Data Analytics, Geostatistics and Machine Learning

Uncertainty Modeling

Lecture outline . . .

- Sources of Uncertainty
- Representing Uncertainty
- Decision Making
- Checking Uncertainty Models

Instructor: Michael Pyrcz, the University of Texas at Austin

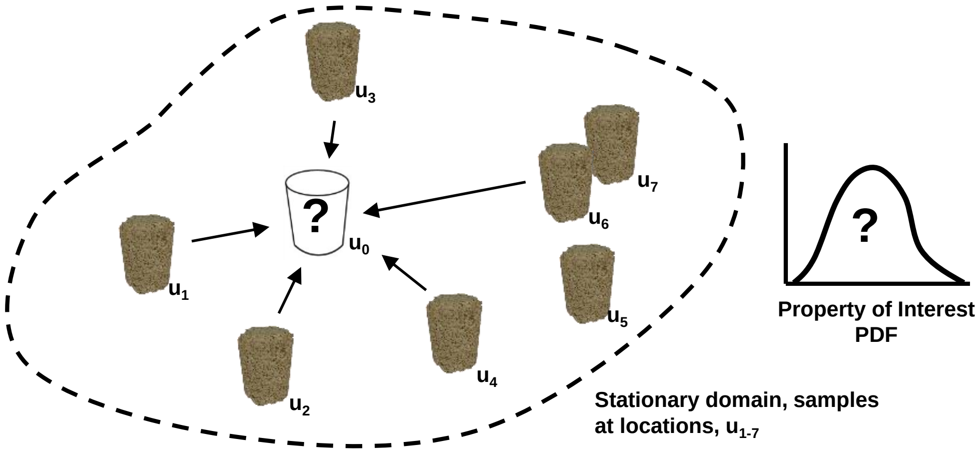
Data Analytics, Geostatistics and Machine Learning
Uncertainty Modeling

Lecture outline . . .

- Sources of Uncertainty

Instructor: Michael Pyrcz, the University of Texas at Austin

Uncertainty
What is uncertainty?



Uncertainty is not an intrinsic property of the subsurface.

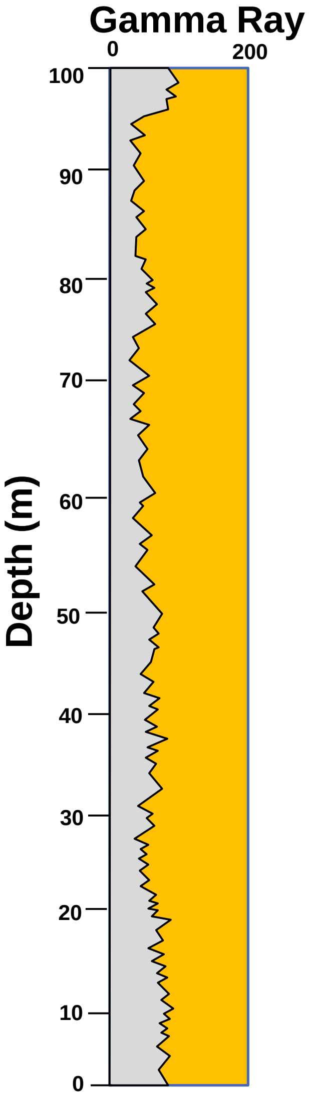
- At every location ($\mathbf{u}_\alpha$) within the volume of interest the true properties could be measured if we had access (facies, porosity etc.).
- **Uncertainty is a function of our ignorance**, our inability to observed and measure the subsurface with the coverage and scale required to support our scientific questions and decision making.

sparsity of sample data + heterogeneity = uncertainty

- If the subsurface was homogeneous, with a few measurements uncertainty would be reduced and estimates resolved to a sufficient degree of exactitude.

Types of Uncertainty

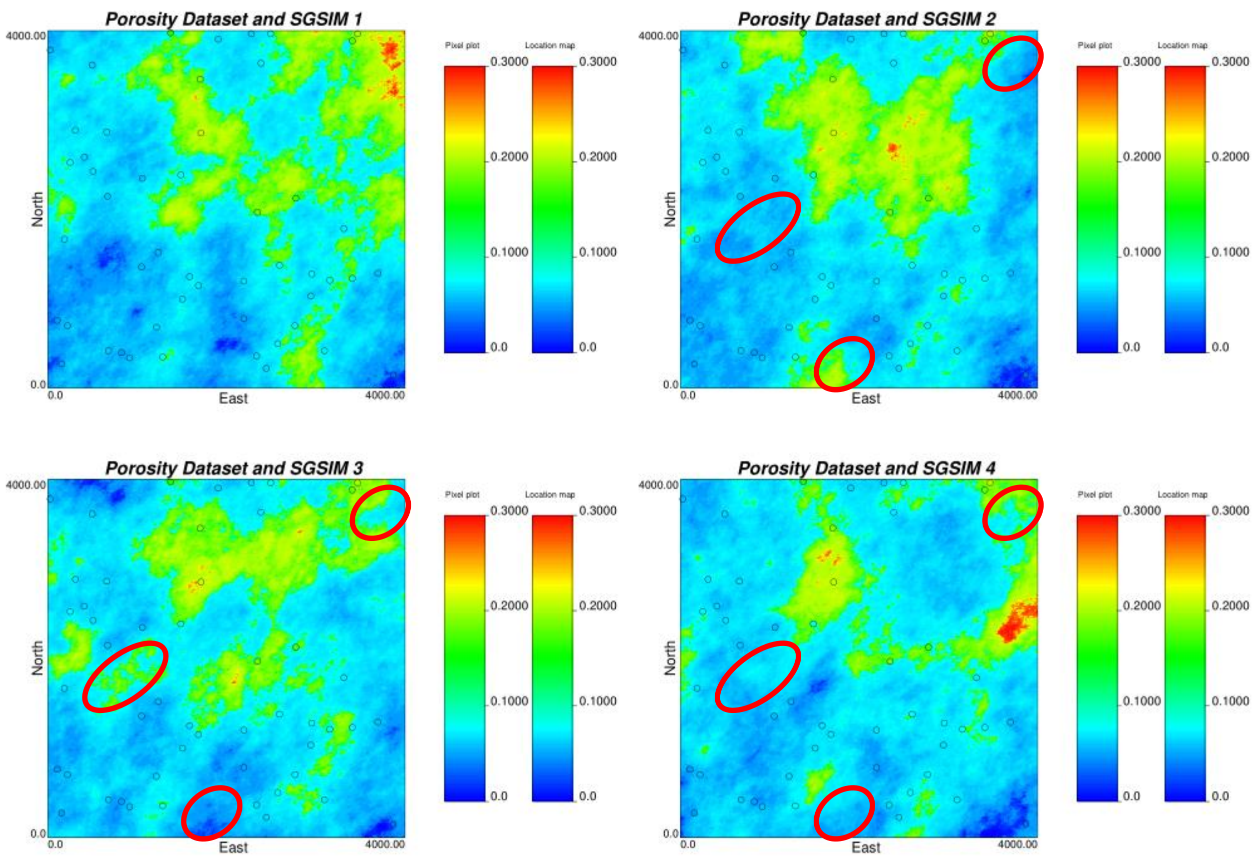
- **Measurement / Interpretation Error.**
  - Formation evaluation – tool tolerance, calibration error, approximations / assumptions
  - Interpreter experience and prior model / assumptions
  - How to integrate it?
    - » Indicator method code as soft inputs
    - » Multiple data realizations in design of experiments



Example well log.

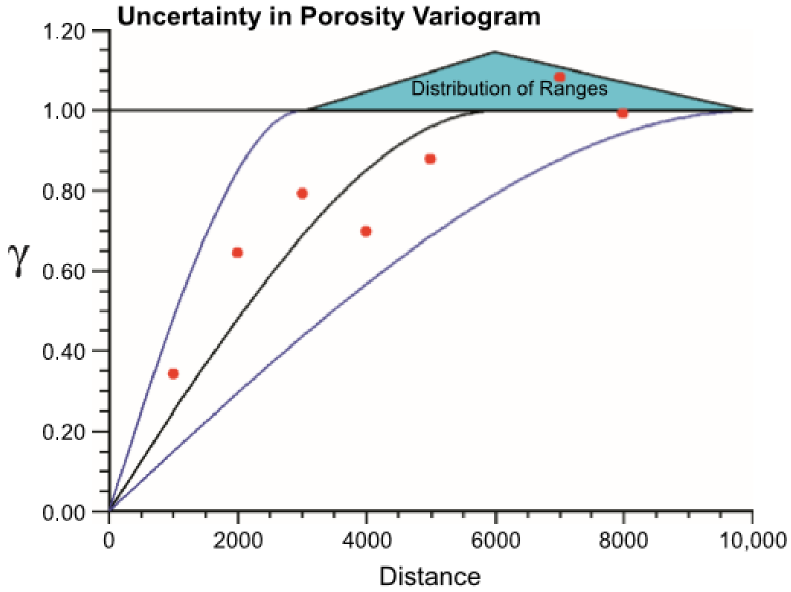
Types of Uncertainty

- **Spatial Uncertainty**
  - Uncertainty due to spatial offset from sampled locations
  - Integrate through multiple local realizations and scenarios

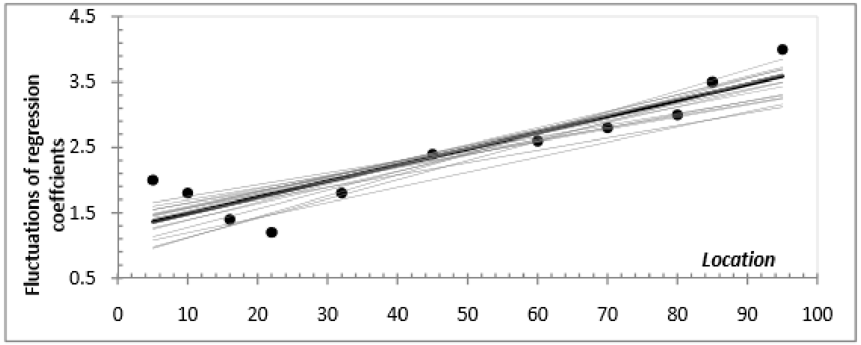


Types of Uncertainty

- **Parameter Uncertainty**
  - Uncertainty in the input statistics to constrain the model area
  - E.g. global reference porosity distribution for simulation
  - Formulate distribution scenarios and bootstrap realizations



Distribution of Variogram Ranges (Pyrcz et al., 2006)



Trend Uncertainty (Villalba and Deutsch., 2010)

Data Analytics, Geostatistics and Machine Learning
Uncertainty Modeling

Lecture outline . . .

