

Statistical Definition: The metrics of interest are invariant under translation over the domain. For example, one point stationarity indicates the that histogram and associated statistics do not rely on location, u. Statistical stationarity for some common statistics:

Stationary Mean: $E\{Z(\mathbf{u})\} = \mathrm{m}, \ \forall \, \mathbf{u}$

Stationary Distribution: $F(\mathbf{u}; z) = F(z), \ \forall \, \mathbf{u}$

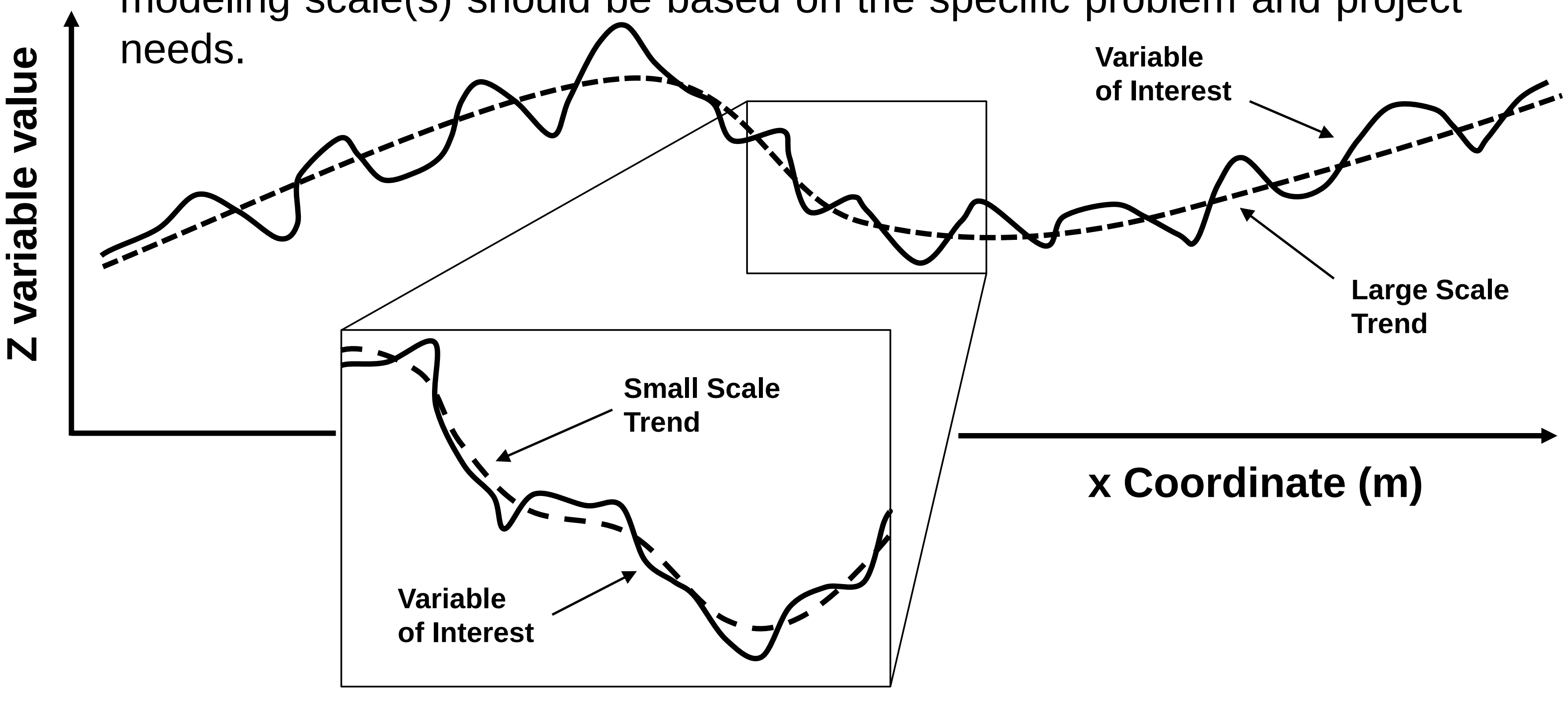
Stationary Semivariogram: $\gamma_z(\mathbf{u}; \mathbf{h}) = \gamma_z(\mathbf{h}), \ \forall \, \mathbf{u}$

May be extended to any statistic of interest including, facies proportions, bivariate distributions and multiple point statistics.

Stationarity
Comments on Stationarity

Stationarity is a decision, not an hypothesis; therefore it cannot be tested. Data may demonstrate that it is inappropriate.

The stationarity assessment depends on scale. This choice of modeling scale(s) should be based on the specific problem and project needs.



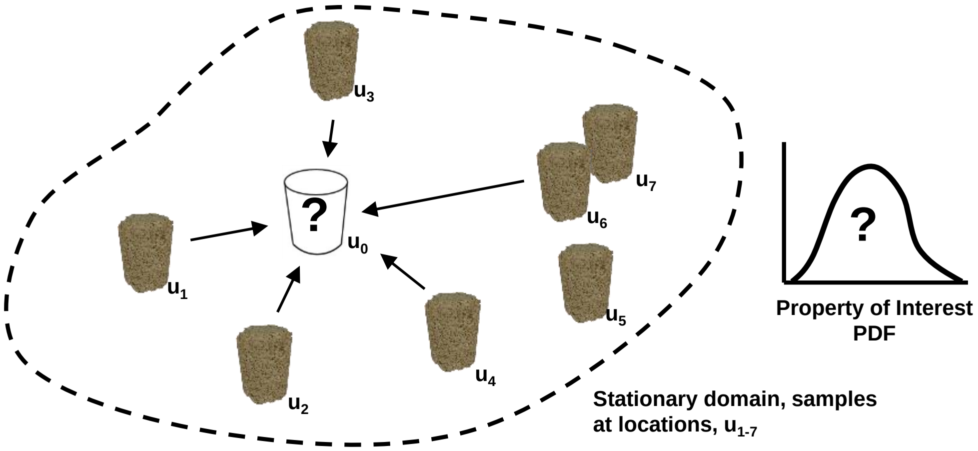
Stationarity
Comments on Stationarity

We cannot avoid a decision of stationarity. No stationarity decision and we cannot move beyond the data. Conversely, assuming broad stationarity over all the data and over large volumes of the earth is naïve. Good geological mapping is essential.

Geomodeling stationarity is the decision (1) over what region to pool data (import license) and (2) over what region to use the resulting statistics (export license).

