Using Data to Define Models

Jerry L. Jensen

University of Calgary

Outline

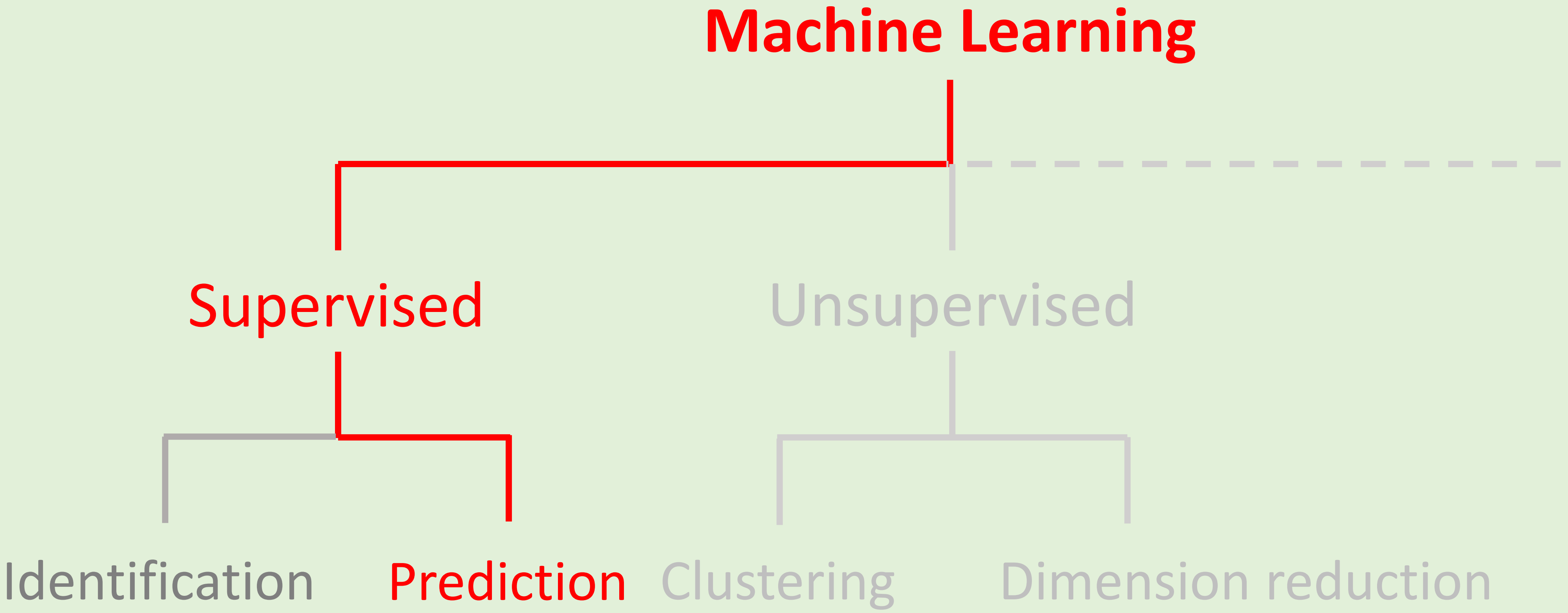
- ❑Prediction and machine learning
- ❑Some statistical “facts of life”
- ❑Linear vs non-linear models
- ❑A reason for non-linearity
- ❑Conclusions

Do I Need Supervision?

Machine Learning

- Supervised
  Needs examples
  - Identification
  - Prediction
- Unsupervised
  No examples needed
  - Clustering
  - Dimension reduction

Today's Topic and ML



Adapted from medium.com

What Is Prediction?

- ❑ Data must be supplied
  - o Values of X (inputs) and Y (outputs)
  - o Serve as ‘examples’ to train or test algorithm
- ❑ Algorithm finds f:

$$Y = f(X)$$

- ❑ The form of f might also be provided

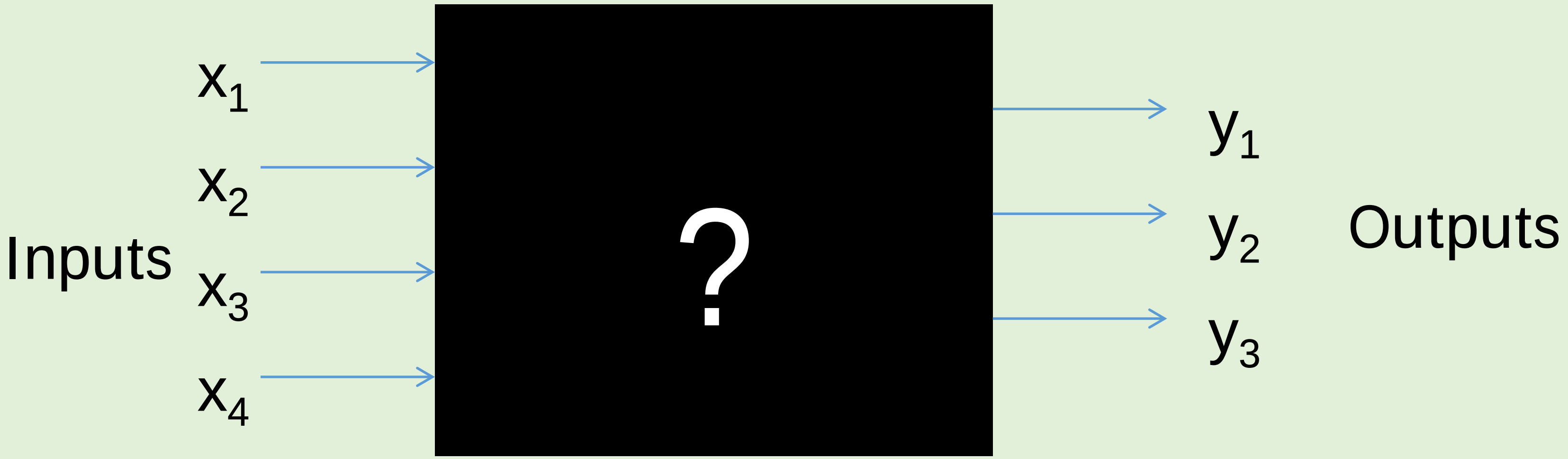
What Is Prediction?

- Example predictors
  - Permeability from well log porosity, gamma ray, …
  - Oil production from frac fluid, proppant, …
  - Fluid contacts from seismic impedance, amplitude, …
- Example procedures
  - Linear regression
  - Non-linear regression
  - Neural networks

Model and Model-Free Approaches to Prediction

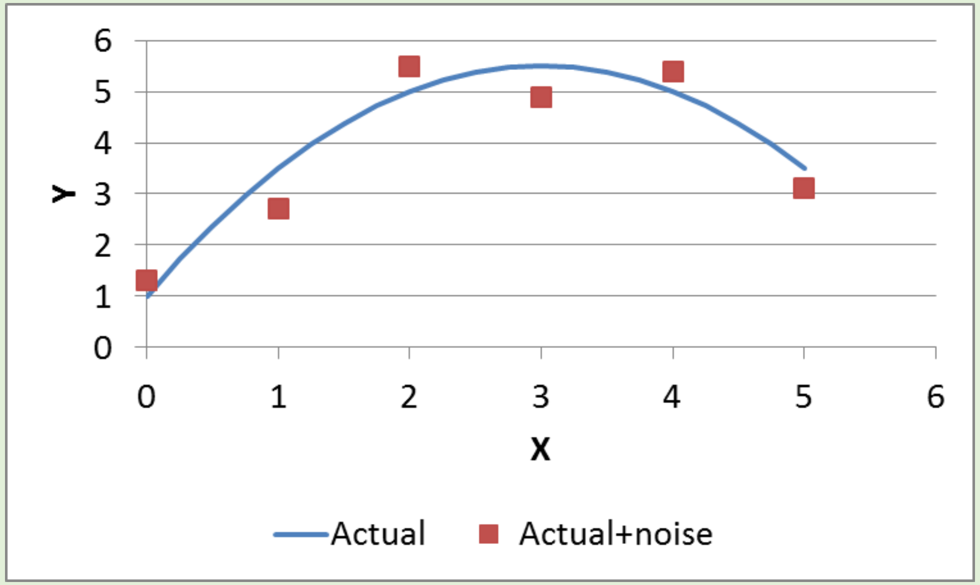
Model-free Approach (1)

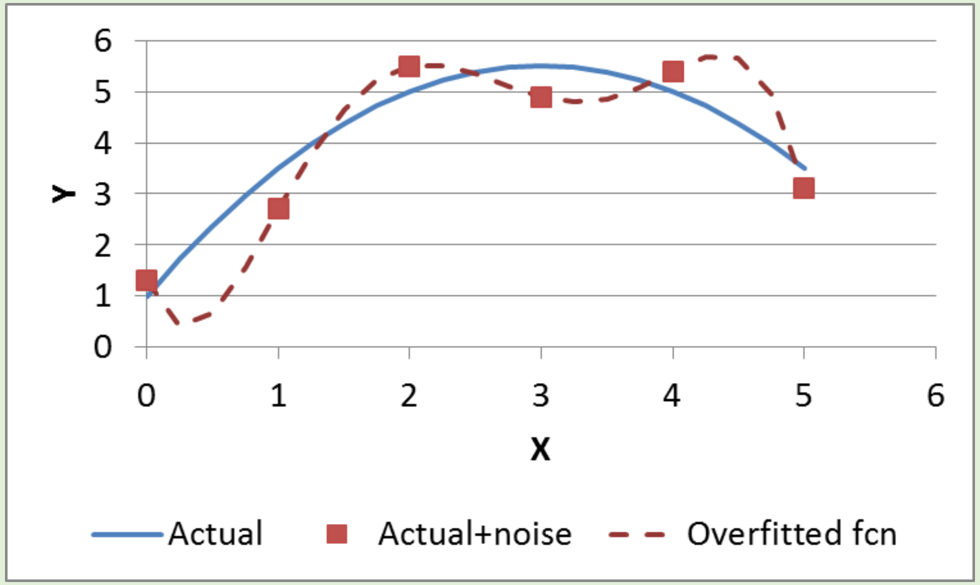
- Artificial neural networks – common method
- Let “data define the relationship”
- Often includes non-linear relationships
- Care is needed



Model-free Approach (2)

- ❑Risk of overfitting
- ❑Results might fit ‘too well’
- ❑Relationship is too specific; poor generalization
- ❑Does the relationship honor the physics?