- Representing Uncertainty

Introduction

Fundamental Concepts

Probability

Data Prep / Analytics

Spatial Continuity / Prediction

Multivariate Modeling

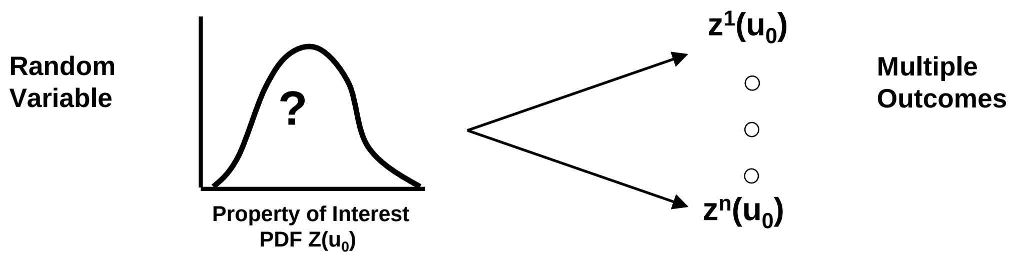
Uncertainty Modeling

Machine Learning

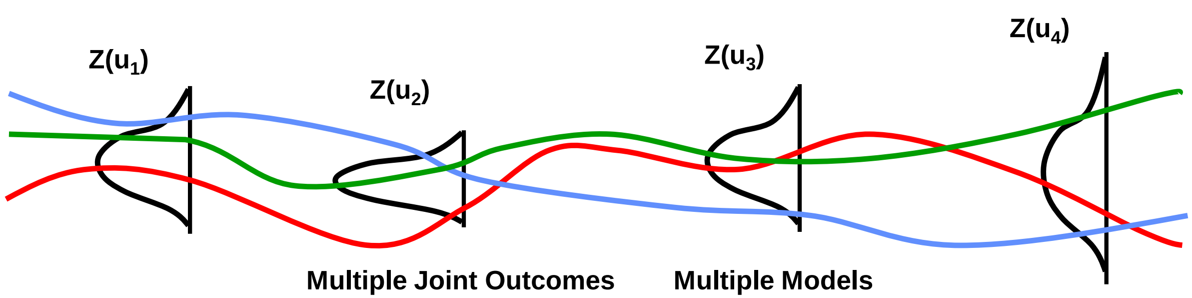
Instructor: Michael Pyrcz, the University of Texas at Austin

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.

Scenarios: when the input decisions and parameters are changed

Captures interpretation and data uncertainty.

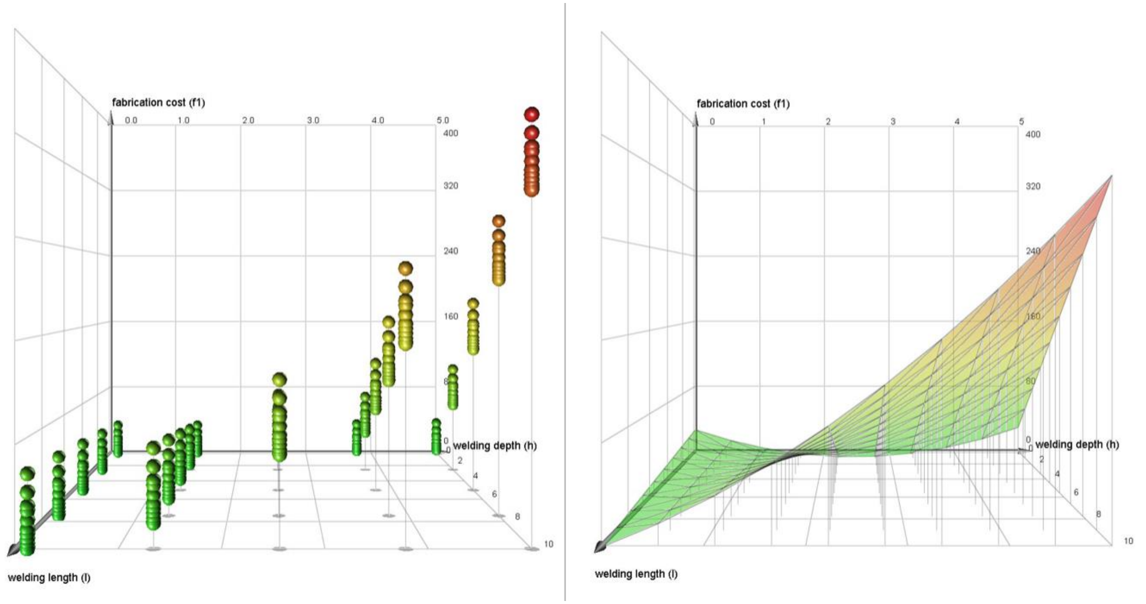
Realizations: when the input decisions and parameters are held constant and only the random number seed is changed

Captures spatial uncertainty.

Working With Multiple Models: It is generally not appropriate to analyze a single or few scenarios and realizations.

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.).

Sampling the Uncertainty Space



(Wikipedia)

- **Exhaustive Sampling:**
  - sample from each of the inputs jointly (accounting for multivariate features) and build a large suite of models
  - the uncertainty space is usually large / curse of dimensionality
    - » consider some form of stratified or direct sampling

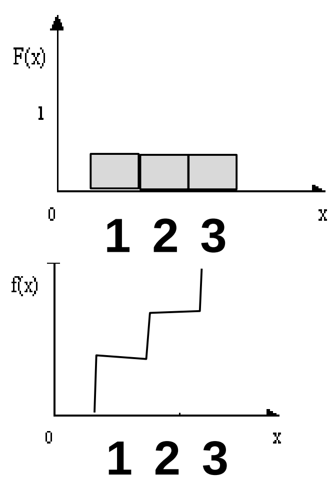
- **Design of Experiments**
  - Sample and model a response surface (Plackett-Burman, full combinatorial)
  - Careful selection of sampling method, prescreening methods
  - Model response surface, statistical function, simulate large set

How is Uncertainty Represented?

- **We have a PDF / CDF of a measure over a volume: therefore our measures of spread / dispersion are our measures of uncertainty.**

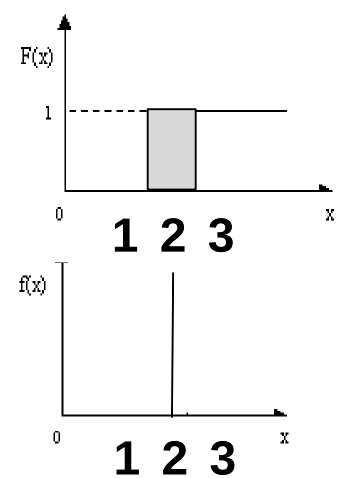
Variance	$\text{Var}(Z) = \int_{-\infty}^{\infty} (z-m)^2 f(z)dz$	Expected squared difference from mean.
Dispersion Variance	$D^2(v,V) = \bar{\gamma}_{V,V} - \bar{\gamma}_{v,v}$	Generalized variance accounting scale and heterogeneity.
Entropy	$H(Z) = -\sum_{p}^{n} P(Z_i) \cdot ln\, P(Z_i)$	Measure of uncertainty for categorical Variables.

Discrete distribution



maximum entropy

Discrete distribution



minimum entropy

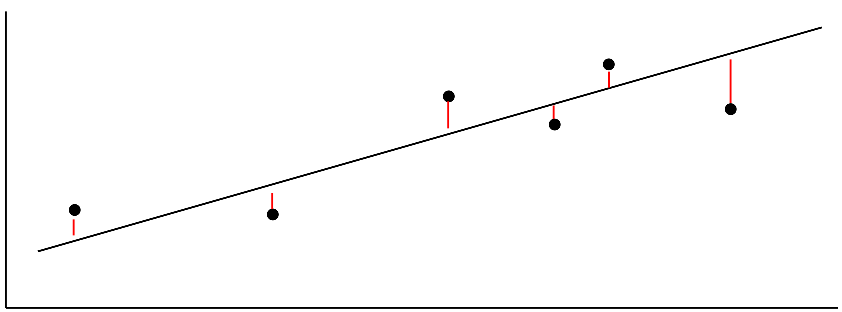
How is Uncertainty Represented?

