# Data Analytics, Geostatistics and Machine Learning

# Fundamamental Concepts

**Fundamental Concepts . .**

- **What is Subsurface Modeling?**
- **Modeling Goals**
- **Modeling Strategies**
- **Workflow Development**

**Instructor: Michael Pyrcz, the University of Texas at Austin**

# Data Analytics, Geostatistics and Machine Learning
# Fundamamental Concepts

**Fundamental Concepts . .**

- **What is Subsurface Modeling?**

**Instructor: Michael Pyrcz, the University of Texas at Austin**

# Statistics and Geostatistics
# 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.

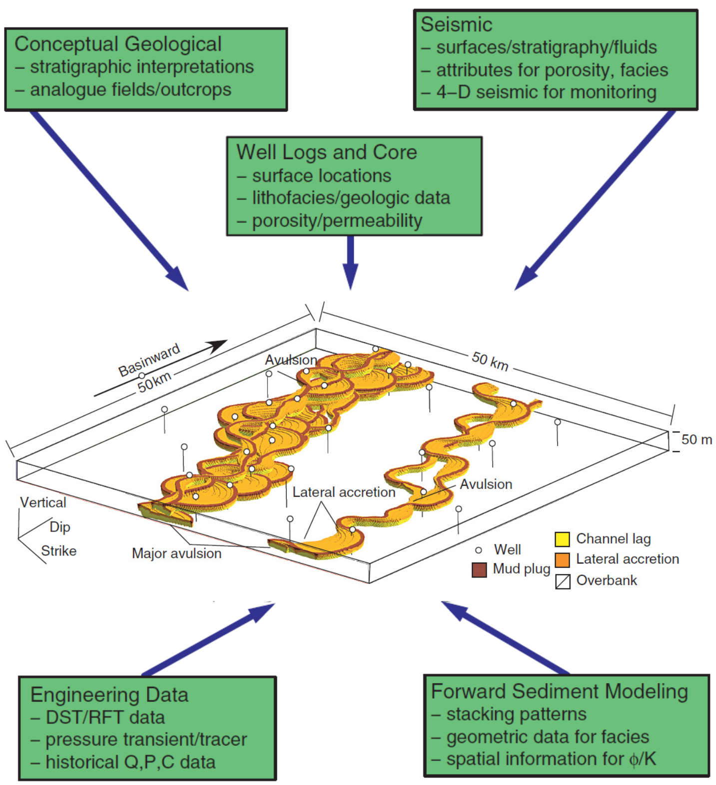
***Geostatistics*** is a branch of applied statistics that emphasizes (1) the geological context of the data, (2) the spatial relationship between data, (3) spatial uncertainty and (4) the different volumetric support and precision of the data.

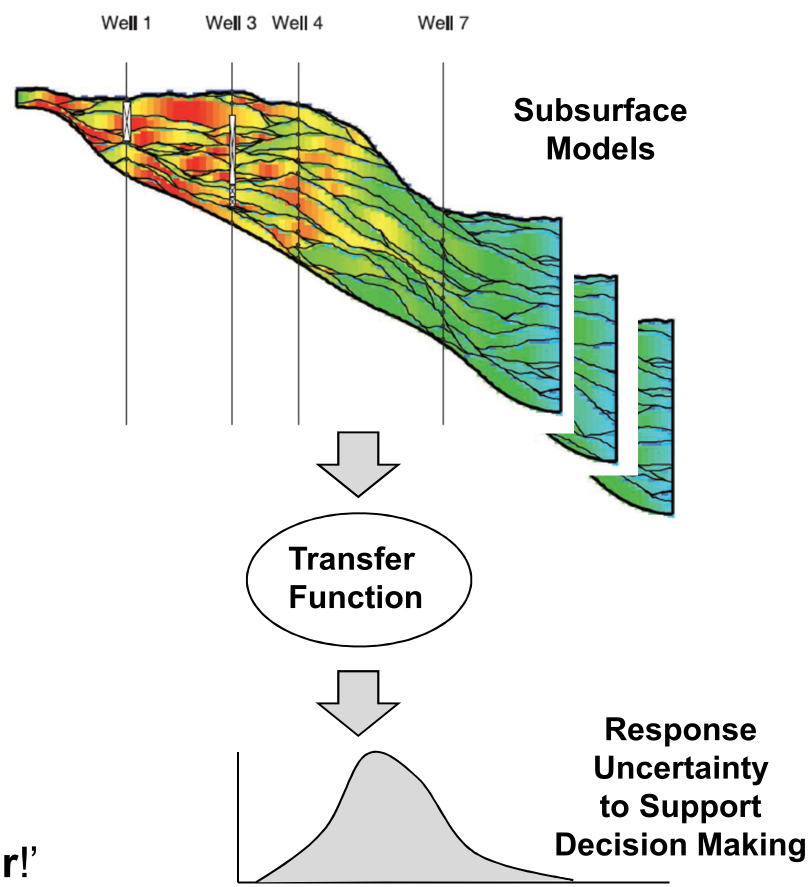
Why do we work with geostatistics in Geosciences?

- ✓ **Geological Context**
- ✓ **Spatial Relationships**
- ✓ **Variable Scale of Data**
- ✓ **Variable Data Precision**
- ✓ **Highly Multivariable**

# What is Subsurface Modeling?

***Reservoir / Subsurface Modeling*** is the integration of all subsurface information to build a suite of models representing uncertainty to support decision making.




**'If it doesn't get in the model, it doesn't matter!'**

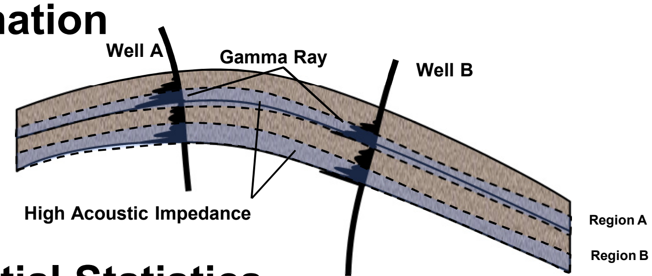
# What is Subsurface Modeling?

### A Numerical Model

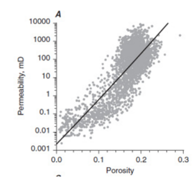
- ***quantification*** – integrate data and concepts, calculate summary spatial statistics and trends over the subsurface volume of interest

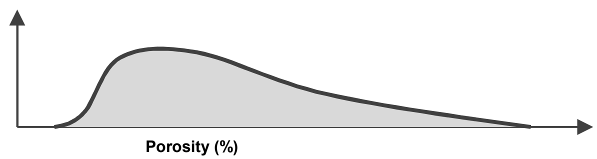
- ***subsurface model*** - spatial reservoir property distributions over the subsurface volume of interest that reproduce the quantification to **support decision making**