Nonstationary trends may be mapped and the remaining stationary residual modelled stochastically, trends may be treated uncertain.

Uncertainty
What is uncertainty?

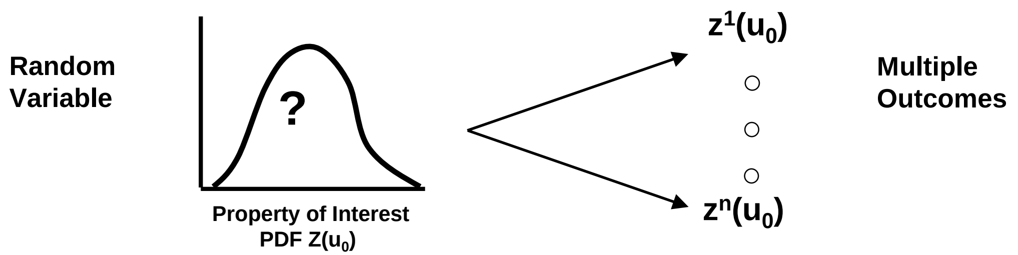


Uncertainty is not an intrinsic property of the subsurface.

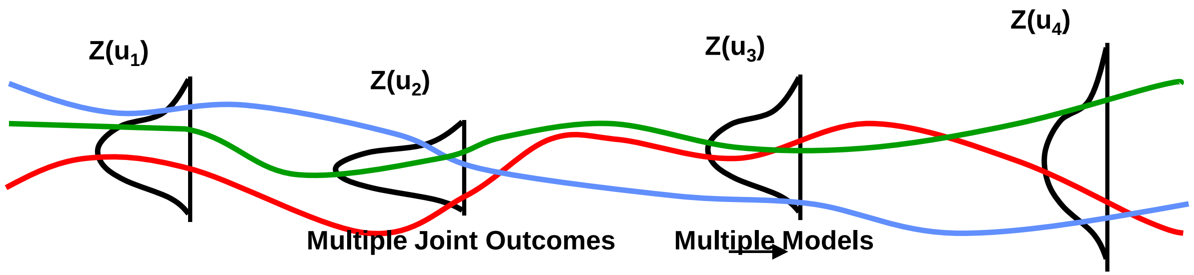
- At every location (u_α) within the volume of interest the true properties could be measured if we had access (facies, porosity etc.).
- Uncertainty is a function of our ignorance, our inability to observed and measure the subsurface with the coverage and scale required to support our scientific questions and decision making.
- This sparsity of sample data combined with heterogeneity results in uncertainty. If the subsurface was homogeneous then with few measurements uncertainty would be reduced and estimates resolved to a sufficient degree of exactitude.

Uncertainty
How do we represent uncertainty?

Random Variables and Functions: A random variable is a property at a location (u_α) that can take on multiple possible outcomes. This is represented by a probability density function (PDF).



If we take a set of random variables at all locations of interest and we impart the correct spatial continuity between them then we have a random function. Each outcome from the random function is a potential model of the subsurface.



Uncertainty
How do we represent uncertainty?

Using Multiple Models: We represent uncertainty with multiple models. It is convenient to assume that each model is equiprobable, but one could assign variable probability based on available local information and analogs. In general, when the input decisions and parameters are changed then these are known as **scenarios** and when the input decisions and parameters are held constant and only the random number seed is changed then these are known as **realizations**.

Working With Multiple Models: It is generally not appropriate to analyze a single or few scenarios and realizations. As Deutsch has recently taught, for decision making, use **all the models all the time** applied to the transfer function (e.g. volumetric calculation, contaminant transport, ore grade scale up, flow simulation etc.).

Uncertainty
Comments on uncertainty

Calculating Uncertainty in a Modeling Parameter: Use Bayesian methods, bootstrap, spatial bootstrap etc. You must account for the volume of interest, sample data quantity and locations, and spatial continuity.

If You Know It, Put It In. Use expert geologic knowledge and data to model trends. Any variability captured in a trend model is known and is removed from the unknown, uncertain component of the model. Overfit trend will result in unrealistic certainty.

Types of Uncertainty: (1) data measurement, calibration uncertainty, (2) decisions and parameters uncertainty, and (3) spatial uncertainty in estimating away from data. Your job is to hunt for and include all significant sources of uncertainty.

Uncertainty
Comments on uncertainty

What about Uncertainty in the Uncertainty? Don’t go there. Use defendable choices in your uncertainty model, be conservative about what you known, document and move on. Matheron taught us to strip all away all defenseless assumptions. Journel warned us to avoid the circular quest of uncertainty in uncertainty in...

Uncertainty Depends on Scale. It is much harder to predict a property of tea spoon vs. a house-sized volume at a location (u_α) in the subsurface. Ensure that scale and heterogeneity are integrated.

You Cannot Hide From It. Ignoring uncertainty assumes certainty and is often a very extreme and dangerous assumption.

Decision Making with Uncertainty. Apply all the models to the transfer function to calculate uncertainty in subsurface outcome to support decision making in the presence of uncertainty.

Facies
What are the Criteria for Facies?

First some general comments:

1. **Facies are a categorical classification of rock.**

2. **Facies / Rock type** is an important decision for subsurface modeling. It should remain a collaborative decision integrating expertise from the project team (Geologists, Reservoir Modelers, Reservoir Engineers, Petro- and Geophysicists).

3. Facies / Rock types **must improve subsurface prediction** away from the data or they do not add value.

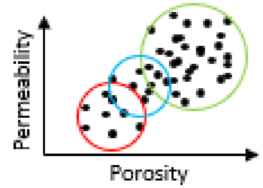
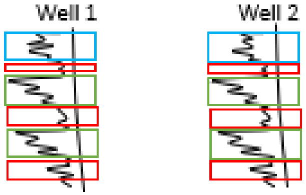
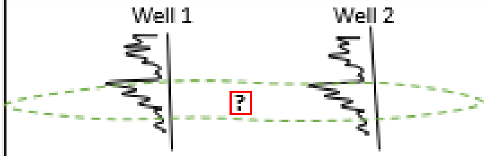
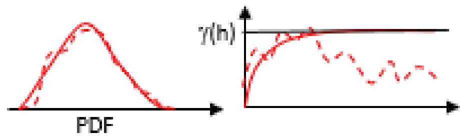
4. **Number of facies** is a balancing act between accuracy of geological concepts and statistical inference, and modeling effort

5. Reservoir modeling is **hierarchical**,
   units *contain* depofacies *contain* lithofacies *contain* por/perm

5. 80-90% of **flow heterogeneity** is captured in the facies models.