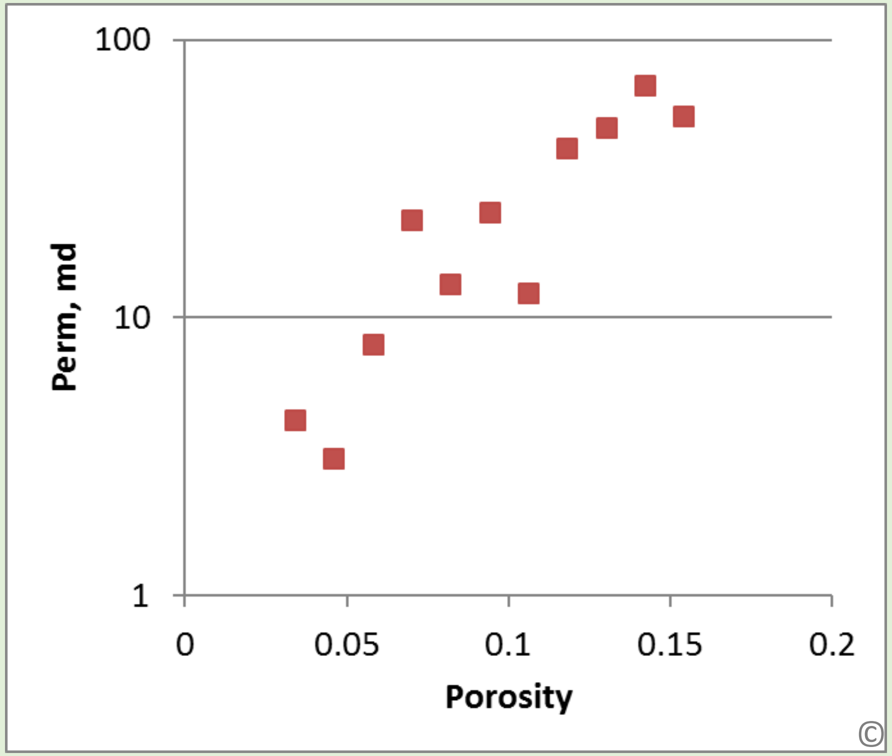


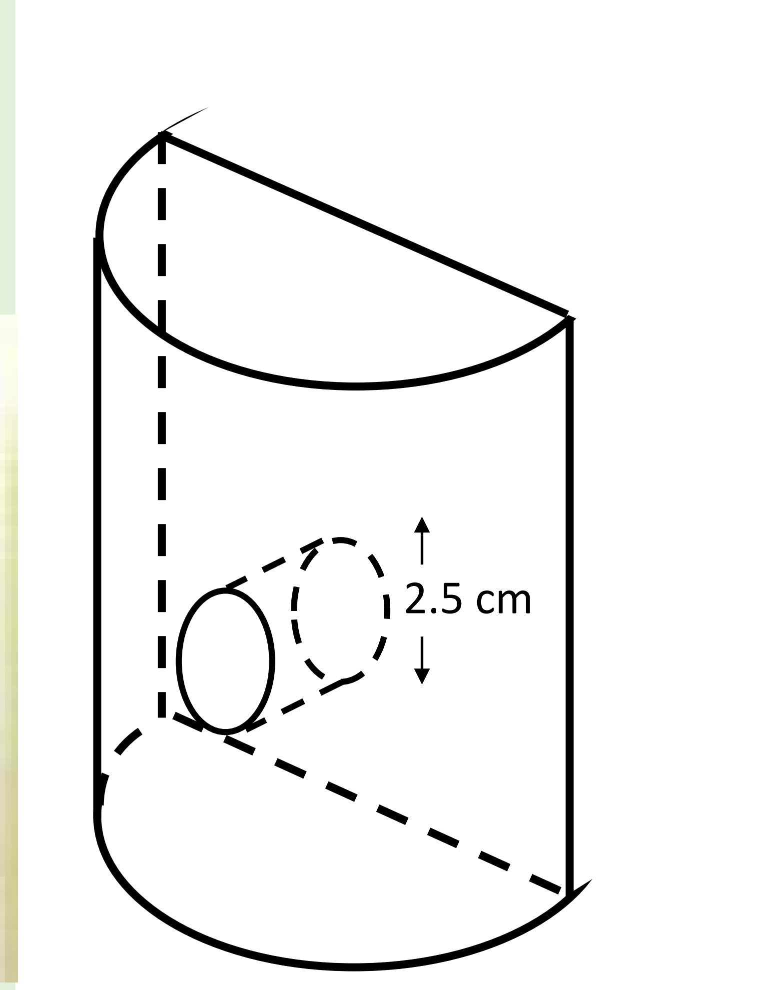
With-Model Approach

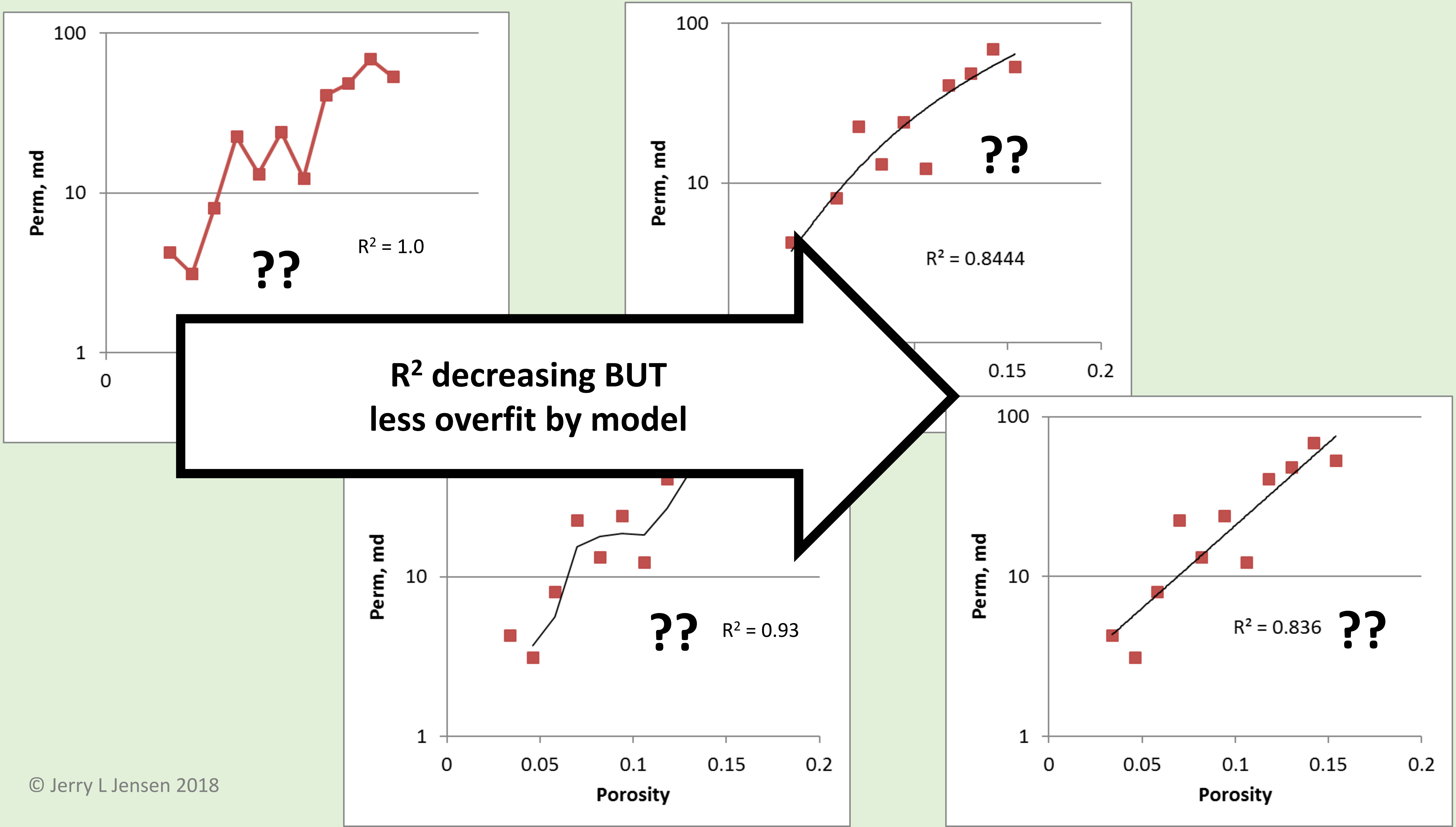
- Regression (linear or non-linear) – common method
- Use data **and model** to define relationship (*f*)
- Example *f*'s where a and b determined by regression
  - $Y = a + bX$
  - $Y = 10^{a + bX}$
- Adding model allows
  - Reduced need for data or reduced uncertainty
  - Control on form of *f*
- Model represents information about *f*

With vs Without Model Example

- ❑Porosity and permeability core plug measurements
- ❑Want to develop perm predictor for new wells

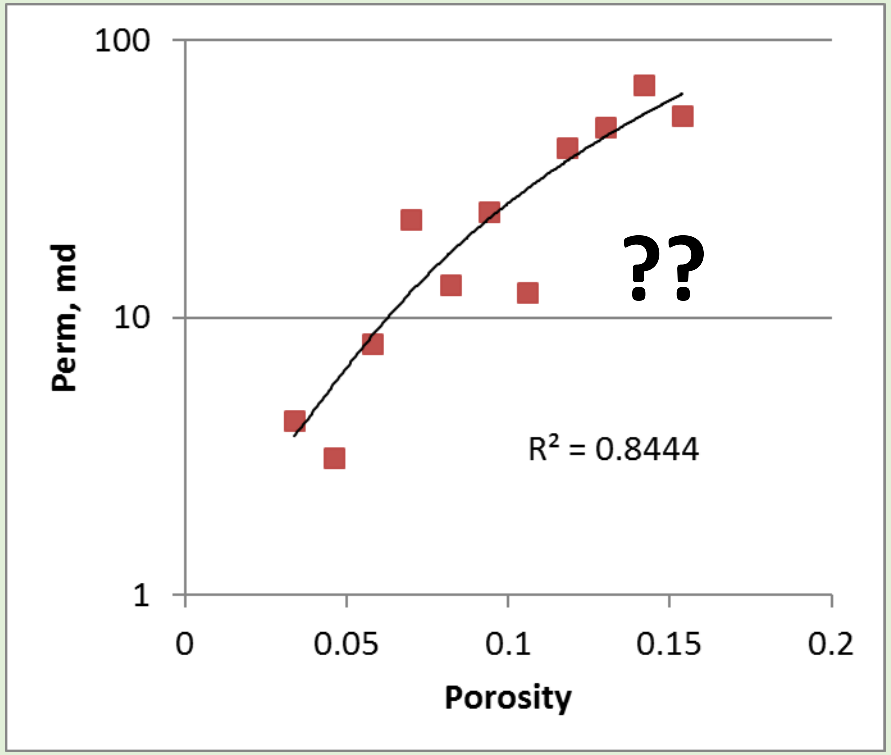


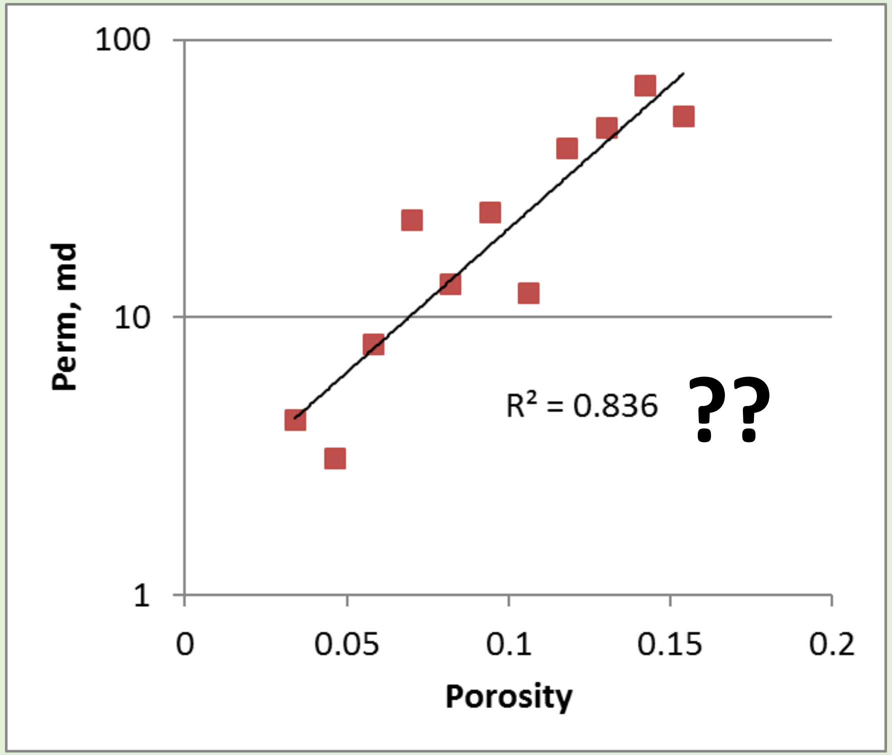



R² decreasing BUT
less overfit by model
R² = 1.0
R² = 0.8444
R² = 0.93
R² = 0.836
??
Perm, md
Porosity
100
10
1
0
0.05
0.1
0.15
0.2

Statistical Perspectives on Prediction (1)

- ❑ Data are important in what they tell us about *f*
- ❑ Model also represents information about *f*
- ❑ Unless we have many data, we do not know what *f* is
- ❑ This leads to **uncertainty**