**Data / Information**

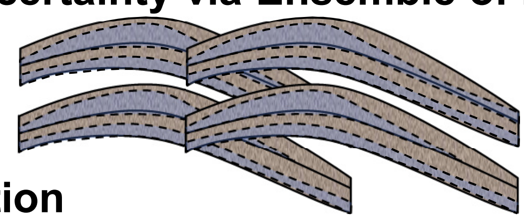



**Trends / Spatial Statistics**



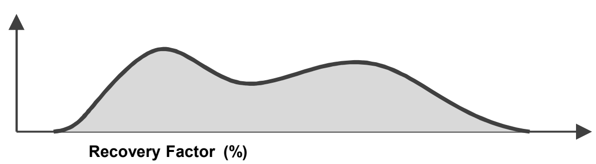

**Subsurface Uncertainty via Ensemble of Models**



**Transfer Function**

**Volumetrics, Flow Simulation**

**Decision Criteria**

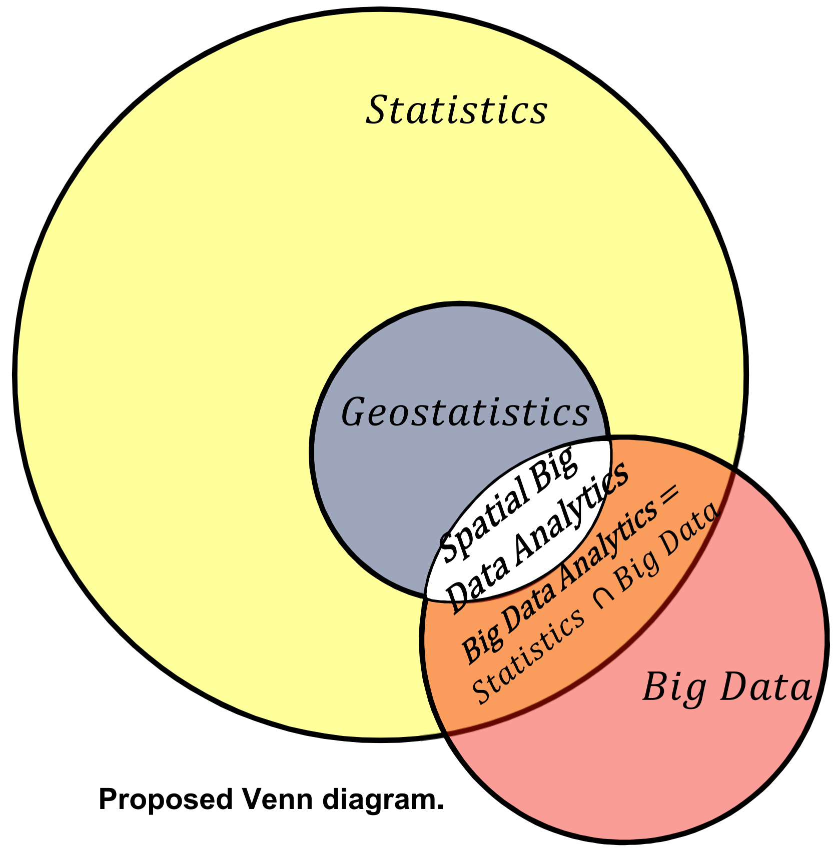


# Big Data, Data Analytics and Geostatistics

***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 spatial (geological) context of the data, (2) the spatial relationship between data, (3) the different volumetric support and precision of the data, and (4) spatial and data uncertainty.

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




**Proposed Venn diagram.**

**Given this:**

**Spatial big data analytics is the expert use of (geo)statistics to learn from our spatial data set.**

# A Data Integration Challenge

| 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*<br>*Geomorphology* | $\simeq \infty$<br>$\simeq \infty$ | none<br>none | Concepts, input statistics<br>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.**

# Why Learn About Geostatistical Subsurface Modeling?

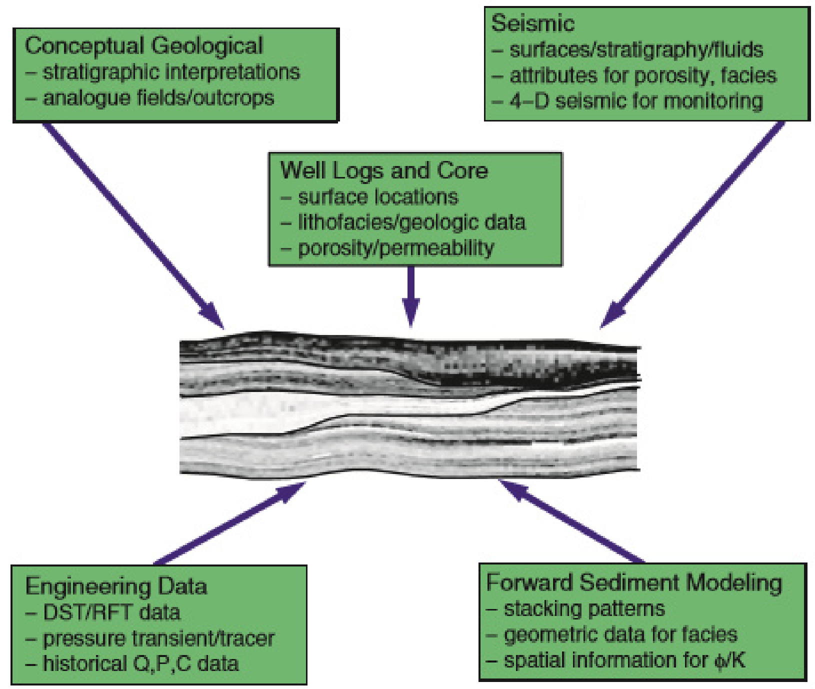
**Why should you have a greater proficiency on reservoir modeling?**

## Level 1: Basic Understanding

Most reservoir asset subsurface teams develop a **stochastic 3D reservoir model**.

• If you work with the subsurface, you will work with stochastic reservoir models!

• **Understand adjacent disciplines** and workflows in you team.




Subsurface asset integration (Pyrcz and Deutsch, 2014).

# Why Learn About Geostatistical Subsurface Modeling?

**Why should you have a greater proficiency on reservoir modeling?**

## Level 2: Improved Communication

Reservoir modeling sits in the middle of the subsurface team and integrates all available engineering, geological and geophysical information.

• Improved reservoir modeling capability results in **improved communication and integration** in the subsurface team.

**TABLE 2.1.** RESERVOIR CONCEPTS AND ASSOCIATED GEOLOGICAL AND GEOSTATISTICAL EXPRESSIONS

| 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 ~~if~~ 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 |

Subsurface concepts, with their geological and geostatitistical expressions (modified from Pyrcz and Deutsch, 2014).

# Why Learn About Geostatistical Subsurface Modeling?

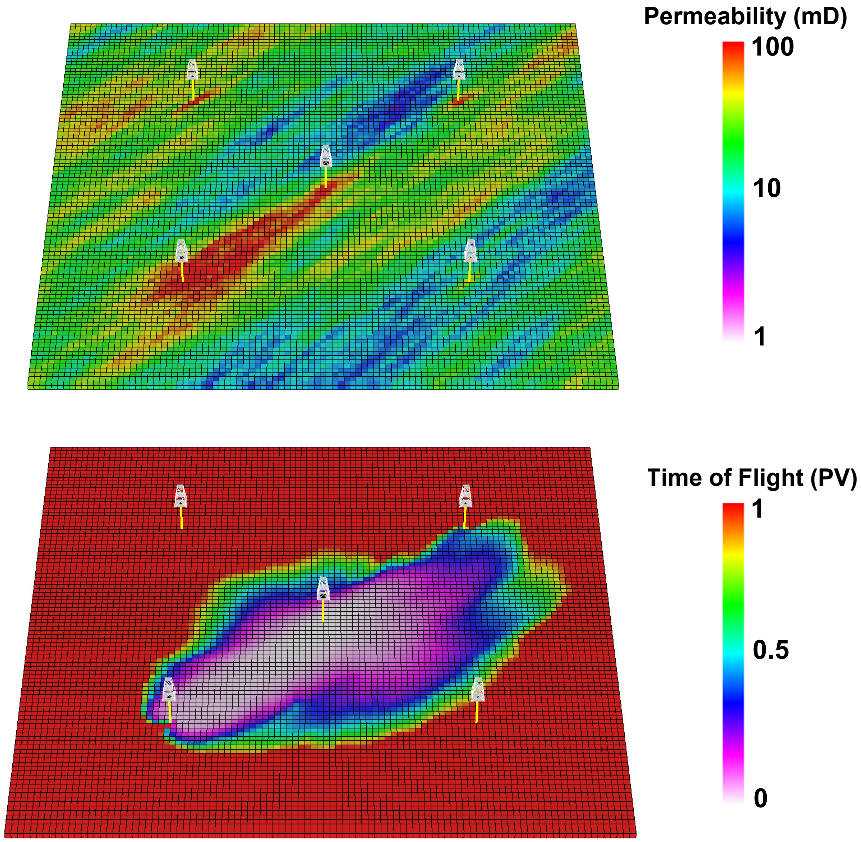
**Why should you have a greater proficiency on reservoir modeling?**

## Level 3: Maximize Your Impact

Reservoir models are directly applied for forecasting that support decision making.

• **Best integration of your knowledge** into the subsurface model.

• If your expertise does NOT impact the model, you may NOT impact the development decision!




Permeabilty heterogeneity and flow response.

