Welcome!

The program will begin soon.

You will not hear audio until we begin.

Webinars

SPE International®

Webinars

SPE Webinars are Sponsored by:

SPE

SPE Foundation

SPE International ®

Research in Data Analytics and Machine Learning for Subsurface Engineering and Geoscience

Michael J. Pyrcz, Ph.D., P.Eng.
Associate Professor

The Hildebrand Department of Petroleum and Geosystems Engineering
Department of Geological Sciences, Jackson School of Geosciences
The University of Texas at Austin

Principal Investigator, Energy Analytics, Core Faculty for the Machine Learning Laboratory, College of Natural Sciences

Associate Director, Center for Subsurface Energy and the Environment, Cockrell School of Engineering, The University of Texas at Austin

Chief Science Officer, daytum
teaching data science skills to energy professionals

Acknowledgements for Examples

Appreciation to the great geoscience and engineering graduate students, working hard to developing important digital capabilities, and novel methods and workflows.

Honggeun Jo, Wen Pan, Javier Santos, Julian Salazar, Eduardo Maldonado, Midé Mabadeje, Wendi Liu, Mahmood Shakiba, Jack Xiao

we are happy to discuss student internships

Appreciation to the excellent faculty from the Department of Petroleum and Geosystems Engineering collaborating and co-supervising these students.

Profs. Carlos Torres-Verdin, Larry Lake, John Foster, Maša Prodanović

we are happy to discus collaboration and partnerships.

Data Analytics and Machine Learning Knowledge Sharing

My Motivation

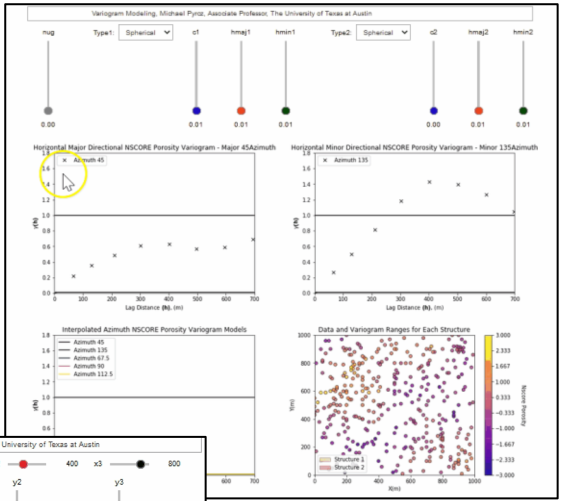
- Energy needs new capabilities and solutions to fully realize value in the digital revolution.
- This requires new skills for engineers and new methods and workflows.
- Engineering and geoscience domain knowledge remains critical.

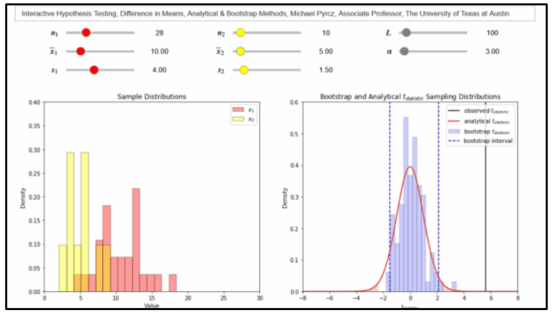
About Michael’s Content (aka GeostatsGuy)

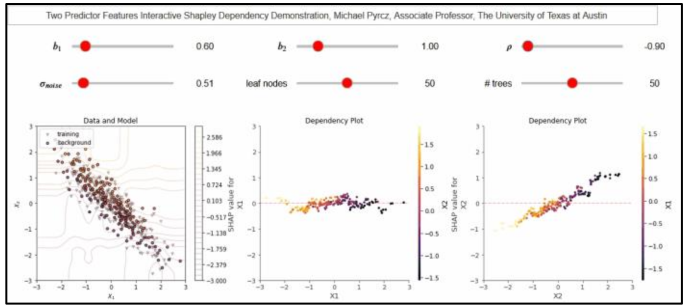
YouTube (>140 lectures) and GitHub (>120 workflows) - accessible, well-documented, interactive educational content to support students and working professionals!

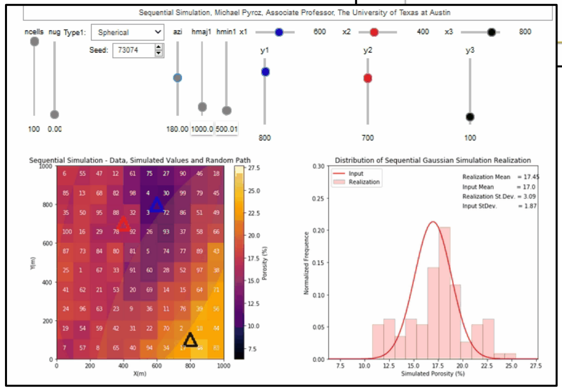
https://www.youtube.com/GeostatsGuyLectures

https://github.com/GeostatsGuy









Topics

- Subsurface Data Analytics and Machine Learning
- Data Analytics and Machine Learning Prerequisites
- Emerging Opportunities in Subsurface Data Analytics and Machine Learning

Topics

- **Subsurface Data Analytics and Machine Learning**
- Data Analytics and Machine Learning Prerequisites
- Emerging Opportunities in Subsurface Data Analytics and Machine Learning

Welcome to the 4th Paradigm of Scientific Discover!

1st Paradigm
Empirical Science

- Experiments

2nd Paradigm
Theoretical Science

- Laws of classical mechanics, electrodynamics

3rd Paradigm
Computational Science
Simulation

- Continuum mechanics for heterogeneous systems
- Computational fluid dynamics

4th Paradigm
Data-driven Science

- Detection of patterns and anomalies in big data
- Artificial intelligence

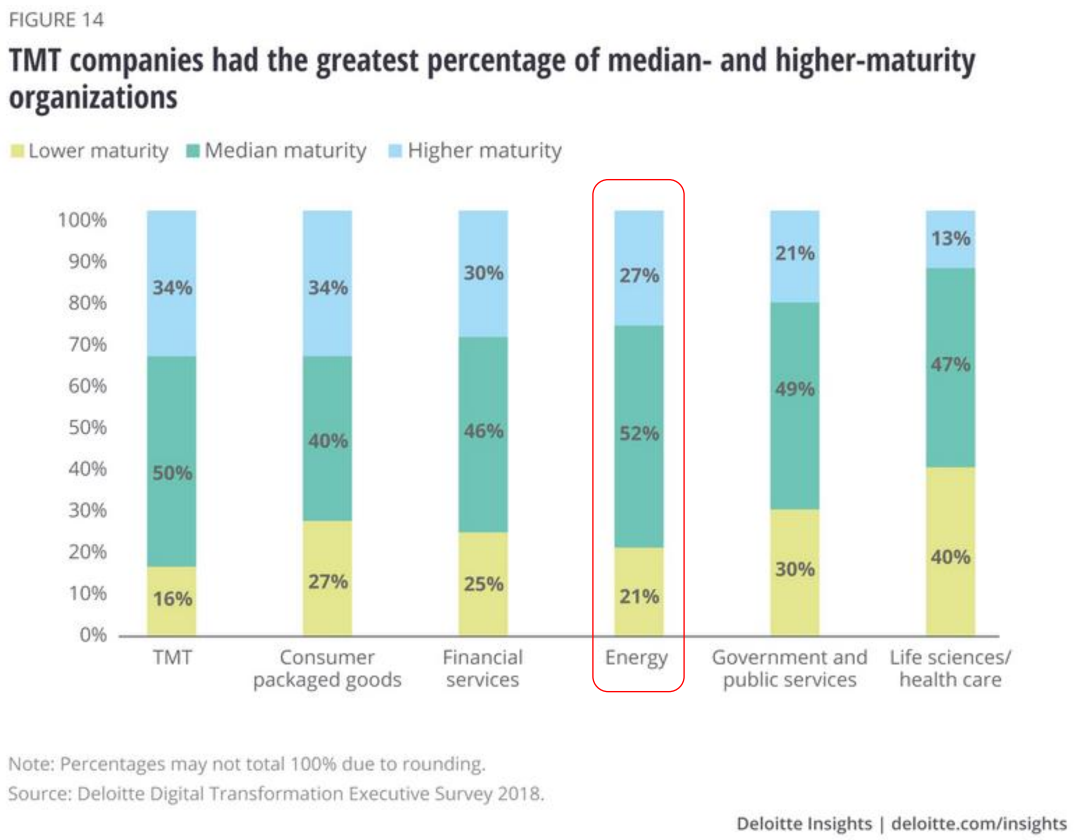
<1600 BCE → **1600s** → **1950s** → **2010s** →

Modified from https://www.nomad-coe.eu/news/147/39/NOMAD-establishes-new-fourth-paradigm-in-computational-materials-science

Energy Digitization

We Are Not Alone

- Digital transformations are underway in all sectors of our economy
- Every energy company that I visit is working on this right now



Digital transformation study by Deloitte, 2019.

Subsurface Digitization

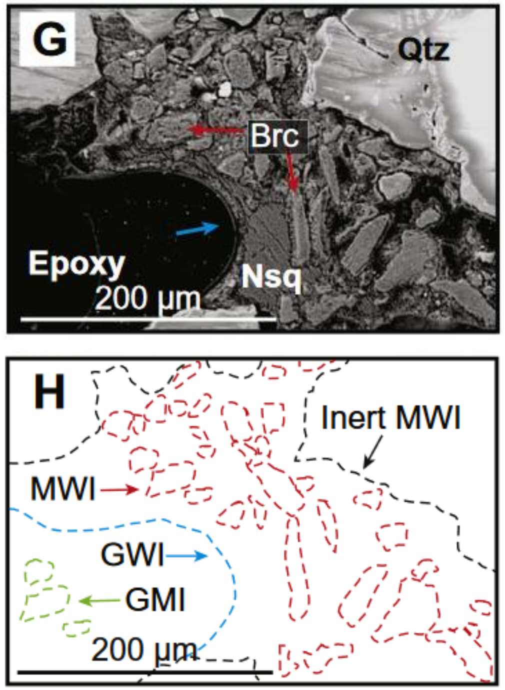
My Biases:

- Opportunities to do more with our data
- Opportunities to teach data analytics and statistical / machine learning methods to engineers and geoscientists for improved capability
- Geoscience and engineering knowledge & expertise remains core to our business

Digital transformation PricewaterhouseCoopers (PwC) panel April 9th, 2019.

Working in the 4th Paradigm!

1st Paradigm
Empirical Science



Microfluidics experiment brucite carbonation experiment (Harrison et al., 2017).