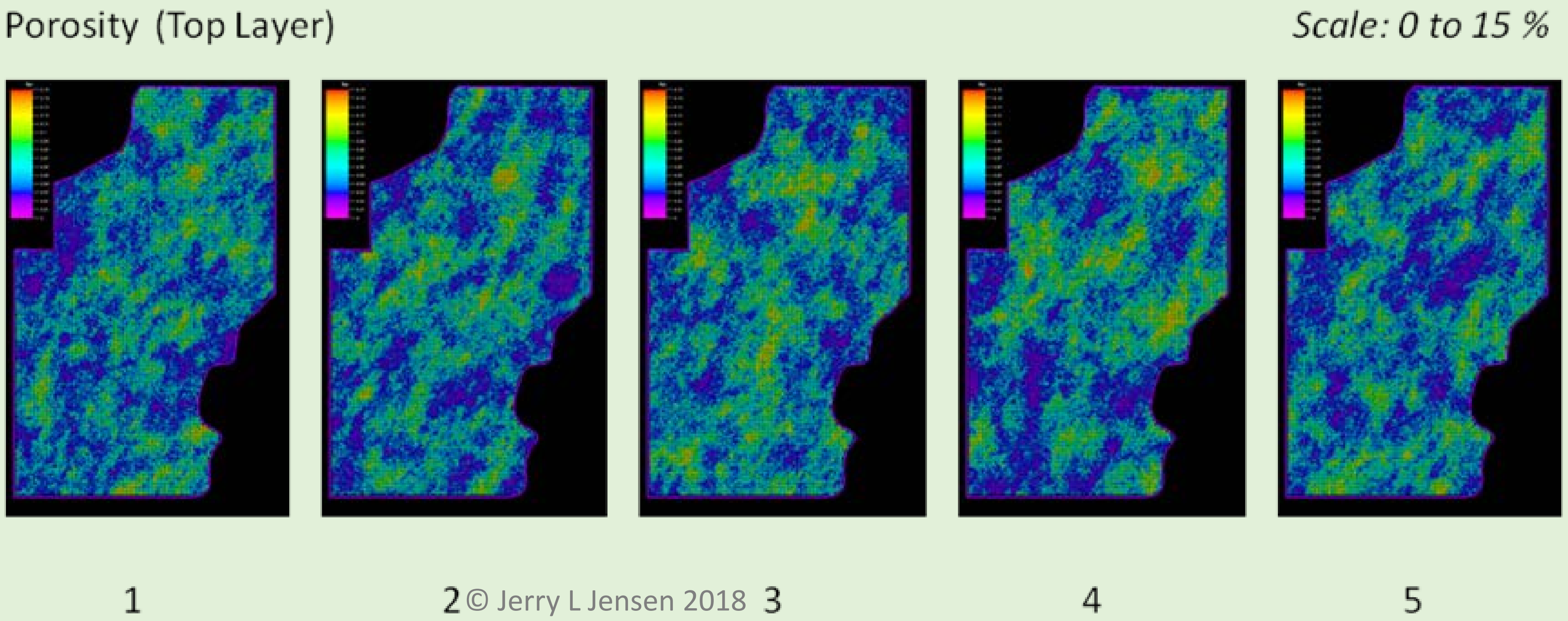




Statistical Perspectives on Prediction (2)

- Every predictor method has uncertainty
- Geostatistics gives equally probable realizations
- Beneficial for understanding uncertainty of any result

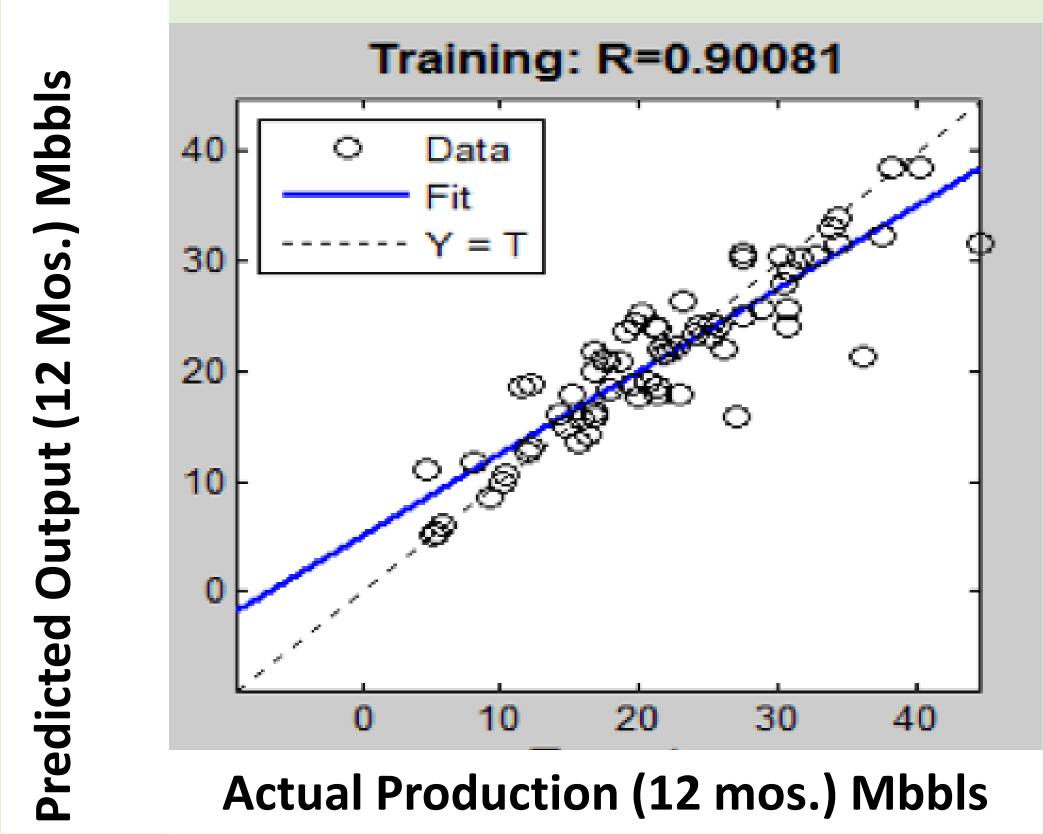
Five realizations of porosity for the Nisku Formation



Statistical Perspectives on Prediction (3)

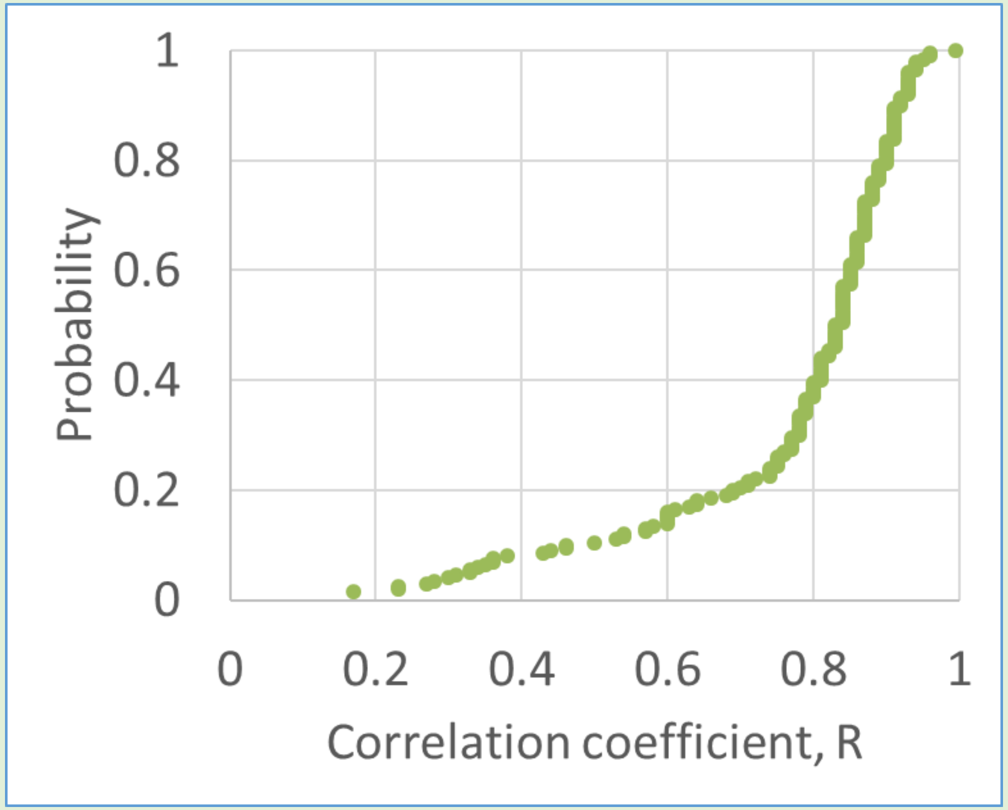
- ❑ If you do not ask, you won’t know what the uncertainty is

Example ANN result



Regression Plot in MATLAB®

200 ANN results for the same X, Y data



Determining Uncertainties

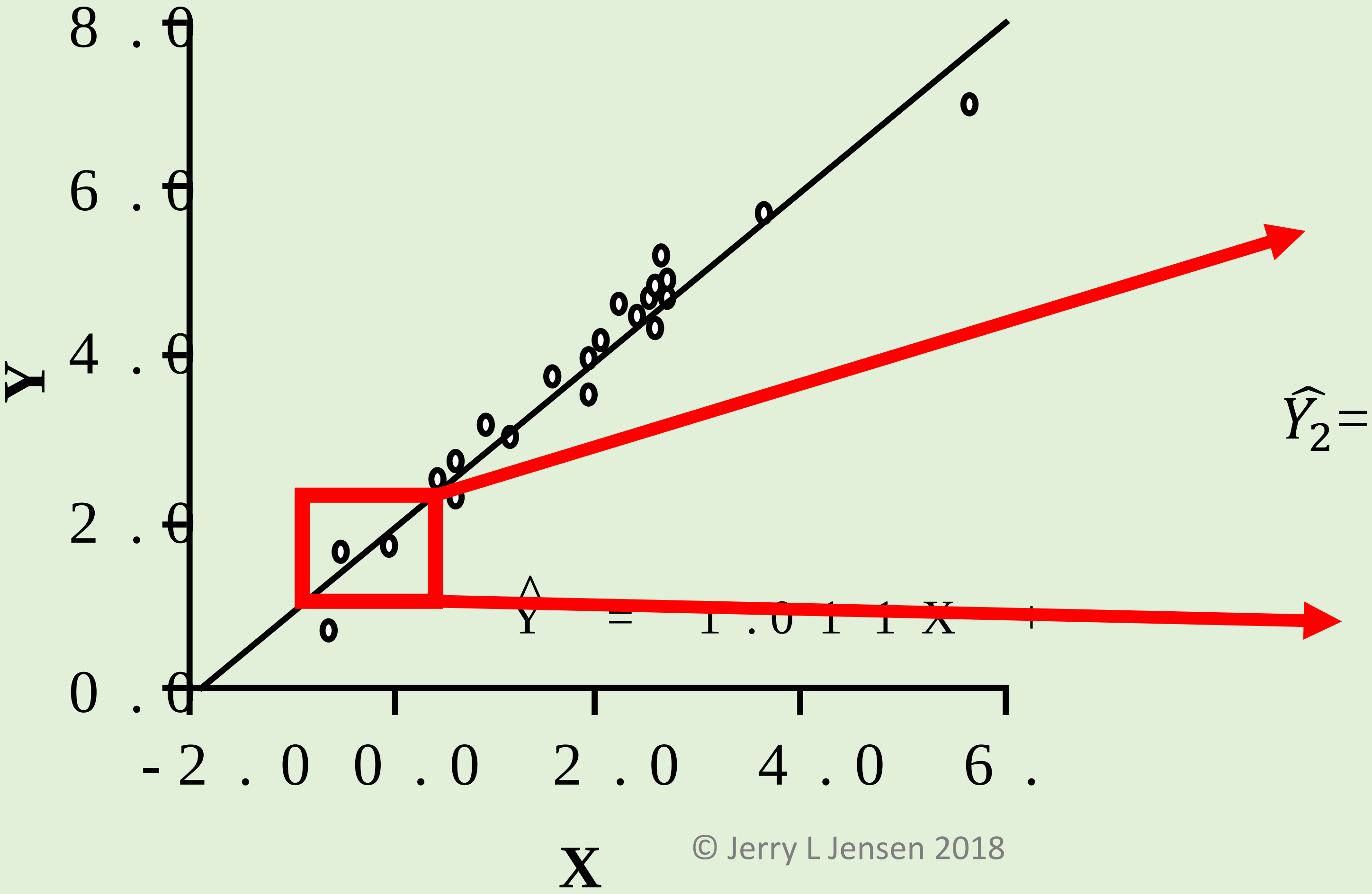
- Computational
  - Recalculate results from different starting points
  - Bootstrap – randomly remove data and recalculate results
  - Combination of bootstrap and recalculation
- Theoretical
  - Regression theory
  - Normal or other distribution assumed
- Adding a model generally decreases uncertainty

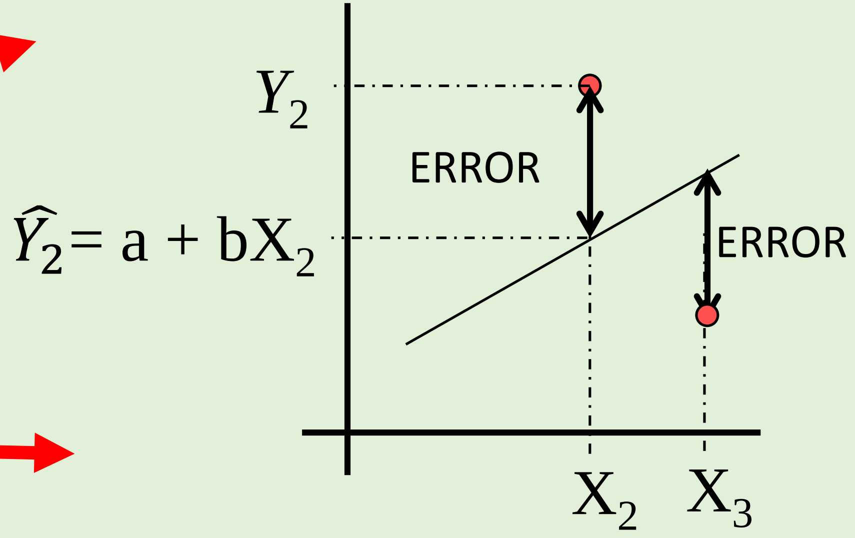
Recommended Modelling Practices

- Avoid removing data **unless you have a good reason**
- Include models where you can
  - Honor physics
  - Include experience
- Start with simple models
- Add complexity if needed
- Examine results critically e.g., use residuals

What Are Residuals?

