What's New for Us in Data Analytics, Geostatistics and Machine Learning

Michael Pyrcz, Associate Professor

P.I. Texas Center for Geostatistics

Hildebrand Department of Petroleum and Geosystems Engineering

Bureau of Economic Geology, Jackson School of Geosciences

The University of Texas at Austin

Who am I?

- New **Professor** in UT PGE, started Fall 2017
- 17 years of experience in consulting, teaching and industrial R&D in **data analytics**, **reservoir modeling** and **uncertainty characterization**, **statistical / machine learning**
- **Associate Editor** for Computers and Geosciences Journal
- **Editorial Board** of Mathematical Geosciences Journal
- **Author** of the textbook "Geostatistical Reservoir Modeling" and > 40 peer reviewed publications, patents etc.
- **Program Chair** for SPE Petroleum Data Driven Analytics Technical section (PD²A)

Michael Not Working

Michael Working

Research

Geostatistics, Spatial Statistics, Statistics
- Multivariate, multiscale, spatial modeling with uncertainty
- Grid-free representations
- Uncertainty models for unconventional reservoirs

Data Analytics
- Data cleaning and bias
- Data preparation, imputation, scaling and updating
- Data visualization

Machine Learning
- Fast proxies for real-time feedback and optimization
- Multivariate, spatiotemporal inference and prediction
- Physics informed machine learning

Motivation

Modeling for Decision Support

we only add value when we impact a decision

Integrated Modeling

if it doesn't get in the model, it doesn't matter

Robust Statistics

Statistical learning needs statistics done well

Deployment

New tools, new workflows to improve practice

Examples

1. Optimal Well Placement

2. Reservoir Multiscale Flow Modeling

3. Production Forecasting

4. Induced Seismicity

5. Improved Geomodeling

Optimal Well Placement

Requires fast iteration of flow forecasting to sample from the large potential solution space

Solution: train and apply a machine learning-based / gradient boosting decision tree (XGBoost Library) for fast proxy for flow forecasting

Standard multivariate machine learning is oblivious of spatial context and connectivity

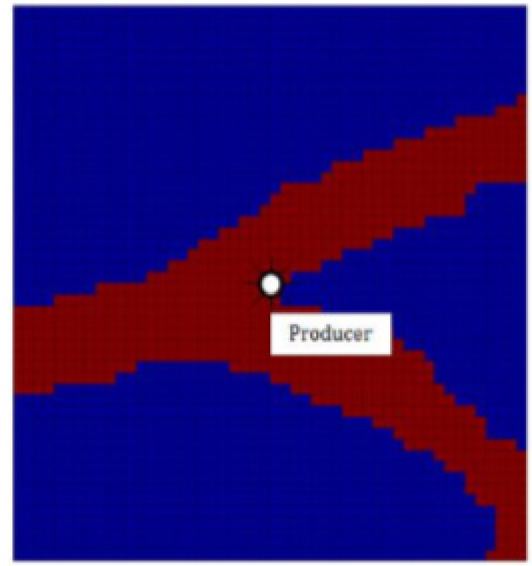
Solution: apply fast marching to calculate an efficient measure of well-to-well connectivity to include as a feature in the machine learning model

Optimal Well Placement

Student Azor Nwachukwu, Co-supervised with Prof. Larry Lake

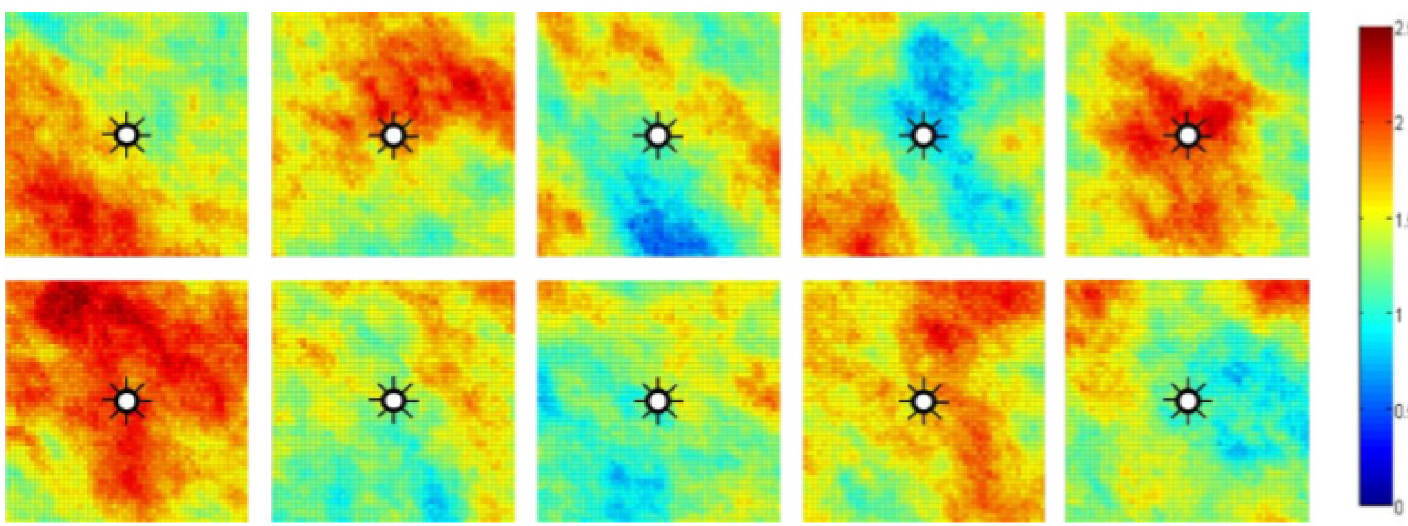
Experiment I:

- Channel facies (30%, 300md)
- Overbank facies (5%, 10 md)
- 2 variable injector locations
- 2000 days of waterflood



Experiment 2:

- Gaussian co-simulation



10 sample models (log of permeability)