# Why Learn About Geostatistical Subsurface Modeling?

**Why should you have a greater proficiency on reservoir modeling?**

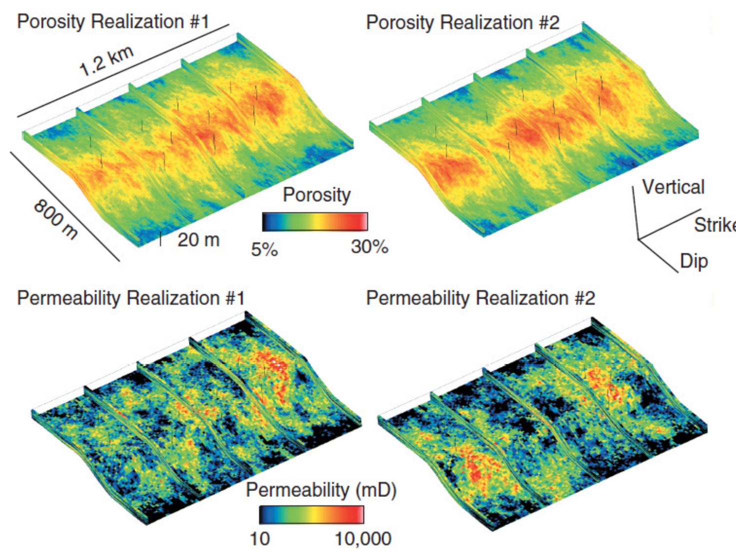
## Level 4: Build Subsurface Models

Most subsurface modelers are geoscientists and engineers that learned on the job.

**Become a subsurface modeler!**

- Black box, uninformed reservoir modeling will result in bad decisions.

- Advanced knowledge unlocks novel workflows to solve difficult subsurface problems.




Subsurface asset integration.

# Data Analytics, Geostatistics and Machine Learning
# Fundamamental Concepts

## Fundamental Concepts . .

- Modeling Goals

Instructor: Michael Pyrcz, the University of Texas at Austin

# Model Goal and Purpose Modeling

## 1. Build a Numerical Model / Common Earth Model

- Platform to integrate all available information, unified understanding of the subsurface
- Establish what is known, unknown and critical risks
- Numerical support for future investigation
- Communication tool

*Is there value is just building a model?*

# Model Goal and Purpose
# Modeling

## 2. Assess Resources

- Compute the gross volume of interest and the associated spatial distribution
- Include some consideration for extraction method and associated scales and thresholds

## 3. Quantify Resources Uncertainty

- Multiple sources and multiple scales of uncertainty
- Results may be used to direct data collection to reduce uncertainty, support development decisions

# Model Goal and Purpose Modeling

## 4. Investigate Geologic Risks

- Consider a wide range of possible subsurface features
- Evaluate sensitivities and risks, down- and up-side outcomes

## 5. Exporting Statistics

- Develop a robust set of statistics from a mature reservoir to apply elsewhere
- This may include trends, distributions, training images etc.
- Common to support early development in sparse data settings.

# Model Goal and Purpose Modeling

## 6. Evaluate the Need for Additional Data

- Determining the local and global uncertainty models
- Assess the value of additional data in reducing these uncertainties

## 7. Assess Reserves

- Calculate the resources that would be extracted after applying economic thresholds and technical limits of the extraction methodology.
- Modeling and calculations are consistent with reporting standards

# Model Goal and Purpose Modeling

**8. Evaluate Different Recovery Processes**

- There are various decisions with respect to primary, secondary and tertiary recovery
- Optimize the recovery method accounting for all information and uncertainty

**9. Make Final Decisions**

- Provide the local best estimates to support well site selection

# Data Analytics, Geostatistics and Machine Learning

## Fundamamental Concepts

### Fundamental Concepts . .

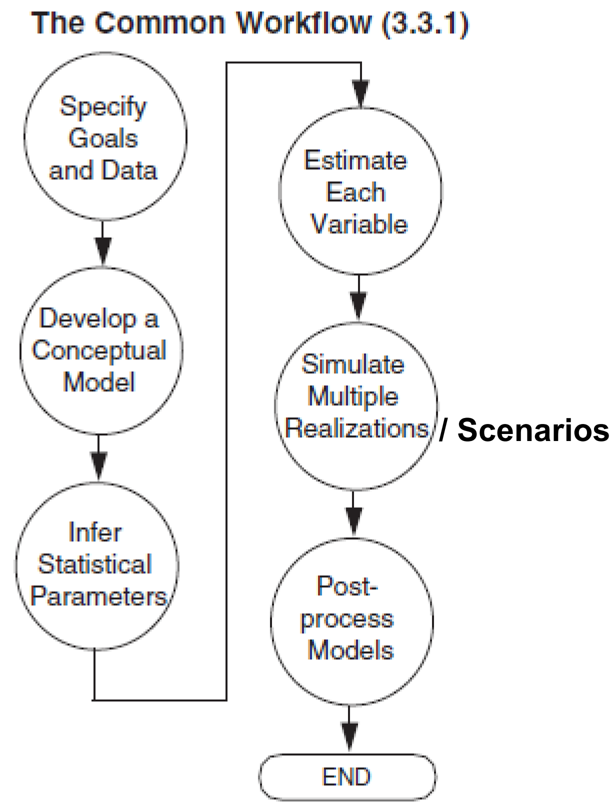
- Modeling Strategies

Instructor: Michael Pyrcz, the University of Texas at Austin

# Modeling Strategies

**Common Subsurface Modeling Workflow**

- Specify modeling goals and available resources
- Develop conceptual model
- Infer statistical parameters
- Build estimation models
- Build simulation models
- Post-process models




The common workflow modified from Pyrcz and Deutsch (2014).

# Modeling Strategies

## Data Integration Workflow

- Integrating wells, seismic, production and concepts
- The results is a consistent numerical representation of the reservoir properties

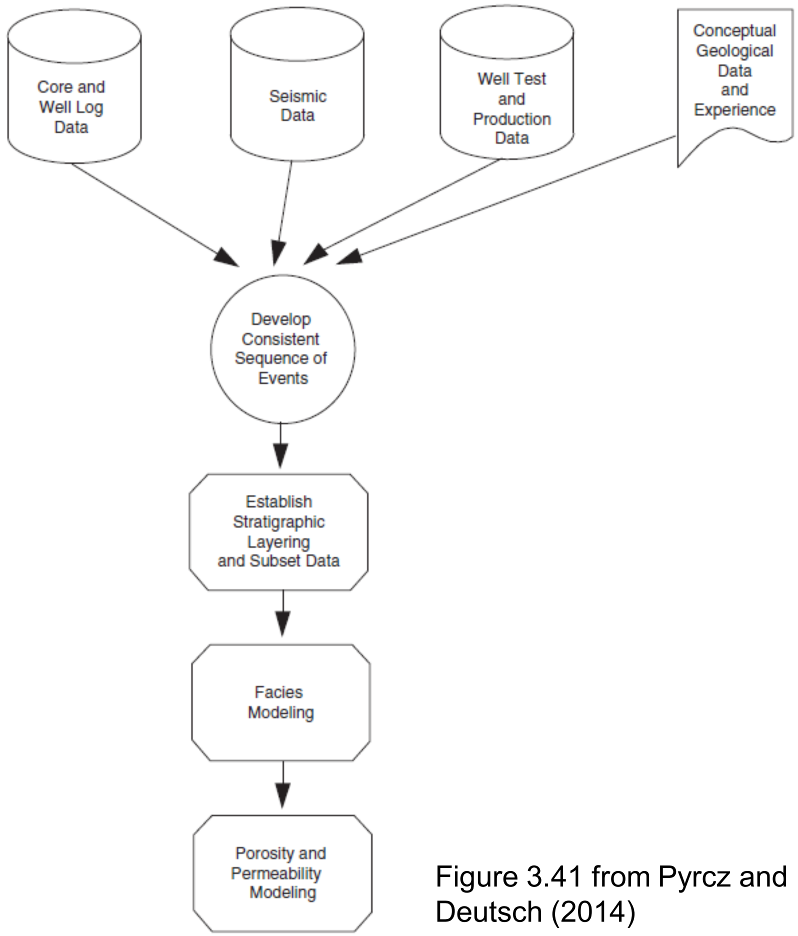
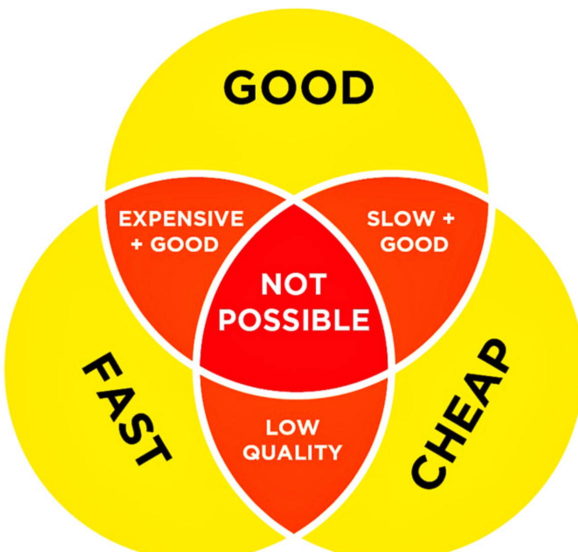