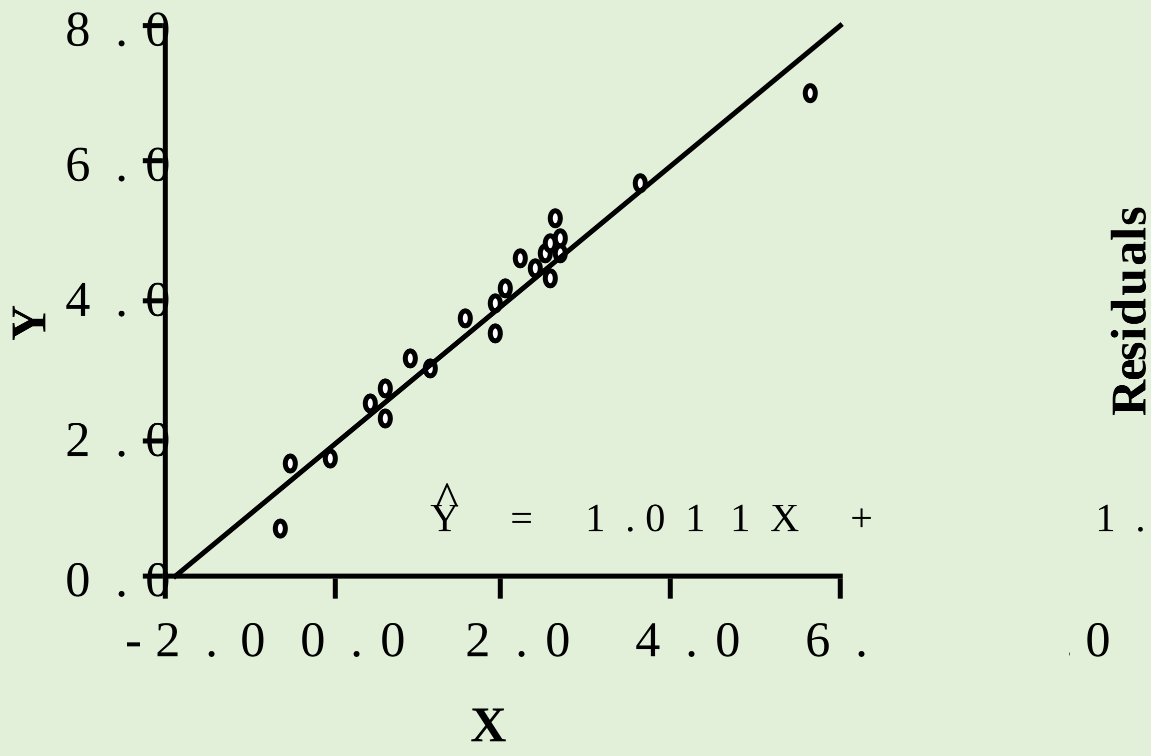
- Residuals are errors between measured and predicted
- Residuals tell us how well the model is working

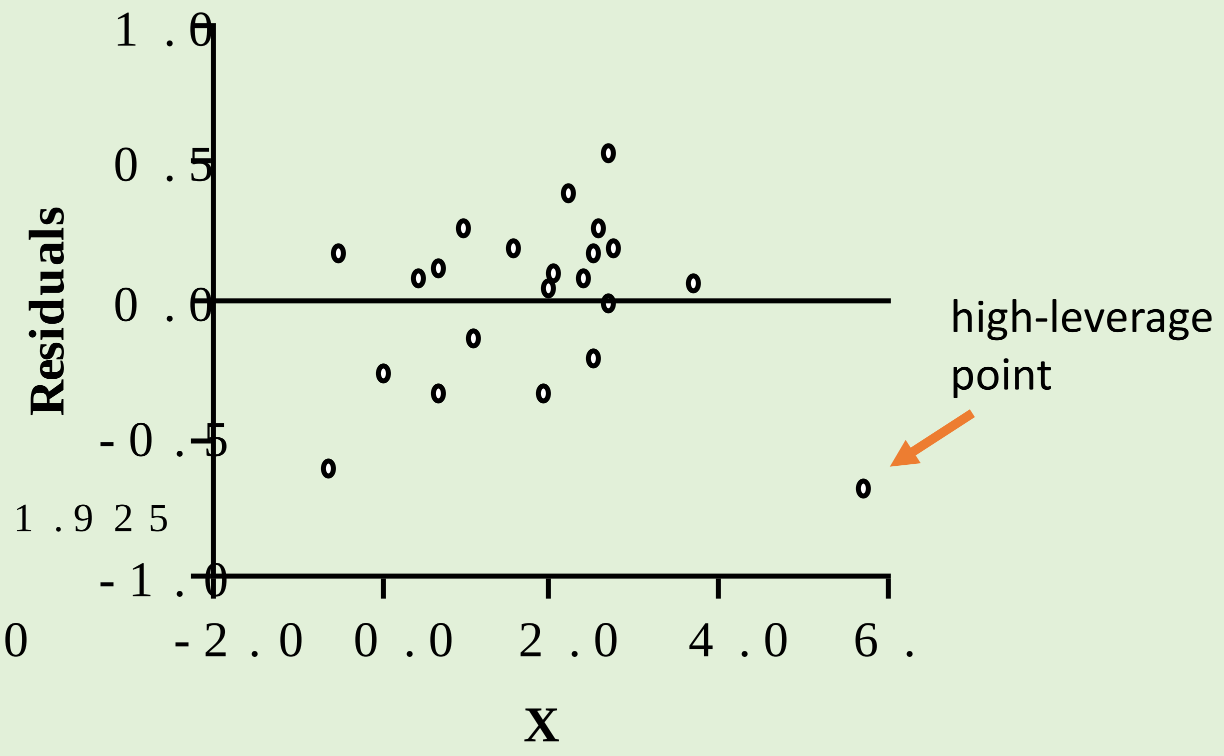




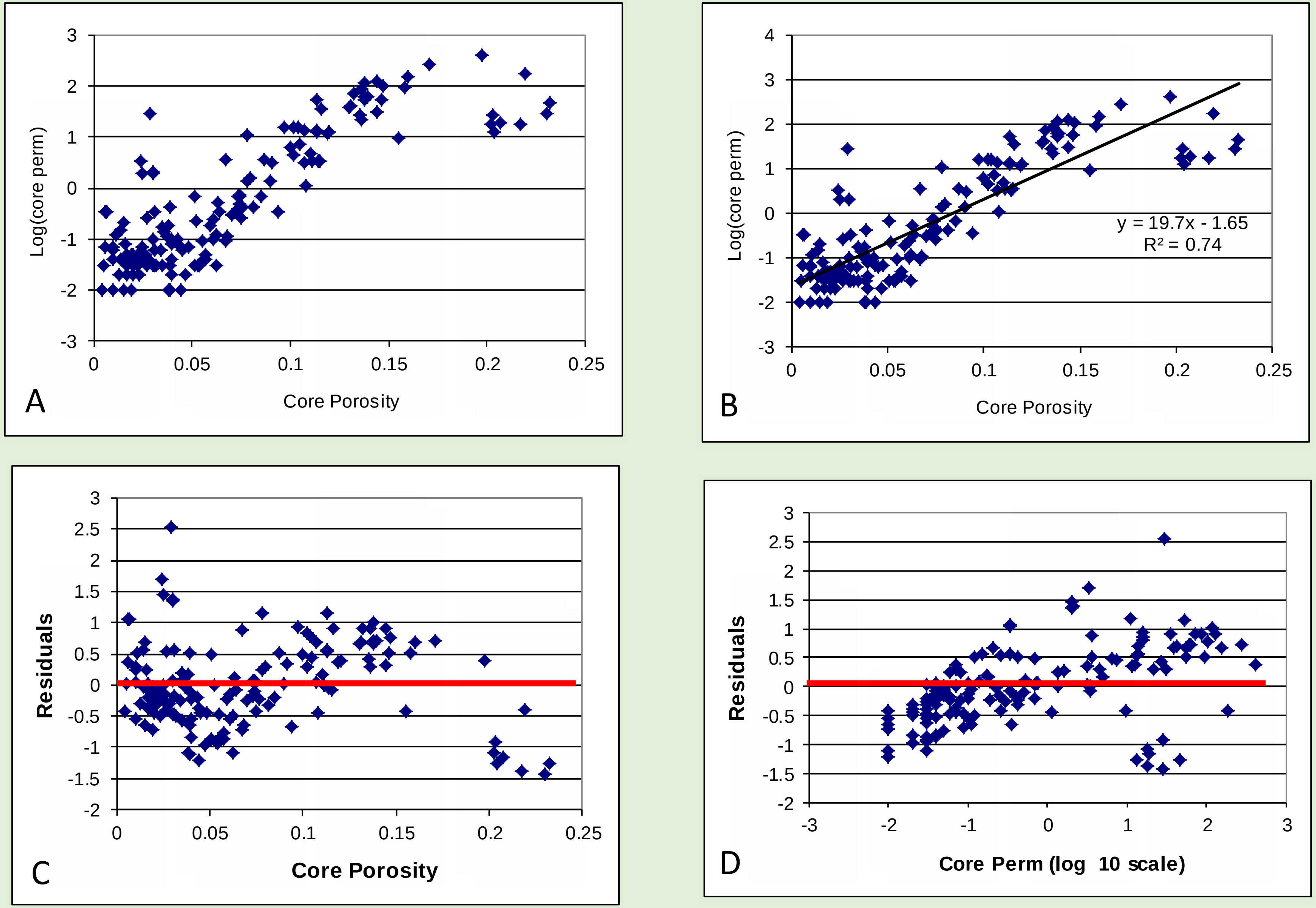
Residuals Analysis

Looks good but the residuals tell us otherwise!

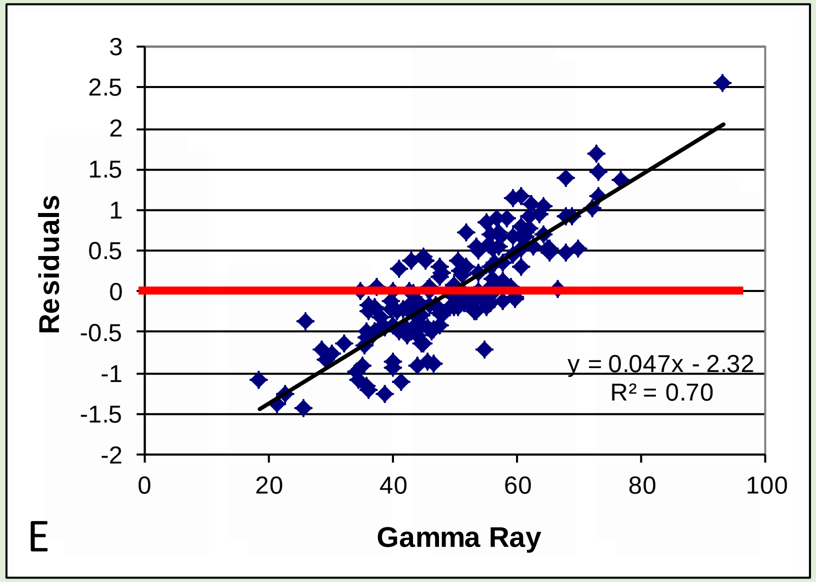


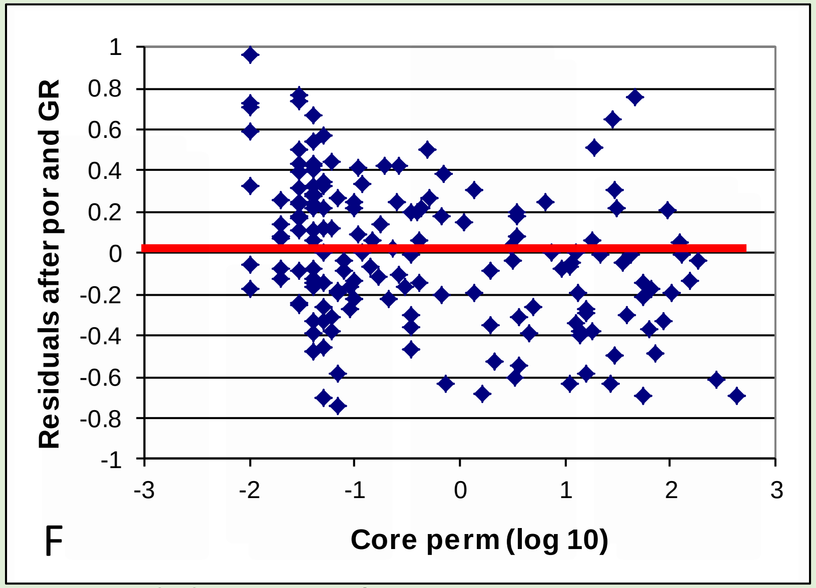


Regression Example Analyzing Residuals (1)