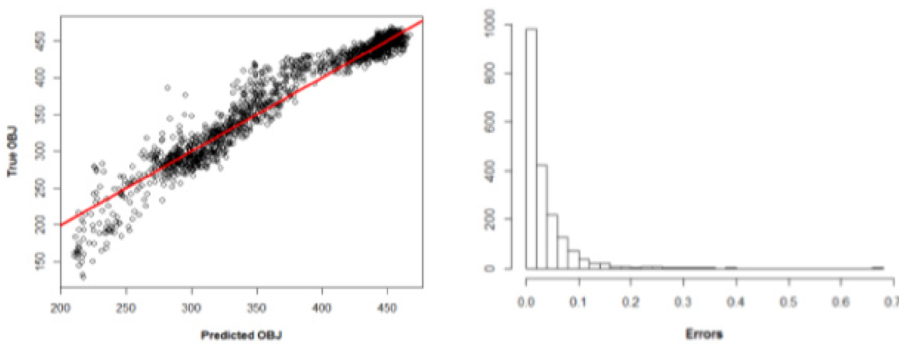
Optimal Well Placement

Student Azor Nwachukwu, Co-supervised with Prof. Larry Lake

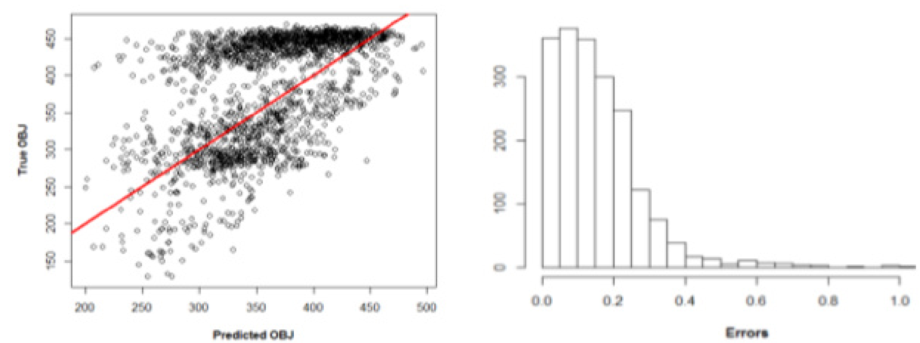
Results:

- Connectivity information is essential for a good machine learning proxy for production rates.

With Connectivity Information

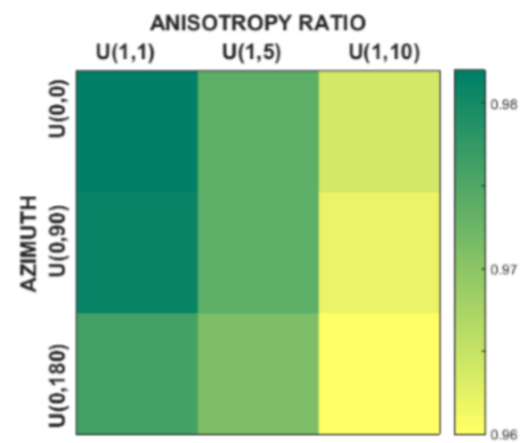


Without Connectivity Information



- Training model diversity is essential for a good machine learning proxy for production rates.

Training Diversity vs. Prediction Accuracy



Well Forecasting

Student Azor Nwachukwu, Co-supervised with Prof. Larry Lake

Requires ability to learn from the past and forecast into the future:

Solution: recurrent neural networks for time series prediction, long short term memory networks. Vector to vector encoder-decoder with retraining when new data is available.

