PGE 337 Lecture 21:
Conclusion to Introduction to Geostatistics

Lecture outline . . .

- **Review**
- **Suggested Applications of Course Material**

Introduction

Probability Theory

Confidence / Hypothesis

Univariate

Bivariate

Spatial Analysis

Machine Learning

This will be my final rant then you can go enjoy your summer!

What You Learned & How you Could Use it?

In my 13 years of industry experience I met **two types of people**:

1. Those who found geostatistical problems and wondered how to solve them.

2. Those who found geostatistical problems and solved them, taught others how to solve them and were widely recognized.

- **Geostatistical problems are everywhere** in subsurface resource characterization, modeling and development.
  - spatiotemporal, univariate and multivariate, sampling, inference and prediction

Rigorous Use of Probability

Calculating a probability of a union of events, remember to account for the intersection!

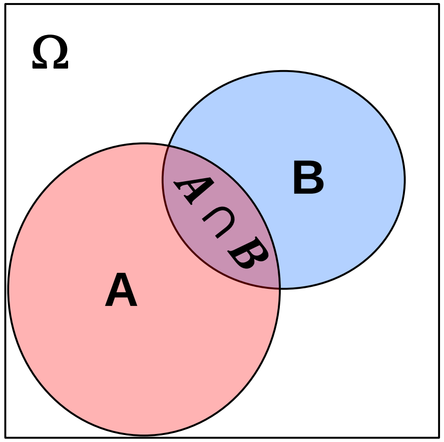
- For example, use the addition rule for unions.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

- If reservoir rock is defined by either of 2 criteria then you can add the probabilities of each, but you must subtract the intersection.

$$P(Net) = P(C_1) + P(C_2) - P(C_1 \cap C_2)$$

- This mistake will significantly exaggerate the probability.

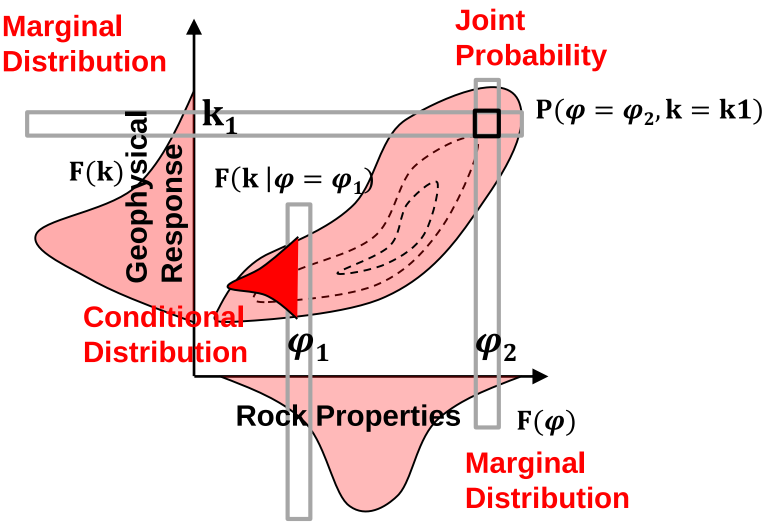


Venn Diagram – illustrating intersection.

When Exhaustively Sampling, May Be Able to Calculate the Joint!

You may already have the answer!

- If you have enough data you can directly calculate the needed marginal, joint and conditional probabilities.



- For example, one may consider Bayesian updating to estimate rock properties from geophysical data.

$$P(\text{Rock Prop}|\text{Geophys. Response}) = \frac{P(\text{Geophys. Response}|\text{Rock Prop})P(\text{Rock Prop})}{P(\text{Geophys. Response})}$$

- But, may be able to **calculate the joint** from the experimental data from 1,000s of runs, no Bayes needed!

Joint Probability

$$P(\text{Rock Prop}|\text{Geophys. Response}) = P(\textbf{Geophys. Response, Rock Prop})/P(\text{Geophys. Response})$$

Check for Independence

It is possible to check for independence in many cases.

- Independence makes a big difference! We have so many things that may be dependent. If independent then the workflows are much easier!