Facies
What are the Criteria for Facies?

These are the **criteria for facies** (or any categories in reservoir models).

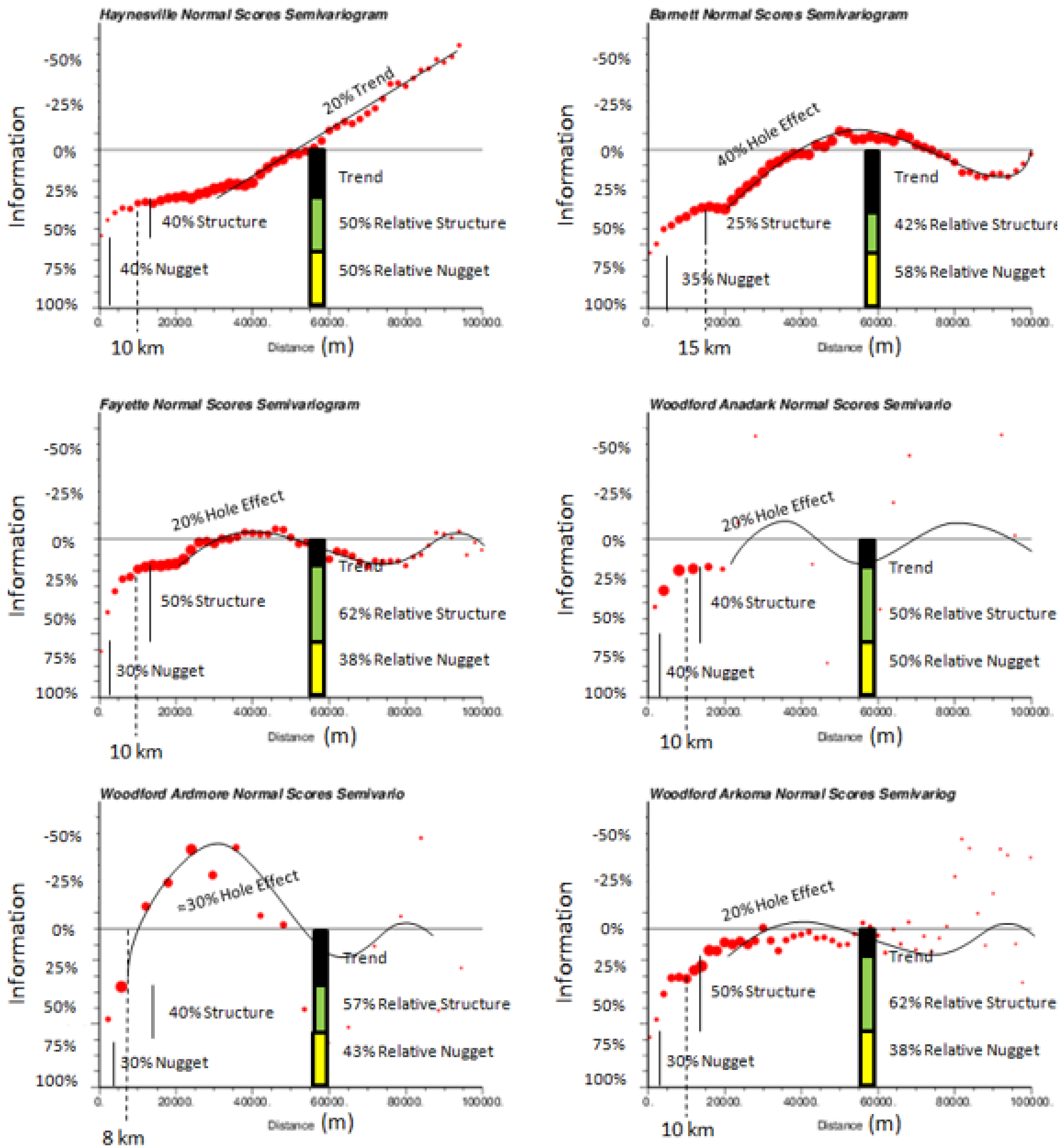
Criteria	Considerations	Example
Separation of Rock Properties	Facies must divide the properties of interest that impact subsurface environmental and economic performance (e.g. grade, porosity and permeability).	
Identifiable in Data	Facies must be identifiable with the most common data available. e.g. facies identifiable only in cores are not useful if most wells have only logs.	
Map-able Away from Data	Facies must be easier to predict away from data than the rock properties of interest directly, facies improves prediction.	
Sufficient Sampling	There must be enough data to allow for reliable inference of reliable statistics for rock properties for each facies.	

Spatial Continuity

A measure of change in a property of interest over distance.

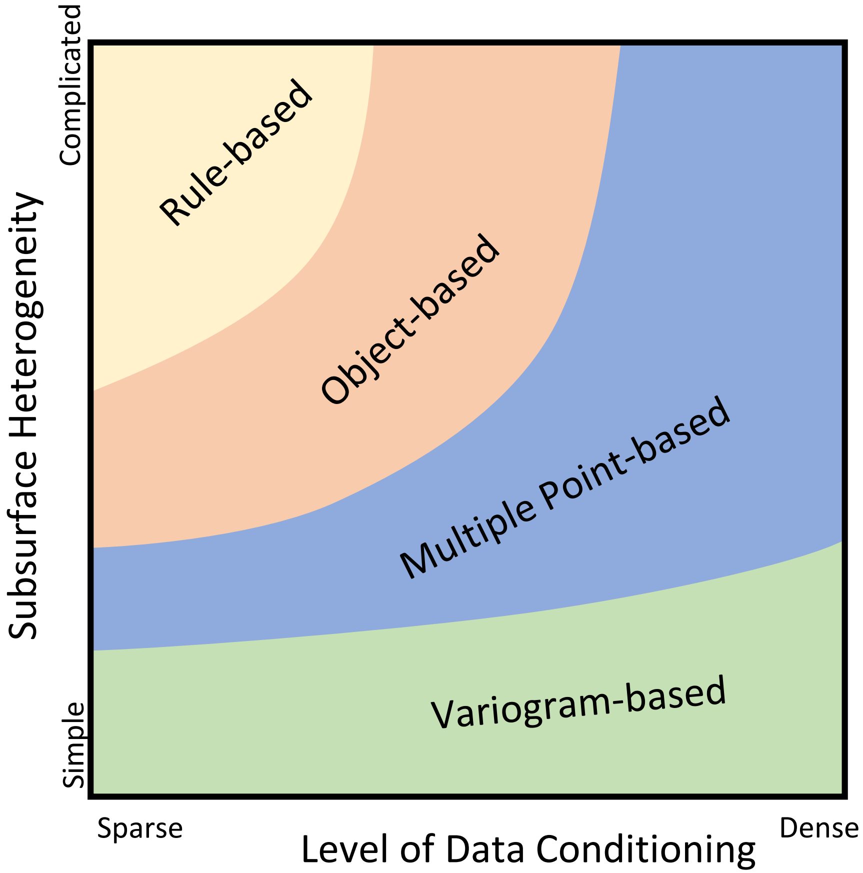
Relationship of Euclidean distance to Geologic distance.