LSTM Design for Short Term Updates to Long Term

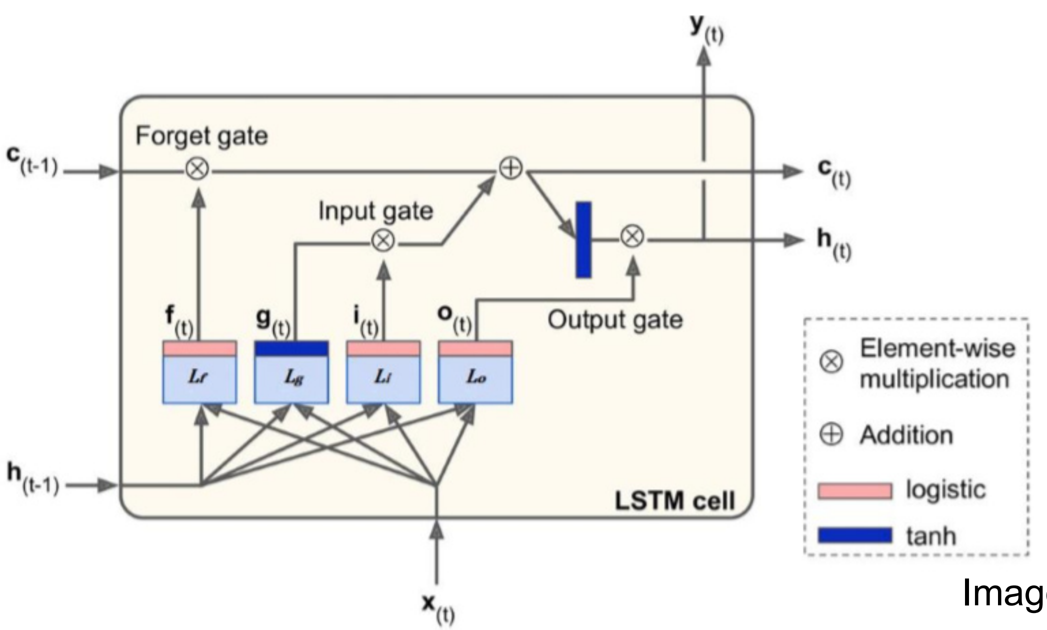


Image from Géron, 2017

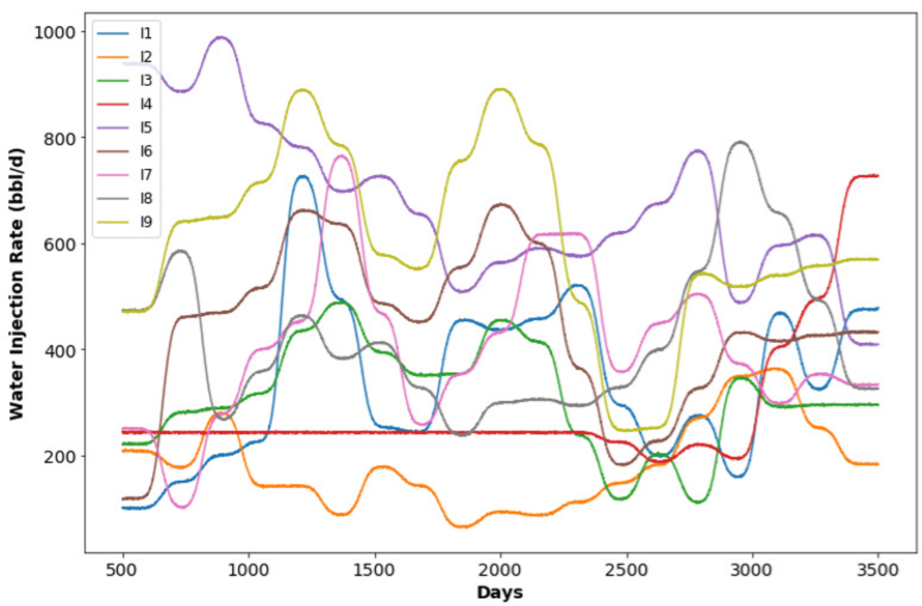
Well Forecasting

Student Azor Nwachukwu, Co-supervised with Prof. Larry Lake

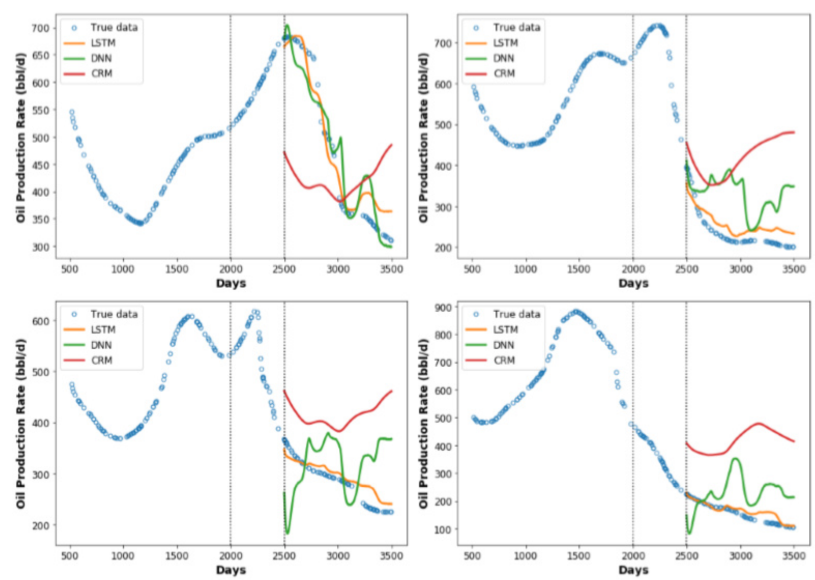
Results:

- Prediction during 3000 days of water flow (500 days of primary not included).
- Inputs included injection and oil rates production rates shifted backward.

Injection Rates Over Train and Test Intervals



Production Over Train and Modeled Over Test



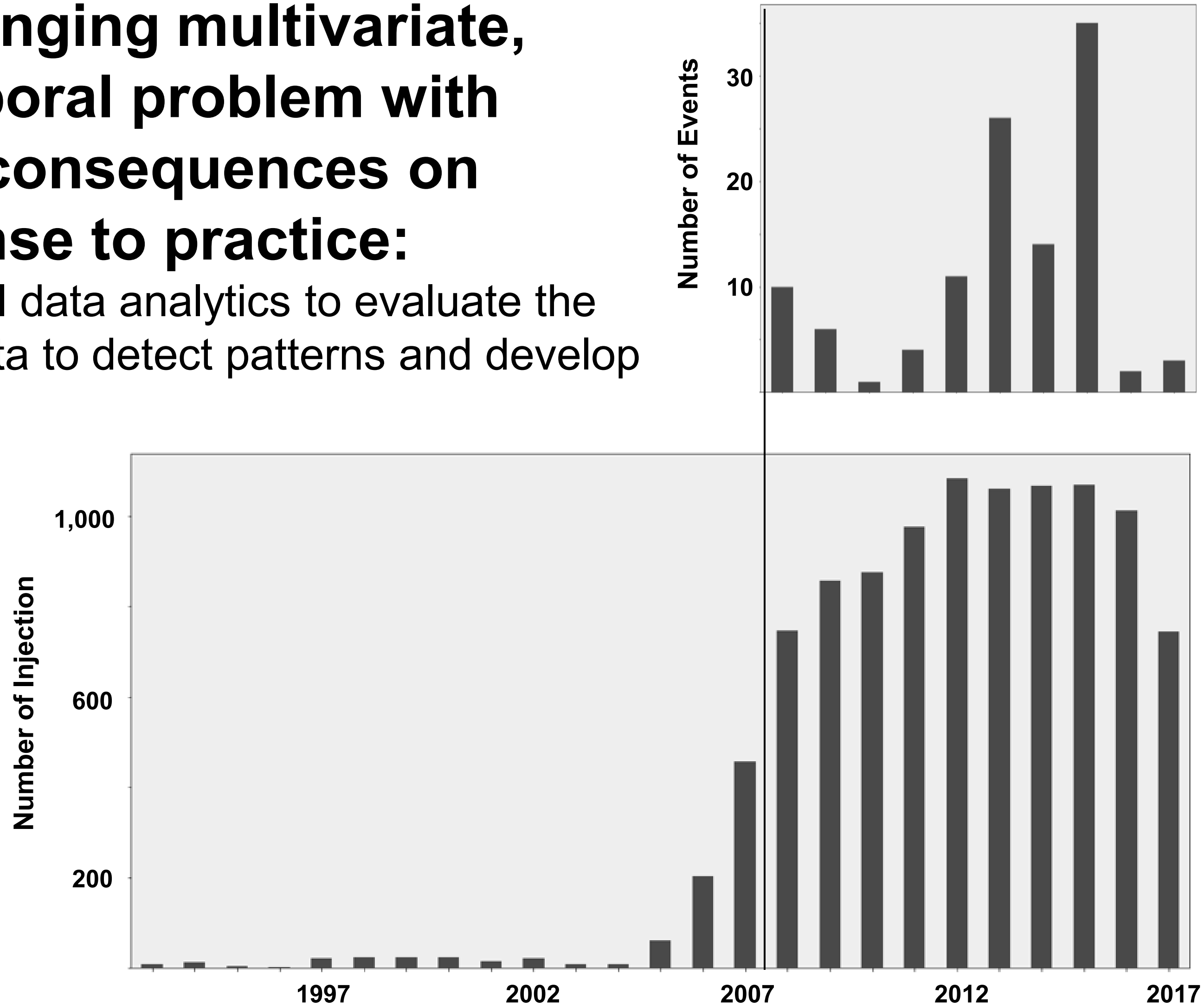
Induced Seismicity

Michael Pyrcz

Very challenging multivariate, spatiotemporal problem with important consequences on social license to practice:

- Apply spatial data analytics to evaluate the available data to detect patterns and develop hypotheses.