- **Variance is partitioned between trend, deterministic, known and residual, stochastic, unknown.**

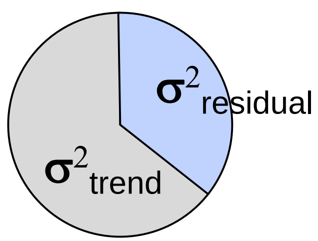
Variance Components

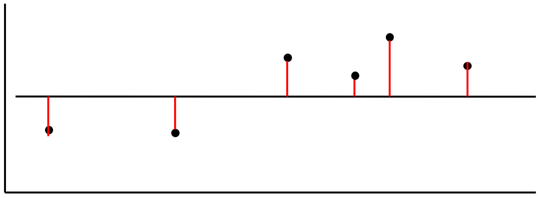
$$\sigma^2 = \sigma_t^2 + \sigma_r^2 + 2\,\cancelto{0}{C_{t,r}(0)}$$

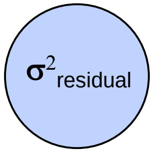
Total variance = Deterministic / Known Variance + Stochastic / Unknown Variance

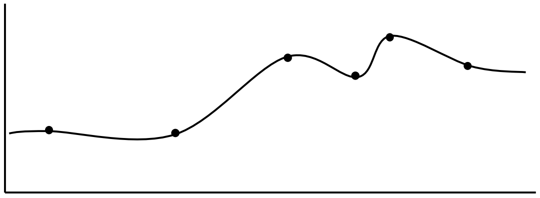


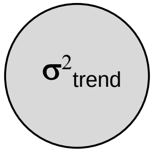
Variance Partitions









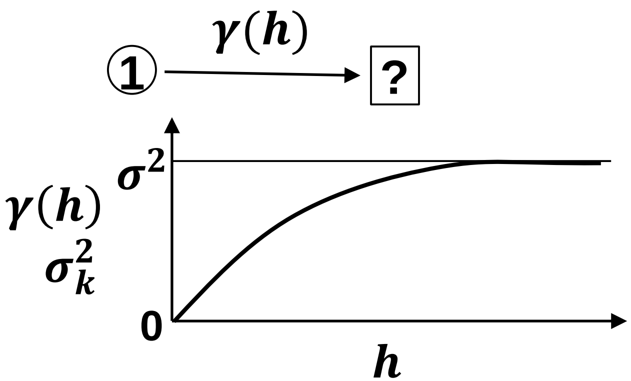


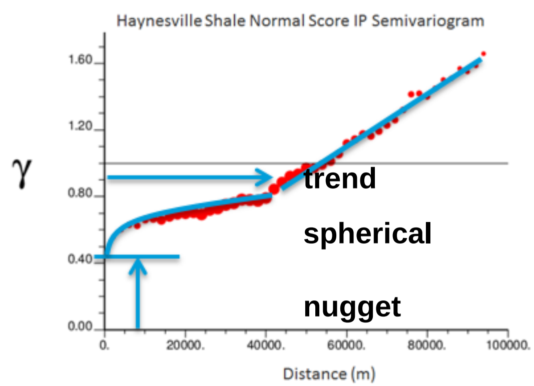
How is Uncertainty Represented?

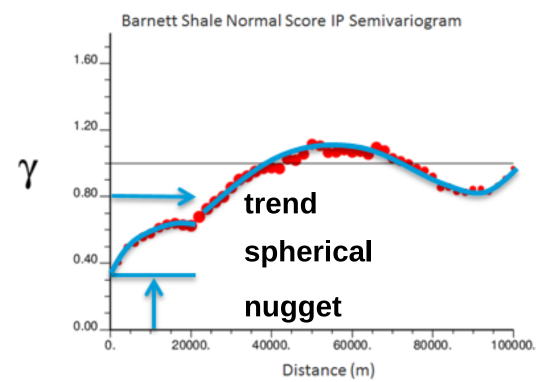
- **Variance is partitioned spatially with variogram features.**

Simple Kriging Est. Var. $\sigma_K^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j C(u_i,u_j) - 2\sum_{i=1}^{n}\lambda_i C(u,u_i) + C(0)$

For one neighboring data: $\sigma_k^2 = \sigma^2 - \frac{C(h)}{\sigma^2} = 1 - C(h) = \gamma(h)$ (if $\sigma^2 = 1$)



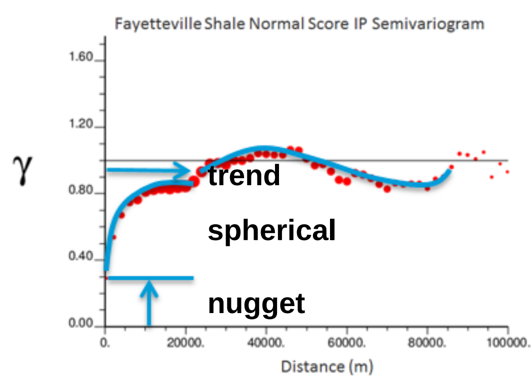




Nested structures each describe spatial uncertainty components.

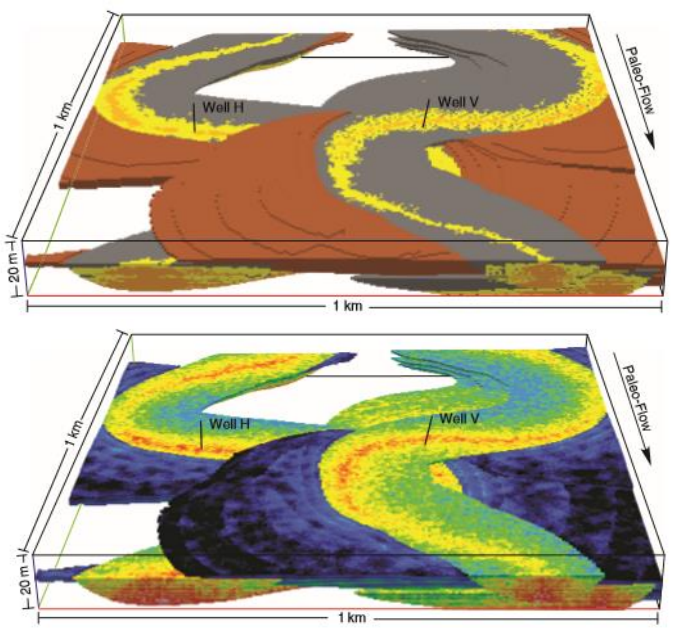
$$\gamma(h) = \gamma(h)_{nugget} + \gamma(h)_{spherical} + \gamma(h)_{trend}$$

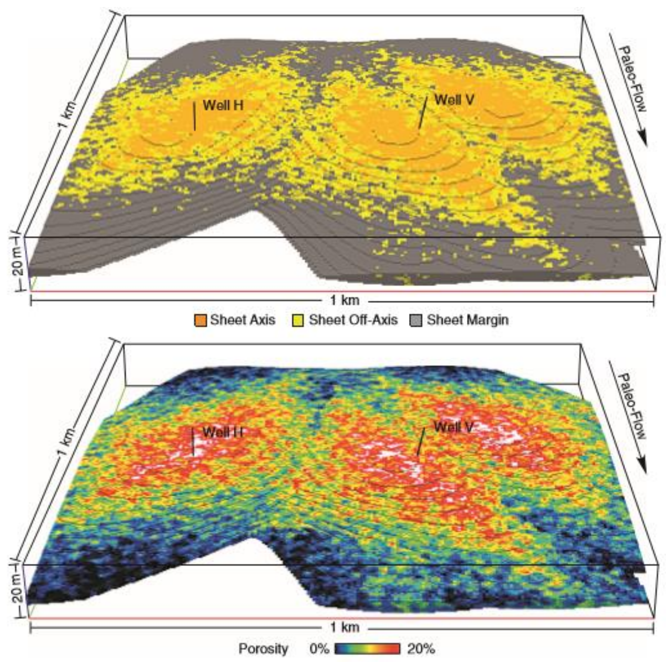
$$\sigma_k^2 = \sigma^2_{nugget} + \sigma^2_{spherical} + \sigma^2_{trend}$$



How is Uncertainty Represented?

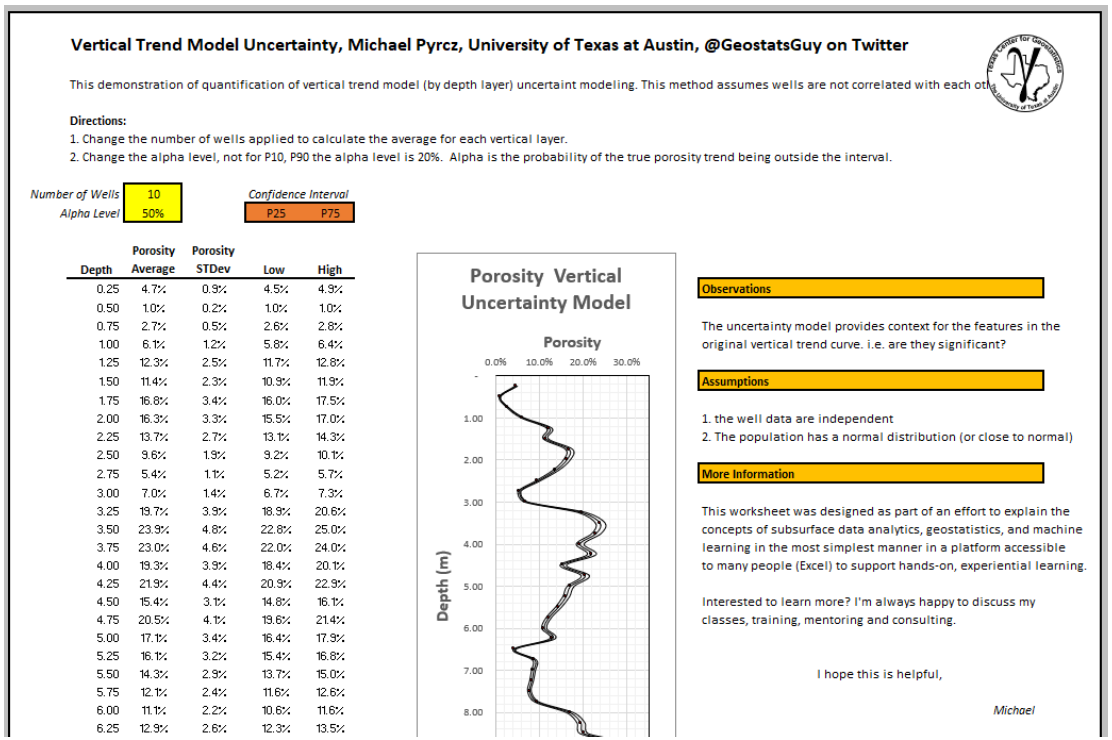
- **Discrete Cases**
  - At times it is NOT possible to represent uncertainty as a continuous distribution.
  - **Discrete scenarios are required**
    - » Porosity compaction trend - yes or no.
    - » Channelized or lobe uncertainty





Vertical Trend
Hands-on in Excel

Experiment with Uncertainty in a Vertical Trend Model:



Vertical Trend Model Uncertainty, Michael Pyrcz, University of Texas at Austin, @GeostatsGuy on Twitter

This demonstration of quantification of vertical trend model (by depth layer) uncertaint modeling. This method assumes wells are not correlated with each ot

Directions:
1. Change the number of wells applied to calculate the average for each vertical layer.
2. Change the alpha level, not for P10, P90 the alpha level is 20%. Alpha is the probability of the true porosity trend being outside the interval.