- We can test for independence by hand (e.g. facies by reservoir units):

Position	Well 1	Well 2	Well 3	Well 4	Well 5	Well 6	Well 7	Well 8	Well 9	Well 10
Top	F3	F2	F2	F1	F1	F1	F2	F2	F1	F1
Middle	F1	F1	F1	F1	F2	F2	F1	F2	F2	F2
Bottom	F2	F2	F2	F3	F3	F3	F3	F3	F3	F2

Event A_1 = F1 is middle facies
Event A_2 = F3 is bottom facies

$\mathrm{P}(\boldsymbol{A_1} \cap \boldsymbol{A_2}) = \mathrm{P}(\boldsymbol{A_1})\mathrm{P}(\boldsymbol{A_2})$ or

$\mathrm{P}(\boldsymbol{A_1}|\boldsymbol{A_2}) = \mathrm{P}(\boldsymbol{A_1})$ and $\mathrm{P}(\boldsymbol{A_2}|\boldsymbol{A_1}) = \mathrm{P}(\boldsymbol{A_2})$

Bayes' Theorem

But when you need Bayes, it is a powerful method!

- Bayes' theorem is powerful for "flipping" conditional probabilities. Get the difficult one from the one you can easily calculate.
- E.g. Probability something is happening, given an indicator it is happening!

$$P\left(\begin{matrix}\text{Something is}\\ \text{Happening}\end{matrix} \middle| \begin{matrix}\text{Looks like}\\ \text{its happening}\end{matrix}\right) = \frac{P\left(\begin{matrix}\text{Looks like}\\ \text{its happening}\end{matrix} \middle| \begin{matrix}\text{Something is}\\ \text{Happening}\end{matrix}\right) P\left(\begin{matrix}\text{Something is}\\ \text{Happening}\end{matrix}\right)}{P\left(\begin{matrix}\text{Looks like}\\ \text{its happening}\end{matrix}\right)}$$

Event A		Event B
You have a disease	←	You test positive for the disease
There is fault compartmentalized	←	Geologist says there's a fault
Low permeability of a sample	←	The laboratory measure is low
A valve will fail	←	X-ray test is positive
You drill a dry well	←	Seismic AVO response looks poor

Statistical Expectation

Know how to work with the expectation operator.

- Projects are evaluated in expected value, given uncertainty distributions.
- This will include the integration of multiple components all with their own uncertainties.
- Use the properties of expectation to solve these problems.

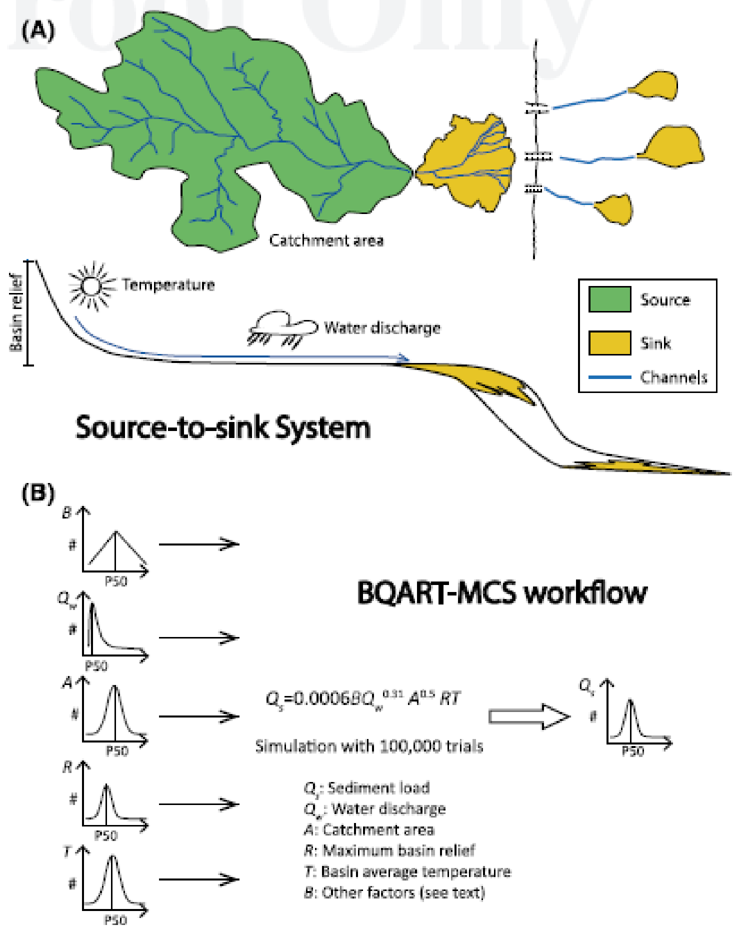
Example Problems	Properties
Expected value if a factor applied to the volume.	$\boldsymbol{E[cX] = cE[X]}$
Expected value of sum of multiple reservoirs.	$\boldsymbol{E[X+Y] = E[X] + E[Y]}$
Low permeability of a sample	$\boldsymbol{E[XY] = E[X]E[Y], if\ X \perp\!\!\!\perp Y}$