Figure 3.41 from Pyrcz and Deutsch (2014)

# Modeling Strategies

**Fit-for-Purpose Modeling**

- Model workflow design considering the goals of the model
- May also consider future needs for the model
- Accounting for resources, time and people (expertise) available




Good, Fast and Cheap is not Possible
image from https://www.purechat.com/blog/fast-cheap-and-good-the-small-business-guide-to-content-creation/

# Modeling Strategies

## Modeling Constraints

- Professional Time
  - work hours available limited by workforce and project timelines schedules

- Organization Capability
  - The skill sets of the professionals that are available

- Computational Facilities
  - The hardware and software available for the project

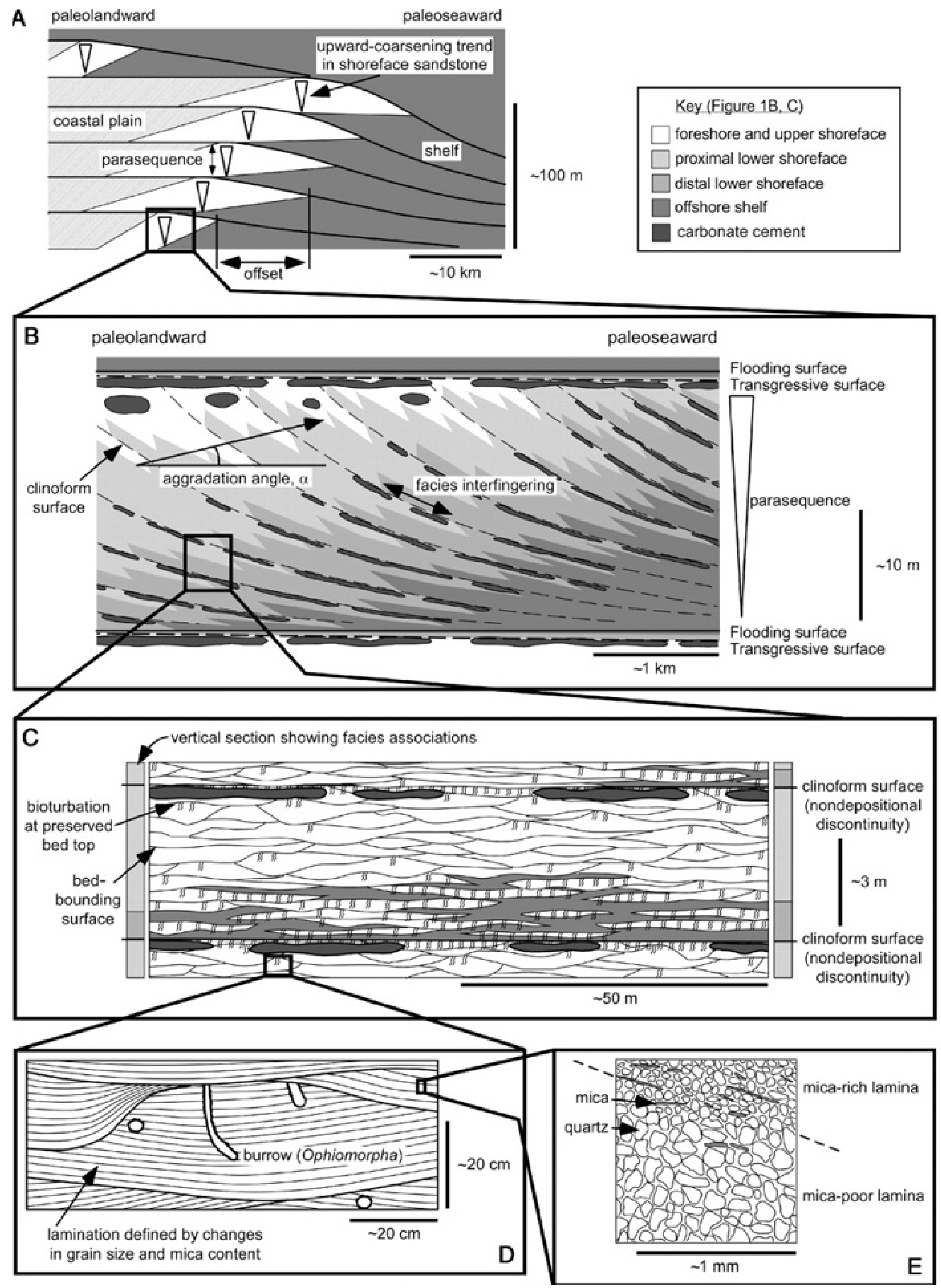
- Total Budget
  - Limiting professional time, computational resources and data collection

*All projects are constrained and require significant prioritization.*

# Modeling Strategies

## Top-down Reservoir Modeling (Williams et al., 2004, Sech et al., 2009)

- Start with the simplest model possible
- Add detail as required, until it doesn't have an impact on the transfer function
- Very efficient for fast initial assessments
- Learn the impact of scale / details

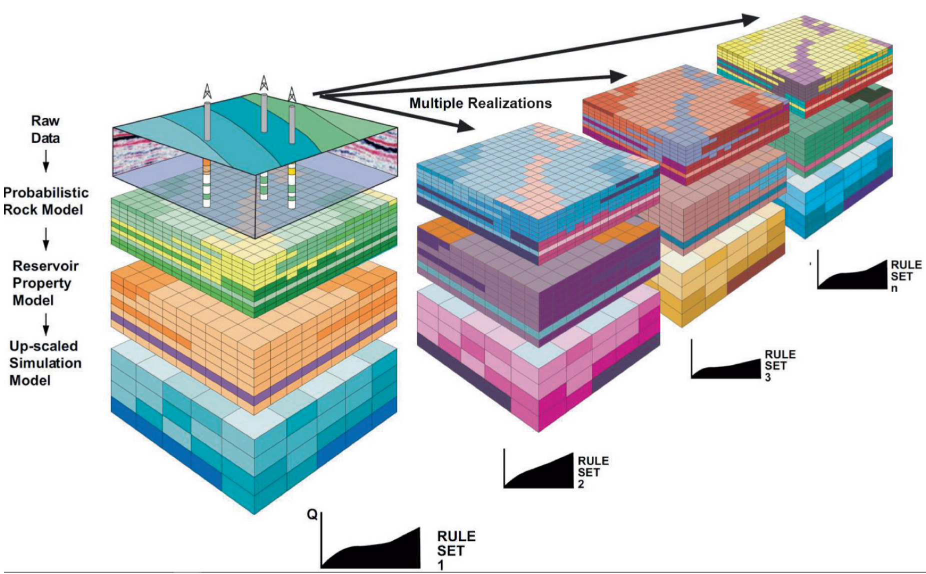



Surface-based top down modeling based on outcrop from Sech et al., 2009

# Modeling Strategies

## Modeling for Discomfort (Bentley, 2015)

- Models become tools for verification of a decision already partially or fully made, this is modeling for comfort!
- Bentley recommends that we model for discomfort! Stress test our current concepts and the decision-making.
- Identify remaining up-side potential and secure against worst case.
- Need to recognize our biases




Multiple deterministic scenarios (Figure 12 of Bentley, 2015).