Injection and Seismic Events Over Time



Frequency vs year bar chart for events (top) and injections (bottom).

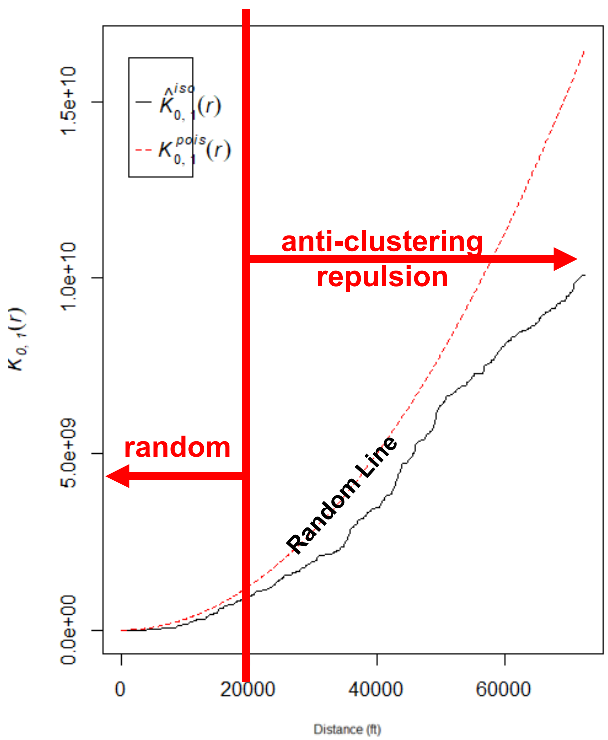
Induced Seismicity

Michael Pyrcz

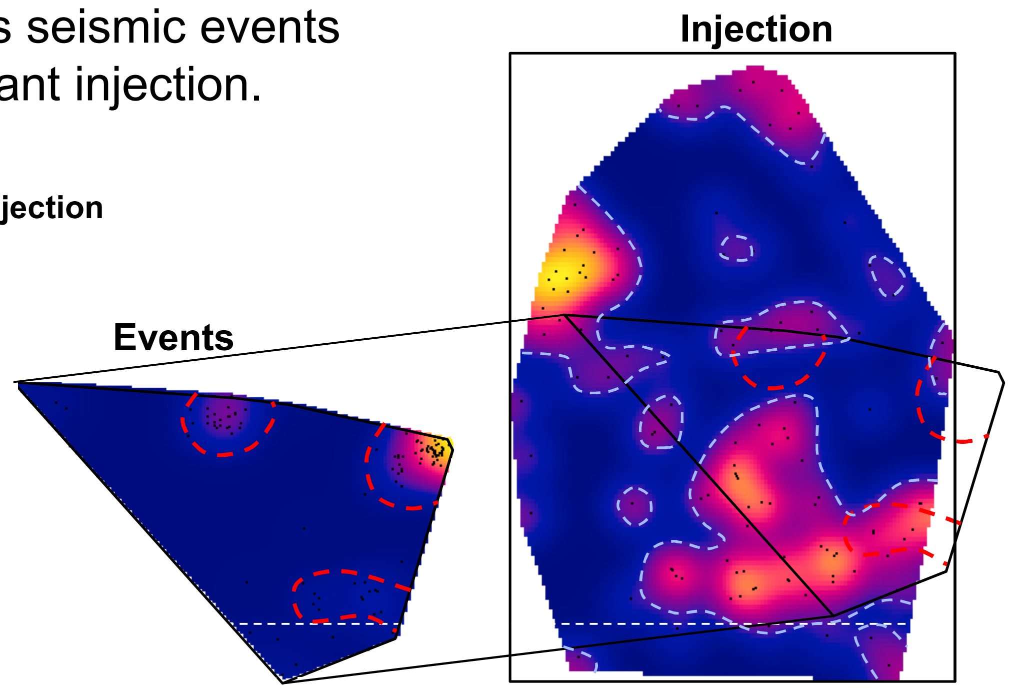
Results:

- Hotspot analysis determines seismic events occur in locations of significant injection.

Ripley's K cross function for events, given injection



Hotspot Analysis of Seismic Events and Injection



- Cross Ripley's K function did not detect higher rates of seismic events near injectors than random.