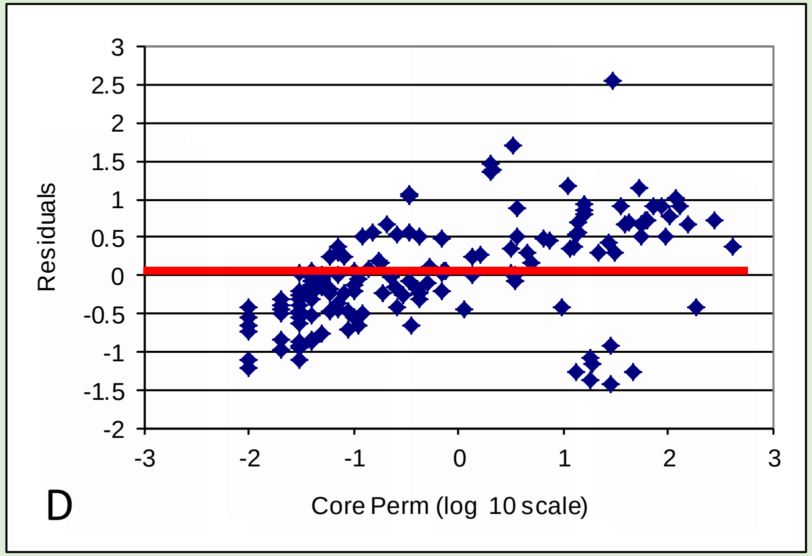
Regression Example Analyzing Residuals (2)





$\text{Log}(k) = 19.7\phi + 0.047GR - 4.0$

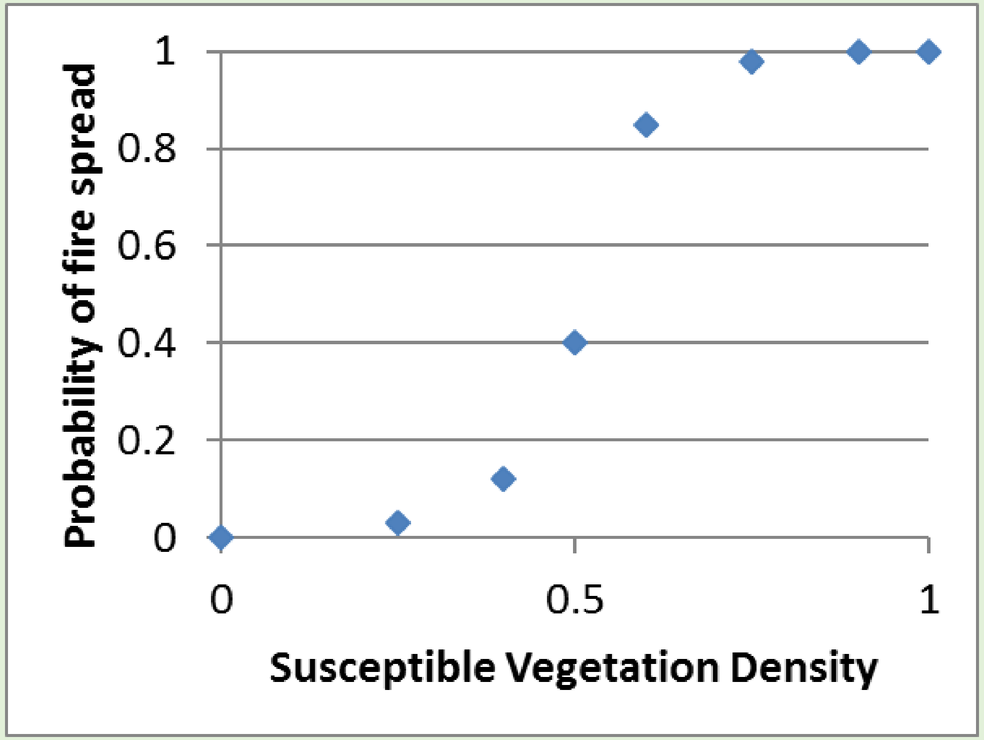
- Line fit with porosity and GR
  - R^2 increases from 0.74 to 0.92
  - Coefficients add
- Plots D vs F
  - F has ½ residuals of D
  - F no obvious trend remaining



Percolation: A Reason for Nonlinear Behavior

What is percolation?

- Connectivity of random systems
- Began in 1940's
- Wide variety of applications
  - Traffic
  - Forest fires
  - Disease
  - Magnetisation
  - Fluid flow



After Turner et al., 1989

Example

- Random distribution of sands and shales
- Assume 1's are sand and 0's are shale
- Possible for left side to "connect" with right side?

IN → | | | | | | | | | | | | | | | | → OUT

0	0	0	1	1	1	0	0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	1	0	0	0	1	1	1	0
1	0	1	1	1	0	0	1	0	1	0	0	0	0	0
0	0	1	0	0	0	1	0	0	1	1	1	1	0	0
1	1	0	0	0	1	0	0	1	1	0	1	1	0	0
1	0	0	1	1	1	1	1	0	0	1	0	0	0	1
1	1	0	0	1	0	0	0	0	0	0	1	1	0	1
0	1	0	0	0	0	1	1	1	1	1	1	1	1	0
1	0	1	0	0	1	0	1	1	0	0	0	1	0	0
0	1	1	1	1	0	1	0	0	1	0	0	1	1	1
1	0	1	1	1	0	1	1	0	0	1	1	0	0	1
0	0	1	0	1	1	1	0	0	0	0	0	1	0	0
1	1	0	0	0	0	1	0	0	1	0	0	0	0	0
1	0	1	1	0	0	0	0	0	0	0	1	0	1	0
0	1	1	0	0	1	1	0	0	0	0	1	0	0	1

40% NTG (Net-to-Gross Ratio)

- ❑Each “cluster” of 2 or more sands is marked
- ❑No cluster connects left to right

0	0	0	1	1	1	0	0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	1	0	0	0	1	1	1	0
1	0	1	1	1	0	0	1	0	1	0	0	0	0	0
0	0	1	0	0	0	1	0	0	1	1	1	1	0	0
1	1	0	0	0	1	0	0	1	1	0	1	1	0	0
1	0	0	1	1	1	1	1	0	0	1	0	0	0	1
1	1	0	0	1	0	0	0	0	0	0	1	1	0	1
0	1	0	0	0	0	1	1	1	1	1	1	1	1	0
1	0	1	0	0	1	0	1	1	0	0	0	1	0	0
0	1	1	1	1	0	1	0	0	1	0	0	1	1	1
1	0	1	1	1	0	1	1	0	0	1	1	0	0	1
0	0	1	0	1	1	1	0	0	0	0	0	1	0	0
1	1	0	0	0	0	1	0	0	1	0	0	0	0	0
1	0	1	1	0	0	0	0	0	0	0	1	0	1	0
0	1	1	0	0	1	1	0	0	0	0	1	0	0	1

biggest cluster

50% NTG

- Still no sand cluster connects left to right

0	0	1	1	1	1	0	0	1	0	0	0	0	0	0
0	0	0	1	0	0	0	1	0	0	0	1	1	1	0
1	0	1	1	1	0	0	1	0	1	0	0	0	0	0
1	0	1	0	0	0	1	0	0	1	1	1	1	0	0
1	1	0	0	1	1	1	0	1	1	1	1	1	0	0
1	0	0	1	1	1	1	1	1	1	1	0	0	0	1
1	1	1	0	1	0	0	0	0	0	0	1	1	0	1
0	1	0	1	0	0	1	1	1	1	1	1	1	1	0
1	1	1	0	0	1	0	1	1	0	0	0	1	0	1
0	1	1	1	1	1	1	0	0	1	0	0	1	1	1
1	0	1	1	1	0	1	1	0	1	1	1	1	1	1
0	0	1	0	1	1	1	0	0	0	0	1	1	0	1
1	1	1	0	0	1	1	0	0	1	0	0	0	0	0
1	1	1	1	1	1	1	0	0	1	1	1	0	1	1
1	1	1	0	0	1	1	0	0	0	1	1	0	0	1

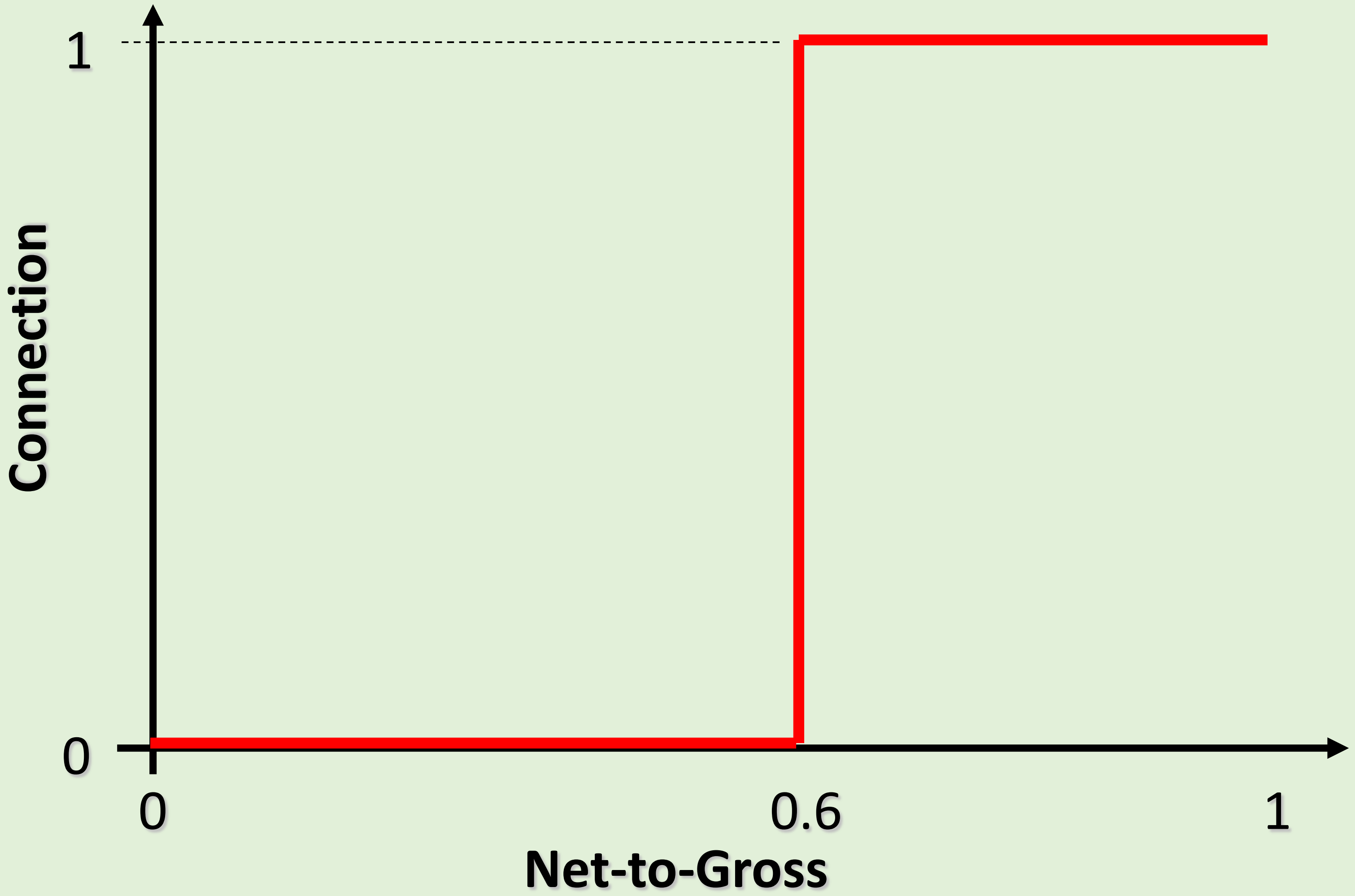
60% NTG….Connection!