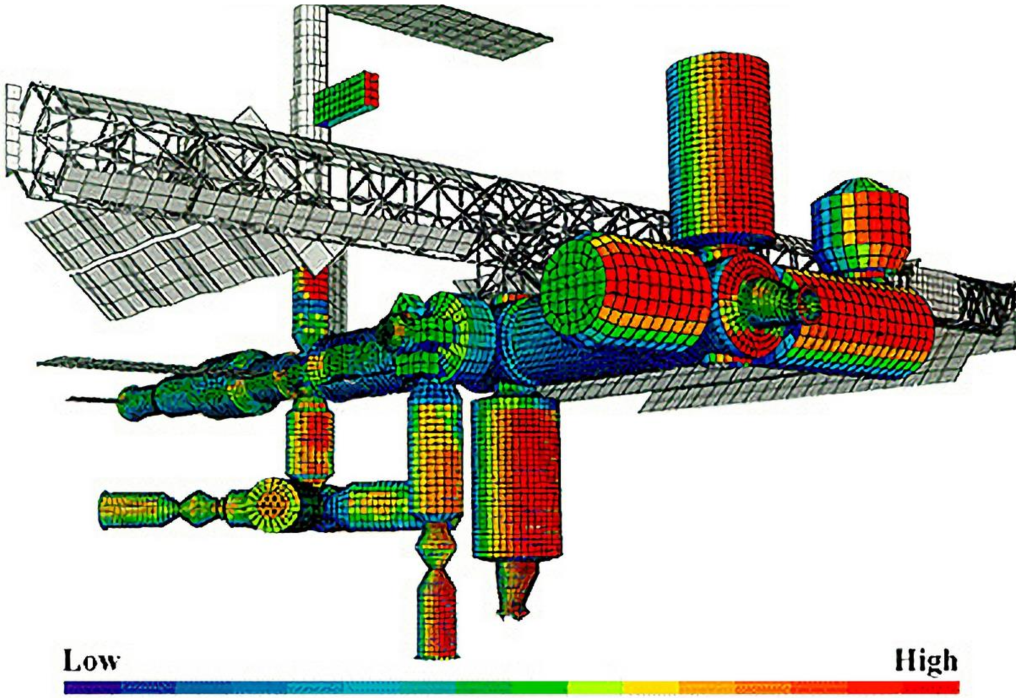
2nd Paradigm
Theoretical Science

$$\boldsymbol{q = -\frac{k}{\mu}\nabla p}$$

q – flux
k – permeability
μ – dynamic viscosity
∇p – pressure gradient

Darcy's law.

3rd Paradigm
Computational Science
Simulation



International space station impact risk from computer simulation. Image from https://en.wikipedia.org/wiki/Risk_management.

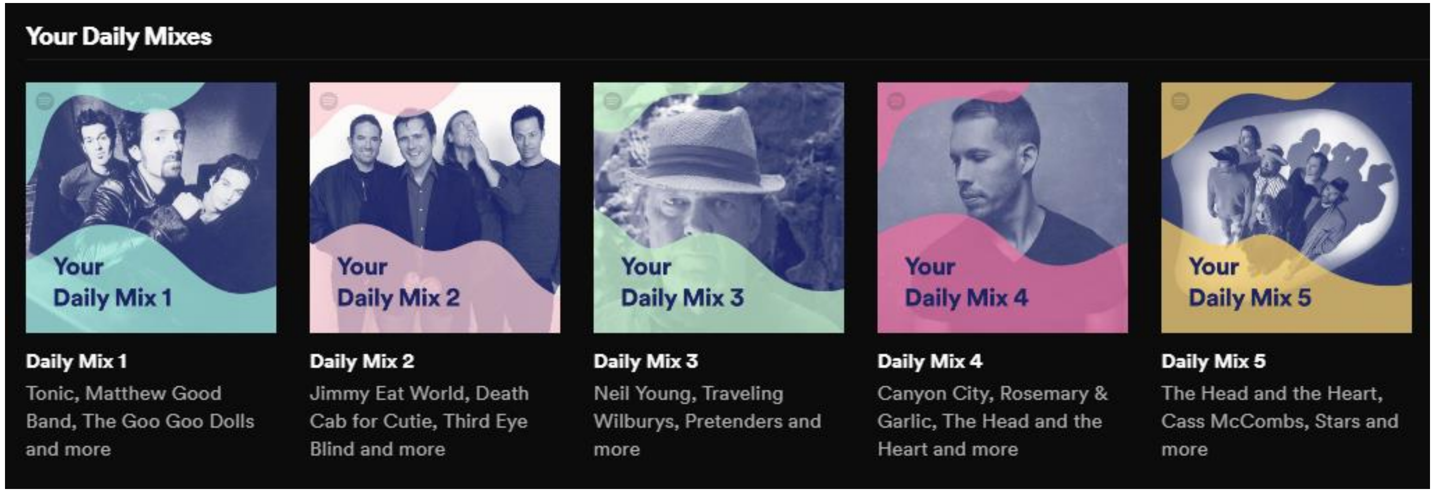
- We augment with new scientific paradigms
- We don't replace older paradigms!

Michael Pyrcz, The University of Texas at Austin

Subsurface is Unique

Subsurface is Different and Needs New Solutions:

- Sparse, uncertain data, complicated and heterogeneous, open earth systems
- High degree of necessary geoscience and engineering interpretation and physics
- Expensive, high value decisions that must be supported



Spotify recommender system from my account summer, 2019.

- We develop novel subsurface data analytics and machine learning solutions that integrate of geoscience and engineering.

What Did We Learn?

Subsurface, energy is unique.

We need unique solutions.

Topics

- Subsurface Data Analytics and Machine Learning
- **Data Analytics and Machine Learning Prerequisites**
- Emerging Opportunities in Subsurface Data Analytics and Machine Learning

Do You Work with Big Data? Yes!

Big Data, you have big data if your data has a combination of these:

Volume: many data samples, difficult to handle and visualize

Velocity: high-rate collection, continuous relative to decision making cycles

Variety: data form various sources, with various types and scales

Variability: data acquisition changes during the project

Veracity: data has various levels of accuracy

"Subsurface has been big data long before tech learned about big data."

– Michael Pyrcz

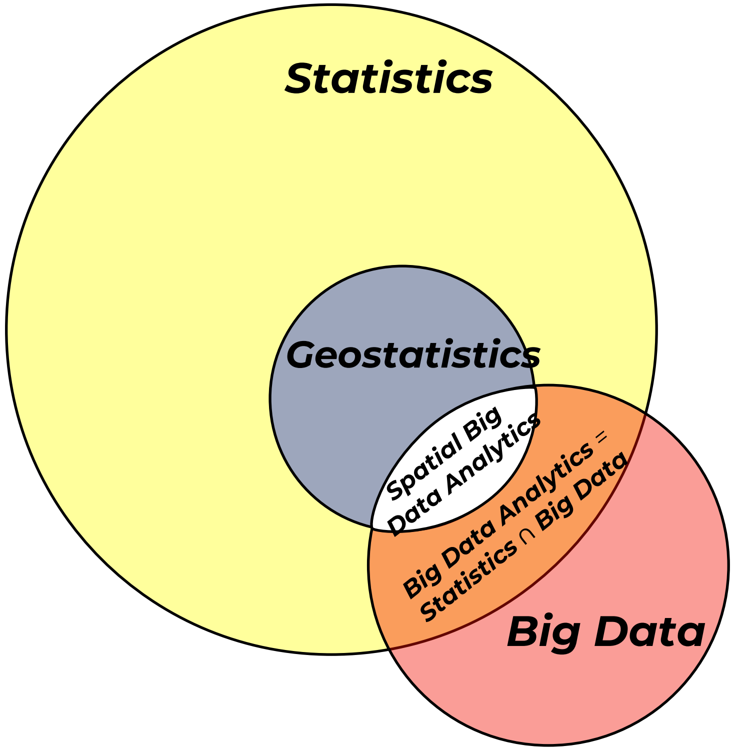
Data Analytics is Statistics

Statistics is collecting, organizing, and interpreting data, as well as drawing conclusions and making decisions.

Data Analytics is the analysis of data to support decision making.

'Data analytics is statistics!'

'Spatial data analytics is geostatistics/spatial statistics!'



Proposed Venn diagram for spatial big data analytics.

Machine Learning

toolkit

"… is the study of algorithms and mathematical models

learning

that computer systems use to

progressively improve their performance on a specific task.

Machine learning algorithms build a mathematical model

training with data

of sample data, known as "training data",

general

in order to make predictions or decisions

without being explicitly programmed to perform the task."

not a panacea

"… where it is

infeasible to develop an algorithm of specific instructions for performing the task."

Machine Learning - Wikipedia

Machine Learning Nuts and Bolts

The Model

Response **Predictors** **Random Error**

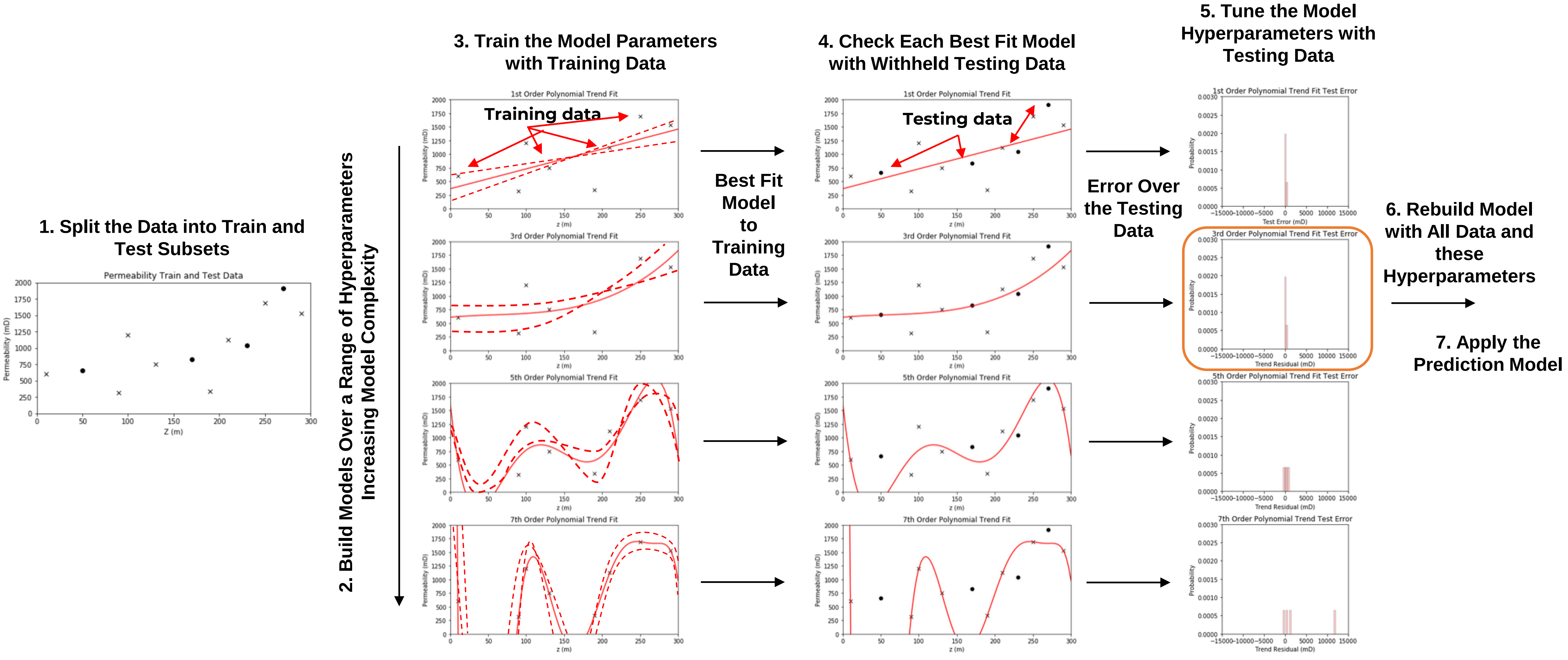
$$Y = f(X_1, \dots, X_m) + \epsilon$$

- Predictors (or Independent) Features (or Variables)
    - the model inputs

- Response (or Dependent) Features (or Variables)
    - the model outputs

- Machine Learning is Statistical Learning (Hastie et al., 2009; James et al., 2013)

Building Predictive Machine Learning Models



The train and test workflow for building predictive machine learning models.