# Modeling Workflows

**Modeling Workflows Based on Model Goals**

Let's discuss:

- 2-D Mapping for Volumetrics
- Regional Mapping
- Mini/Micro Modeling
- Reservoir Modeling

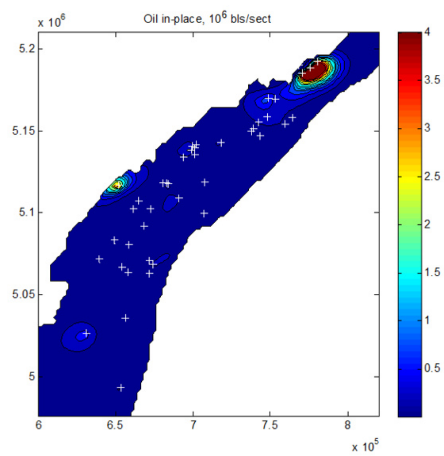
# Modeling Workflows

## 2-D Mapping for Volumetrics

- Goal: Produce a map of remaining resource in place
- Properties: thickness, vertically averaged porosity and saturation, seismic attributes
- Model: estimation model for smoothly varying properties between wells, may integrate physics
- Calculate resources in place, e.g.

$$OIP = \sum_{\alpha=1}^{n} t(\mathbf{u}_\alpha) \cdot \bar{\phi}(\mathbf{u}_\alpha) \cdot \bar{s_o}(\mathbf{u}_\alpha)$$

where $t$ is thickness, $\bar{\phi}$ and $\bar{s_o}$ are vertically averaged porosity and oil saturation.



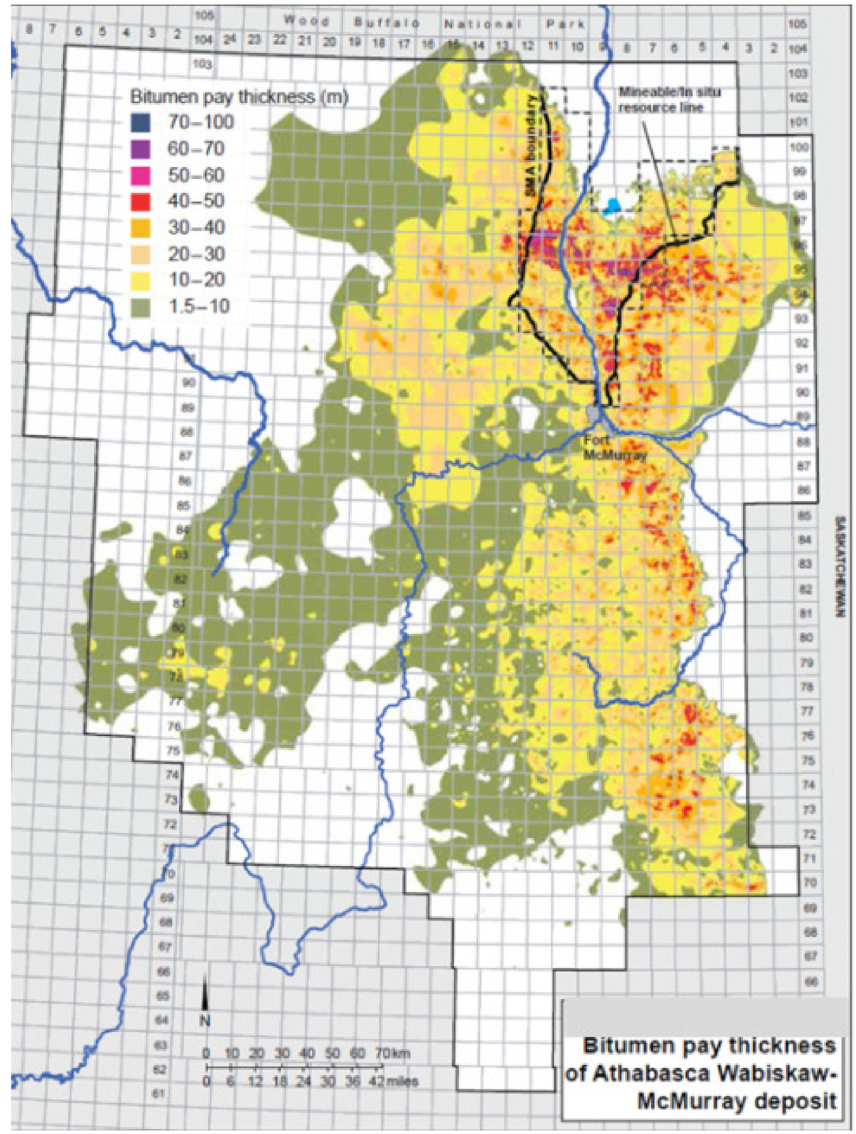

Kriged remaining resource map for Utica Shale, Quebec by the Geological Survey of Canada

(https://www.researchgate.net/publication/263818399_GEOLOGICAL_SURVEY_OF_CANADA_OPEN_FILE_7606_Geological_Characteristics_and_Petroleum_Resource_Assessment_of_Utica_Shale_Quebec_Canada/figures?lo=1&utm_source=google&utm_medium=organic)

# Modeling Workflows

## Regional Mapping

- Goal: Understand the spatial distribution of the resource to evaluate different lease areas, sequencing development, and layout of facilities
- Properties: thickness, permeability height
- Model: very large area (1000’s of km$^2$) with large model cells (low detail), vertical details collapsed to major heterogeneities and model may include 10’s of variables




Regional map of bitumen pay thickness (m) for NW Alberta, Canada from Hein, (2015).

https://www.sciencedirect.com/science/article/pii/B9780444635297000183

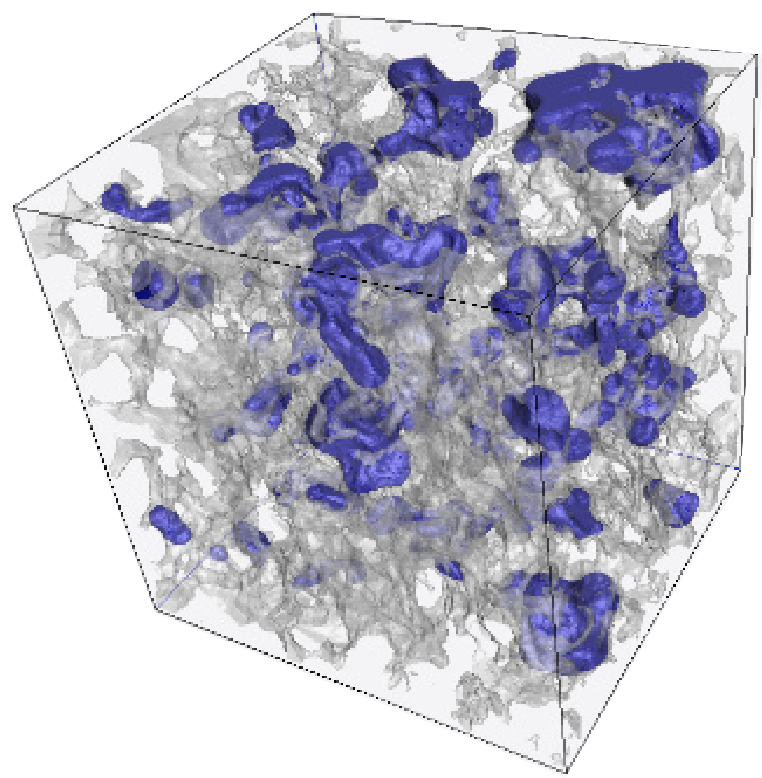
# Modeling Workflows

## Micro / Mini Modeling

- Goal: Build models pore-scale to transfer their influence to reservoir models for forecasting.