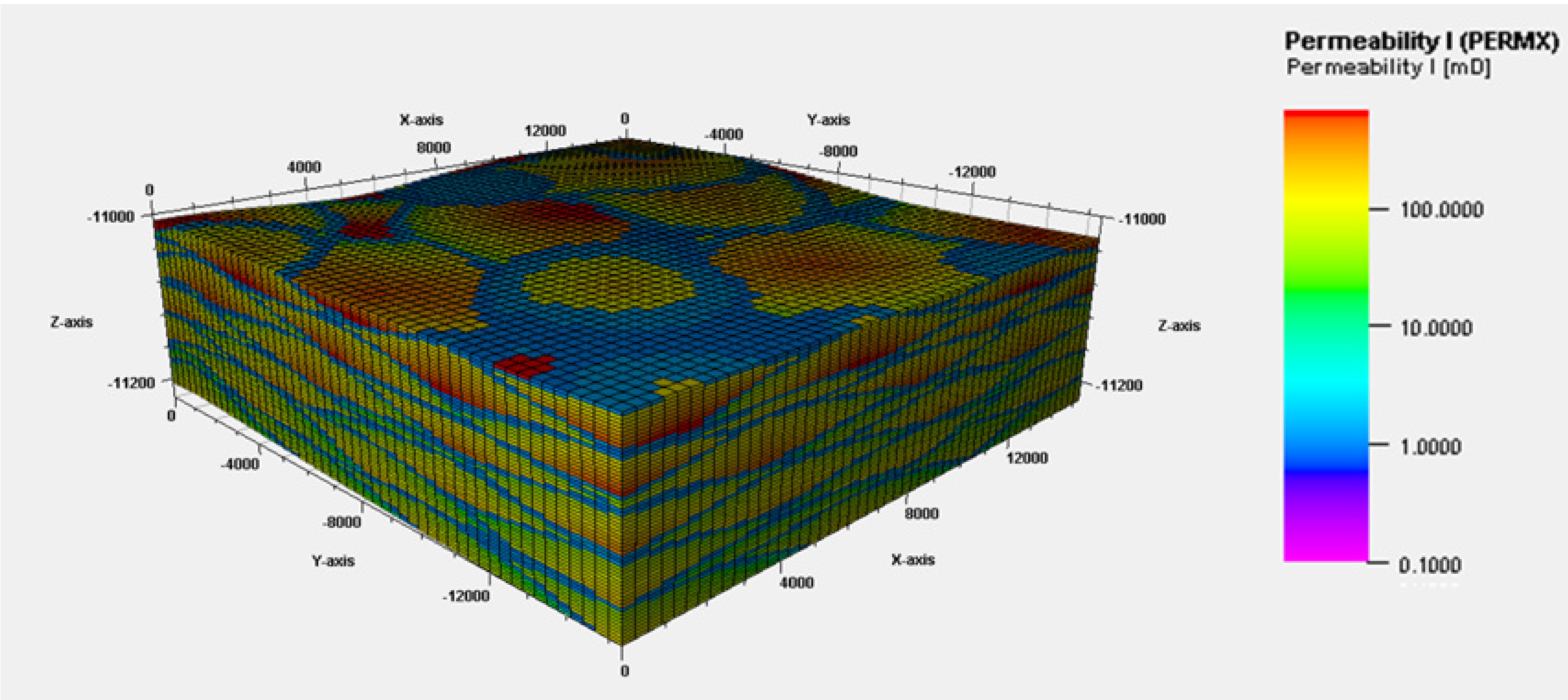
Improved Geomodelling

Student Honggeun Jo

Improved reservoir models are required to capture complicated heterogeneity fields:

- New process-mimicking methods that condition to concepts and data
- Rules calibration to geologic observation, robust, artifact free features

Geologically Calibrated, Artifact Mitigated Process-mimicking Model



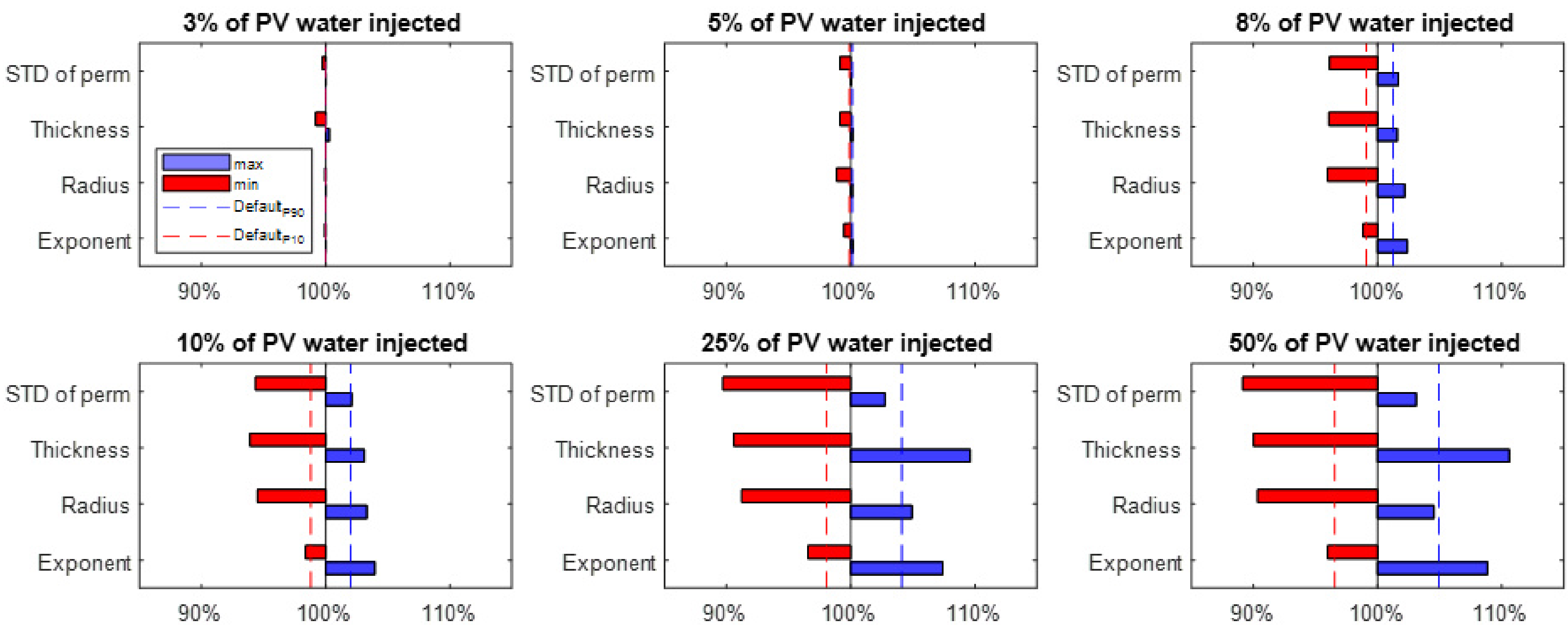
Improved Geomodelling

Student Honggeun Jo

Value and impact of architectural complexity in models:

- Standard reservoir models omit architectural complexity
- Inform reservoir modeling and to motivate new methods adoption.