Michael Pyrcz, The University of Texas at Austin

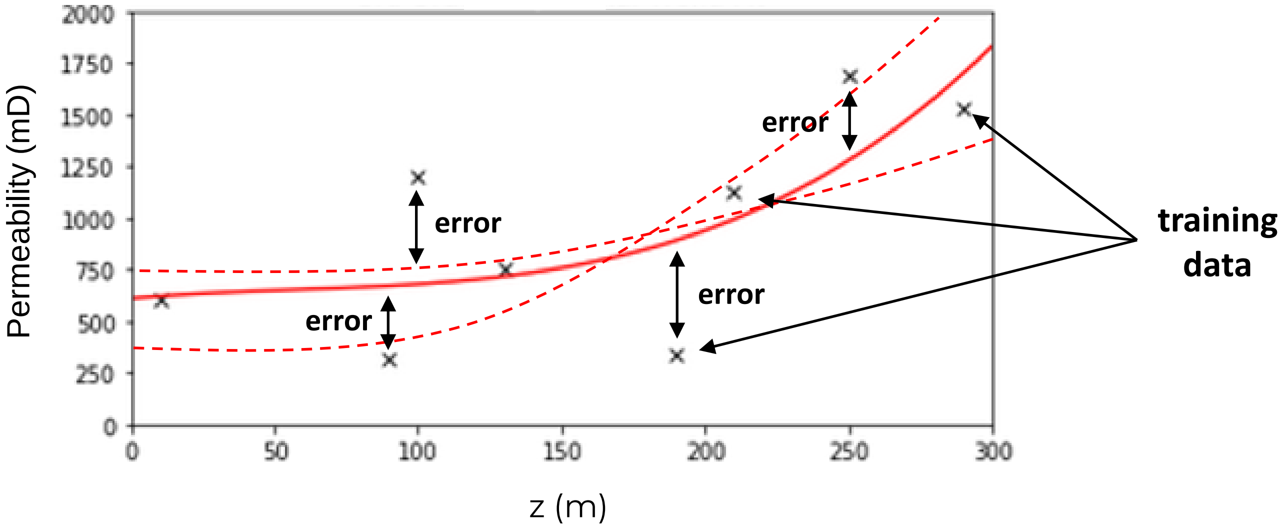
Machine Learning Model Parameters

Model Parameters

- Fit during training phase to minimize error at the training data
- For this 3rd order polynomial:

$$k = b_3 z^3 + b_2 z^2 + b_1 z + c$$

Parameters:
b_3, b_2, b_1 **and** c



Setting model parameters to minimize the error relative to training data.

Machine Learning Model Hyperparameters

Model Hyperparameters

- Constrain the model complexity to avoid overfit.
- Select hyperparameters that maximize accuracy with the testing data.

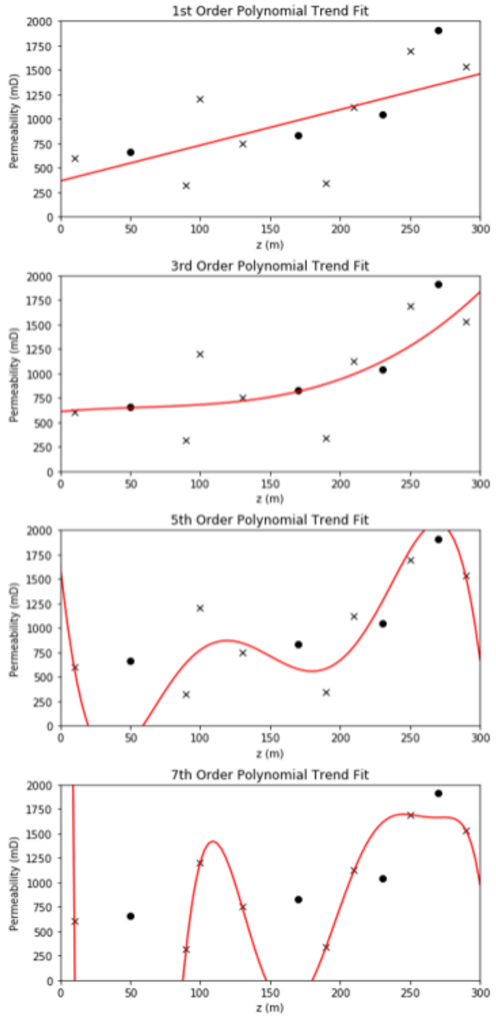
- For a polynomial model:

Increasing Complexity

3rd Order: $k = b_3 z^3 + b_2 z^2 + b_1 z + c$

2nd Order: $k = b_2 z^2 + b_1 z + c$

1st Order: $k = b_1 z + c$

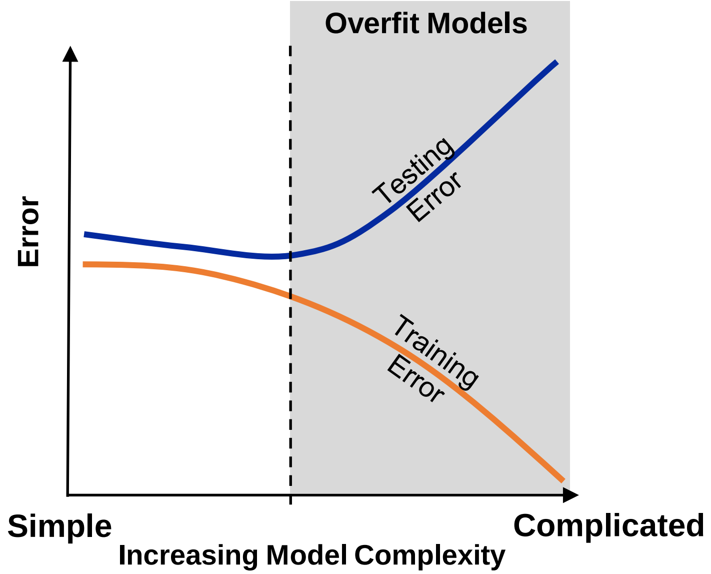


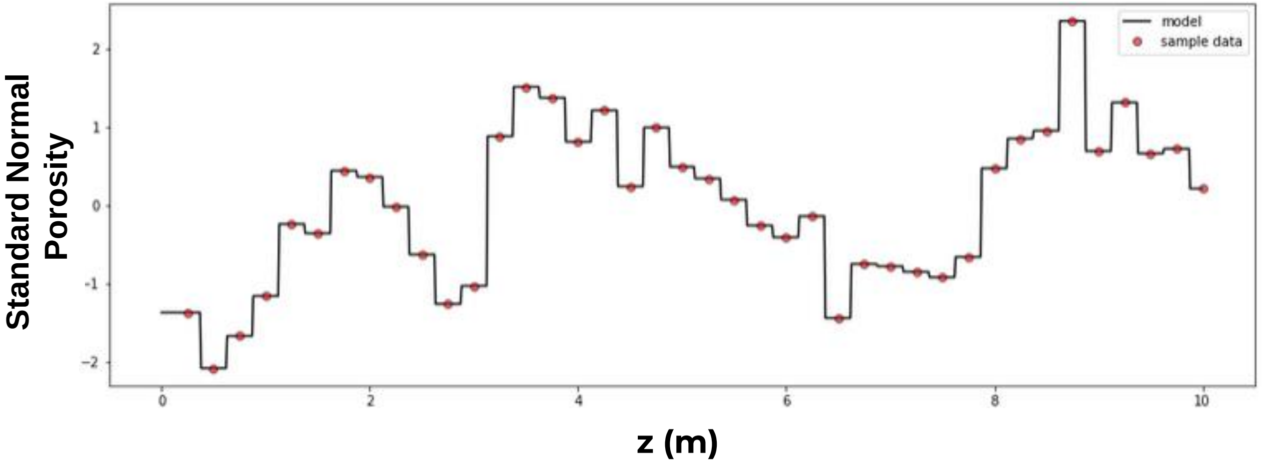
Varying the model complexity, model hyperparameter, to maximize fit with testing data.

Machine Learning Model Overfit

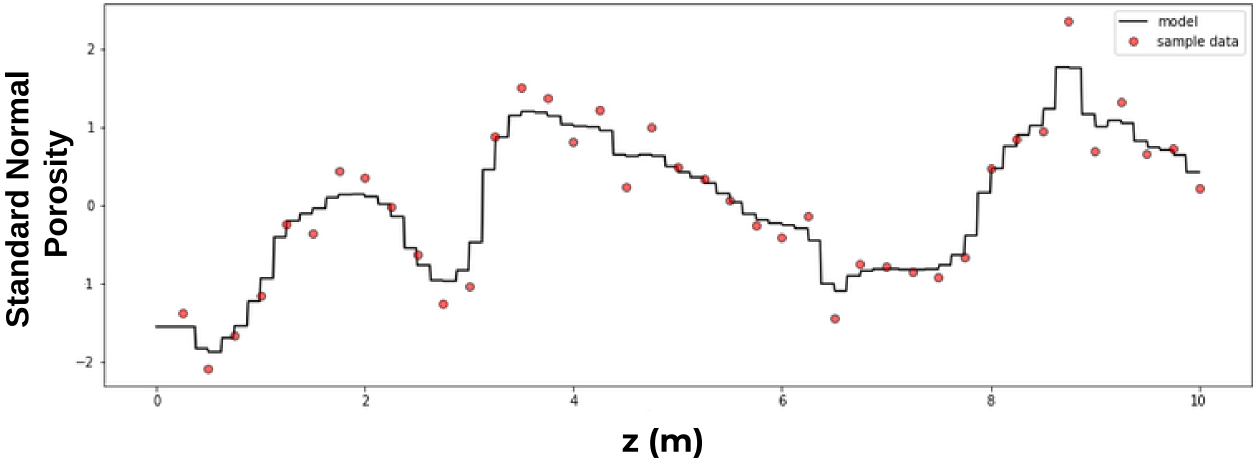
Model Overfit

- Fitting data noise / data idiosyncrasies
- Increased complexity will generally decrease error with respect to the training dataset
- but, may result in increase error with testing data





Overfit model to training data.

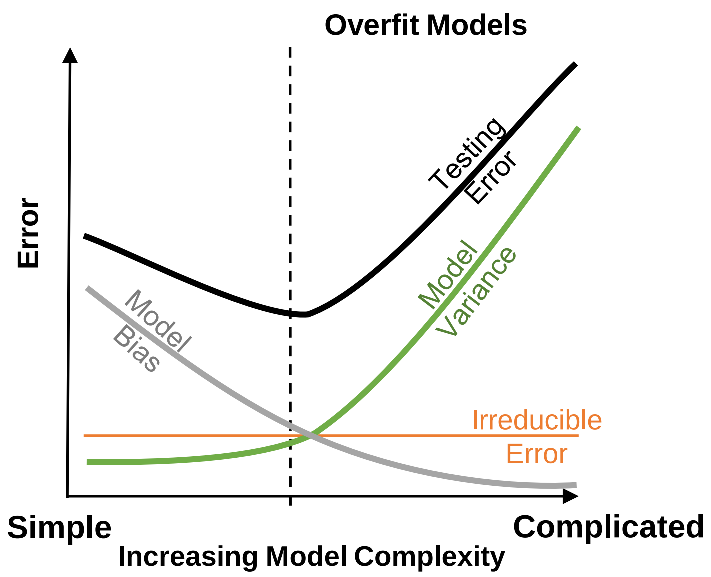


A more balanced fit model to training data.

The Variance-Bias Trade-Off

The Components of Testing Error, Real-World Use

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



- **Model Variance** is the error in the model predictions due to sensitivity to the data (what if we used different training data?)

- **Model Bias** is error in the model predictions due to using an approximate model / model is too simple

- **Irreducible error** is error in the model predictions due to missing features and limited samples can't be fixed with modeling / entire feature space is not sampled

What Did We Learn?

We work with big data.

Data analytics is statistics.

Machine learning parameter training and hyperparameter tuning to build models that can predict away from the data.

Testing error is the addition of model bias, model variance and irreducible error.

Topics

- About Michael
- Subsurface Data Analytics and Machine Learning
- **Emerging Opportunities in Subsurface Data Analytics and Machine Learning**

New Powerful Tools, New Ways for Engineers and Geoscientists to Add Value

Machine Learning to Assist with Data Interpretation.