Number of Wells 10
Alpha Level 50%

Confidence Interval P25 P75

Depth	Porosity Average	Porosity STDev	Low	High
0.25	4.7%	0.9%	4.5%	4.9%
0.50	1.0%	0.2%	1.0%	1.0%
0.75	2.7%	0.5%	2.6%	2.8%
1.00	6.1%	1.2%	5.8%	6.4%
1.25	12.3%	2.5%	11.7%	12.8%
1.50	11.4%	2.3%	10.9%	11.9%
1.75	16.8%	3.4%	16.0%	17.5%
2.00	16.3%	3.3%	15.5%	17.0%
2.25	13.7%	2.7%	13.1%	14.3%
2.50	9.6%	1.9%	9.2%	10.1%
2.75	5.4%	1.1%	5.2%	5.7%
3.00	7.0%	1.4%	6.7%	7.3%
3.25	19.7%	3.9%	18.9%	20.6%
3.50	23.9%	4.8%	22.8%	25.0%
3.75	23.0%	4.6%	22.0%	24.0%
4.00	19.3%	3.9%	18.4%	20.1%
4.25	21.9%	4.4%	20.9%	22.9%
4.50	15.4%	3.1%	14.8%	16.1%
4.75	20.5%	4.1%	19.6%	21.4%
5.00	17.1%	3.4%	16.4%	17.9%
5.25	16.1%	3.2%	15.4%	16.8%
5.50	14.3%	2.9%	13.7%	15.0%
5.75	12.1%	2.4%	11.6%	12.6%
6.00	11.1%	2.2%	10.6%	11.6%
6.25	12.9%	2.6%	12.3%	13.5%

Observations

The uncertainty model provides context for the features in the original vertical trend curve. i.e. are they significant?

Assumptions

1. the well data are independent
2. The population has a normal distribution (or close to normal)

More Information

This worksheet was designed as part of an effort to explain the concepts of subsurface data analytics, geostatistics, and machine learning in the most simplest manner in a platform accessible to many people (Excel) to support hands-on, experiential learning.

Interested to learn more? I'm always happy to discuss my classes, training, mentoring and consulting.

I hope this is helpful,

Michael

Things to try:

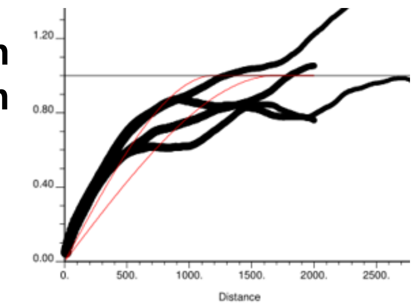
1. Increase and decrease the number of wells.
2. Change the alpha level.

The file is Vertical_Trend_Uncertainty_Demo.xlsx at https://git.io/fhAHX

How is Uncertainty Represented?

- **Ergodic fluctuations – Can't leave uncertainty to the algorithm!**
  - In the "old days" ergodic fluctuations were treated as a comprehensive uncertainty model in input statistics.
  - Function of variogram range, data conditioning and size of model area.
  - Can be predicted:

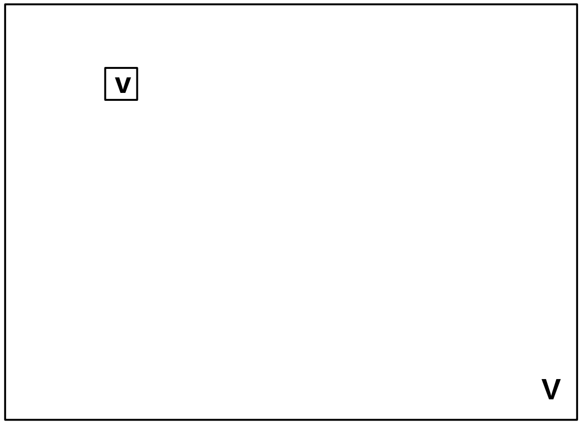
Fluctuations in Variogram



$$D^2(v,V) = E\left\{\left(\underset{\text{Support v}}{z_i} - \underset{\text{Support V}}{m_i}\right)^2\right\}$$

Generalize Variance Definition

$$D^2(v,V) = \bar{\gamma}(V,V) - \bar{\gamma}(v,v)$$

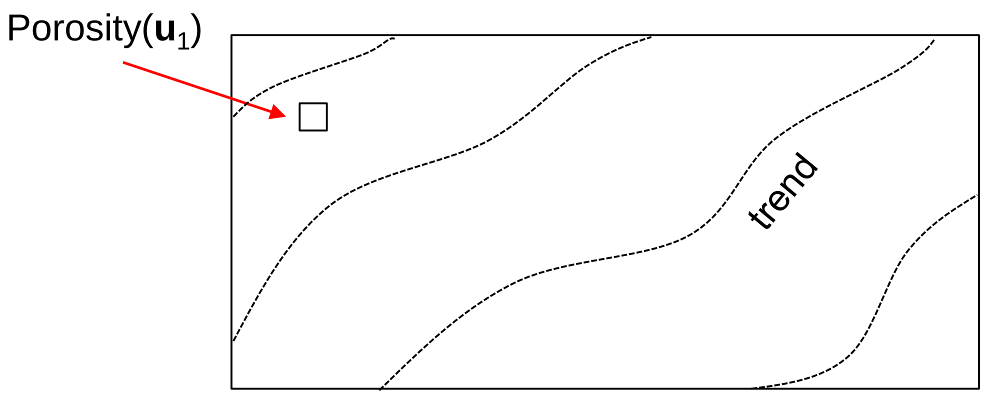


- Does not completely account for uncertainty in inferring input statistics.
- We must take ownership of our inputs and associated uncertainty models.

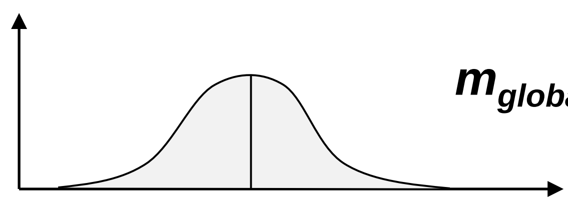
How is Uncertainty Calculated?

Simple example:

– **Uncertainty in porosity estimate at location $\mathbf{u_1}$.**



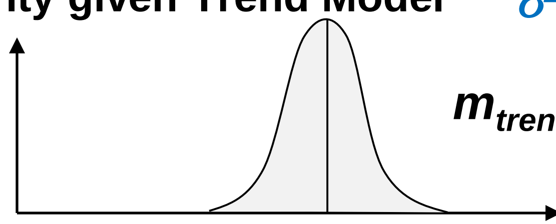
1. **Global Uncertainty from Scenario: Depositional Setting Deepwater Lobe**



$$m_{global}, \sigma^2_{global}\ (\mathrm{Case}_{\mathrm{Lobe}})$$

2. **Local Uncertainty given Trend Model**

$$\sigma^2_{residual}(\mathbf{u}_1) = \sigma^2_{global}(\mathbf{u}_1) - \sigma^2_{trend}(\mathbf{u}_1)$$

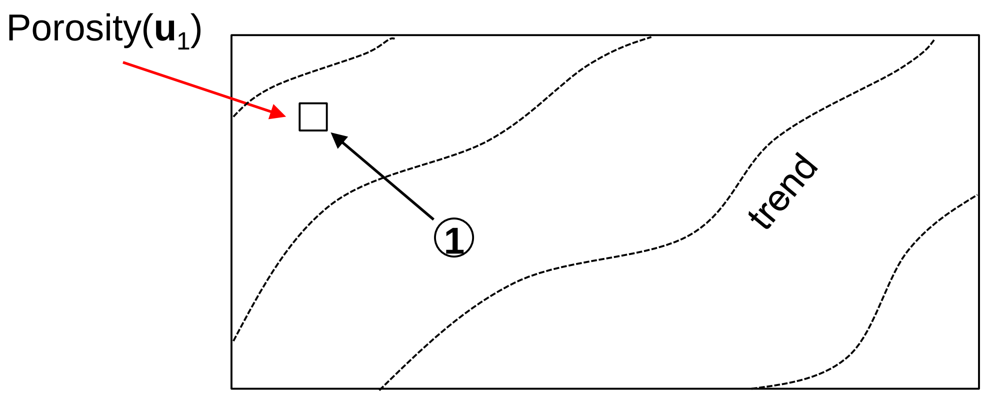


$$m_{trend}(\mathbf{u}_1),\ \sigma^2_{residual}(\mathbf{u}_1)$$

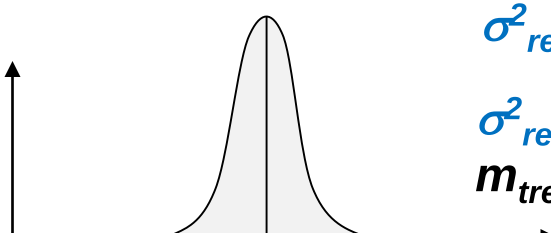
How is Uncertainty Calculated?

Simple example:

– **Uncertainty in porosity estimate at location u_1.**



3. Local Uncertainty given Trend Model and Conditioning Data



$$\sigma^2_{residual\,|\,data}(u_1) \sim \sigma^2_{residual}(u_1) - C(h)$$

$$\sigma^2_{residual\,|\,data}(u_1) = \sigma_k^2$$

$$m_{trend\,|\,data}(u_1) = Z^*$$

$$m_{trend\,|\,data}(u_1),\ \sigma^2_{residual\,|\,data}(u_1)$$

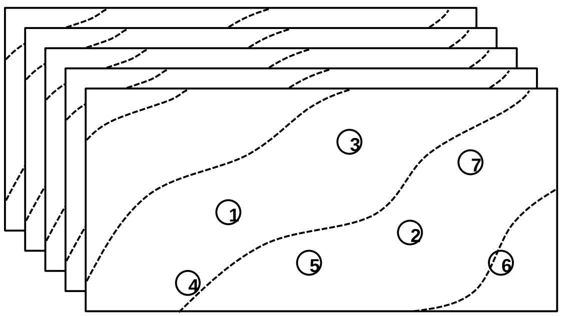
Uncertainty at u_1 for porosity given $\text{Case}_{\text{Lobe}}$**, trend model and conditioning data.**

How is Uncertainty Calculated?