Example for production from shale assets.



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

A toolbox



Geostatistics includes a tool box of subsurface modeling methods, here is a scheme for selecting between them based on the subsurface heterogeneity and the level of data conditioning, also consider project goals and resources.

- New methods that model the subsurface with data integration, uncertainty, conditional to data. **Add it to the box.**

Big Data, Machine Learning Wave!

Big Data criteria:

- Volume
- Velocity
- Variety
- Veracity

We've been big data before there was big data.

Machine Learning

Training a computer detect features, find complicated relationships with complicated, multivariate, large datasets

We've been doing that too.

We are working on these.

Resource, Check out Our Webinar

Webinar - Big Data Analytics for Petroleum Engineering: Hype or Panacea?

December 8, 2017: Big Data Analytics for Petroleum Engineering: Hype or Panacea? Little Data + Simple Model = Big Data?

http://www.cpge.utexas.edu/?q=node/385

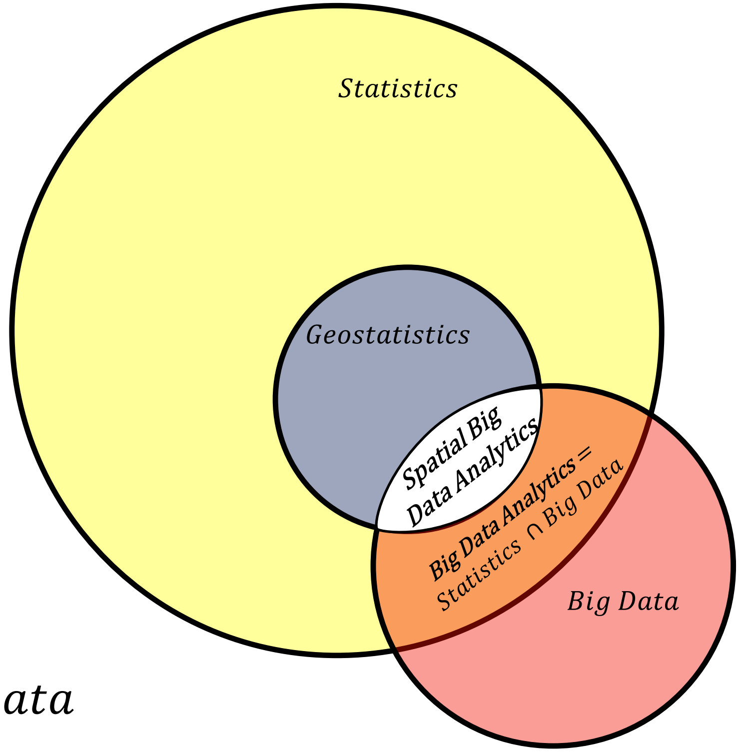
Big Data, Machine Learning Wave!

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

Geostatistics is a branch of applied statistics: (1) the spatial (geological) context, (2) the spatial relationships, (3) volumetric support, and (4) uncertainty.

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

$$\textit{Spatial Big Data Analytics} = Geostatistics \cap Big\ Data$$



Proposed Venn diagram for spatial big data analytics.

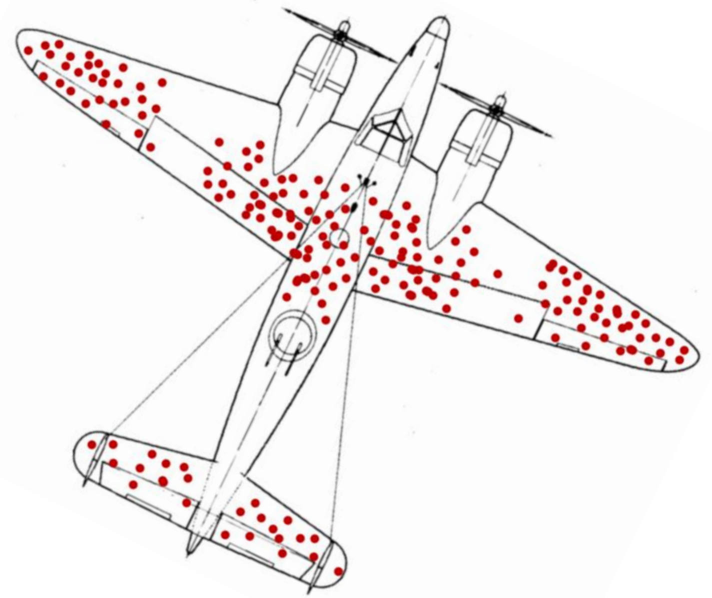
Rigorous Statistics Mitigate Bias

We must guard against bias.

Survivorship Bias:
a form of selection bias resulting from selecting samples that "survived" some previous selection process. This often leads to false conclusions.

For example, in 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. The plans shot in critical locations did not return to base. Wald suggested reinforcement of locations that were not damaged in planes that safely returned to base! (https://en.wikipedia.org/wiki/Survivorship_bias#In_the_military)



Hypothetical dataset of aircraft damage for planes that returned to based. Source https://en.wikipedia.org/wiki/Survivorship_bias#/media/File:Survivorship-bias.png

Is there preselection in our subsurface datasets?