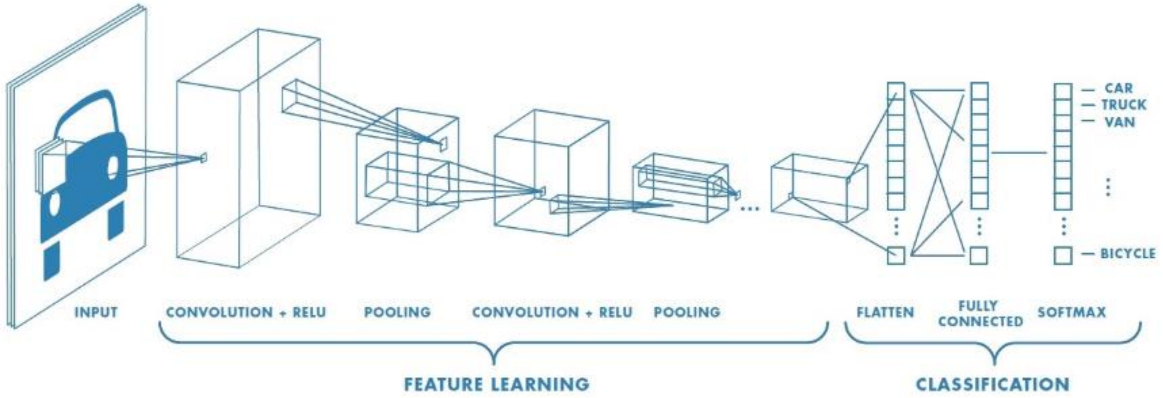
Well Log Pattern Extraction

Wen Pan, Carlos Torres-Verdin and Michael Pyrcz

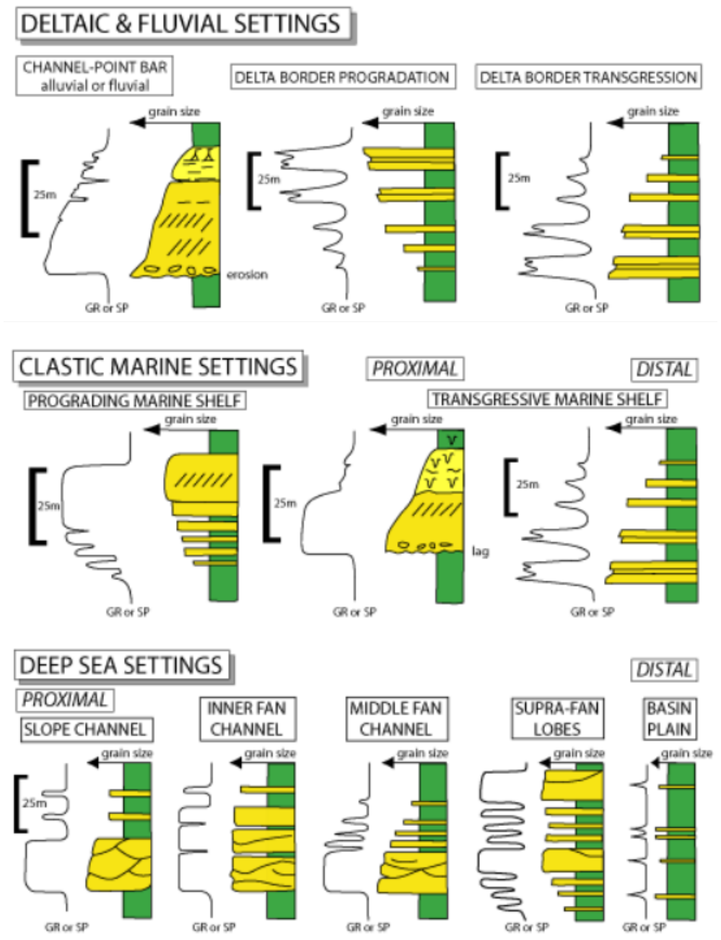
Convolutional Neural Networks for Automatic Well Log Pattern Extraction

Augmented with New Constraints:

1. Depositional environment (pattern) and spatial continuity.
2. Correlate logs and petrophysical properties in pattern space.



Convolutional neural network illustration by Prabhu, 2019.



Gamma ray log response and depositional setting, Kendall 2003 modified from Rider, 1999.

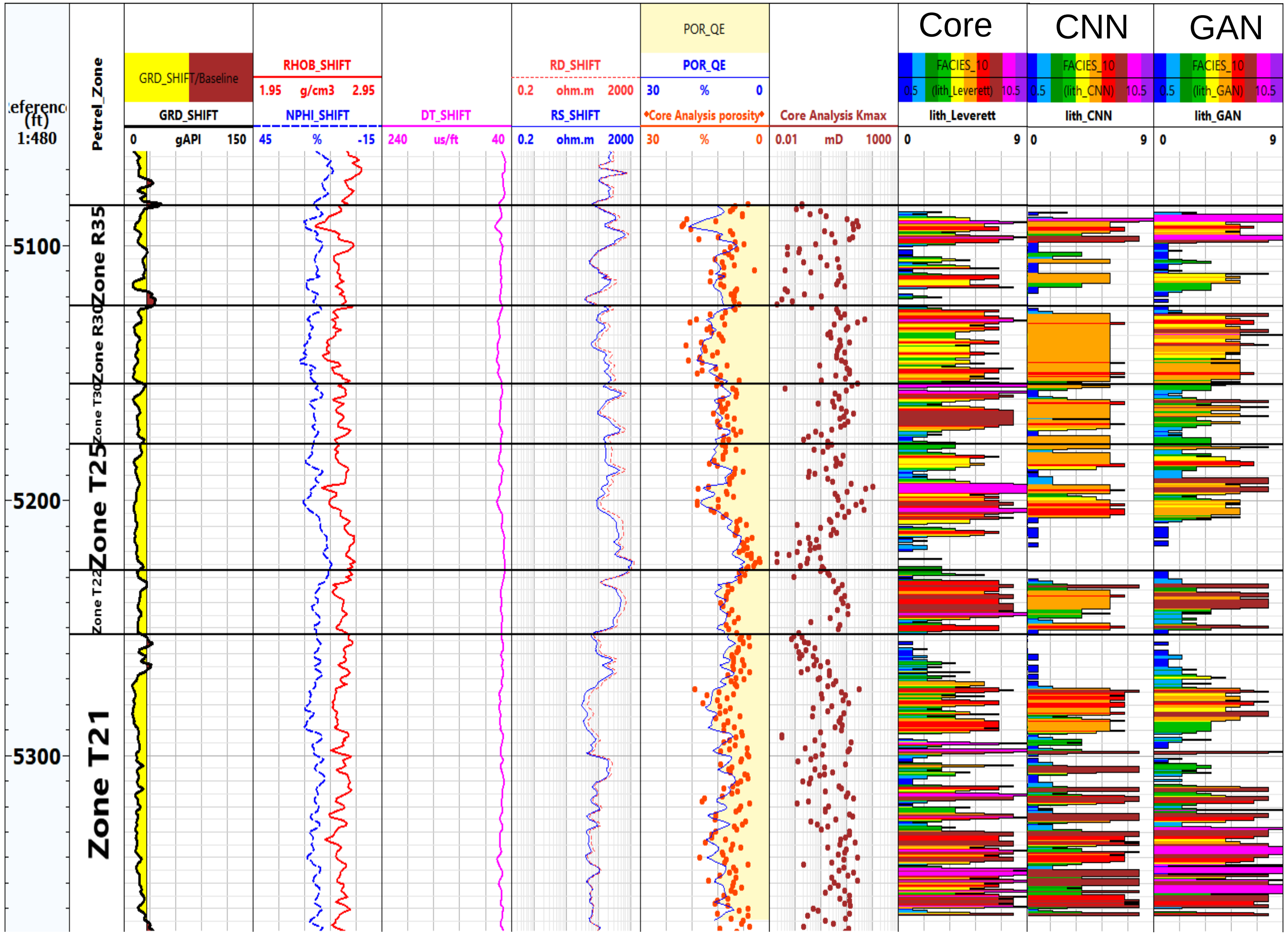
Well Log Pattern Extraction

Wen Pan, Carlos Torres-Verdin and Michael Pyrcz

Convolutional Neural Networks for Automatic Well Log Pattern Extraction

Feature Engineering

1. Deconvolution to remove measurement smoothing
2. By-zone workflow within stratigraphic framework
3. Predictions with CNN and GAN methods.



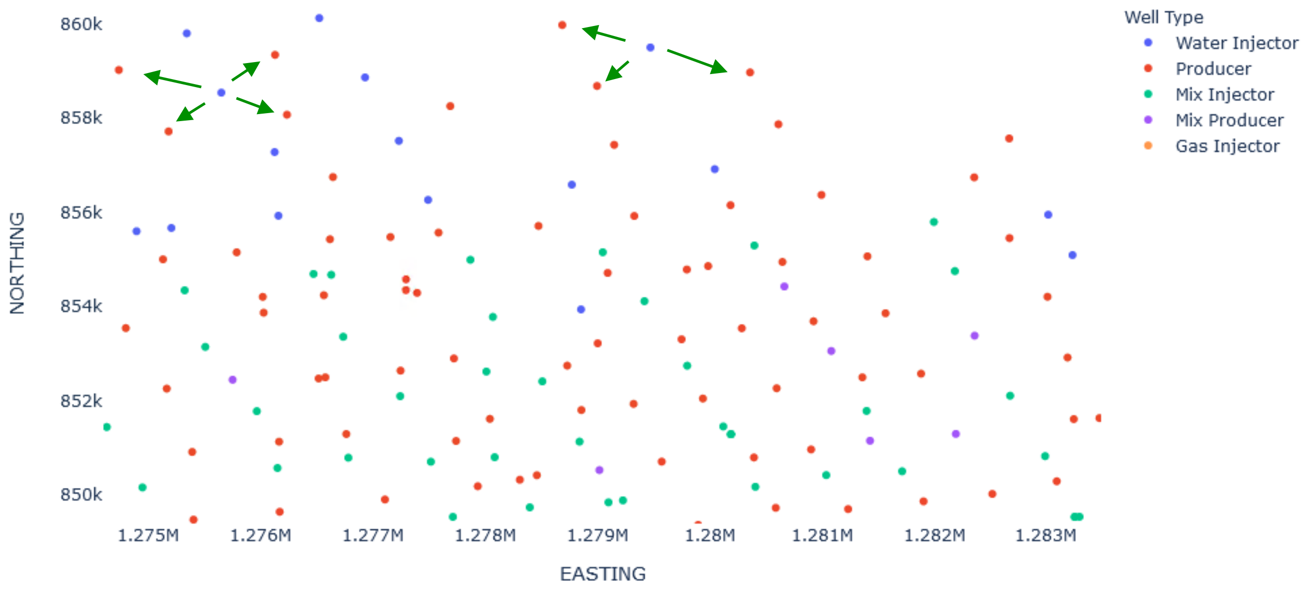
Machine learning predictions with gamma ray log response and depositional setting, Kendall 2003 modified from Rider, 1999.

Dynamic Time Warping for Well Connectivity

Jose Hernandez and Michael Pyrcz

Gap and Opportunity

- Production and injection rates are the most abundant data available during water flooding.
- Determine the communication between injector and producing wells.
- Optimize waterflooding operations and economics by maximizing oil recovery.



Interpolate the well-to-well connectivity framework.