- One cluster connects left to right and top to bottom

0	0	1	1	1	1	0	1	1	0	0	0	0	0	0
0	0	0	1	0	0	0	1	1	1	0	1	1	1	0
1	0	1	1	1	1	0	1	0	1	0	0	0	0	1
1	0	1	0	0	0	1	0	0	1	1	1	1	0	0
1	1	0	1	1	1	1	0	1	1	1	1	1	0	0
1	0	0	1	1	1	1	1	1	1	1	1	0	0	1
1	1	1	0	1	0	0	0	0	0	0	1	1	0	1
0	1	1	1	0	0	1	1	1	1	1	1	1	1	0
1	1	1	1	0	1	0	1	1	0	0	0	1	0	1
1	1	1	1	1	1	1	1	0	1	0	0	1	1	1
1	0	1	1	1	0	1	1	0	1	1	1	1	1	1
1	0	1	1	1	1	1	1	0	0	0	1	1	1	1
1	1	1	0	0	1	1	0	0	1	1	0	1	0	0
1	1	1	1	1	1	1	1	0	1	1	1	0	1	1
1	1	1	0	0	1	1	0	0	0	1	1	0	0	1

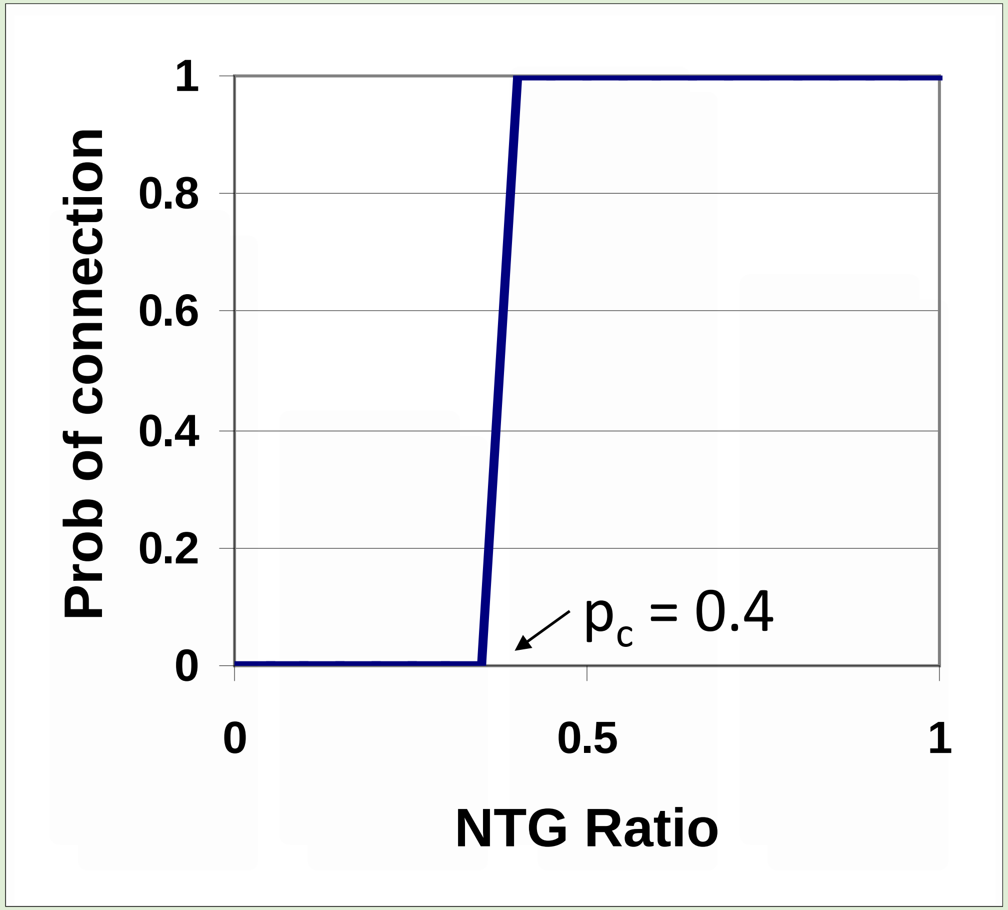
connecting cluster

Connection and NTG



Features of Ideal Percolating Systems

- Infinite size network
- Regular arrangement of sands
  - Square
  - Hexagonal
  - ….
- Non-linear relationship
- "Minimum" NTG for sands to connect – **percolation threshold**
  - p_c = 0.4 to 0.7 for 2D
  - p_c = 0.1 to 0.25 for 3D



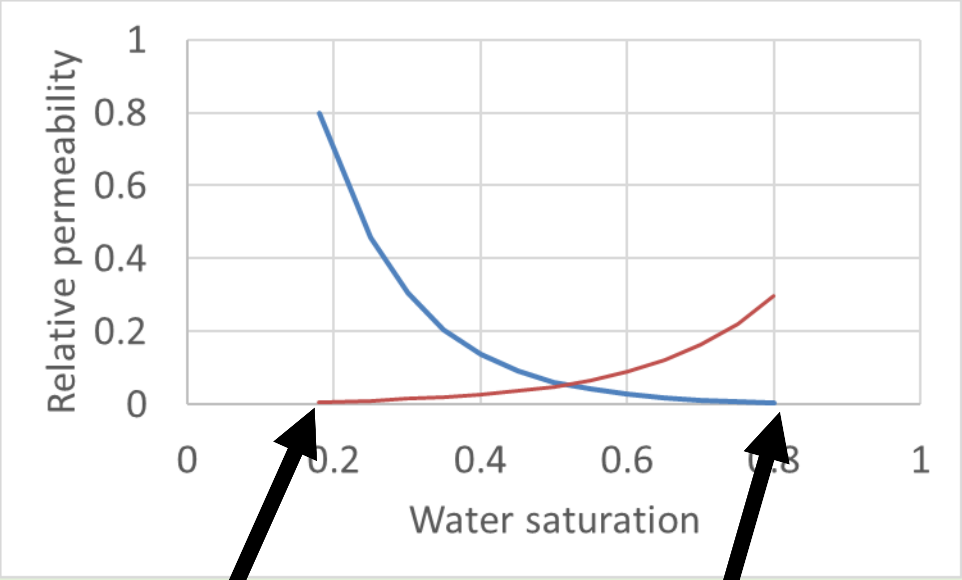
Petrophysics and Percolation

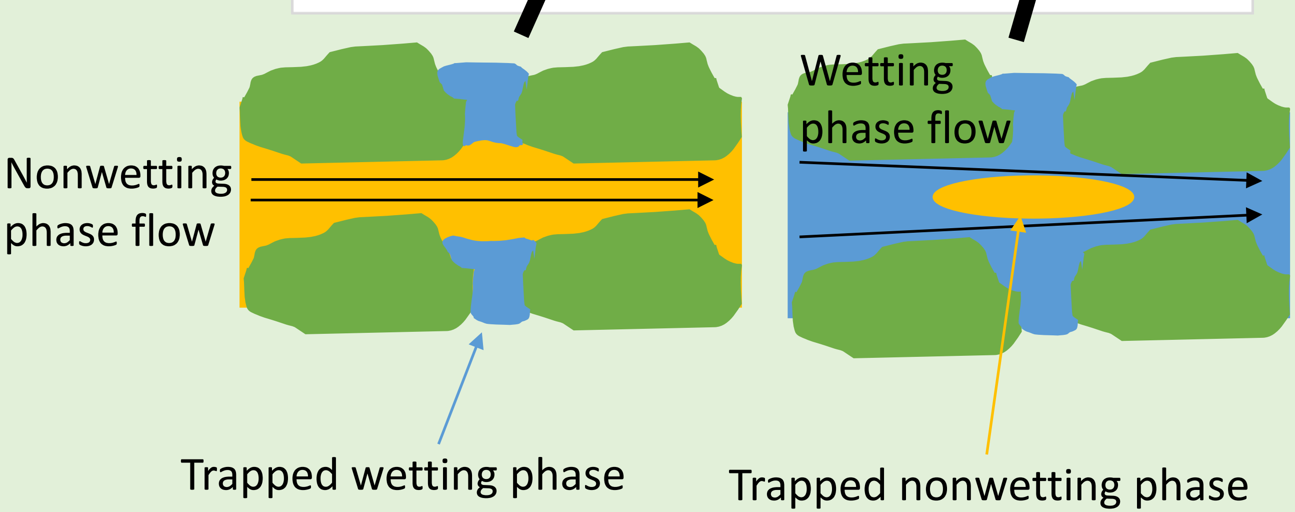
- Despite
  - Simplicity of concept
  - And reservoir behavior is complex
  - Percolation can be seen in many phenomena
- Examples
  - Endpoint saturations
  - Connectivity

Example 1
Endpoint Saturations

Endpoint Saturations

- Two-phase perms viewed as percolation phenomenon
  - Trapping vs flowing = non-percolating vs percolating
  - Typical endpoint sat values similar to 3D p_c values
- Percolation viewpoint helps evaluate
  - Variability of endpoint values
  - Effects of heterogeneity