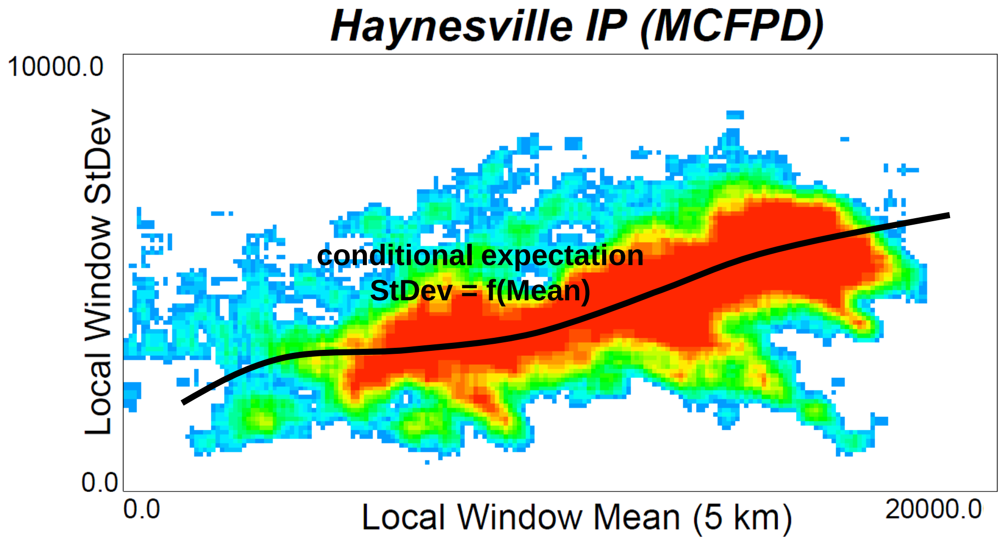
Motivational (Geo)statistics Examples

Why Use (Geo)statistics?

Summarization / Comparison

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

Robust measure of significance of differences



Abstraction allows for efficient characterization and leads to insights.

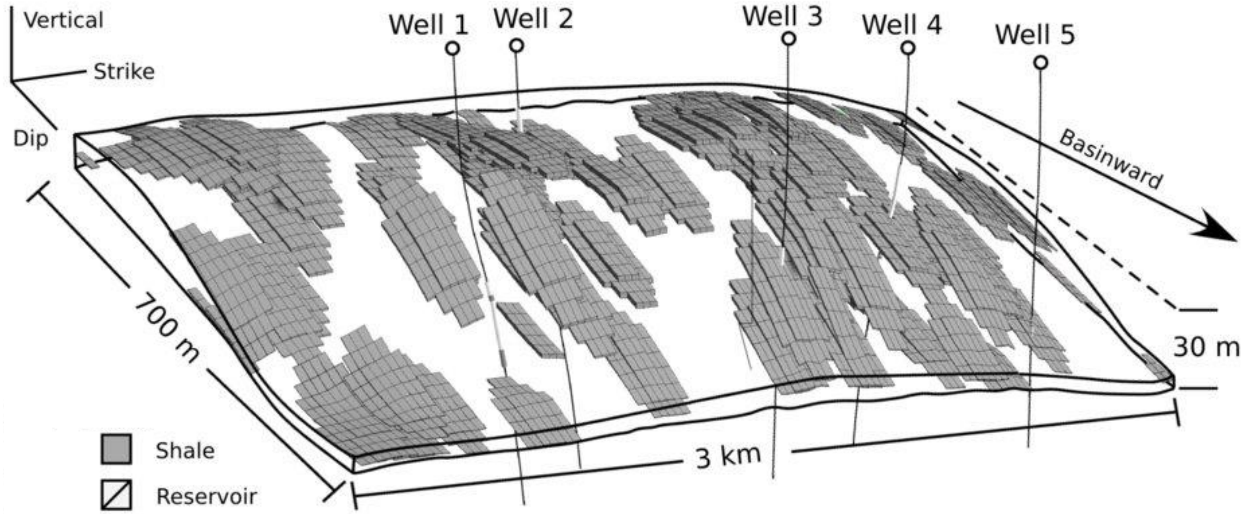
Motivational (Geo)statistics Examples

Why Use (Geo)statistics?

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.



Can't know exactly where the shales are from 5 wells and given the shale discontinuity.

Motivational (Geo)statistics Examples

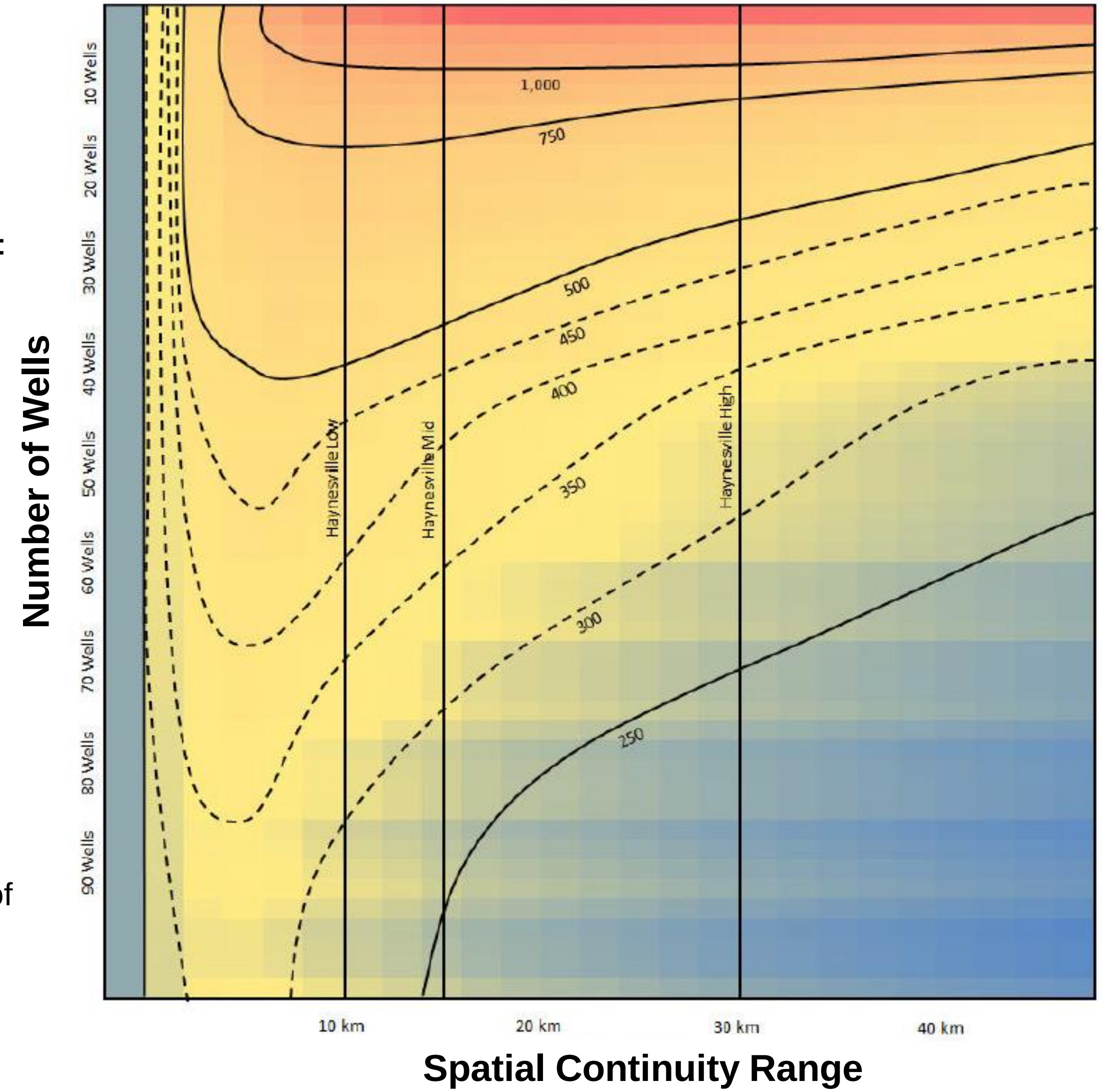
Why Use (Geo)statistics?