Sample uncertainty through modeling scenarios and realizations:

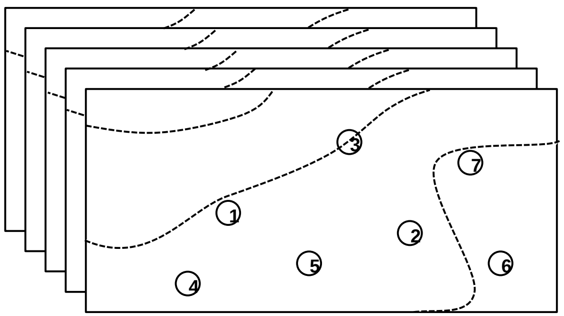
– We can ask any question of the model by considering all scenarios and realizations jointly.

Scenario 1



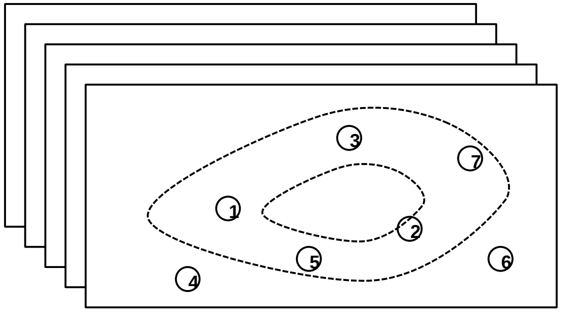
1,…,L realizations (all inputs the same)

Scenario 2



1,…,L realizations (all inputs the same)

Scenario 3



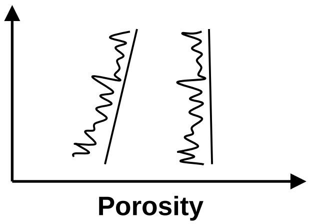
1,…,L realizations (all inputs the same)

How is Uncertainty Calculated?

Modeling Subsurface Uncertainty Without Scenarios

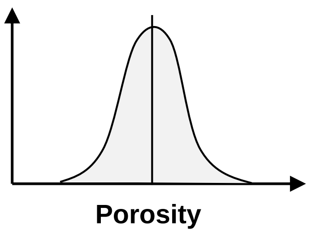
Implicitly assume perfect knowledge of:

Primary and Secondary Data



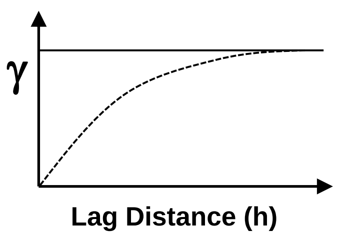
conditioning data - results in reduced local uncertainty

Univariate Distributions



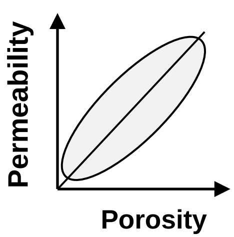
mean porosity - results in reduced volumetric uncertainty

Spatial Continuity



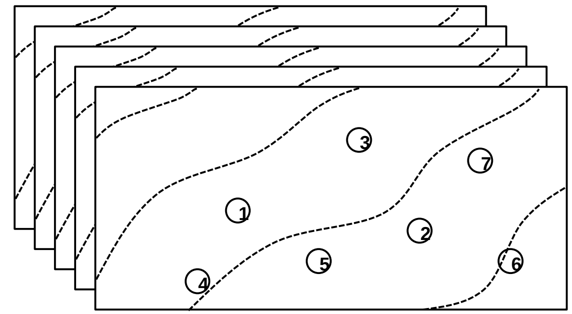
spatial heterogeneity - results in reduced recovery uncertainty

Bivariate Distributions



bivariate relationships - results in reduced local uncertainty and recovery uncertainty

Scenario 1



1,…,L realizations (all the same inputs)

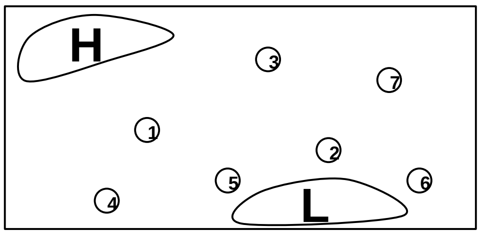
Relying on ergodic fluctuations to characterize uncertainty in the input statistics.

How is Uncertainty Calculated?

Working without realizations?:

– Implicitly assume perfect knowledge:

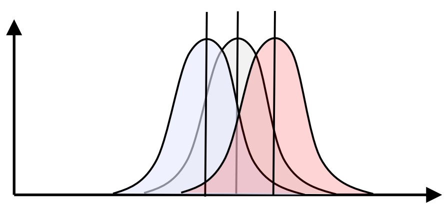
Spatial Uncertainty



Porosity Realization

Spatially, away from data. Freeze stochastic islands away from data!

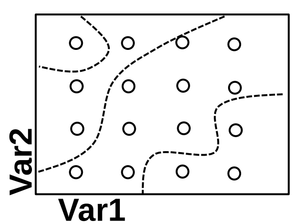
Input Statistic Fluctuations



Porosity Output Distribution

Ergodic fluctuations from target statistics.

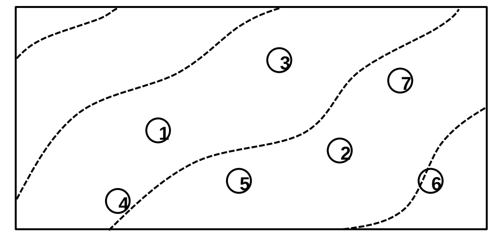
Response Surface



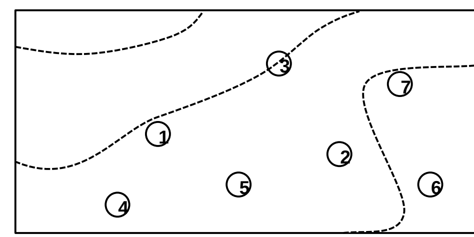
Response Surface

Ergodic fluctuations transfer to response surface creating additional undulations.

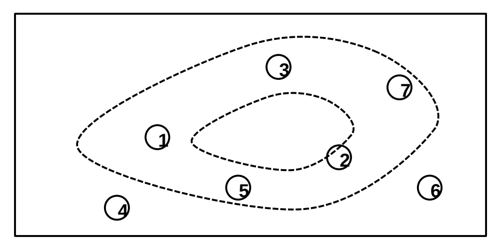
Scenario 1



Scenario 2



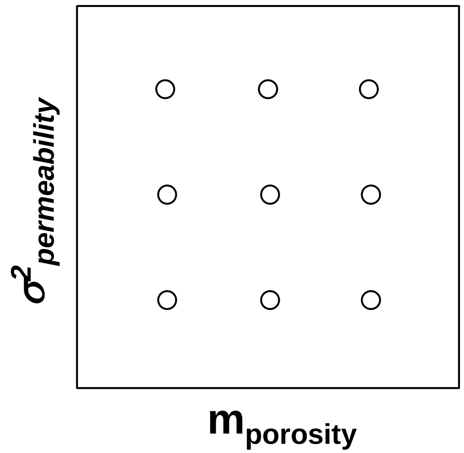
Scenario 3



How is Uncertainty Calculated?

Uncertainty Space is Vast!

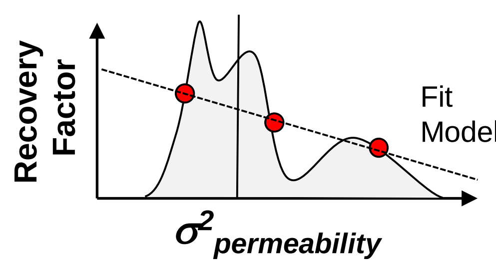
- Consider typical design



3 level design, 2 variables
level^{var} = 9 scenarios

- Consider there are typically are around 10 or more uncertain variables.
  - » 3^{10} = 59,049 scenarios x 10 realizations of each scenario = 590,490 models
  - » Variable screening is important!
  - » 3 level may still poor sampling

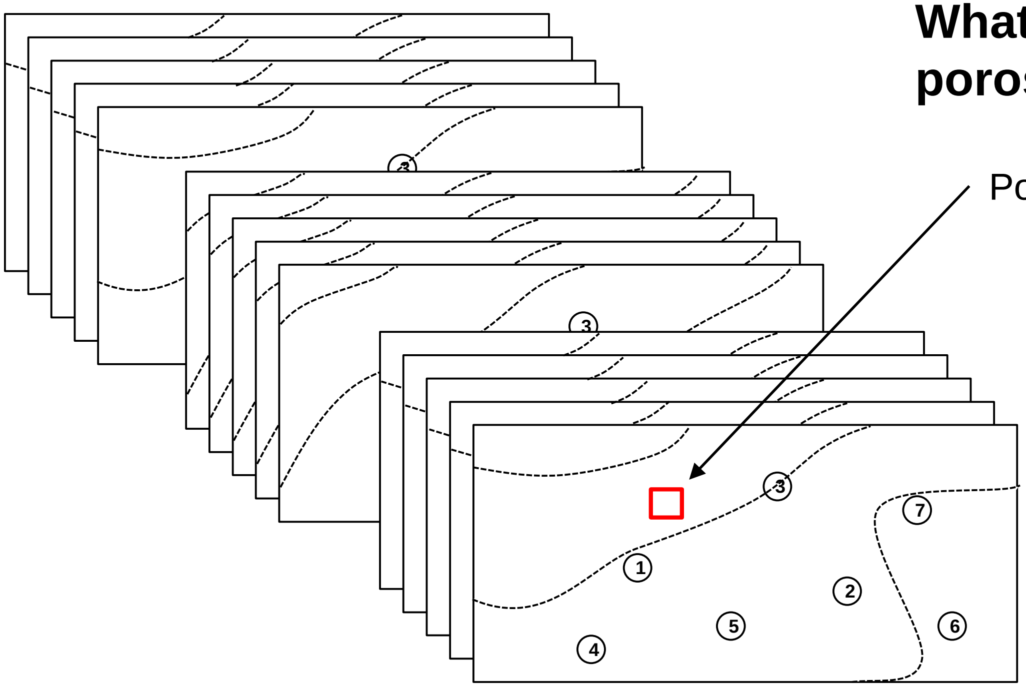
Univariate Response



How is Uncertainty Calculated?