Monte Carlo Simulation

You can use Monte Carlo Simulation On Anything!

- Extremely powerful. Any distributions, any equations.
- Use Monte Carlo Simulation to propagate uncertainty from inputs to output(s).
- For example sediment flux delivered to deep basins, given variables describing the sediments source.

Figure 1. (A) Sketches of the map view and the cross section of a source-to-sink system including the upstream characteristics used in the BQART model (modified from Helland-Hansen et al., 2016). (B) Workflow of BQART-MCS method. # = frequency; MCS = Monte Carlo simulation; P50 = 50th percentile.



(Zhang, Covault, Pyrcz, Sharman, Carvajal and Milliken, 2018)

Measures of Heterogeneity

Heterogeneity is Critical and You Can Measure It.

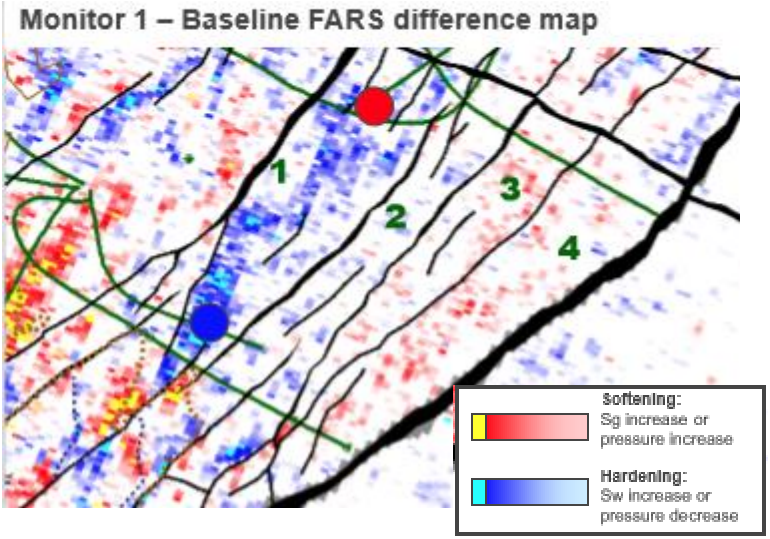
- Extremely important. major source of uncertainty
- Measures of heterogeneity provide early indicators
- Even univariate measures are meaningful:

permeability median

Dystra-Parsons

$$V = \frac{k_{50} - k_{16}}{k_{50}}$$

$\boldsymbol{F_K^{-1}}$**(0.16)**



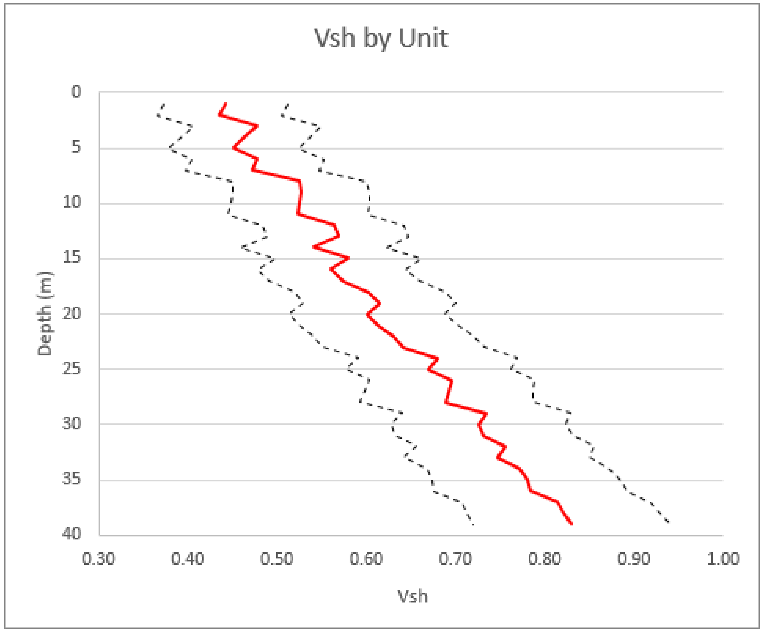
4D seismic demonstrating early break though Enfield Oil Field by (Hamson, 2012)

Confidence Intervals on Every Plot!

Every Measure Has Uncertainty, Include this on Your Visualizations

- Are the wiggles important?
- What should we focus on?
- Confidence interval provide context for interpretation!
- For example, for a mean (known standard deviation, Gaussian distributed):

$$\bar{y}_{C.I.} = \bar{y} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$



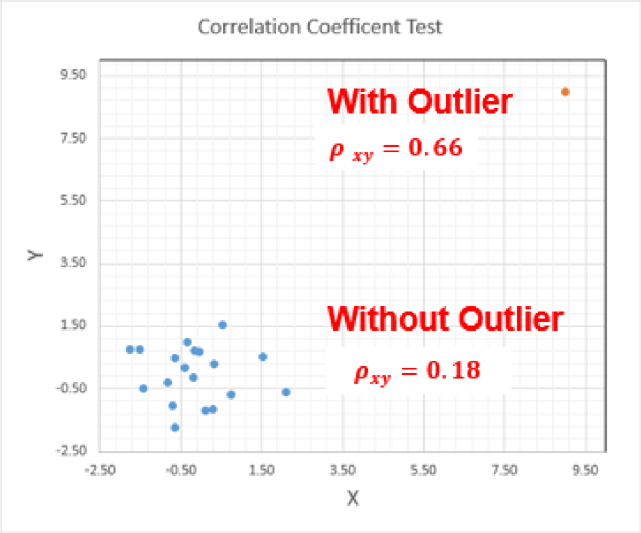
Vsh trend by depth with confidence intervals.

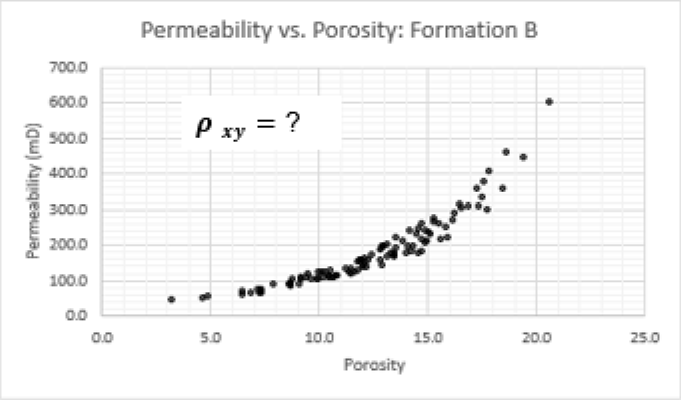
Limits of Correlation Coefficient

Understand the limits of correlation coefficients

- Correlation coefficients get used a lot along with correlation matrices. Know their limits.