Proxy Models

sparse data + simple model = big data

(Lake, personal communication)

construction of a statistical model that represents the system for fast evaluation, uncertainty modeling



Proxy model for uncertainty in production rate as a function of well density and geological complexity / discontinuity.

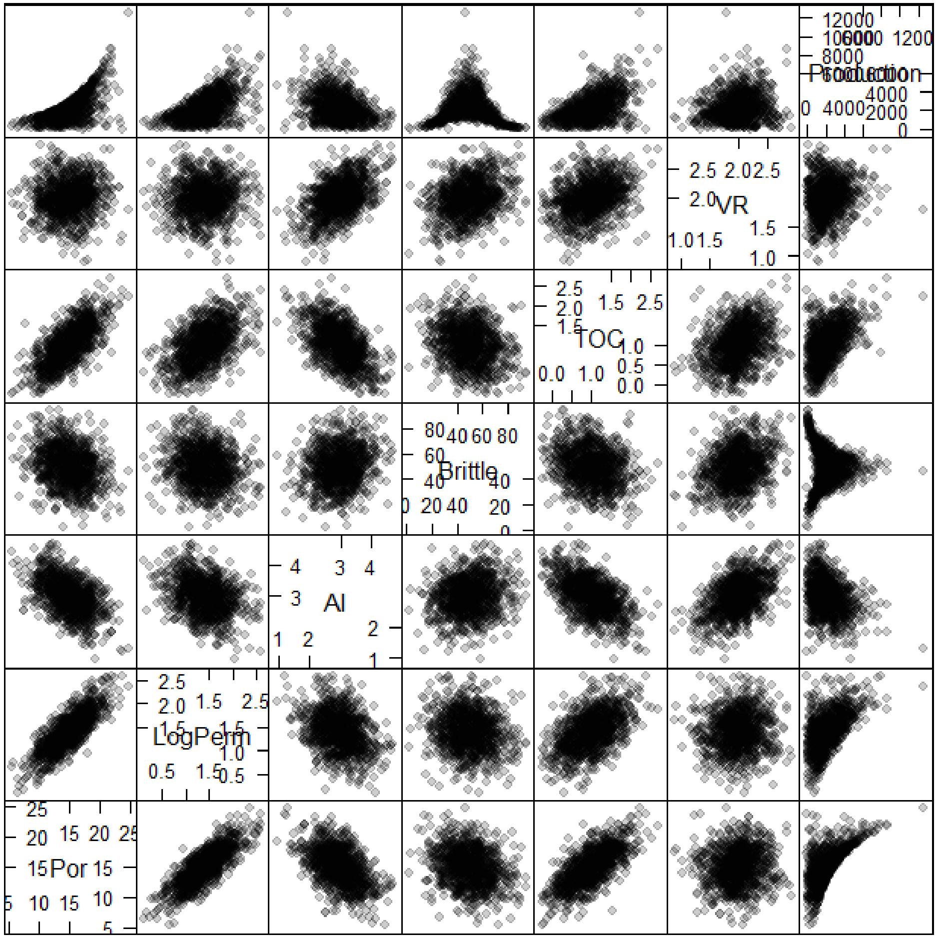
Motivational (Geo)statistics Examples

Why Use (Geo)statistics?

Too Big / Too Complicated

- Massive Multivariate
- Due to the curse of dimensionality we may not have enough data to characterize the system, need to used a statistical multivariate model