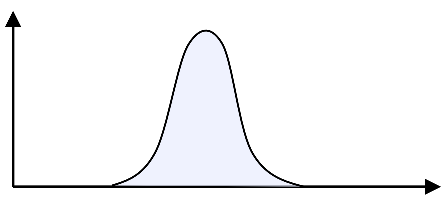
We have represented the "uncertainty model" through scenarios and realizations:

- **We can ask any question of the model by considering all scenarios and realizations jointly.**



What is the uncertainty in porosity at location u_1?

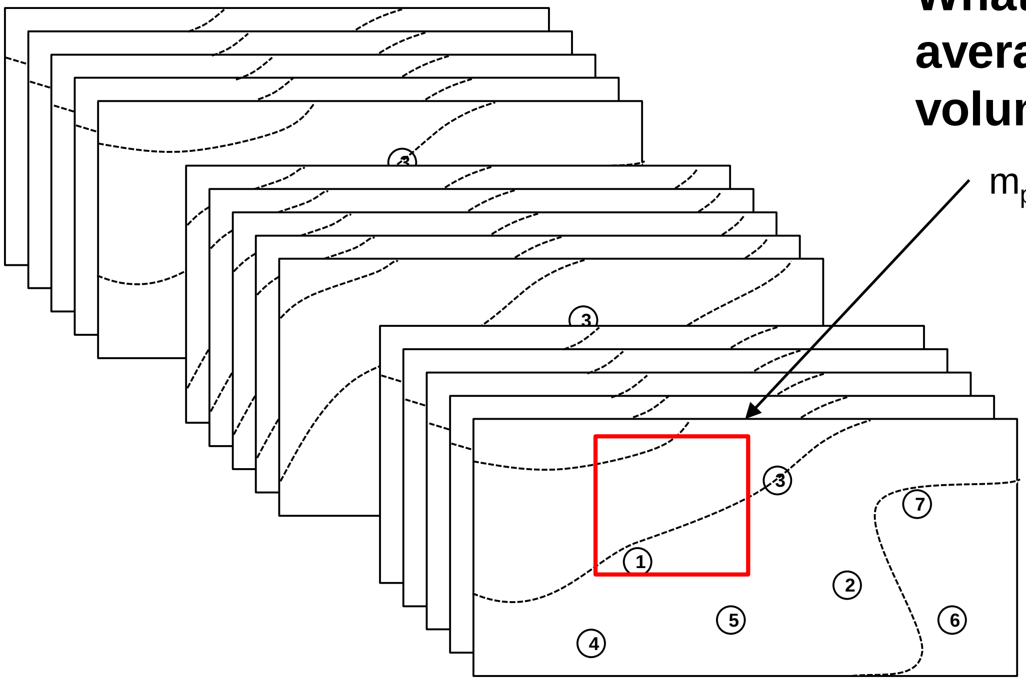
Porosity($\mathbf{u}_1$) = F(Φ; $\mathbf{u}_1$)



How is Uncertainty Calculated?

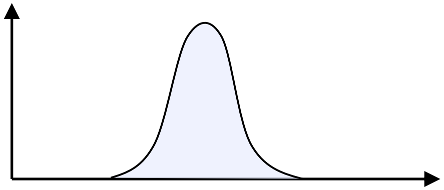
We have represented the "uncertainty model" through scenarios and realizations:

- **We can ask any question of the model by considering all scenarios and realizations jointly.**



What is the uncertainty in average porosity over volume, v?

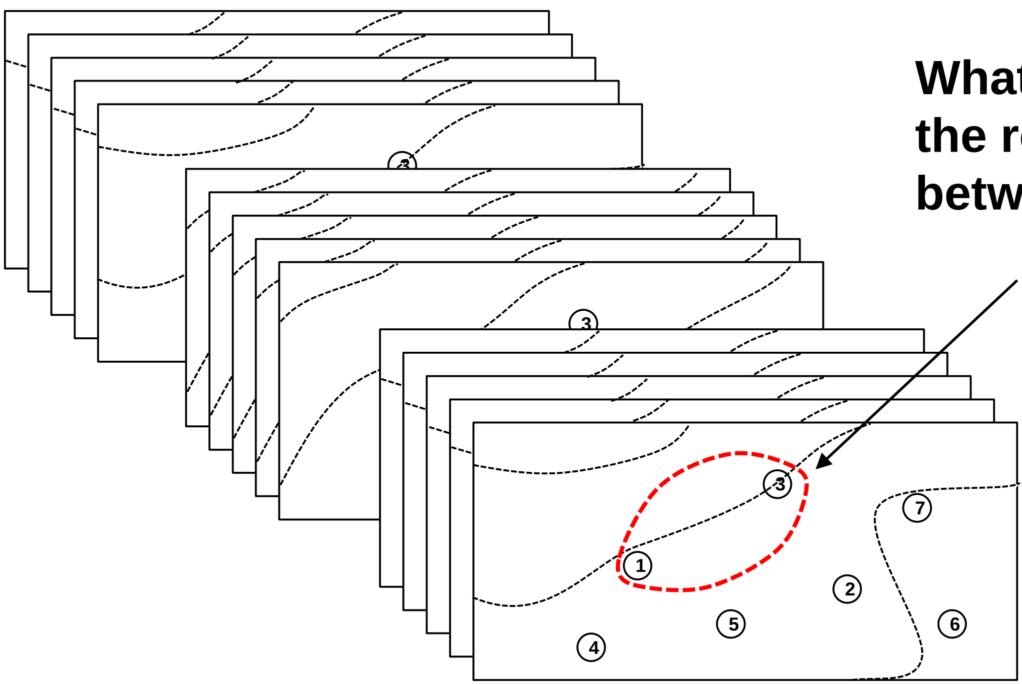
$m_{porosity}(\mathbf{v}) = F(\bar{\Phi}; \mathbf{v})$



How is Uncertainty Calculated?

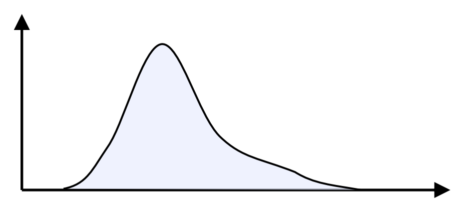
We have represented the "uncertainty model" through scenarios and realizations:

- **We can ask any question of the model by considering all scenarios and realizations jointly.**



What is the uncertainty in the recovery factor between Inj_1 and Prod_3?

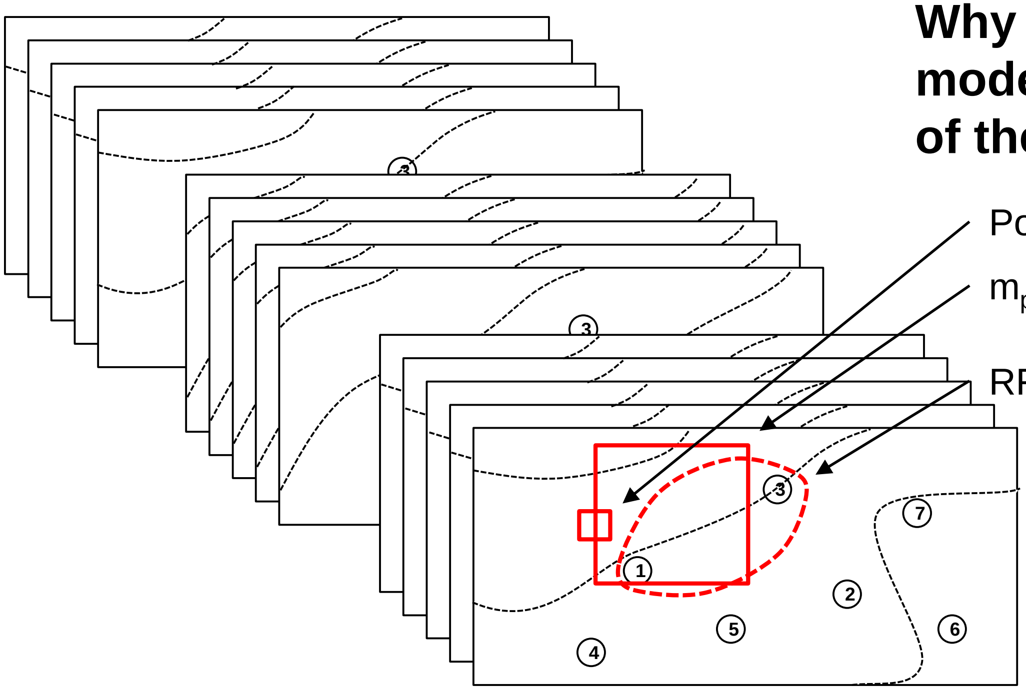
RF(**1-3**) = F(RF; **1-3**)



How is Uncertainty Calculated?

We have represented the "uncertainty model" through scenarios and realizations:

- **Models could be ranked as Pxx for a specific question.**
- **But for every new question, the calculation of rank must be RERUN!**



Why would a single model be the P90 for all of these questions?