- Properties: rock (minerology and fluid models

- Model: for micro the scale of a single core plug / core is typical, for mini the scale of a single reservoir modeling cell



Direct simulation of residual phase (disconnected blobs in blue) in Berea Sandstone (imaged based pore grain surface shown in transparent gray).

https://www.digitalrocksportal.org/

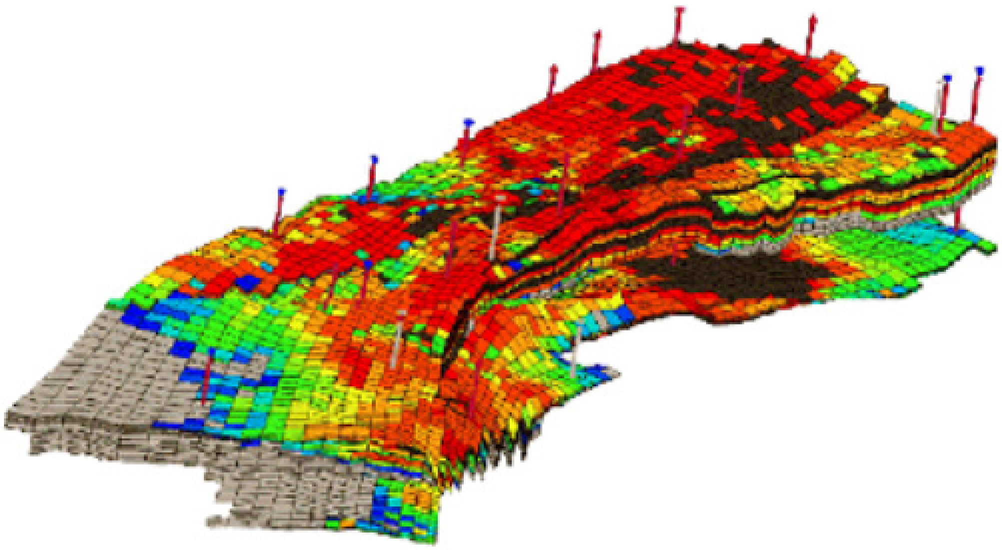
# Modeling Workflows

## Reservoir Modeling

- Goal: Input for connectivity calculations and flow simulation.
- Properties: facies, porosity, permeability, saturation, seismic attributes, pressure and production rates.
- Model: cells of 10’s m areal x 0.25 – 1.0 m vertical extent, over 10’s km x 10’s m.

Reservoir model with color indicating oil saturation (hot colors have more oil).

https://wiki.seg.org/wiki/Reservoir_simulation



# Modeling Workflows

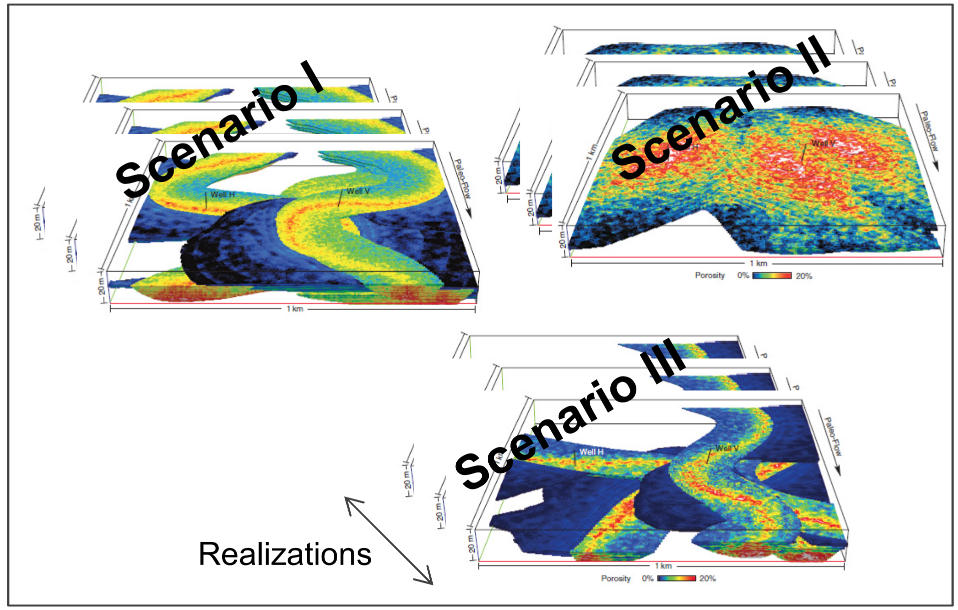
## Reservoir Modeling

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

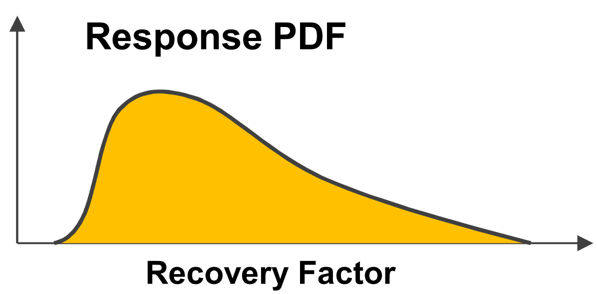
2. Apply all the models to the transfer function

3. Assemble the distribution of the outcomes

4. Make the decision accounting for this uncertainty model.




**Transfer Function**



# Subsurface Modeling Take-aways

| Topic | Application to Subsurface Modeling |
|---|---|
| **Modeling Goals** | Modeling for integration? Flow forecasting? Reserves assessment?<br><br>*There are a wide variety of modeling goals.* |
| **Model Workflows** | Regional, reservoir, minin and micro modeling.<br><br>*Utilize a variety of modeling workflows for a variety of goals.* |
| **Modeling for Discomfort** | Model to test and validate hypothesis<br><br>*Model for discomfort, actively attempt to disprove current theories to prove upside and to mitigate downside risk.* |

# Data Analytics, Geostatistics and Machine Learning
# Fundamamental Concepts

**Fundamental Concepts . .**

- **Workflow Development**

**Instructor: Michael Pyrcz, the University of Texas at Austin**

# Basics of Workflow Design

## Design a Set of Steps to Accomplish the Goal, common steps include:

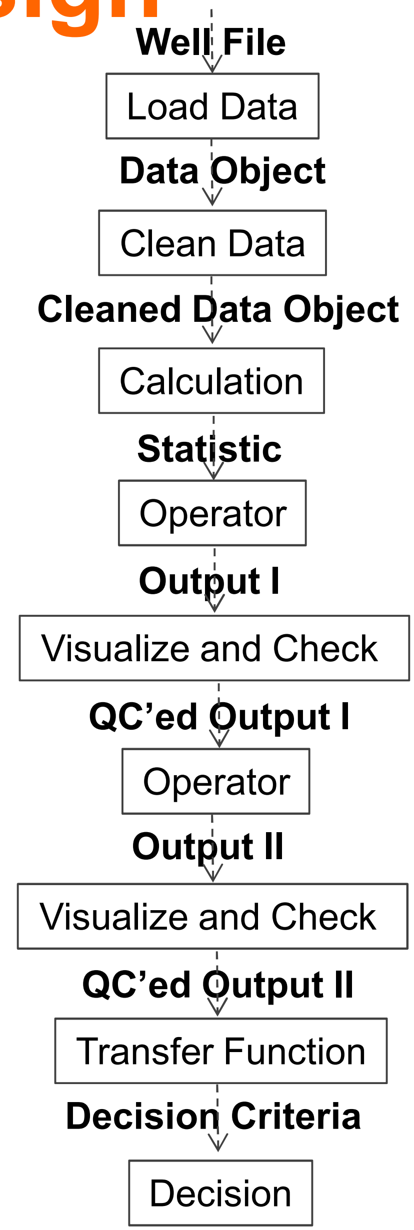
a) Load Data
- Load data into a data structure that we can work with

b) Format / Check / Clean Data
- Get the data ready for the workflow

c) Run Statistical Calculation / Visualization
- Histograms, location maps, variogram, trend, conditional probabilities, data mining

d) Run Operator
- Declustering, spatial continuity, spatial estimation and simulation, model post-processing