Unconventional Dataset



Scatter Plot Matrix

Multivariate modeling accounting for complicated relationships

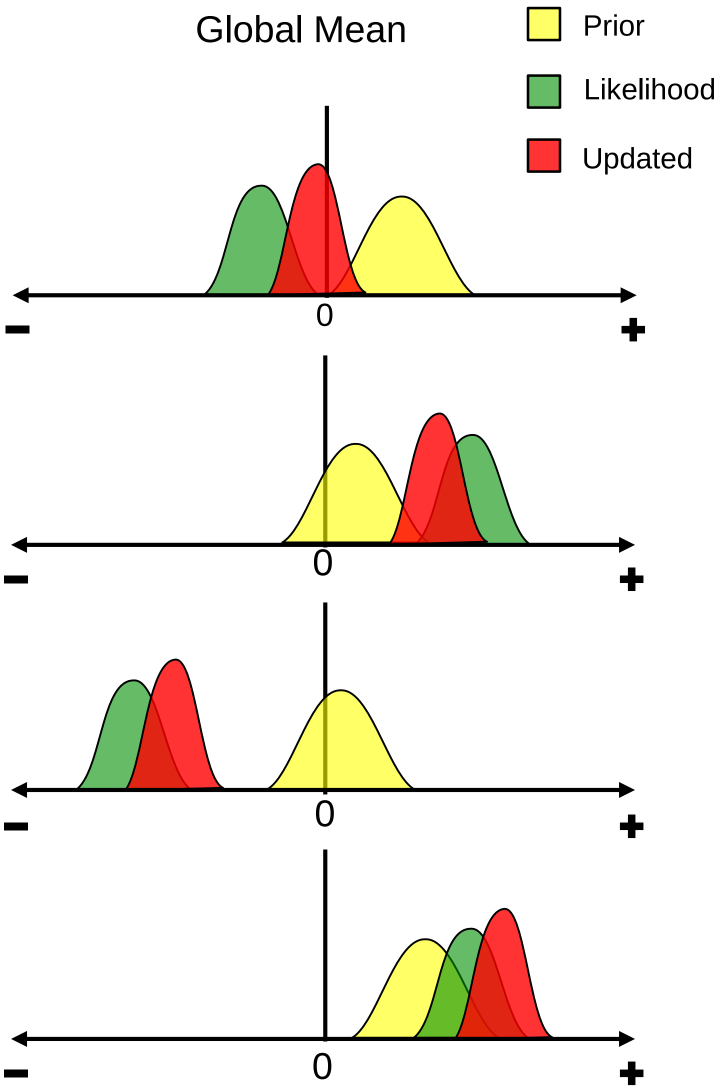
Motivational (Geo)statistics Examples

Why Use (Geo)statistics?

Combining / Updating with New Information

Bayesian Updating

need statistical models to describe data redundancy



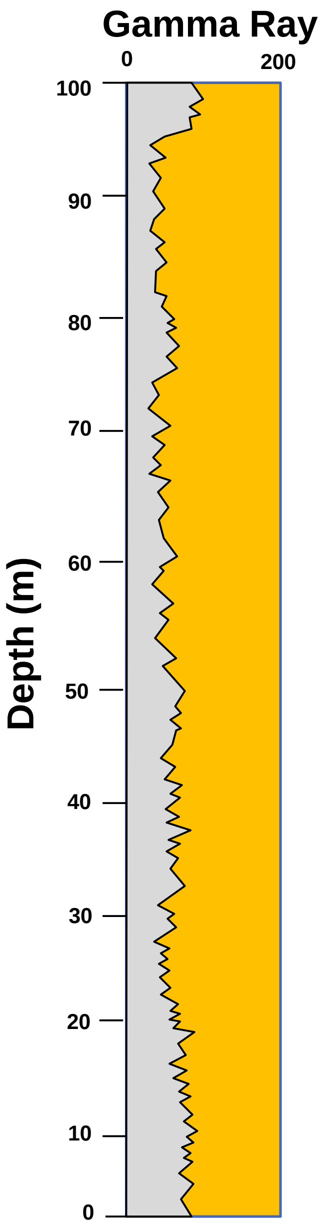
Bayesian updating under the assumption of Gaussianity.

Motivational (Geo)statistics Examples

Why Use (Geo)statistics?

Accounting for Scale

- Pores to Production
- Statistical models for change of support size



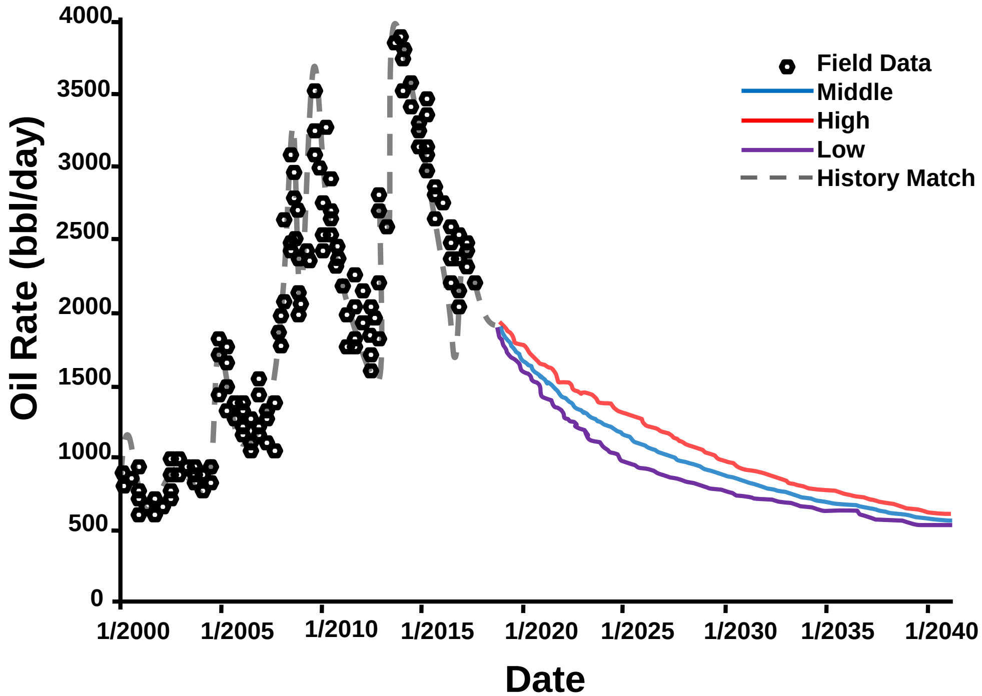
Example well log.

Motivational (Geo)statistics Examples

Why Use (Geo)statistics?

Forecasting / Decision Making

- Decision Support
- Need all the previous information to build forecast uncertainty models to optimize very expensive project decisions



Reservoir forecasting with uncertainty.

Learn More?