Physics-Informed Dynamic Time Warping

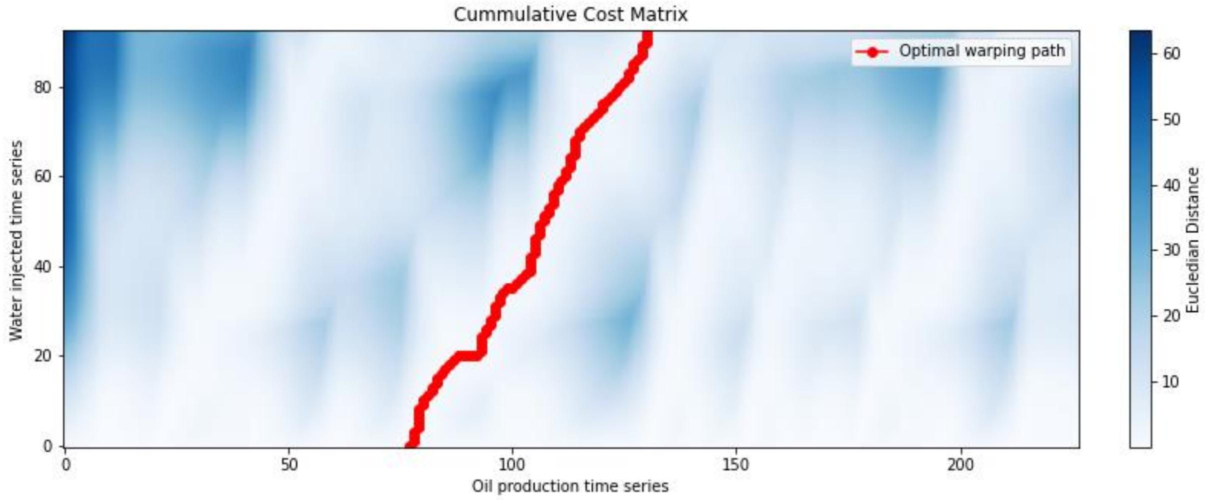
The objective is to temporally align water injection with oil production time series under physical fluid flow constrains.

Dynamic Time Warping for Well Connectivity

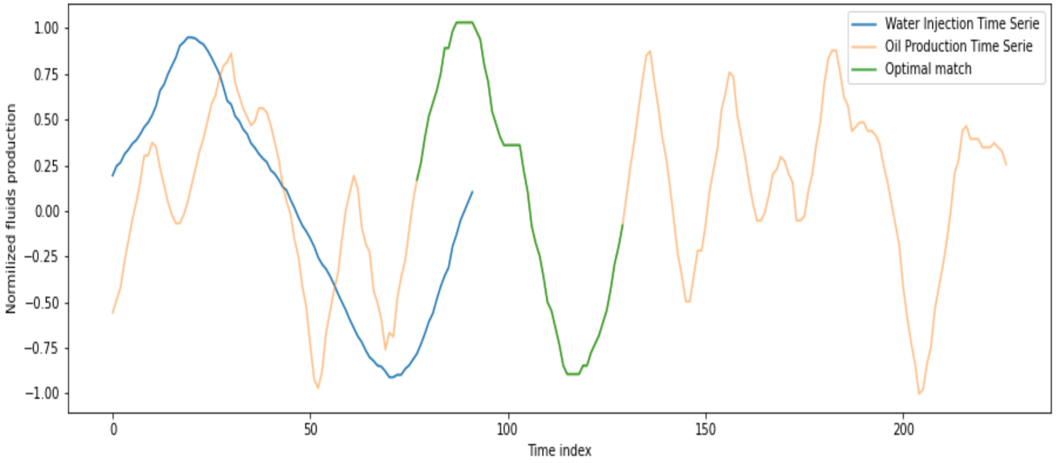
Jose Hernandez and Michael Pyrcz

Results

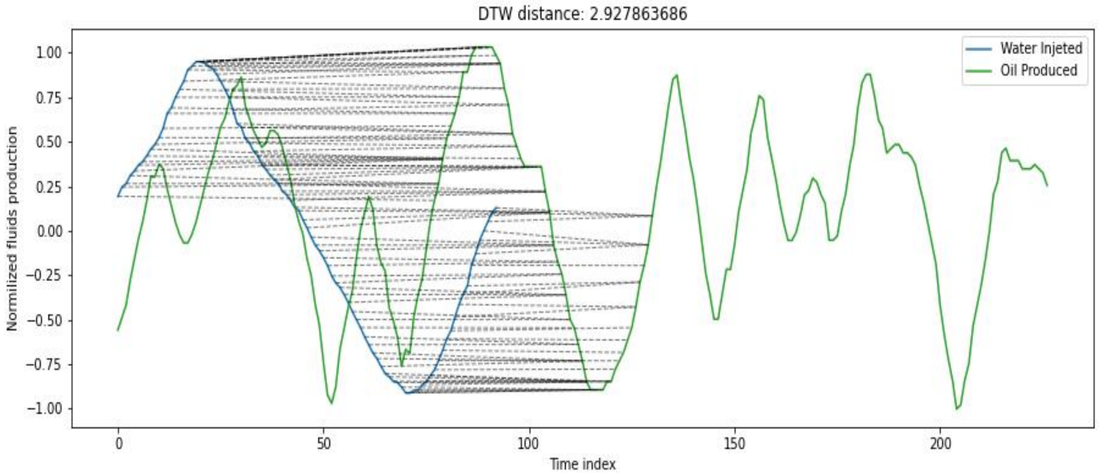
- Novel method to estimate well connectivity between injectors and producing wells.
- The algorithm output will determine the optimal match between water injected and oil production time series.
- The algorithm can identify the impact of different injecting wells around a producing well.
- Hernandez and Pyrcz, (in prep)



Physically constrained optimum path through the cumulative cost matrix.



Infector and producer time series information.



Infector and producer connectivity model.

Garbage In, Garbage Out.

Bias In, Bias Out?

Spatial Sampling Bias in Machine Learning Predictions

Wendi Liu and Michael Pyrcz

Sampling bias is ubiquitous and will propagate to ML prediction.

Method:

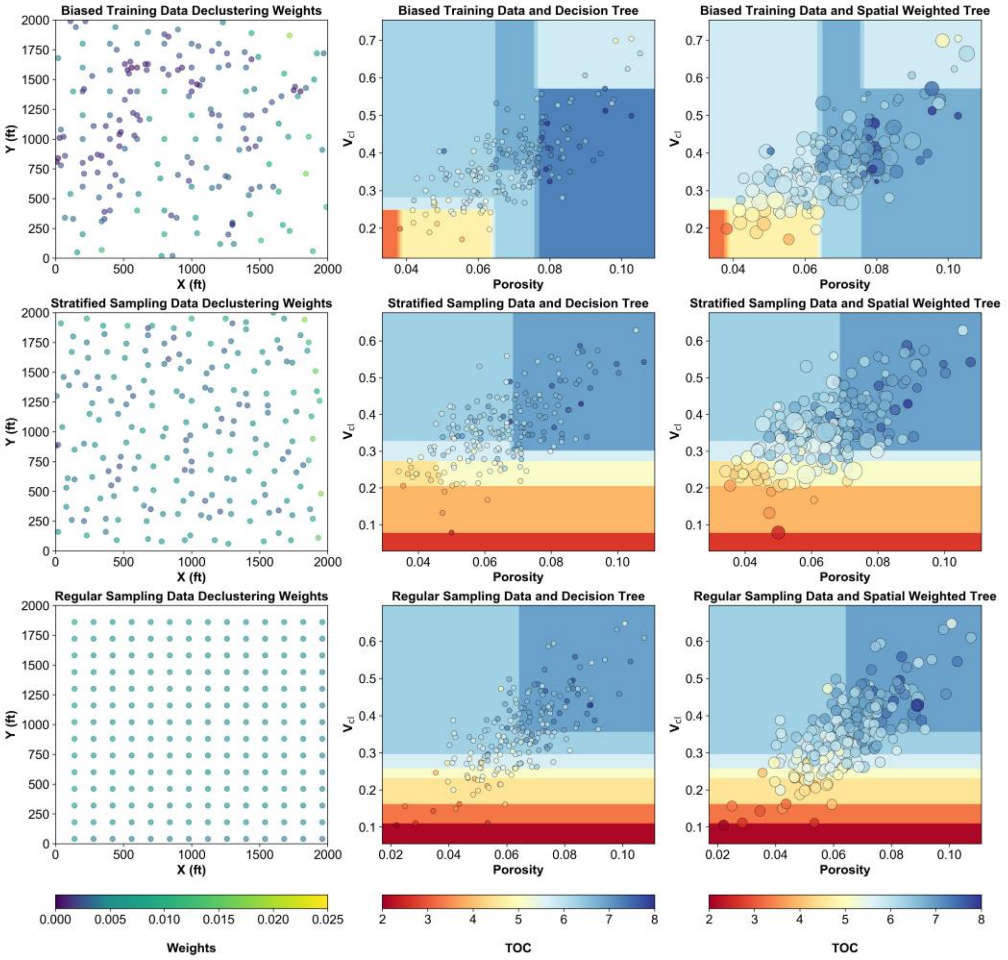
- **Polygonal declustering**
  - Data weighting based on local sample density
  - $w_i = \frac{A_i}{\sum_{j=1}^{n} A_j}$
- **Declustering weights**

  – Assign to the biased training data when maximizing RSS before splitting to remove the influence of bias from model construction

$$RSS = \sum_{j=1}^{J} \sum_{i \in R} (y_i - \widehat{y_{R_j}})^2$$

  – Apply to the estimates of each terminal node

$$\widehat{y_{R_j}} = \sum_{i=1}^{n} w_i \cdot z_i$$



Biased training data and representative training data sampled from stratified sampling and regular sampling with conventional and weighted tree models predictions. (Liu et al. 2020)

Spatial Sampling Bias in Machine Learning Predictions

Wendi Liu and Michael Pyrcz

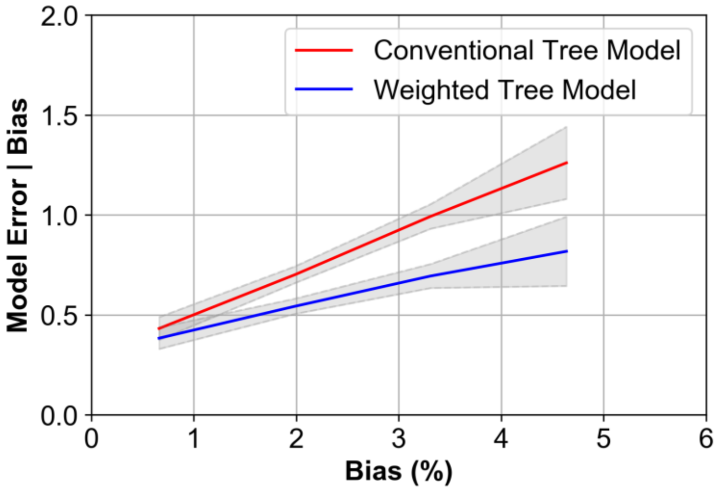
- Iterate over 625 realizations with various degree of bias.
- Quantify the expected model error reduction conditional to degree of bias over multiple sample sets

$$bias = \frac{|\mathbb{E}(Z) - \mathbb{E}(Z^*)|}{\mathbb{E}(Z)} \times 100\%$$

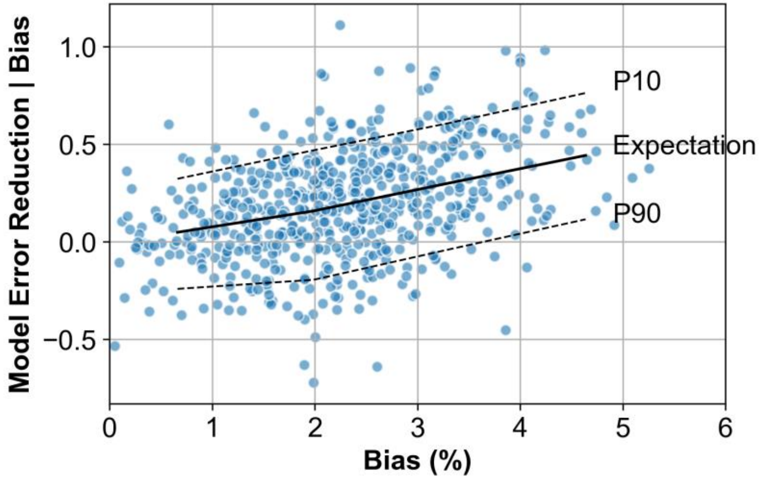
$$e = \frac{|\mathbb{E}(Z) - \mathbb{E}(\hat{Z})|}{\mathbb{E}(Z)} \times 100\%$$

$$Error\ reduction = e_{naive} - e_{spatial\ weighted}$$

- **Conditional expectation:** $\mathbb{E}\,(X|Y=y) = \frac{\sum_{\omega\in\{Y=y\}} X(\omega)}{|\{Y=y\}|}$
- Model accuracy improvement applies to most cases.
- As degree of bias increases, the model error reduction is more significant.