Impact of Architectural Complexity Varies with Production Stage



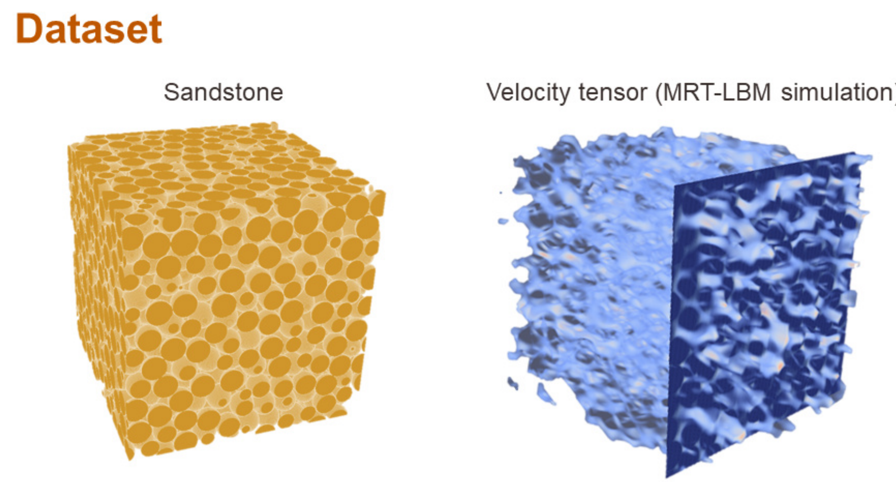
Improved Geomodelling

Student Javier Santos, Co-supervised by Prof. Prodanovic

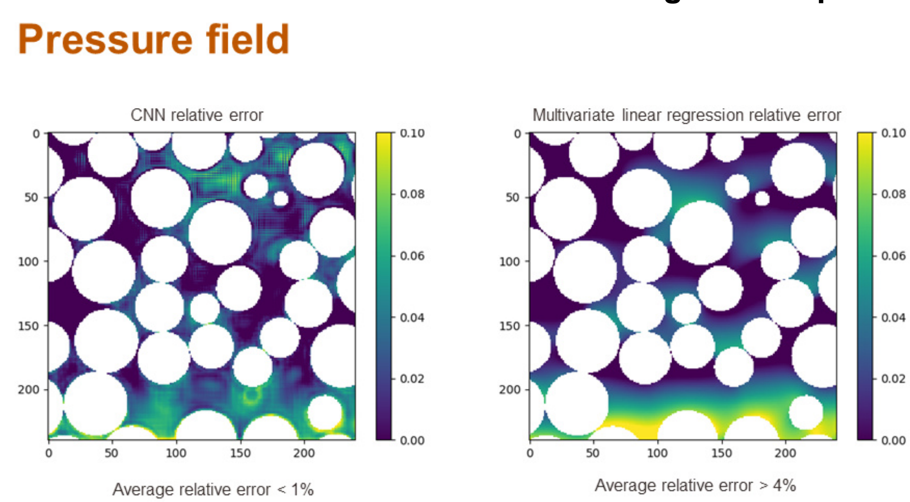
Integrating pore scale flow behavior into reservoir scale models:

- Train fast machine learning models to predict velocity and pressure
- Image segmentation into training sets
- Train convolution neural net with various measures e.g. distance transform

Pore Scale Model and Flow Model



Pressure Field Prediction Error Machine Learning and Simple Regression

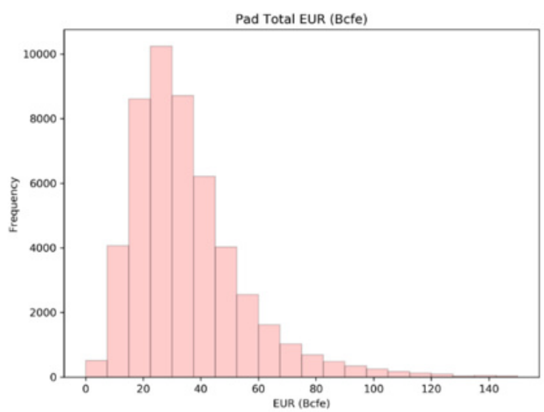


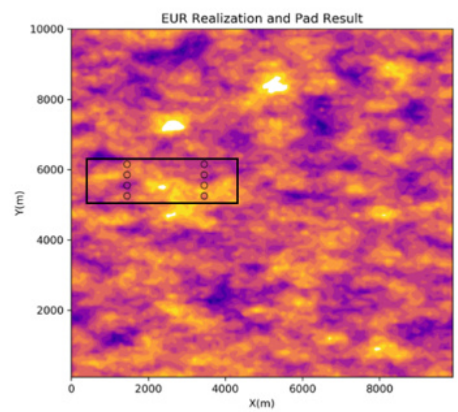
Uncertainty for Unconventionals

Michael Pyrcz

Uncertainty models are needed for unconventional reservoirs:

- Empirical model resampling for spatially constrained uncertainty

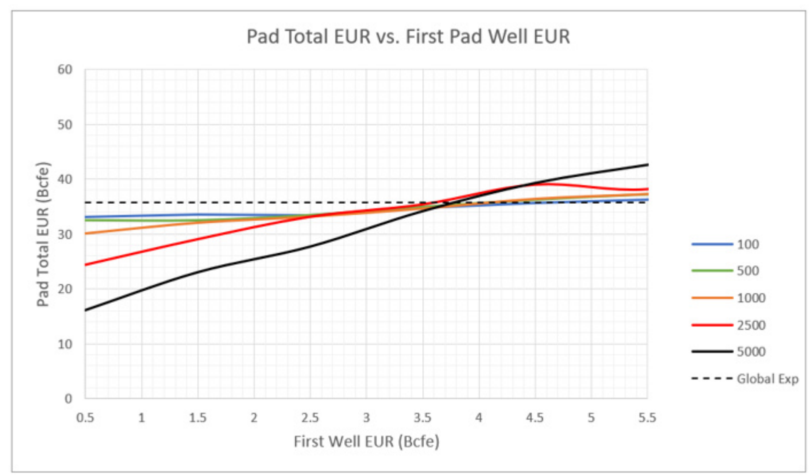




Well EUR realization and candidate pad location and aggregate performance

- Decision criteria with pad early indicator

Pad Abandonment First Well Decision Criteria



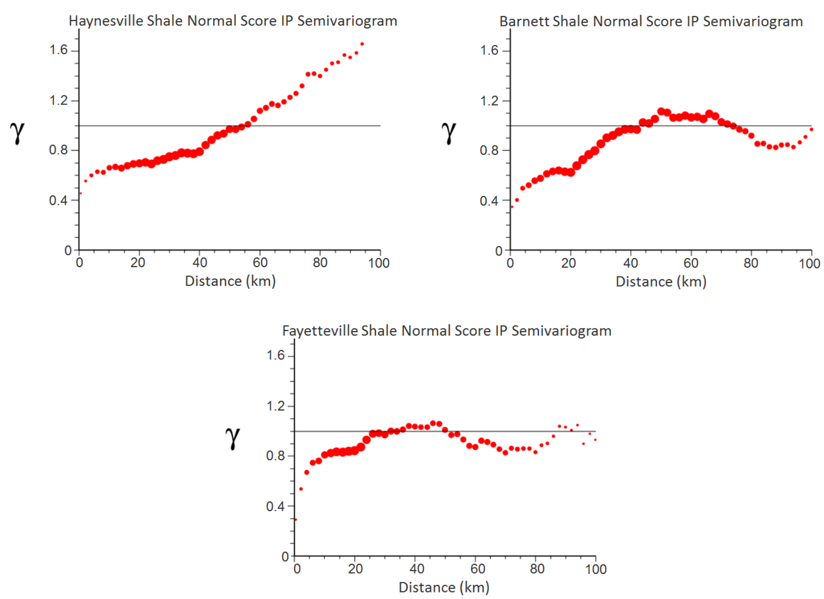
Data Analytics for Unconventionals

Michael Pyrcz

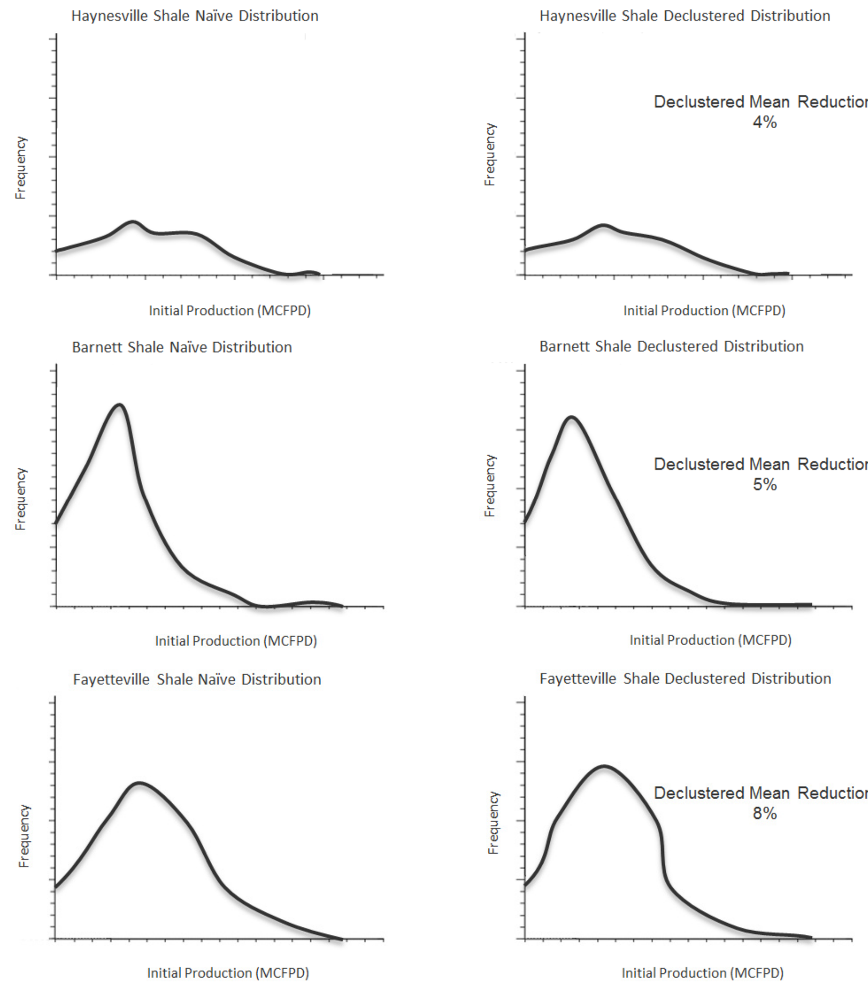
Formulation of representative statistics for unconventional data analytics and machine learning:

- Spatial declustering and continuity measures.

Spatial Correlation for US Domestic Shale Gas Production



Declustered Distributions for US Domestic Shale Gas Production



Subsurface Data Analytics and Machine Learning Consortium

New Industrial Affiliates Proposal:

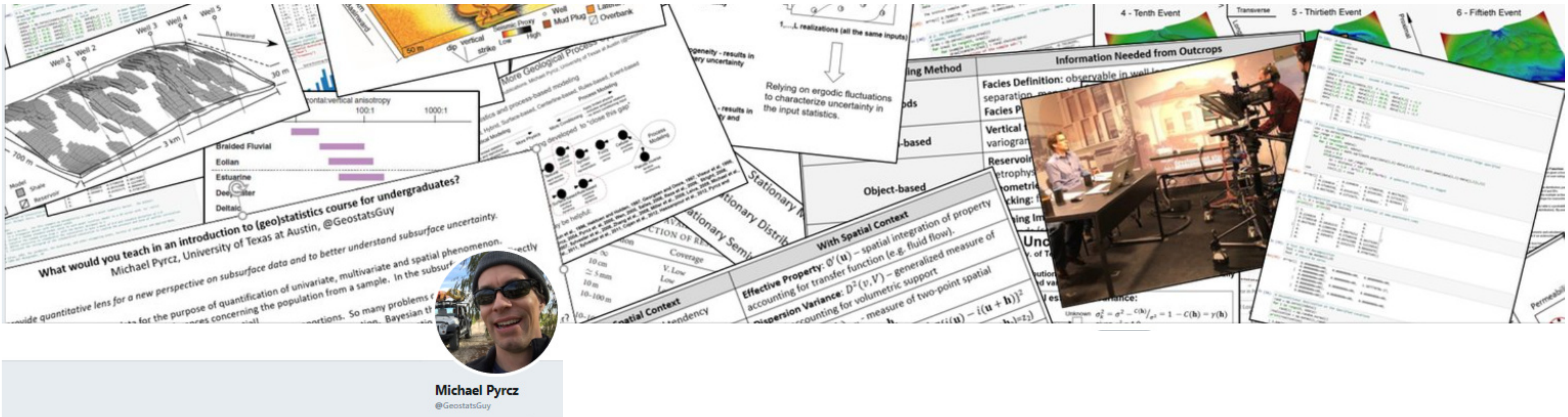
- Maximizing the integration of deterministic engineering, geological description, target oriented drilling, geophysical measurements, borehole formation evaluation, and core data while preserving expert geoscience information to construct high-resolution reservoir models subject to production forecasts.