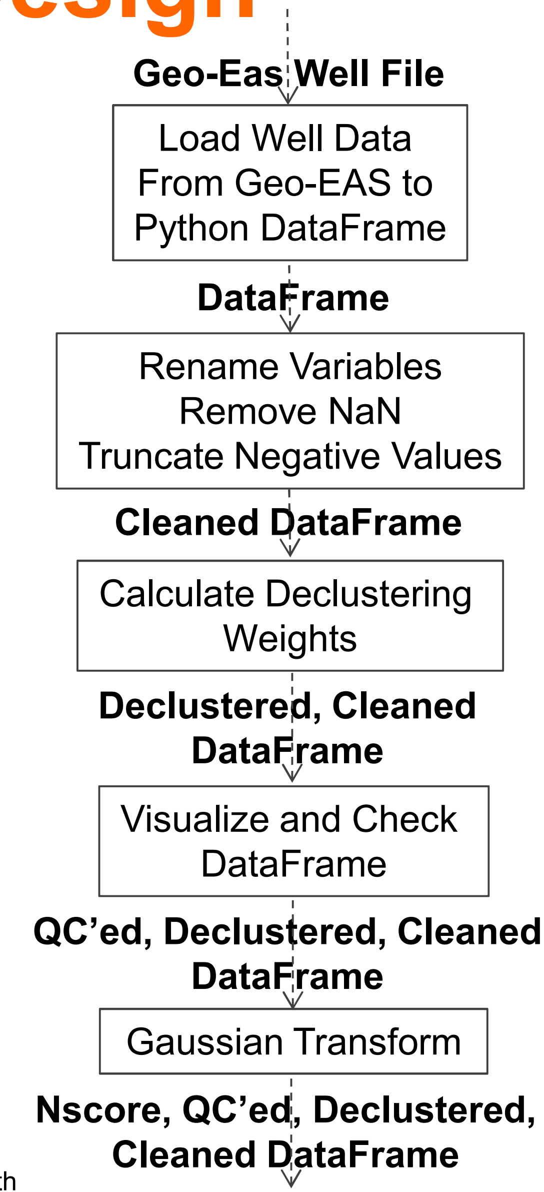
e) Transfer Function
- Any type of summary of the model to support decision making, such as volumetric calculation, connectivity analysis, trend modeling, flow simulation



# Basics of Workflow Design

## 2. Design a Set of Steps to Accomplish the Goal, workflows:

- The data, statistics, and models flow from step to step.
- The well data (data table) and seismic data and other mapped data along with models (gridded) are passed from step to step.
- Complete subsurface modeling workflows are often complicated, many interrelated steps
- The process is often sequence dependent, e.g. data cleaning and transformations before variogram calculation
- Visualization and checking steps are required after every operation




Example subset of a workflow with data flowing over multiple steps.

# Basics of Workflow Design

## 2. Design a Set of Steps to Accomplish the Goal, documentation:

- Every step includes expert decisions.
- **Writing script / code for your workflow provides a very useful audit trail / documentation.**
- Takes 2-3 times longer, but given most subsurface modeling workflows are iterated multiple times it is well worth it.
- It is essential to document the many modeling decisions.
- You are writing to your future self or your replacement.

**Documentation**

Wells provided by ______ on Date _______. Includes the first ____ wells and excludes _______. Porosity by neutron density log calibrated to core measures.

Values < ____ are considered below measurement threshold and set to 0.001 porosity.

Cell size for declustering weights selected to be 200m based on nominal well spacing.

Weights have range [0.01,2.5] and are spatially rational.

Transform with tail assumption of [0.001, 0.30] linear extrapolation.

Example subset of a workflow with data flowing over multiple steps with documentation.

**Geo-Eas Well File**

Load Well Data From Geo-EAS to Python DataFrame

**DataFrame**

Rename Variables
Remove NaN
Truncate Negative Values

**Cleaned DataFrame**

Calculate Declustering Weights

**Declustered, Cleaned DataFrame**

Visualize and Check DataFrame

**QC'ed, Declustered, Cleaned DataFrame**

Gaussian Transform

**Nscore, QC'ed, Declustered, Cleaned DataFrame**

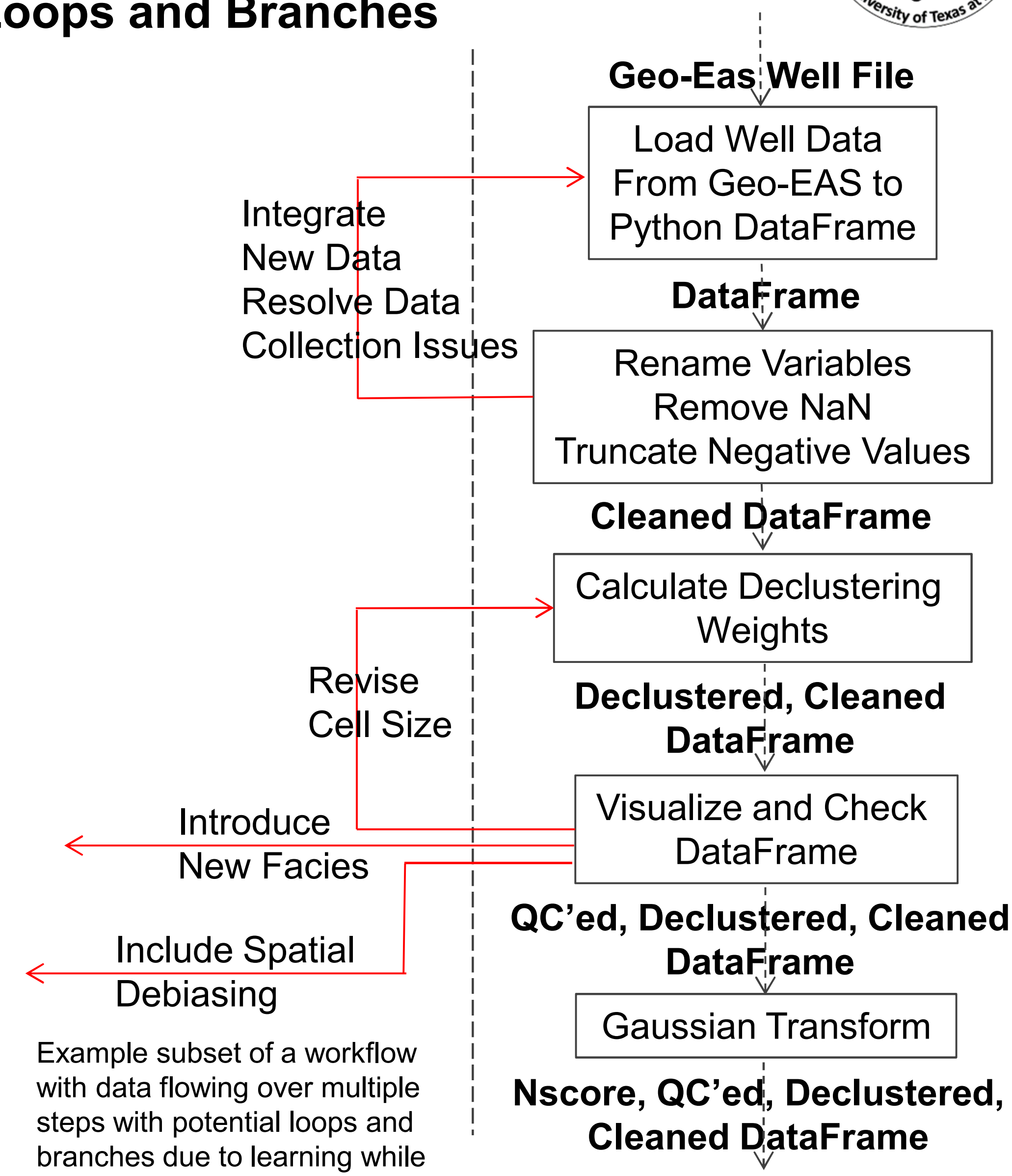
# Basics of Workflow Design

**Loops and Branches**

## Design a Set of Steps to Accomplish the Goal, branches, loops and iteration:

It is practical to represent modeling as a linear workflow

- This is not best practice
- We learn new things with every step – **learning while modeling**
- We may refine / modify the workflow due to new information
- We may cycle back to a previous step




Example subset of a workflow with data flowing over multiple steps with potential loops and branches due to learning while modeling.

# Basics of Workflow Design

## Design a Set of Steps to Accomplish the Goal, uncertainty:

- Data, statistical summary, models all have uncertainty
- Significant uncertainty sources must be investigated and integrated into the workflow
- The fundamental methods are :
  - *scenarios* – change the modeling decisions / inputs
  - *realizations* – hold modeling decisions and inputs constant and just change the random number seed

**Uncertainty**

**Porosity mean and variance P10, P50, P90**

**Imputed missing data as random variables.**

**Optimum declustering hyper parameters**

**Representative porosity mean, variance**

**Geo-Eas Well File**

Load Well Data From Geo-EAS to Python DataFrame

**DataFrame**

Rename Variables
Remove NaN
Truncate Negative Values

**Cleaned DataFrame**

Calculate Declustering Weights

**Declustered, Cleaned DataFrame**

Visualize and Check DataFrame

**QC'ed, Declustered, Cleaned DataFrame**

Gaussian Transform

**Nscore, QC'ed, Declustered, Cleaned DataFrame**

Example subset of a workflow with data flowing over multiple steps with uncertainties.