Conditional expectation of conventional and spatial weighted tree model error vs. bias with 95% confidence interval (Liu et al. 2020)



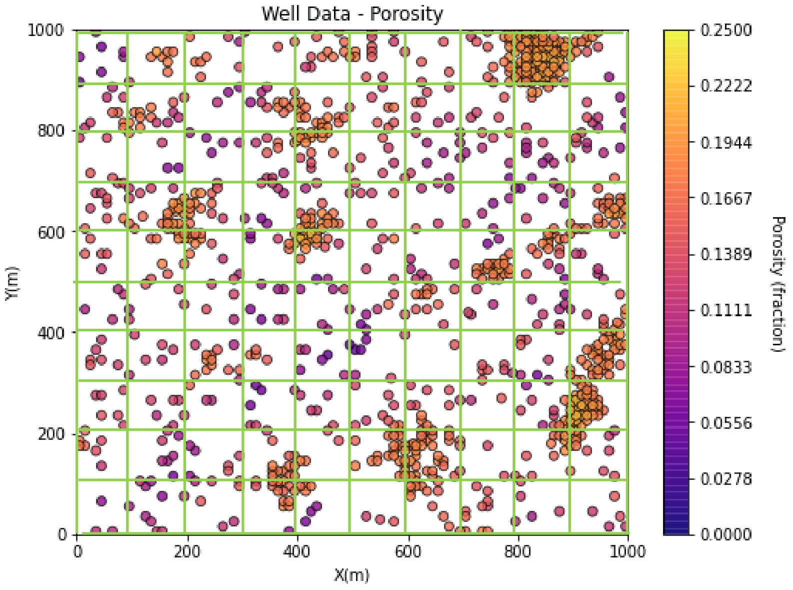
Conditional statistics of model error reduction to bias (Liu et al. 2020)

Spatial Data Analytics to Support Declustering Approaches

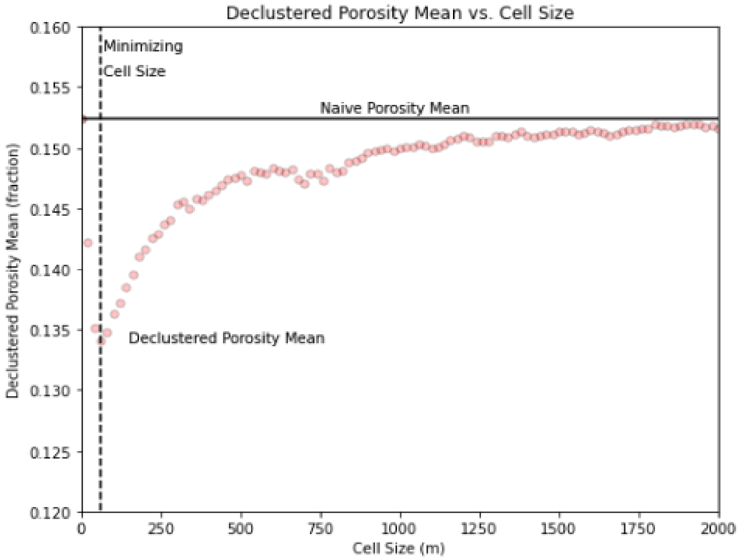
Midé Mabadeje and Michael Pyrcz

Proposed Workflow

- Common method is cell-based declustering, and it is very sensitive to cell size choice.
- The current method for cell size selection is irrational and ad hoc.
- Develop a workflow that rationally and objectively selects the cell size used for robust, spatial data analytics informed cell-based declustering.



An example from the 2000 data realizations.



Determination of declustered mean porosity using the minimizing cell size from current approach.

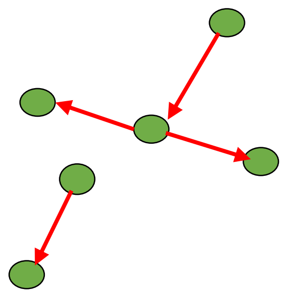
Spatial Data Analytics to Support Declustering Approaches

Midé Mabadeje and Michael Pyrcz

Proposed Workflow

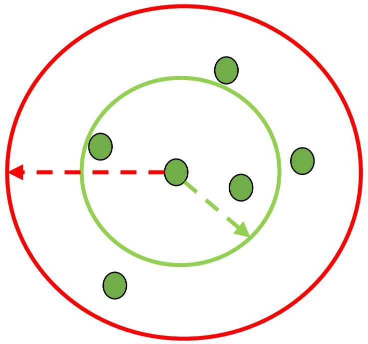
- The cell size selection should be informed by the 'nominal spacing' of the data, but this is difficult to determine from ocular inspection with real data sets.
- Use spatial data analytics, spatial statistics to determine nominal spacing.

Nearest Neighbor

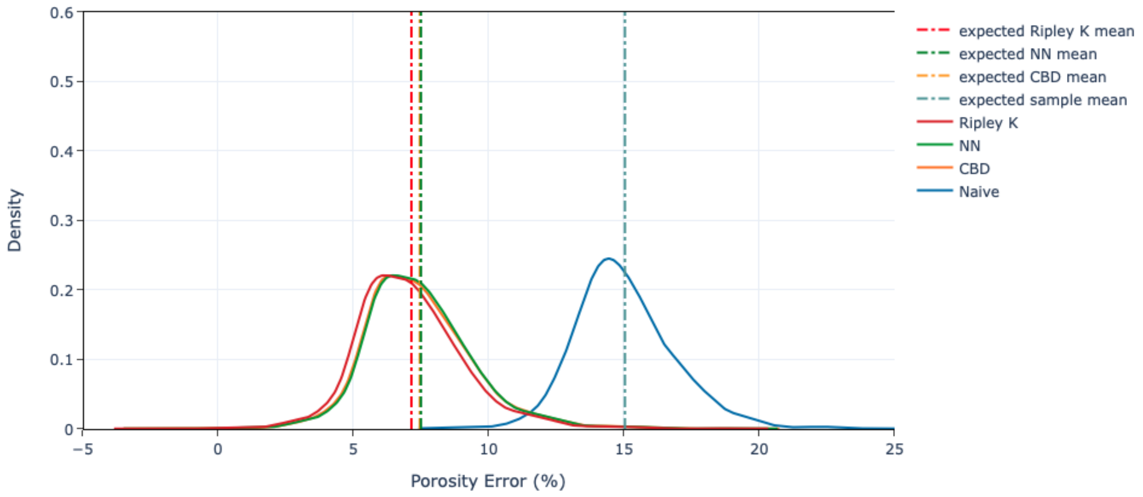


Each data point gets its own neighbor then, a distribution of the average distance to each other provides insight about spacing between the points.

Ripley's K



Count the number by radius and compare it to the expectation from complete stationary randomness (CSR).



Large study of performance for determination of cell size by current method (CBD), nearest neighbor (NN) and Ripley's K. Note, NN and CBD are superimposed.

Results

- Improved interpretability, information integration and performance for more accurate, debiased statistics!

Anomalies Are Impactful, Spatial Anomalies?

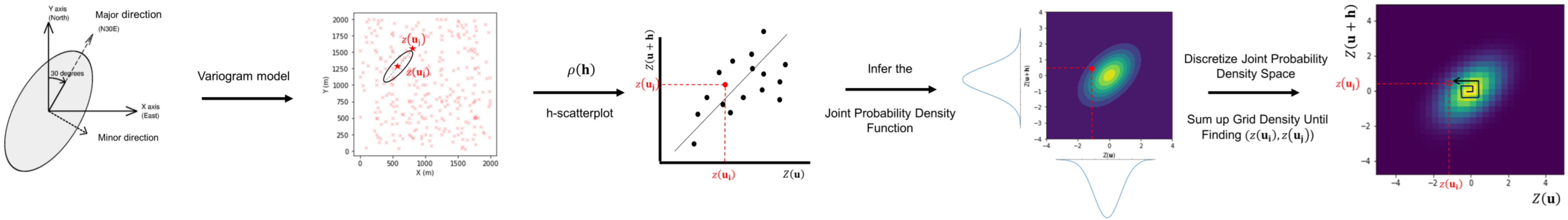
Spatial Correlation Anomaly Detection Method

Wendi Liu and Michael Pyrcz

Opportunity

- A useful automation tool to identify anomalous samples, region, segmentation boundaries or facies transition zones in big spatial data to focus professional attention.
- A spatial correlation anomaly detection method integrating domain expertise knowledge with sparse, spatial data.

Method



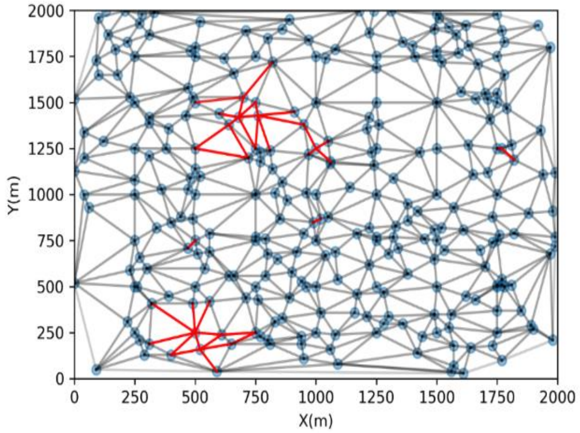
Anomaly detection workflow schematic (Liu and Pyrcz 2020).

Spatial Correlation Anomaly Detection Method

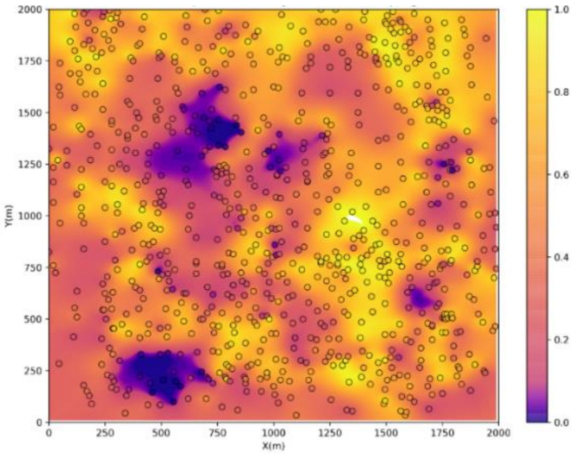
Wendi Liu and Michael Pyrcz

Results

- Tessellation to construct sample pairs to obtain spatial correlation (covariance)
- K-nearest neighbor sampling was compared with tessellation. It turns out tessellation covers area of interest better.
- Useful for segmenting reservoir model to reduce uncertainty and improve flow simulation efficiency
- The proposed spatial correlation anomaly detection method is able to integrate domain expertise knowledge through trend and correlogram models as well as conditioning spatial sparse well data
- Flexibility of accounting for feature magnitude dependence



Tessellation anomaly map of wells samples (Liu and Pyrcz, 2020)



Anomaly probability map (Liu and Pyrcz 2020)