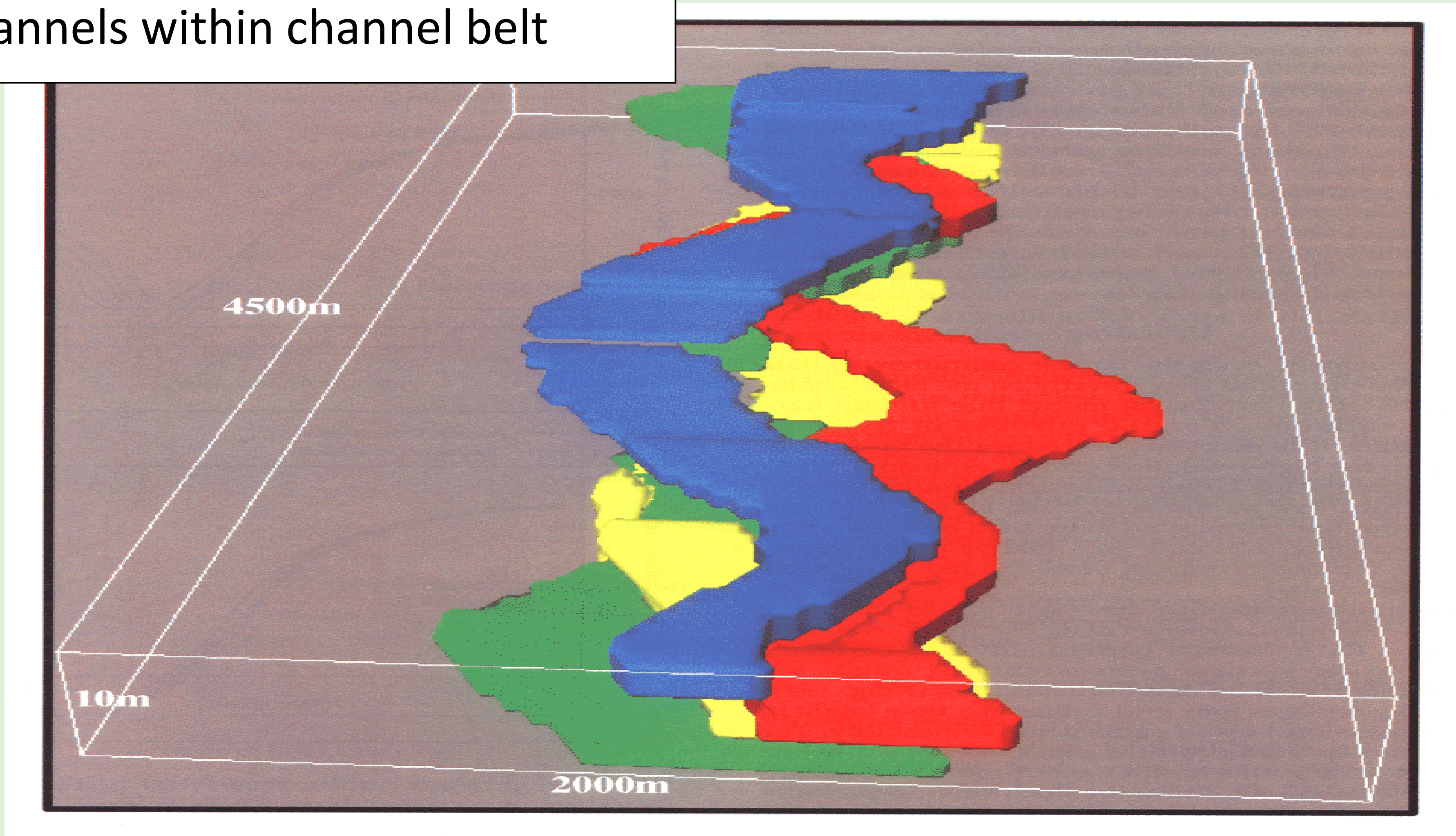


After Stegemeier, 1977

Example 2
Reservoir Connectivity
Fluvial Models

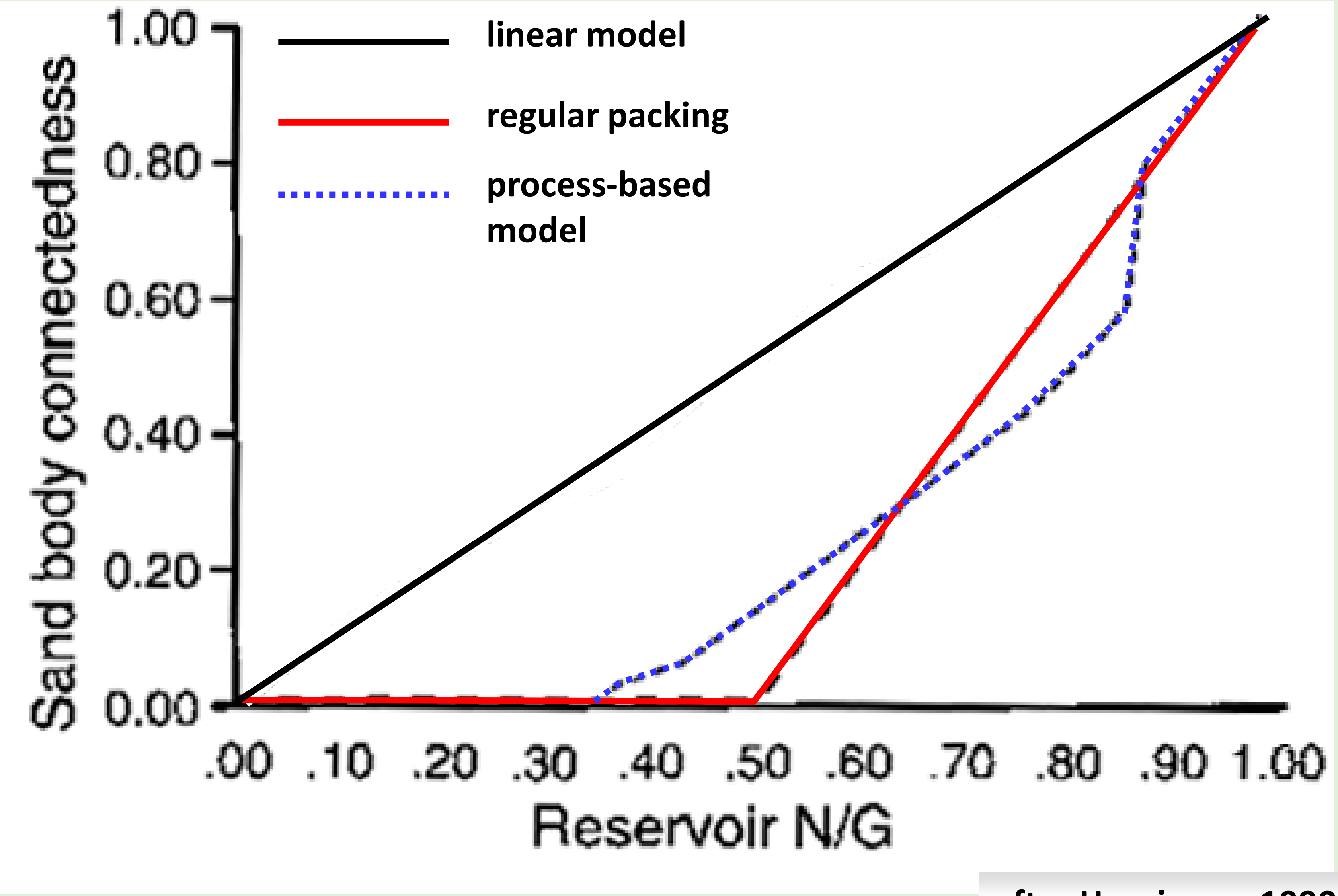
Model output from meandering fluvial simulation

channels within channel belt



Tyler et al., 1994

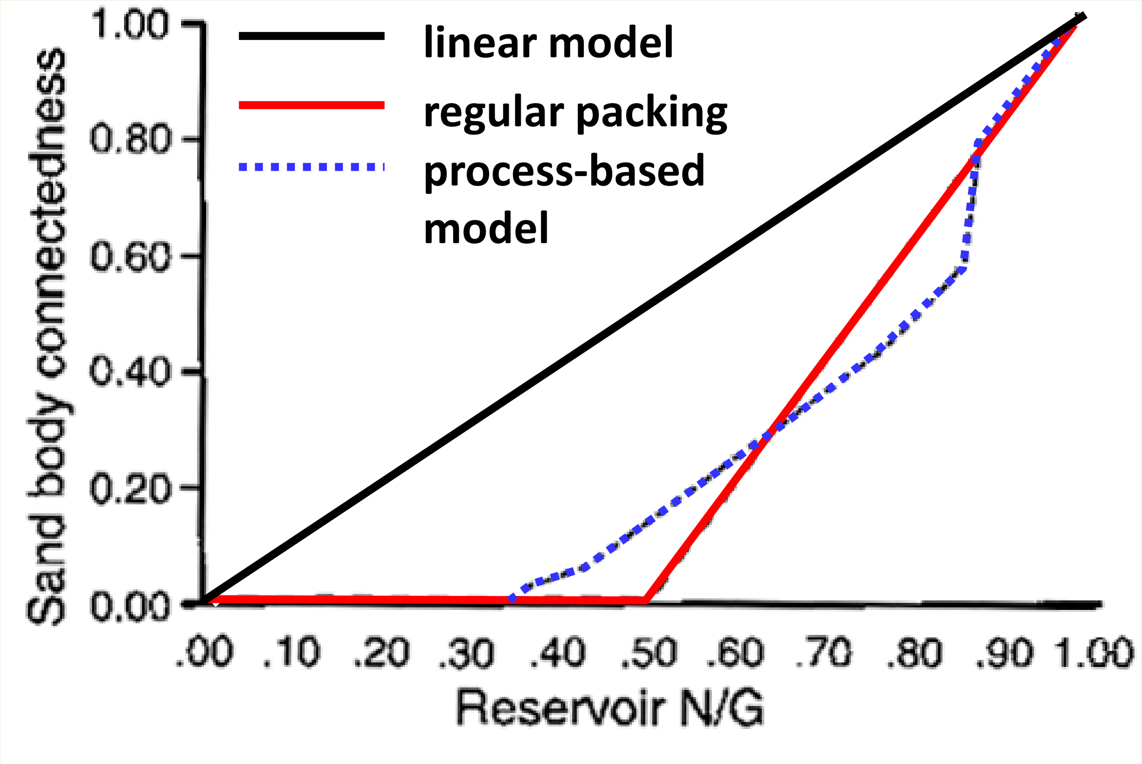
Channel Placement and Connectivity



after Henriquez, 1990

Connectivity and Percolation

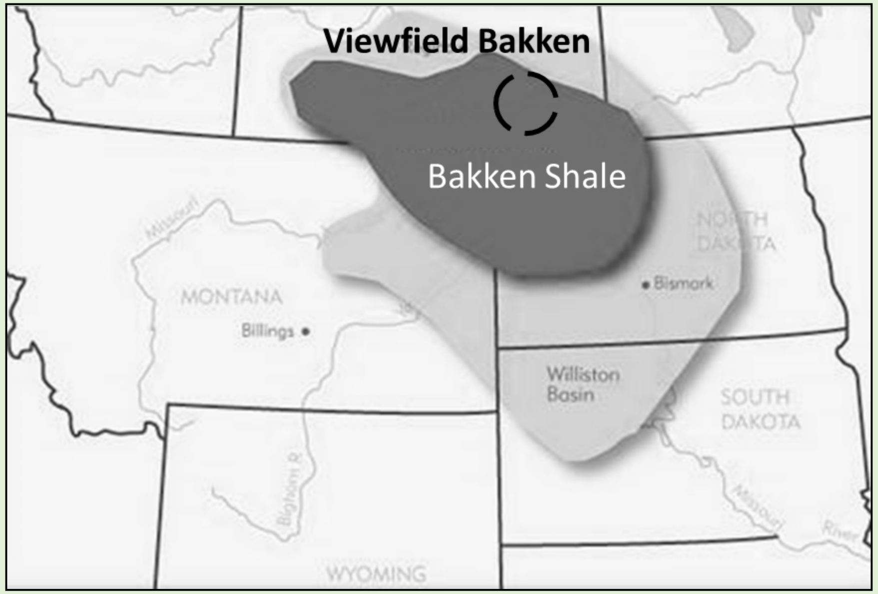
- ❑ Linear model may not work
- ❑ Below critical net-to-gross
  - o Careful model choice
  - o Detailed evaluation required
- ❑ Above critical net-to-gross
  - o Insensitive to model
  - o Less evaluation needed
  - o Large NTG's give similar results