# Coding to Support Modeling Workflow Development

## More on Software / Coding:

- This is not a coding / software workshop.
- I can't teach Python in 1 day.
- We will demonstrate well-documented workflows in Python.
- We will focus on the steps, inputs and outputs.
- Don't be concerned if you don't completely understand the code.

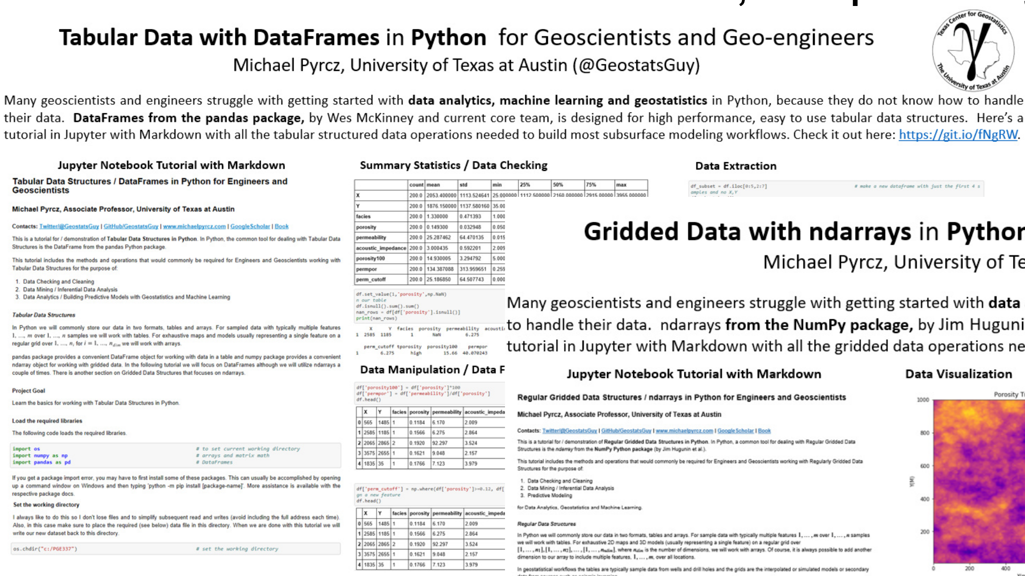
## But if coding to solve subsurface problems is your interest:

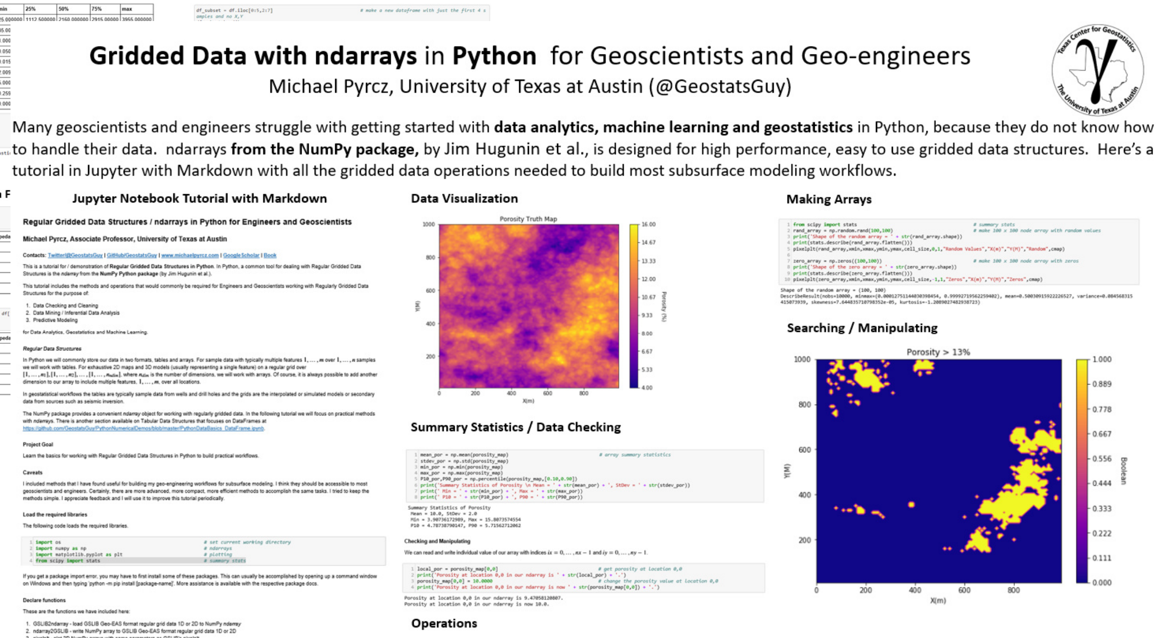
- I teach that also!
- I have a lot of basic tutorials and example workflows to solve practical subsurface problems
- Will conduct a Data Science Bootcamp starting this summer with Professor John Foster from the Institute for Computational Engineering and Science (ICES) at the University of Texas at Austin.

# Example Python Coding Demonstrations / Tutorials

## Python Basics

- Numpy for arrays (ndarrays) for gridded data
- Pandas for tabulated data (DataFrames) for well data
- Tutorials available on GitHub, I'm planning to cover in class?



# Aside: Benefit of Coding

## Reasons All Geoscientists and Engineers Should Learn to Code

**Transparency** – *no compiler accepts hand waiving!* Coding forces your logic to be uncovered for any other scientist or engineer to review.

**Reproducibility** – *run it, get an answer, hand it over, run it, get the same answer.* This is a main principle of the scientific method.

**Quantification** – *programs need numbers.* Feed the program and discover new ways to look at the world.

**Open-source** – *leverage a world of brilliance.* Check out packages, snippets and be amazed with what great minds have freely shared.

**Break Down Barriers** – *don't throw it over the fence*. Sit at the table with the developers and share more of your subject matter expertise for a better product.

**Deployment** – *share it with others and multiply the impact*. Performance metrics or altruism, your good work benefits many others.

**Efficiency** – *minimize the boring parts of the job*. Build a suite of scripts for automation of common tasks and spend more time doing science and engineering!

**Always Time to Do it Again!** – *how many times did you only do it once?* It probably takes 2-4 times as long to script and automate a workflow. Usually worth it.

**Be Like Us** – *it will change you*. Users feel limited, programmers truly harness the power of their applications and hardware.

# Aside: Benefit of Coding

**Reasons All Geoscientists and Engineers Should Learn to Code**
**Some Caveats**

1. Any type of coding, scripting, workflow automation matched to your working environment is great. We don't all need to be C++ experts.

2. I respect the experience component of geoscience and engineering expertise. This is beyond coding and is essential to workflow logic development, best use of data etc.

3. Some expert judgement will remain subjective and not completely reproducible. I'm not advocating for the geoscientist or engineer being replaced by a computer.

# The Subsurface Modeling Steps

- We could spend more time together! I do a lot of training, e.g. 3 day course:
  - Some Prerequisites
  - Data Preparation
  - Univariate and Multivariate Analysis
  - Spatial Analysis
  - Estimation and Trend Modeling
  - Stochastic Simulation
  - Uncertainty Analysis
  - Model Checking
  - Decision Making
- For each section
  - Lectures and demos
  - Subsurface inference and modeling
  - Completion of project update documentation and presentation

Tailored to geoscientists, engineers, managers etc. Talk to the Midland, TX Geoscientists.

# Data Analytics, Geostatistics and Machine Learning
## Fundamamental Concepts

**Fundamental Concepts . .**

- **What is Subsurface Modeling?**
- **Modeling Goals**
- **Modeling Strategies**
- **Workflow Development**

**Instructor: Michael Pyrcz, the University of Texas at Austin**