Spatial Feature Engineering to Integrate Scale and Heterogeneity

Heterogeneity Metric for Spatial Feature Engineering

Wendi Liu and Michael Pyrcz

Opportunity and Proposed Method

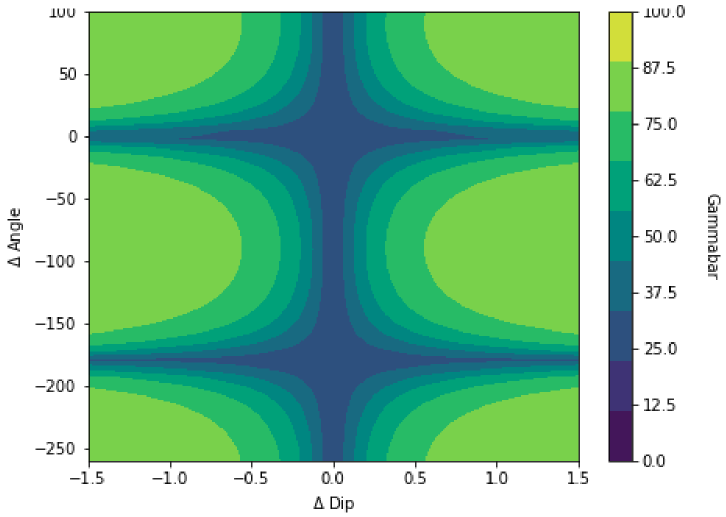
- Reservoir heterogeneity: a vital factor for flow performance and hydrocarbon recovery prediction
- New heterogeneity metric accounting for multiscale, spatial sparse data for spatial feature engineering to assist ML predictions

Method

- A generalized form of variance that accounts for the volume calculated from $\bar{\gamma}$
- $D^2(\cdot, well)$ is the variation within one well. $D^2(well, reservoir)$ is the variation between wells within the area of interest.
- **Krige's relation:**

$$D^2(\cdot, reservoir) = D^2(\cdot, well) + D^2(well, reservoir)$$

Fixed well **length** $D^2(\cdot, well)$



One example of dispersion variance map with azimuth and dip deviating from major direction

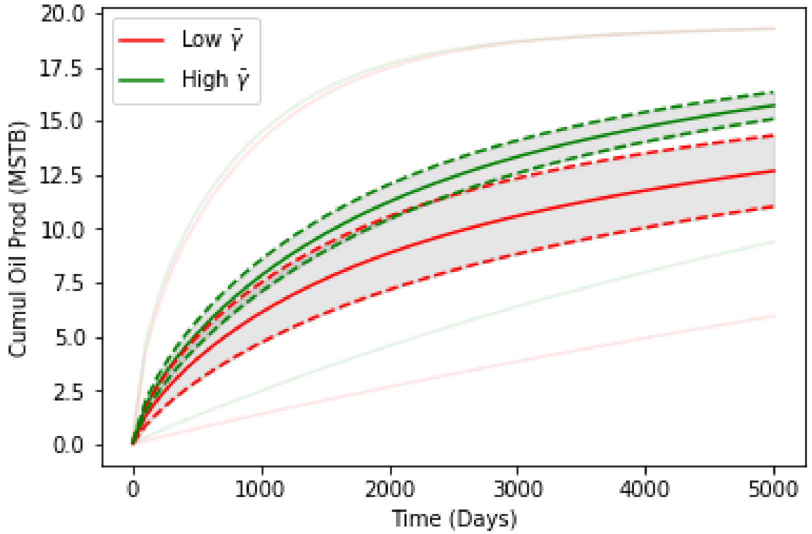
Dispersion Variance quantifies the variability over volume of the drainage radius

Heterogeneity Metric for Spatial Feature Engineering

Wendi Liu and Michael Pyrcz

Results

- Fixed flow simulation parameters and well length, pressure depletion; multiple reservoir model realizations with different Dykstra-Parsons coefficient
- Changing well trajectory → different $\bar{\gamma}$ value → indicating impact of heterogeneity on production
- $\bar{\gamma}$ can be used as a spatial feature accounting for heterogeneity in ML prediction feature engineering



Cumulative oil production expectation curve (solid line), 95% confidence interval (dash line) and max/min (faded) grouped by high and low $\bar{\gamma}$ threshold with DP=0.779.

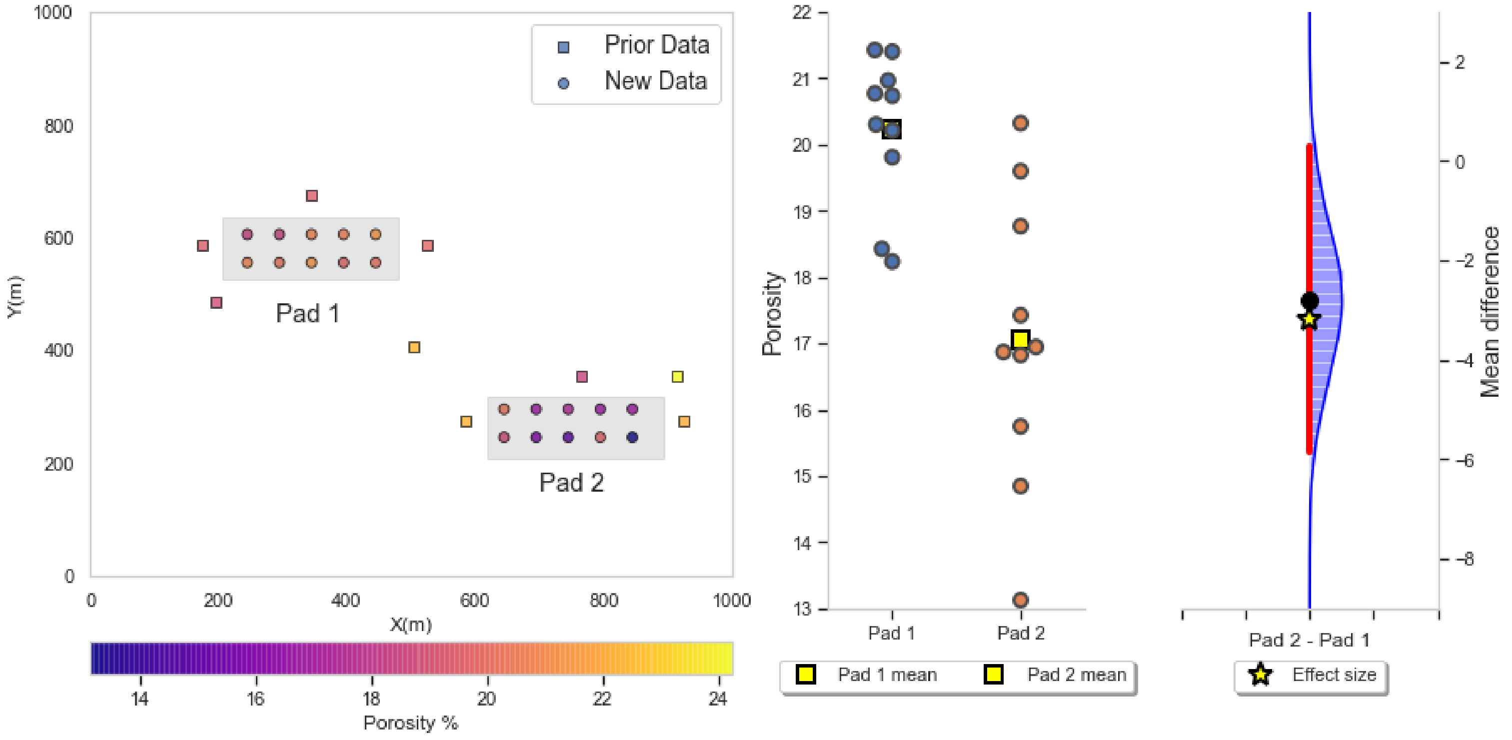
Is It Really Different? Geostatistical Significance

Geostatistical Significance

Jose J. Salazar, Larry Lake and Michael Pyrcz

Inferential statistics applied to spatial problems

- Spatial data violate the assumption of identically and independently distributed (i.i.d.) data.
- Geostatistical significance evaluates statistical significance while integrating the spatial context in sparse and heterogeneous spatial datasets.
- You compare the magnitude of the difference both empirically and visually
- (Salazar and Pyrcz, in press)



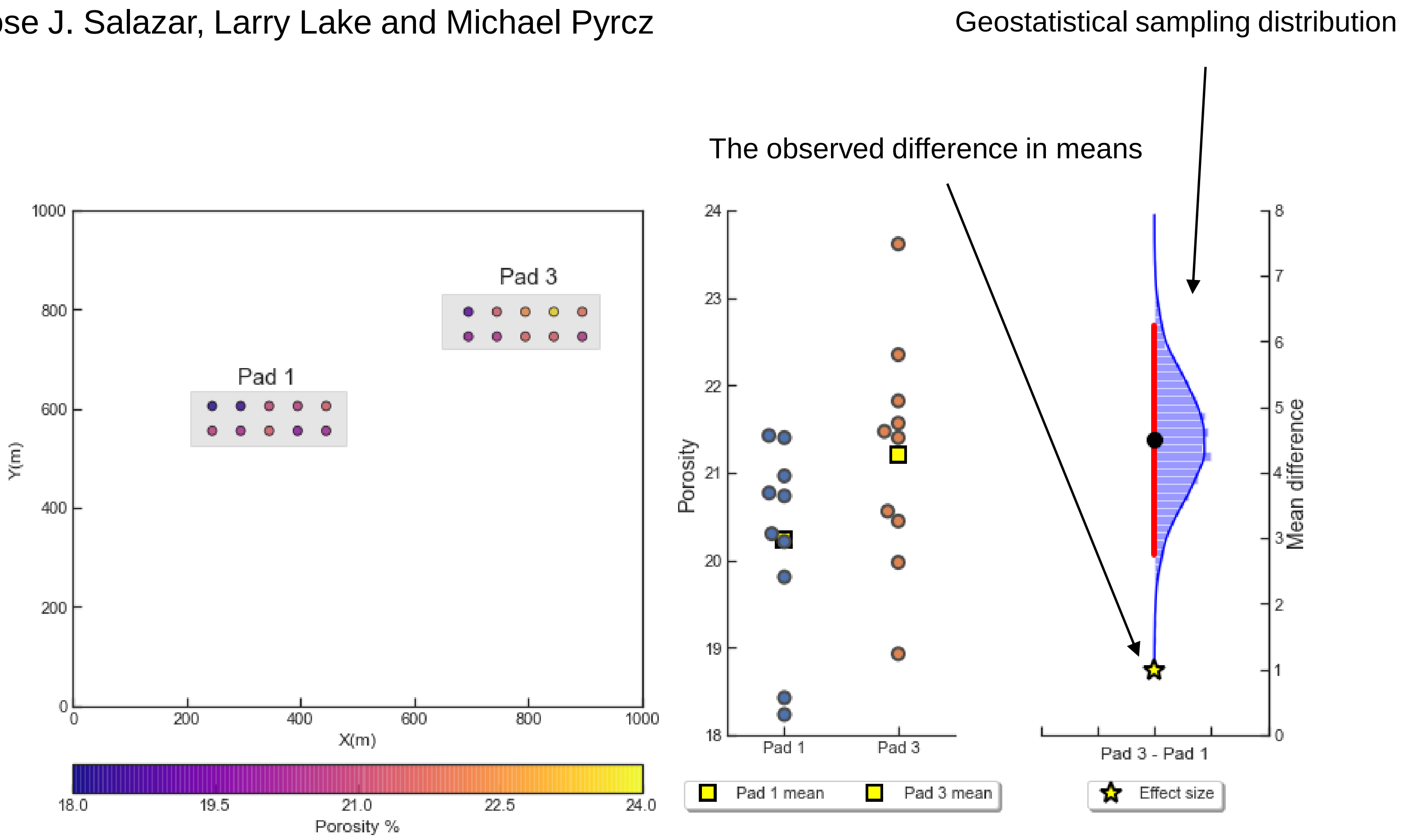
Modified Gardner-Altman plot. Left: porosity values. Center: Distribution of porosity from well pads. Right: Sampling distribution of mean differences, (Salazar and Pyrcz, in press)

Geostatistical Significance

Jose J. Salazar, Larry Lake and Michael Pyrcz

More of Geostatistical significance

- It avoids dichotomous judgment and thresholds (p-value < 0.05)
- Integrates geoscience and engineering data and expertise!
- The use of Monte Carlo p-value, Bayes factor bound, and graphics offers a more robust approach.
- The method avoids erroneously concluding the significance of variation in spatial phenomena (i.e. the law of small numbers).