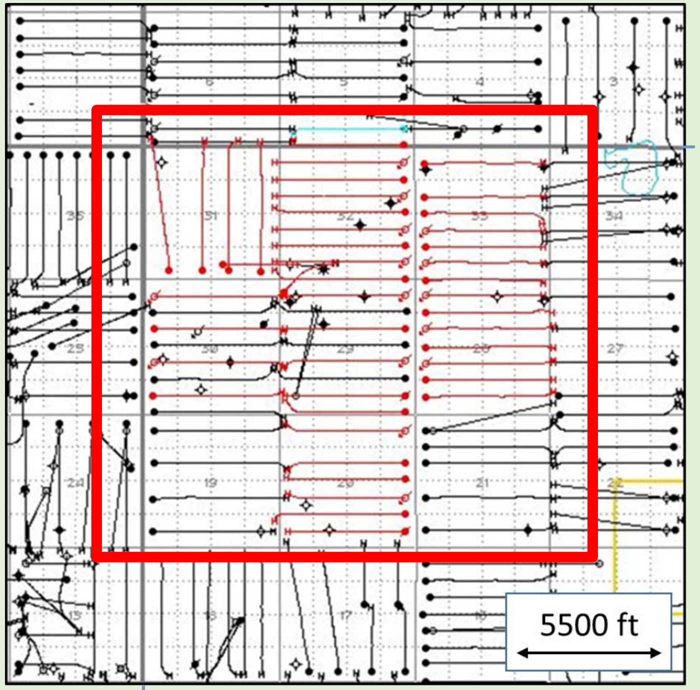


Example 3
Reservoir Connectivity
in Tight Oil

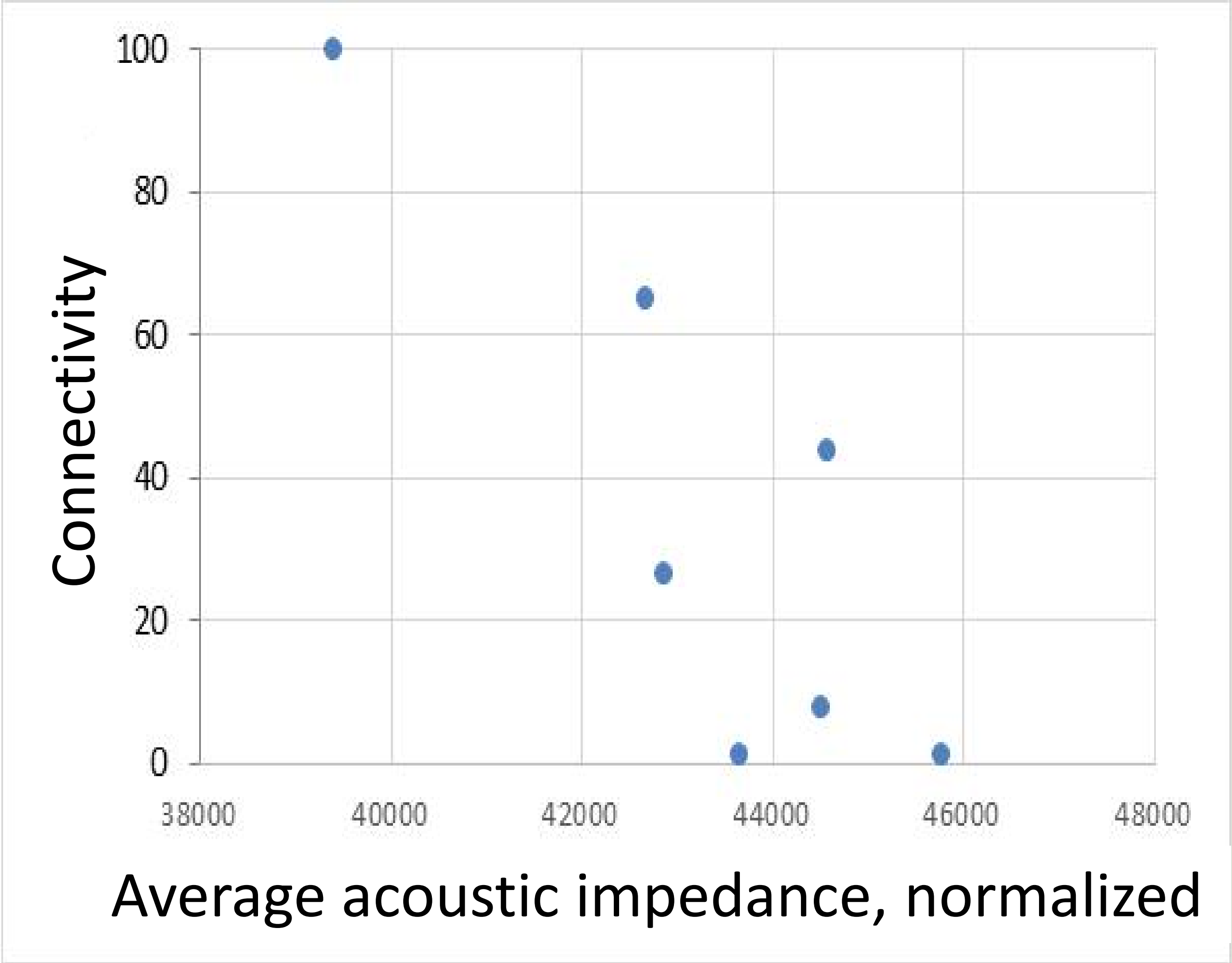
Bakken Waterflood

- Williston Basin
- 1μd – 1 md perm
- Complex geology
  - Faults & fractures
  - Aquifer (Lodgepole) above
  - Silts and shale
- Data include
  - Inj + prod flow rates
  - 3D seismic
- Details in Mirzayev et al. (2017)



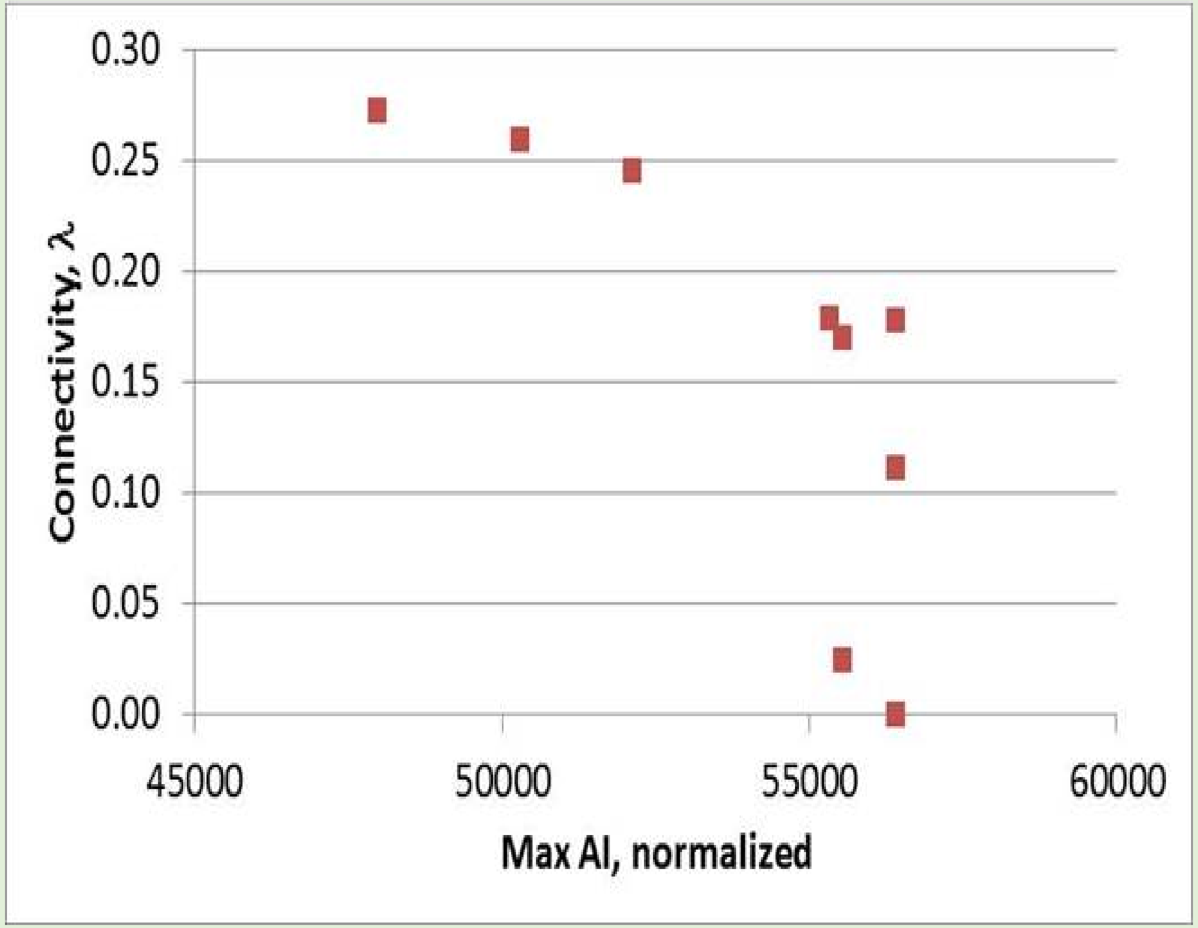


Evidence of faults & fractures on longer distance (non-neighboring wells) connectivity



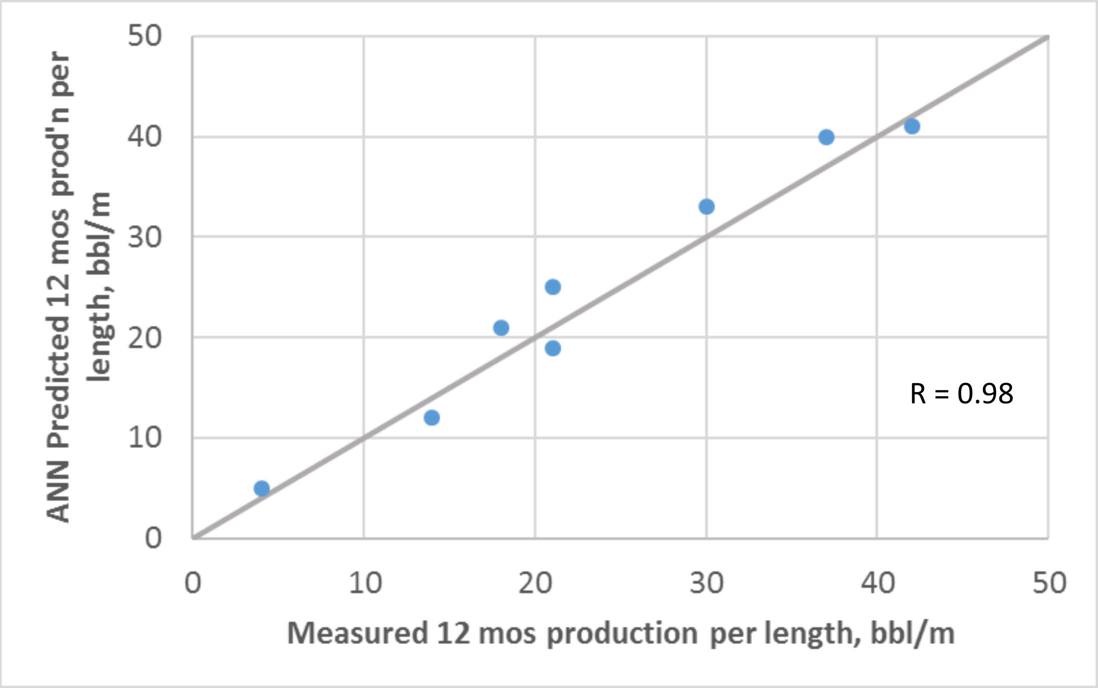
Rethinking the Acoustic Impedance Results

- ❑ Max AI vs connectivity shows percolation threshold
- ❑ Connectivity controlled by <u>least</u> porous event(s)



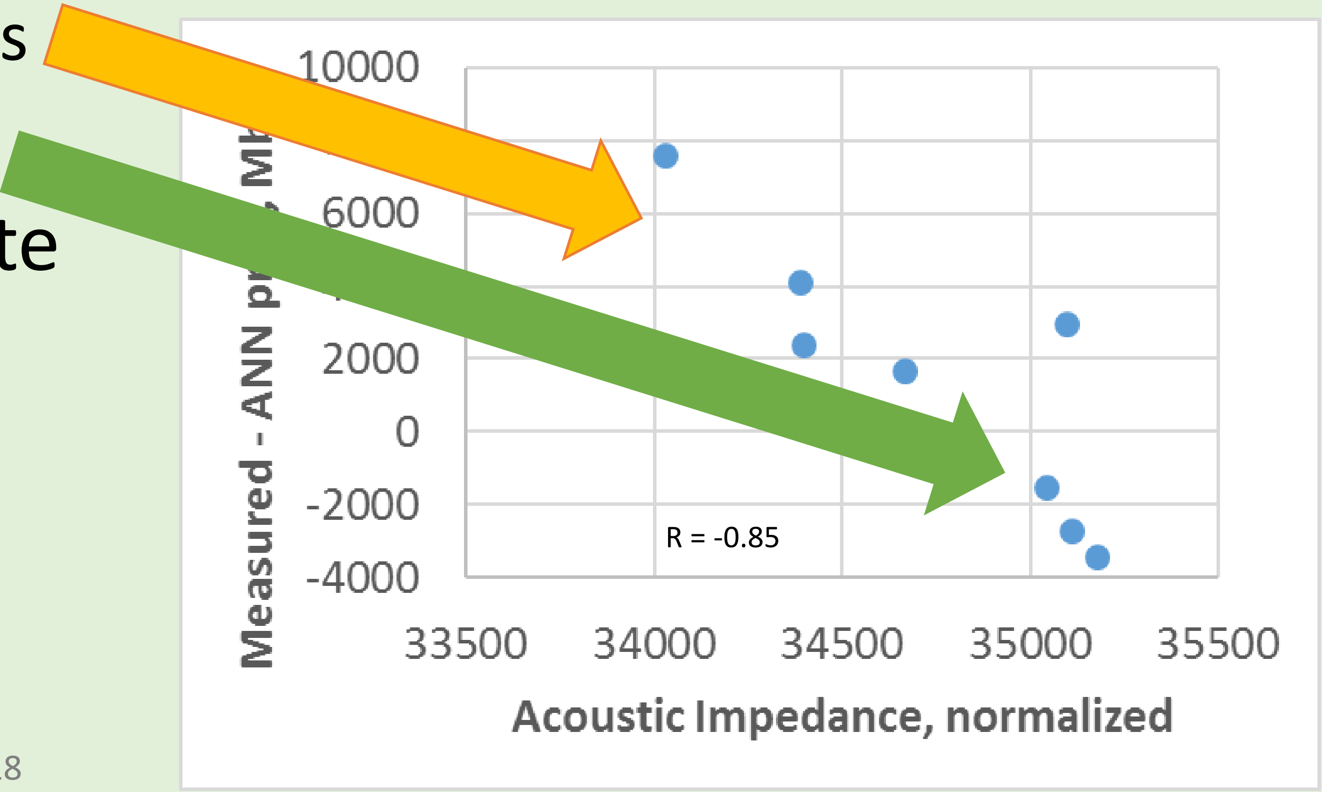
Does Connectivity Matter?

- AI relates to connectivity; does conn affect production?
- Examine 8 wells outside of analysis area
- ANN to predict production per meter well length
  - Does **not** include AI
  - Removes effects of frac company, proppant amt etc



Does Connectivity Matter?

- Compare ANN predictions to measured 12 months production
- Measured – ANN predicted (= residuals)
  - If zero, ANN has all variables and models prod'n accurately
  - If > 0, NN underpredicts
  - If < 0, NN overpredicts
- Residuals & AI correlate



Conclusions on Predictions

- Uncertainty is present, no matter what method we use
- Models + Data usually give less uncertainty than data alone
- Model building
  - Use simple models first
  - Spend time to examine residuals; don’t depend on R^2
- Percolation - good model for many reservoir characteristics

Acknowledgements

- Students
  - Kush Kakar
  - Yu Sun
- Colleagues
  - Per Pedersen
  - Chris Clarkson
  - Danial Kaviani