Porosity($\mathbf{u}_1$) = F(Φ; $\mathbf{u}_1$)

$m_{porosity}$($\mathbf{v}$) = F(**Φ**; $\mathbf{v}$)

RF(**1-3**) = F(RF; **1-3**)

Data Analytics, Geostatistics and Machine Learning

Uncertainty Modeling

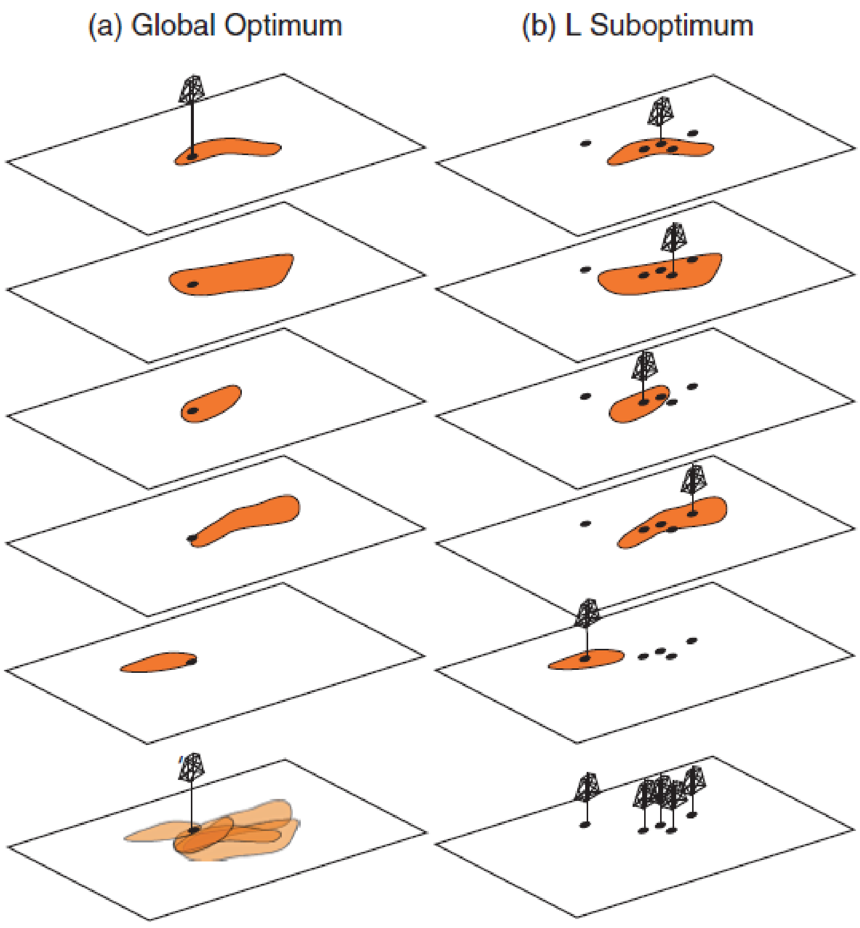
Lecture outline . . .

- Decision Making

Instructor: Michael Pyrcz, the University of Texas at Austin

Types of Decisions

- **Well site selection methods:**
  - integer programing
  - optimization combined with simulating one potential well at a time
  - experimental design and response surface methodology
- **Optimize the profitability of reservoir production**
  - complicated by large number of parameters and timing
  - locations of injector and producers
  - injection, well completions



Global optimum vs. L suboptimum (Pyrcz and Deutsch, 2014).

Decision Making By Maximizing Value

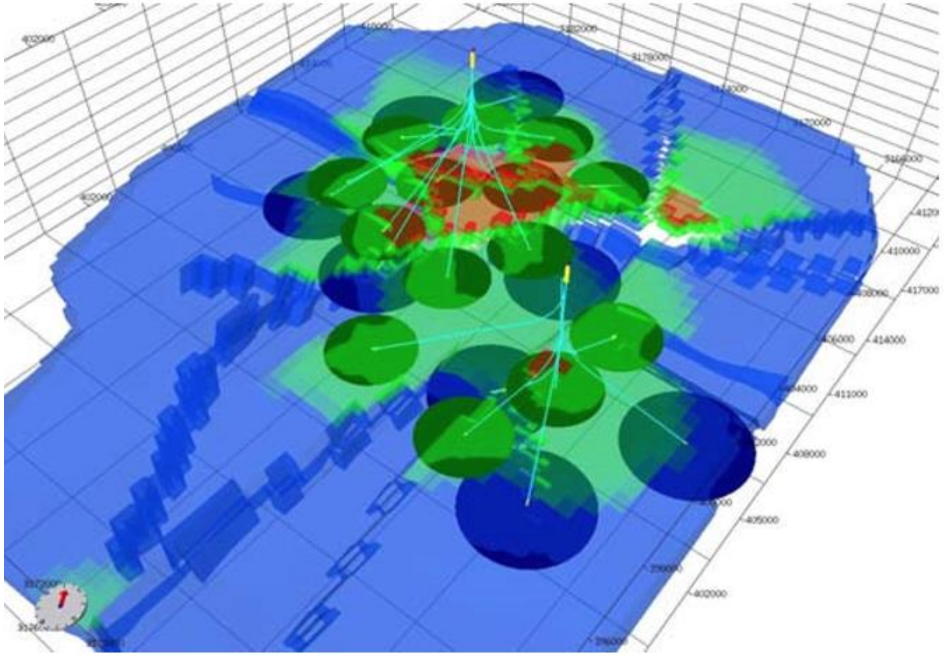
- **Expected Profit Workflow:**
  1. calculate L realizations / reservoir scenarios (the uncertainty model)
  2. establish S, development scenarios (includes all development decisions, could be many)
  3. establish profit metric (accounts for all relevant factors)
  4. calculate profit for each L reservoir realization / scenario and for each S development scenario.

$$P_{s,l}, s = 1, \dots, S, l = 1, \dots, L$$

  5. calculate the expected profit for each development scenario

$$\overline{P_s} = \frac{1}{\sum_{l=1}^{L} \lambda_l} \sum_{l=1}^{L} \lambda_l \cdot P_{s,l}$$

  5. define the optimal development scenario as the one with highest expected profit

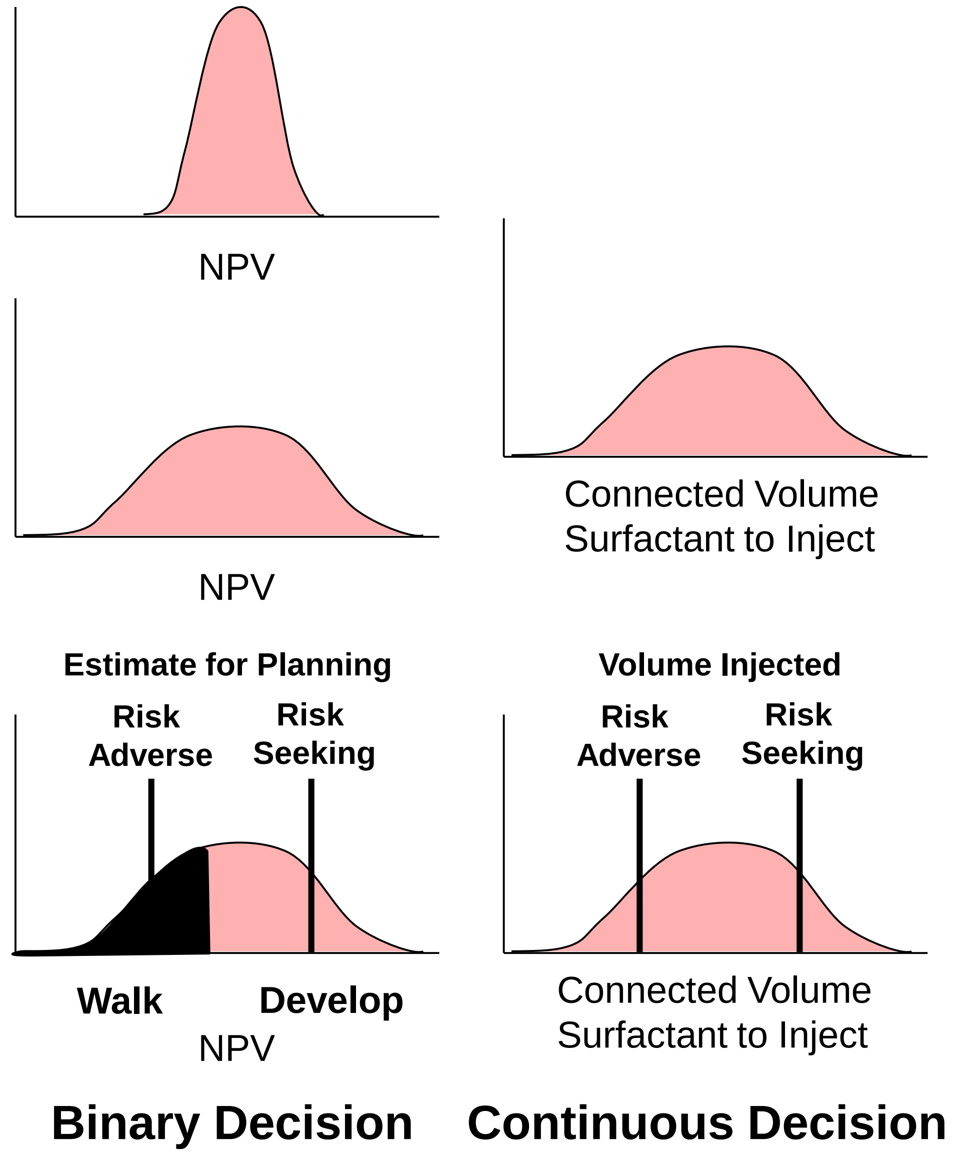


Production target, green, and peripheral water injection targets blue (Cullick et al., 2005).

Decision Making
Impact of Uncertainty

What Distribution is Best?