- Use rank correlation coefficients when needed. Don't use if non-linear. Don't use at all if non-linear scatter plot

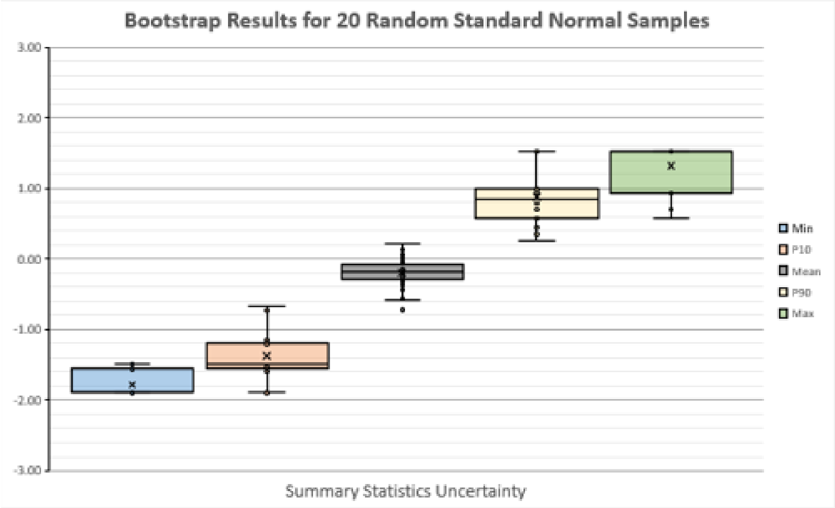




You Can Bootstrap to Get Uncertainty in Any Sample Statistic

Got a sample statistic, need to add uncertainty to it?

- Bootstrap is powerful.
- Correlation coefficient, vertical trend, P10, P90, mean etc.
- Add uncertainty to any measure, but assumes independence!



Keep Stationarity in Your Back Pocket

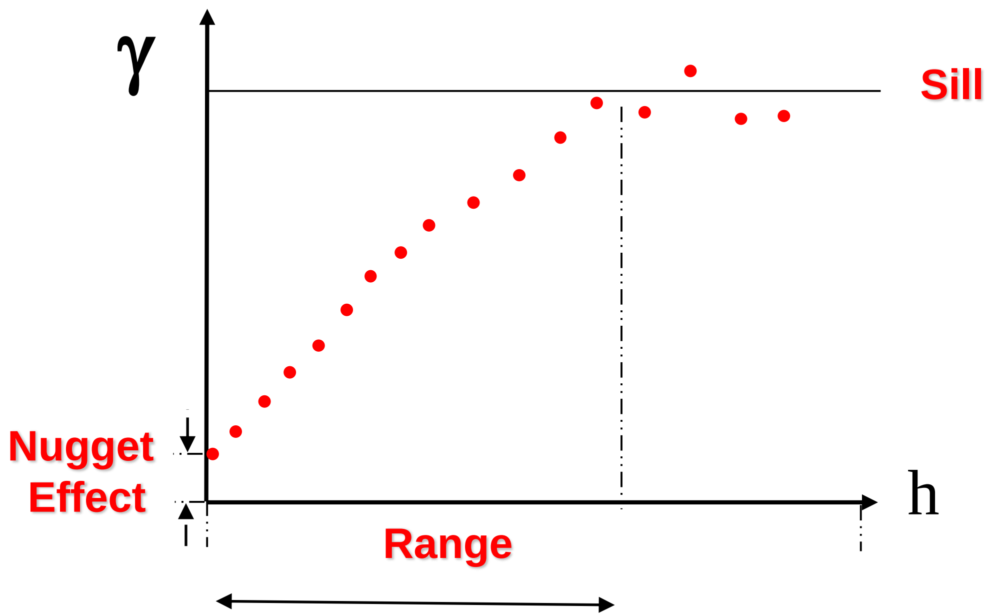
See a statistic sampled over an interval – ask about import and export license!

- **Import:** what data did you pool and **Export:** where do you want to use it?
- Any statistic has this decision: entire distribution, mean, variance, variogram…