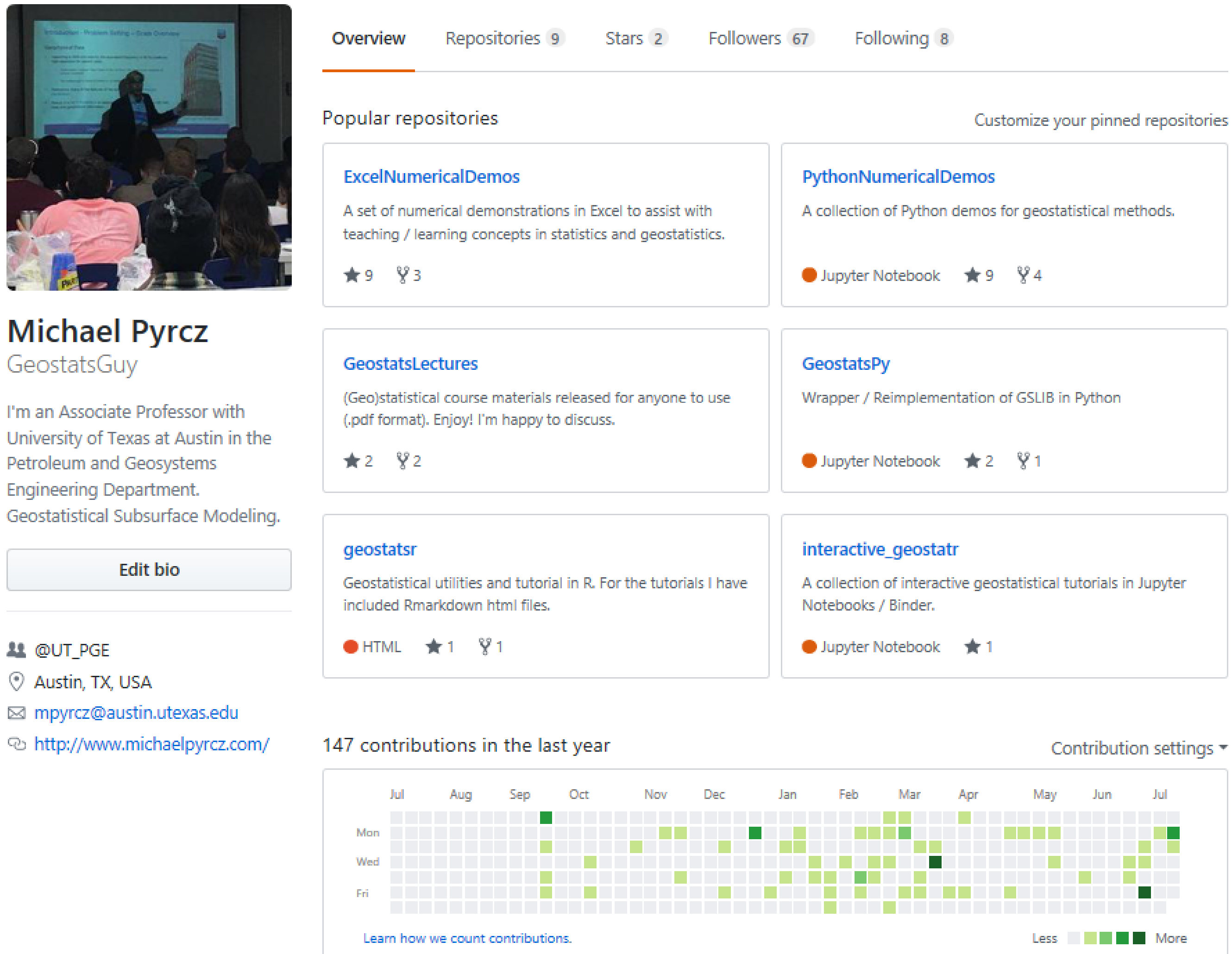
For tweets with Geostatistical resources, follow @GeostatsGuy

Learn More?

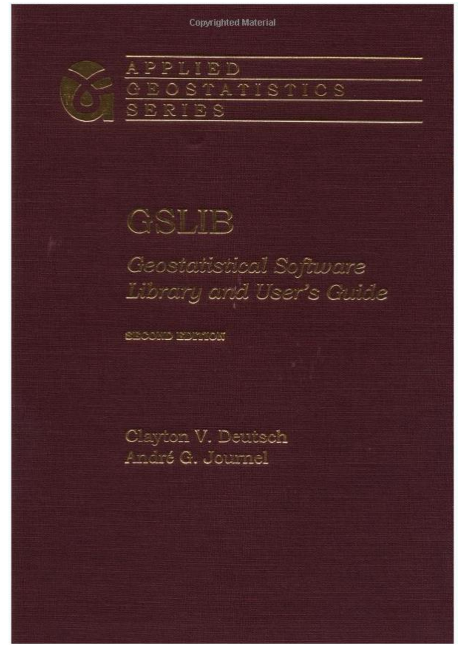
@GeostatsGuy

Excel, R and Python
- Distributions
- Bootstrap
- Cellular Automata
- Hypothesis Testing
- Lorenz Coefficient
- Decision Making
- Bayesian Updating
- Kriging
- Simulation
- Volume-variance

- Python Implementation of GSLIB

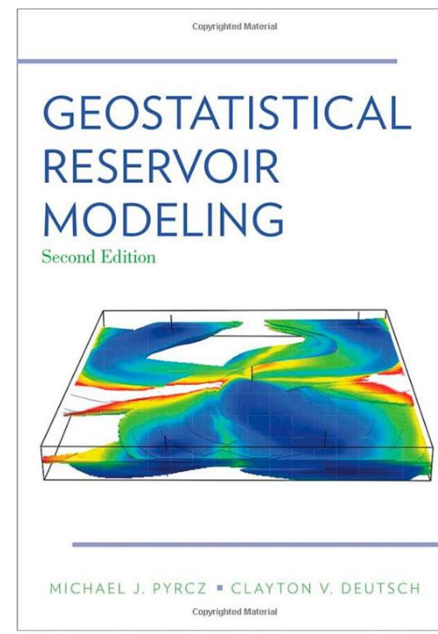


Books



GSLIB: Geostatistical Software Library and User's Guide

- A Practical Set of Geostatistical Tools and User's Guide
- The software is available online at www.GSLIB.com.
- Check it out.



Geostatistical Reservoir Modeling

- Theory and fundamental concepts
- Methods and workflows
- Best practice

and so it begins…

Topic	Application to Subsurface Modeling	Example
Probability Theory	Methods to infer marginal, conditional and joint probabilities for subsurface events with frequentist and Bayesian frameworks.	Porosity and permeability relationships Facies trends from seismic
Representative Sampling	Workflows for identifying and mitigating spatial sampling bias and accounting for scale.	Well declustering Seismic debiasing
Spatial Data Analysis	Quantification of spatial continuity Interpretation and modeling of spatial continuity	Facies definitions Trend modeling Characterize heterogeneity
Spatial Estimation	Modeling and utilizing spatial continuity make predictions away from wells and at sub-seismic scale.	Spatial estimation Value of well data Shale block uncertainty
Spatial Simulation	Calculation of model parameter uncertainty, stochastic simulation methods	Predrill assessments Stochastic realizations
Uncertainty Management	Post-processing subsurface realizations and decision making in the presence of uncertainty.	NTG uncertainty from wells Summarizing spatial uncertainty
Machine Learning	Working with many variables, modeling complicated phenomenon.	Dimensional reduction and modeling for unconventionals.

Let's learn some useful stuff!

Spatial Modeling with Geostatistics

Introduction

Lecture outline . . .

- **Basic Concepts**
- **Motivation**
- **Overview**