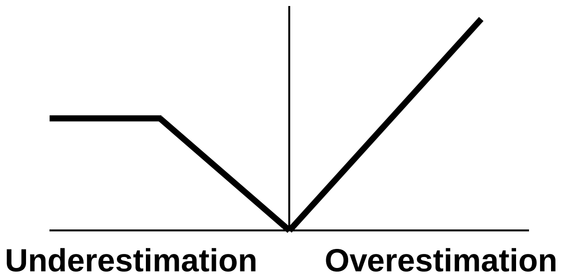
- Is narrower variance better?
- Higher variance has more downside and more upside.
- e.g. in shale play well product has a high variance, lognormal distribution and there are large well counts
- A few boomers can pay for the operation
- To make an optimum estimate / decision from a distribution we need a function that describes the:
  - **loss due to under and overestimation.**

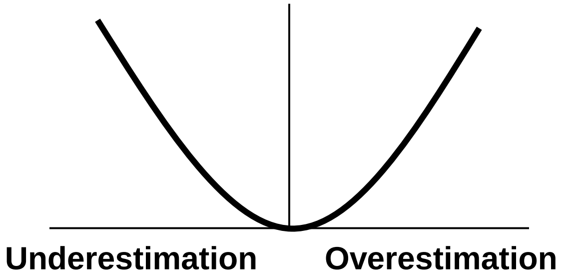


Decision Making With Uncertainty

Loss Functions

- To make a decision in the presence of uncertainty we need to quantify the loss function
- Loss due to over and underestimation of the true value.
- For example, overestimation is more costly than underestimation and at some threshold underestimation any further has no more cost.
- Note: the estimating with the mean minimizes the parabolic loss function

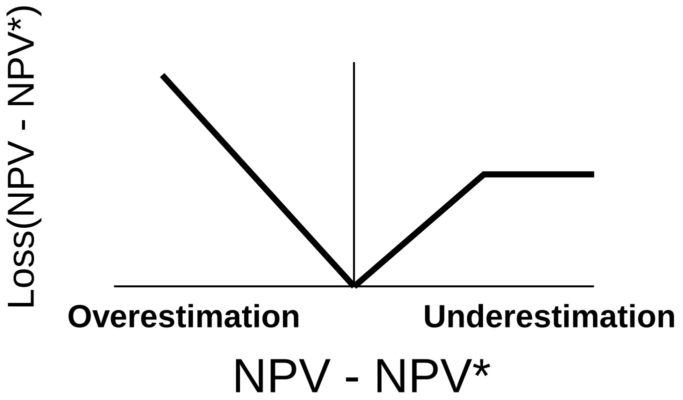




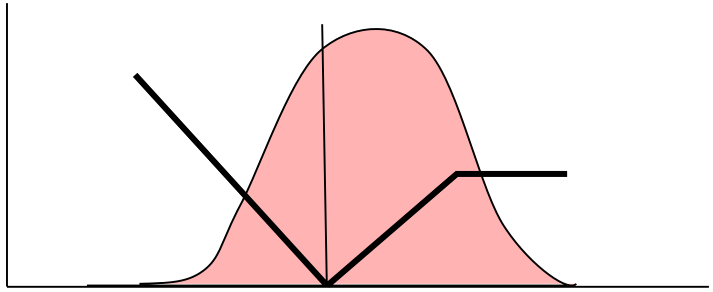
Decision Making With Uncertainty

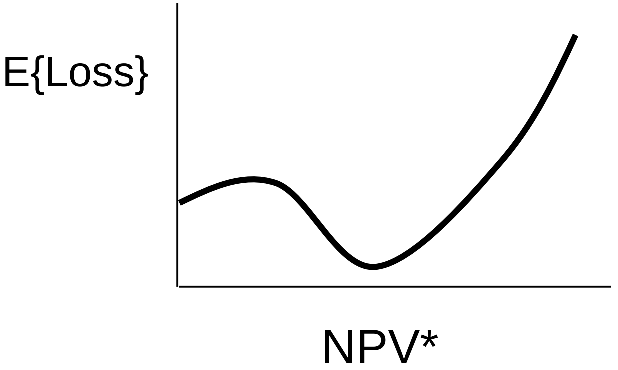
- **Decision Making in the Presence of Uncertainty :**
  - quantify cost of over and underestimation in a loss function
  - apply the loss function to the random variable of interest for a range of estimates
  - calculate the expected loss for each estimate
  - make decision that minimizes loss

Loss Function



NPV Uncertainty PDF





E{Loss(NPV*)} = $\int_{-\infty}^{\infty} L(NPV - NPV^*) \cdot g(NPV)\, dNPV$

Decision Making With Uncertainty Hands-on

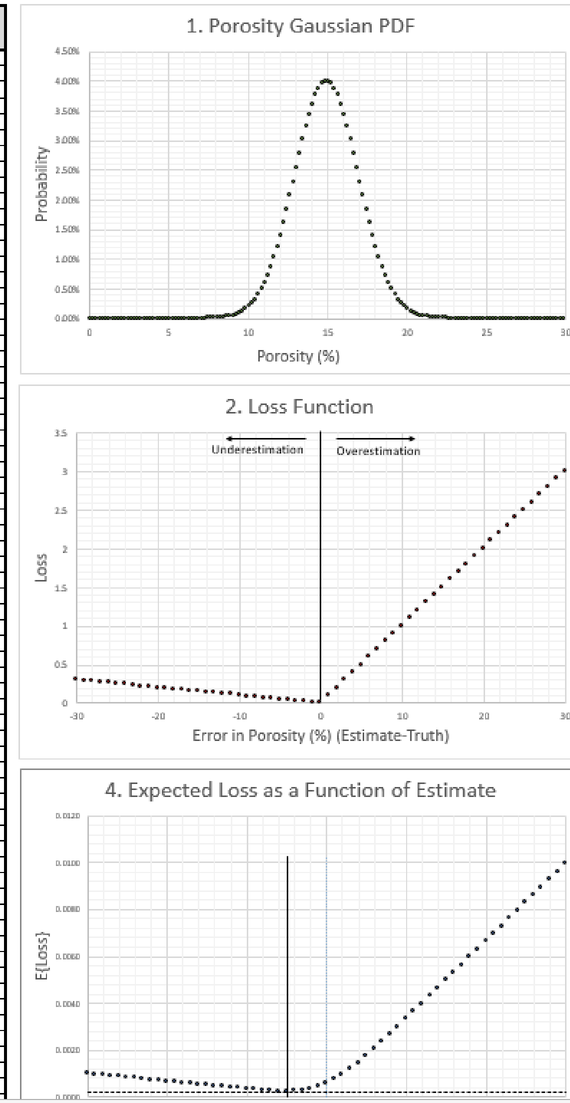
Things to Try Out:

- Try increasing and decreasing the cost of overestimation and observe the estimate.
- Try increasing and decreasing the standard deviation of the uncertainty model.
- What does ‘risk aversion’ do to decision making?

The File is Decision_Making_with_Loss_Function.xlsx at https://git.io/fNgBx.

1. Gaussian Distribution Parameters		
Mean	15	
Standard Deviation	2	

Porosity (%)	Probability	Probability Normalized
0.1	1.8E-13	0.00%
0.3	3.7E-13	0.00%
0.5	7.7E-13	0.00%
0.7	1.6E-12	0.00%
0.9	3.2E-12	0.00%
1.1	6.5E-12	0.00%
1.3	1.3E-11	0.00%
1.5	2.5E-11	0.00%
1.7	5.0E-11	0.00%
1.9	9.6E-11	0.00%
2.1	1.8E-10	0.00%
2.3	3.5E-10	0.00%
2.5	6.6E-10	0.00%
2.7	1.2E-09	0.00%
2.9	2.2E-09	0.00%
3.1	4.1E-09	0.00%
3.3	7.4E-09	0.00%
3.5	1.3E-08	0.00%
3.7	2.3E-08	0.00%
3.9	4.1E-08	0.00%
4.1	7.1E-08	0.00%
4.3	1.2E-07	0.00%
4.5	2.1E-07	0.00%
4.7	3.5E-07	0.00%
4.9	5.8E-07	0.00%
5.1	9.5E-07	0.00%
5.3	1.6E-06	0.00%
5.5	2.5E-06	0.00%
5.7	4.0E-06	0.00%
5.9	6.4E-06	0.00%
6.1	1.0E-05	0.00%
6.3	1.6E-05	0.00%
6.5	2.4E-05	0.00%
6.7	3.6E-05	0.00%
6.9	5.5E-05	0.00%
7.1	8.2E-05	0.00%
7.3	1.2E-04	0.00%
7.5	1.8E-04	0.00%
7.7	2.6E-04	0.01%
7.9	3.7E-04	0.01%
8.1	5.2E-04	0.01%
8.3	7.3E-04	0.01%
8.5	1.0E-03	0.02%
8.7	1.4E-03	0.03%
8.9	1.9E-03	0.04%
9.1	2.6E-03	0.05%
9.3	3.4E-03	0.07%
9.5	4.5E-03	0.09%
9.7	6.0E-03	0.12%
9.9	7.7E-03	0.15%
10.1	9.9E-03	0.20%
10.3	1.3E-02	0.25%
10.5	1.6E-02	0.32%
10.7	2.0E-02	0.40%
10.9	2.4E-02	0.49%
11.1	3.0E-02	0.60%
11.3	3.6E-02	0.72%
11.5	4.3E-02	0.86%
11.7	5.1E-02	1.02%
11.9	6.0E-02	1.20%
12.1	7.0E-02	1.39%
12.3	8.0E-02	1.60%
12.5	9.1E-02	1.83%
12.7	1.0E-01	2.06%
12.9	1.1E-01	2.30%
13.1	1.3E-01	2.54%



Uncertainty Comments

Calculating Uncertainty in a Modeling Parameter: Use Bayesian methods, 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.

Be an uncertainty detective! Discover and evaluate various sources.

Uncertainty Comments

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.

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.

Ignoring uncertainty is assuming certainty.

Uncertainty New Tools

Topic	Application to Subsurface Modeling
Sources of Uncertainty	Seek out and integrate all significant sources of uncertainty. *Uncertainty in the data measures based on data realizations combined with spatial uncertainty with multiple spatial realizations.*
Scenarios	Capture uncertainty in model parameters and decisions through multiple scenarios. *Construction of 3 set of realizations for low, mid and high case spatial continuity models.*
Decision Making with Uncertainty	*The estimate from an uncertainty distribution for decision making depends on the cost of under and over estimation.* *Construct a loss function and determine the OIP estimate for a reservoir from the OIP uncertainty distribution accounting for the cost of under and over estimation of OIP.*