Calculate Spatial Continuity

The variogram, covariance function

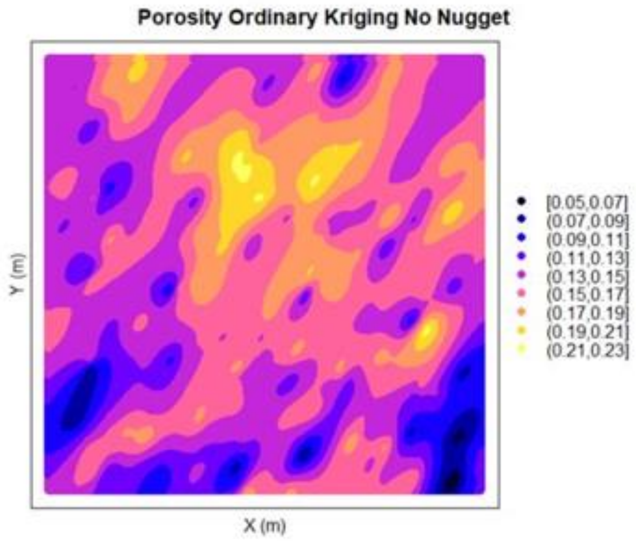
- Also, try it out in time series analysis (autocorrelation)
- Interpret the spatial frequencies, use the range to describe correlated length

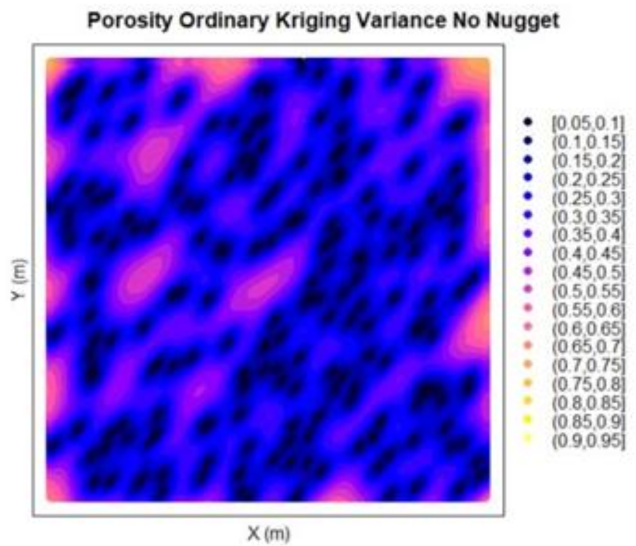


Make Spatial Estimates

Make predictions in space and provide uncertainty

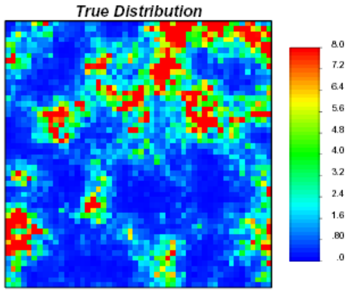
- Be careful if you put it in a map
- Identify new drilling opportunities
- Assess volumetrics and mapped “resources”
- e.g. Predrill estimates with uncertainty!

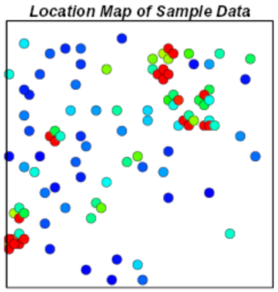




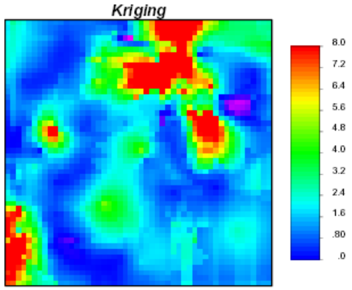
Build Simulated Realizations

Always use simulation to model the subsurface for input into flow simulation



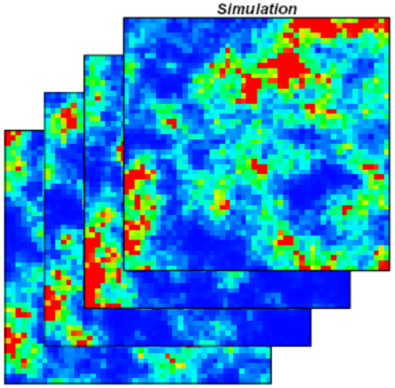


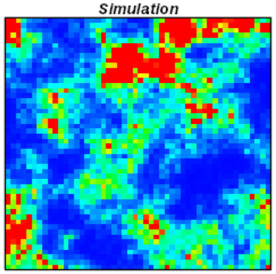
- Need simulation, unless heterogeneity may be imposed from wells and spatial uncertainty is very low
  - data spacing < heterogeneity scale



- Retain all the realizations for each question!