- Development of spatial data analytics and machine learning methods and workflows to support geomodeling and reservoir forecasting.

- Automated, expert systems, improved visualization and decision support

- Dr. Carlos Torres-Verdin – geophysics, data analytics and machine learning
- Dr. Eric van Oort – geomechanics, data analytics and machine learning
- Dr. John Foster – numerical simulation, flow and data analytics
- Dr. Michael Pyrcz – geostatistics, reservoir modeling, data analytics and machine learning

Learn More - Twitter



For tweets with Subsurface Geostatistical, Data Analytics and Machine Learning resources -

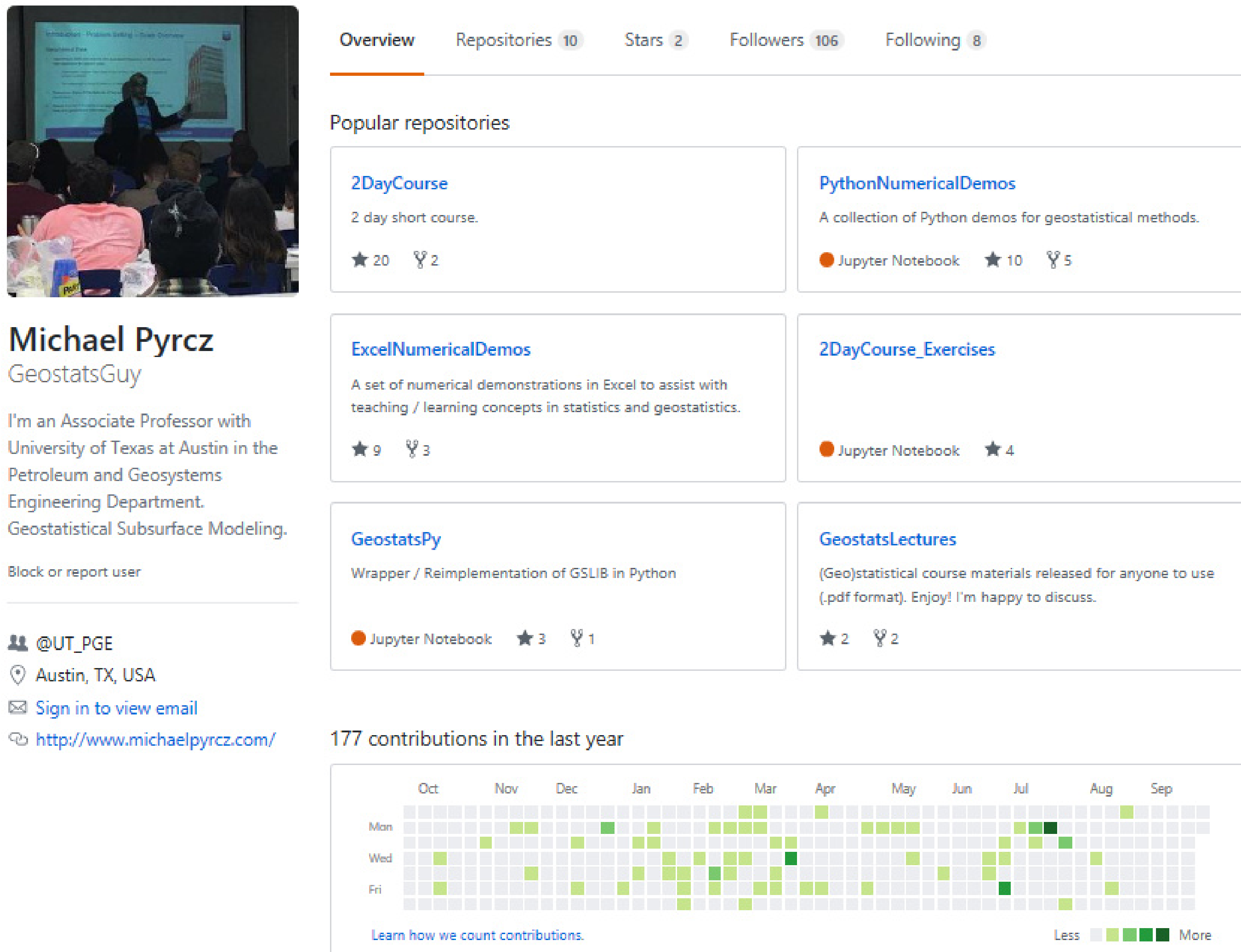
follow **@GeostatsGuy**

Learn More - GitHub

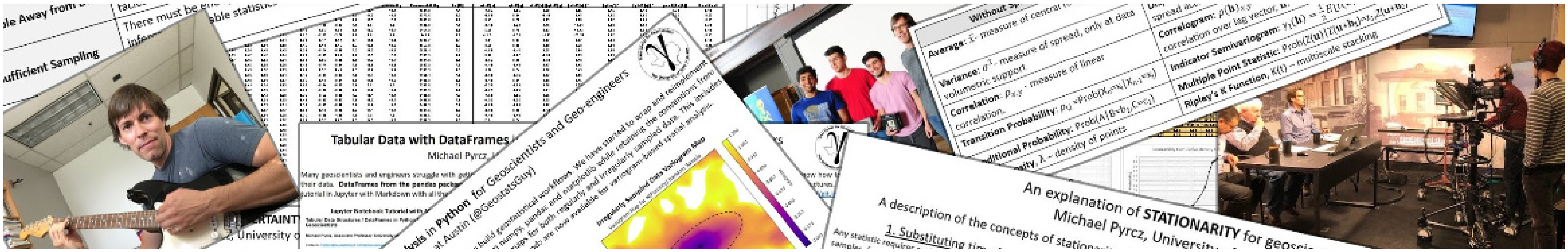
GitHub GeostatsGuy

Excel, R and Python

- Distributions
- Bootstrap
- Cellular Automata
- Hypothesis Testing
- Lorenz Coefficient
- Decision Making
- Bayesian Updating
- Kriging
- Simulation
- Volume-variance



Learn More - YouTube



For my lectures check out my YouTube Channel, 'GeostatsGuy Lectures'.

Example Topics:

- probability theory
- frequentist vs. Bayesian statistics
- binomial distribution to model exploration success