- Avoid ranking is possible





Know About Machine Learning

Understand how they work, many of them are based on intuitive natural examples:

- Decision Trees – a branching tree for hierarchical, binary decision making
- Artificial Neural Net – the neurons in a brain
- Swarm Optimizations – swarm intelligence
- Simulated Anneal – metallurgical annealing

Still statistical! Understand limits, assumptions (e.g.):

Expected Test Mean Square Error

$$\mathrm{E}\left[\left(y_0 - \hat{f}(x_1^0, \dots, x_m^0)\right)^2\right] = \underbrace{Var\left(\hat{f}(x_1^0, \dots, x_m^0)\right)}_{\text{Model Variance}} + \underbrace{\left[Bias\left(\hat{f}(x_1^0, \dots, x_m^0)\right)\right]^2}_{\text{Model Bias}} + \underbrace{Var(\epsilon)}_{\text{Irreducible}}$$

Now What?

Stay Current!

- There is so many resources available
- Open Source Code, Python and R
- Online Courses (Coursera, Udemy, Udacity, GeostatisticalLessons, Big Data University)
- Conferences, Journals
- Podcasts, Meet-ups, Blogs, Social Media

Network

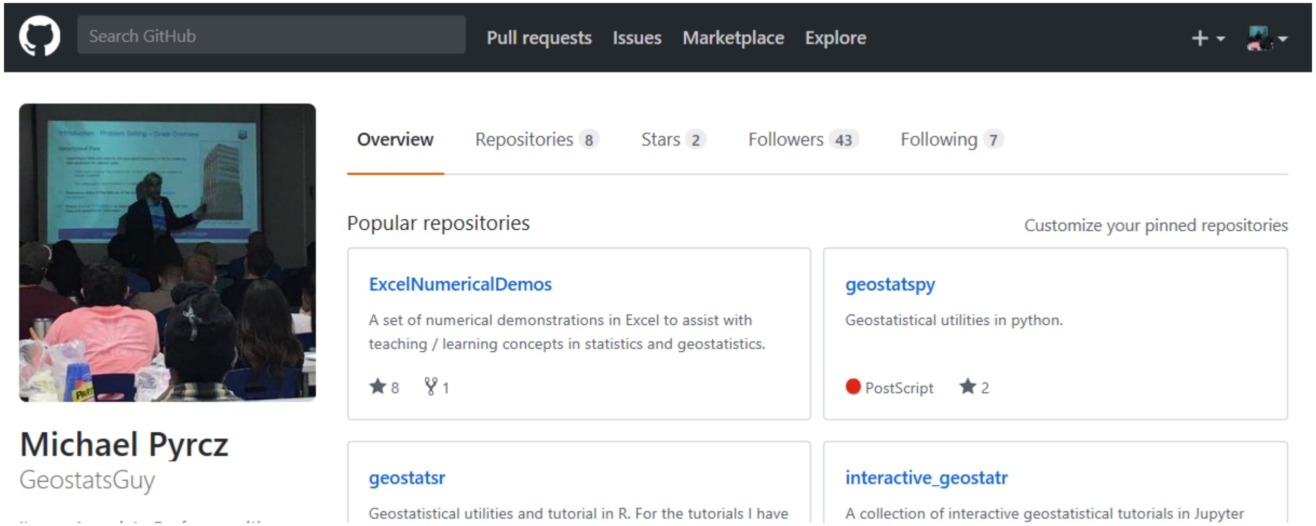
We are in it together!

- Great network here at UT and beyond, keep building it.

- During this class, my door was open, it stays open!
  - Got a geostats problem next year, senior project?
  - Interested in an undergraduate research project?
  - Want to do grad school?
  - Want to bounce a stats question from your office in Houston, Calgary, Starvanger etc.

- It has been my sincere joy to be able to work with all y'all!
  - Thank you for being understanding
  - Providing great feedback and ideas
  - Participating in this class

I’ll Keep Sharing Stuff



“So it begins.”

at the battle of Helms Deep, King Théoden,
Lord of the Rings, The Two Towers