The geostatistical significance is large and suggests well pads 1 and 3 are not in the same reservoir (Salazar and Pyrcz, in press).

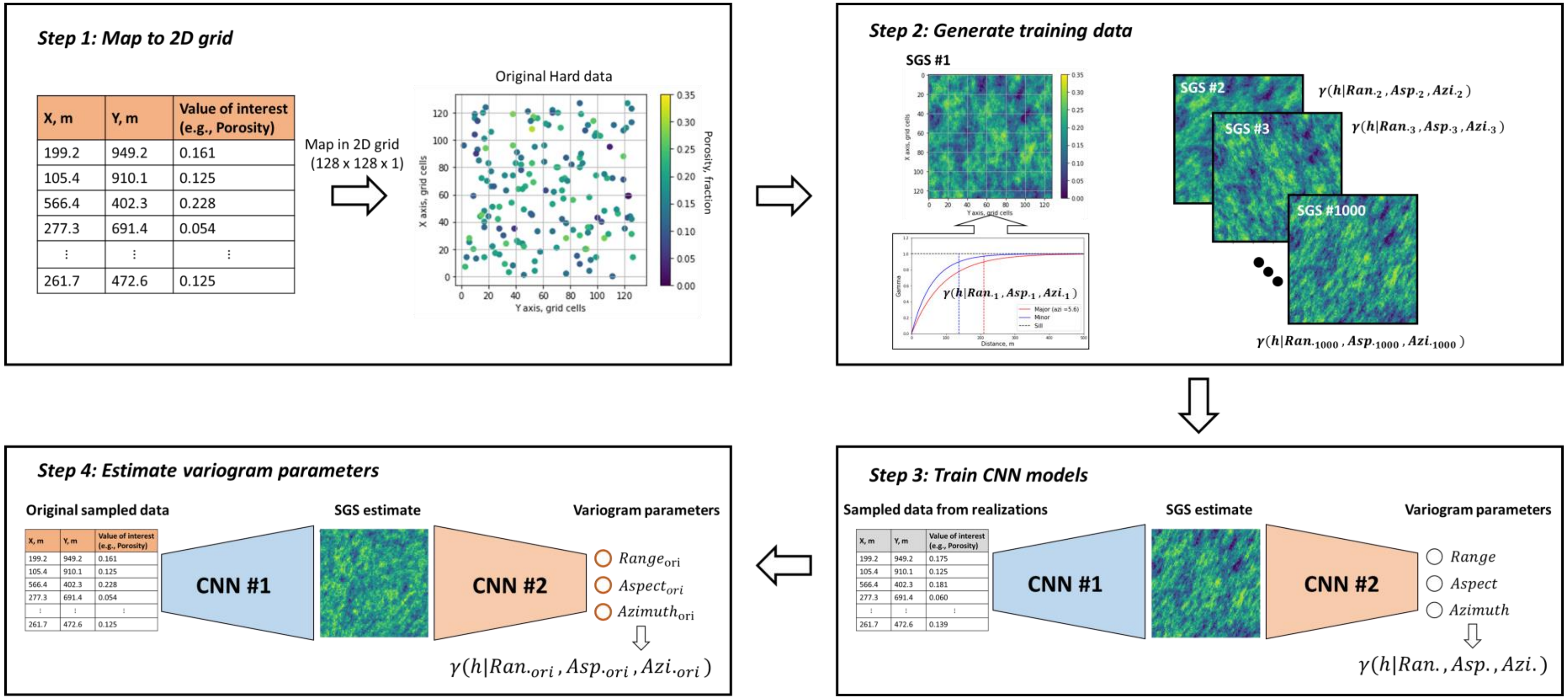
Spatial Continuity, Essential Inference. Can Machine Learning Assist?

Spatial Continuity Quantification

Honggeun Jo and Michael Pyrcz

Convolutional Neural Network Workflow to Automate Variogram Modeling

- First CNN data to model, second CNN model to variogram parameters



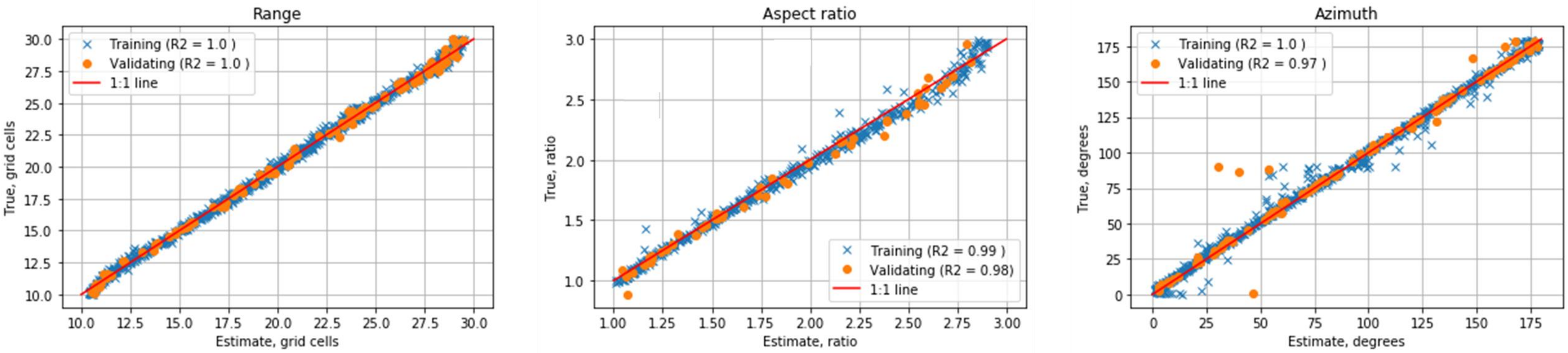
Automated semivariogram modeling with convolutional neural networks workflow (Jo and Pyrcz, in press).

Spatial Continuity Quantification

Honggeun Jo and Michael Pyrcz

Convolutional Neural Network Workflow to Automate Variogram Modeling

- Very accurate estimates of azimuth, range, and aspect ratio / anisotropy ratio



Prediction of major range **Prediction of aspect ratio** **Prediction of azimuth**

Cross validation of convolutional neural networks prediction of variogram model parameters from spatial data.

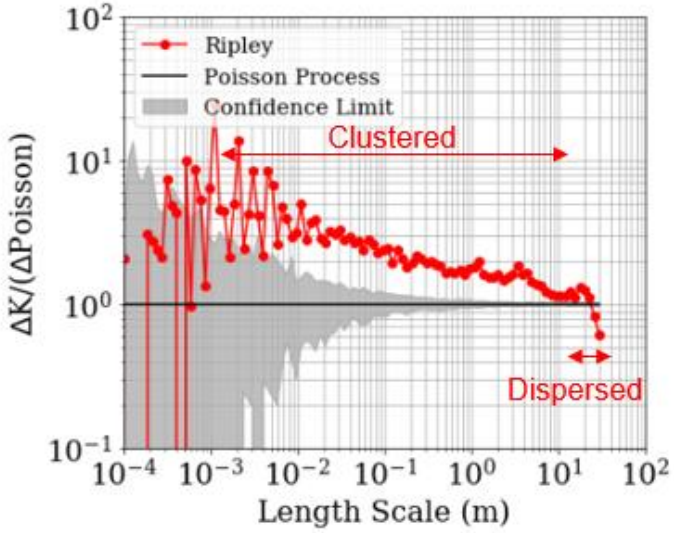
Spatial Point Processes for Fractures?

Multiscale Spatial Characterization of Fracture Pattern

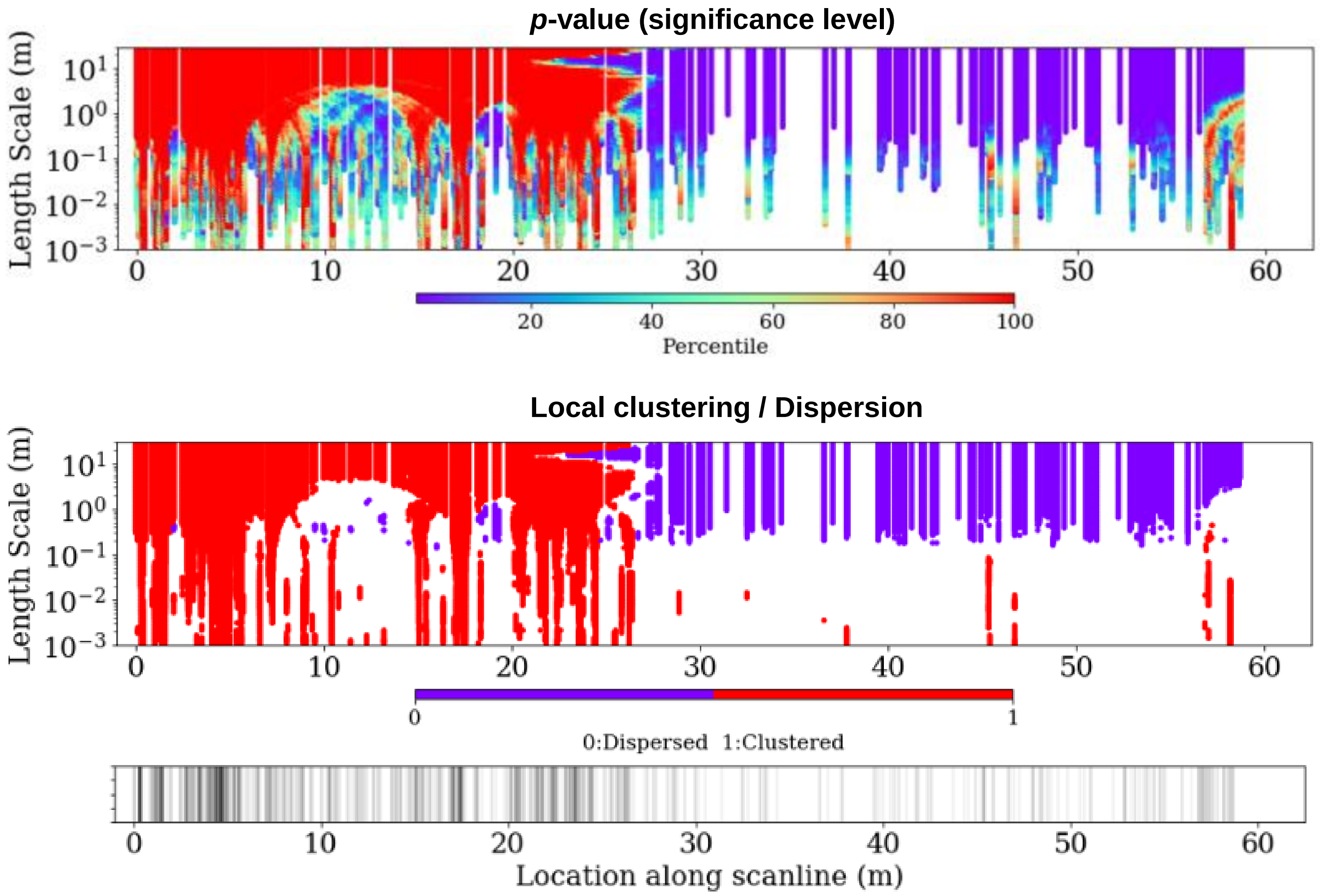
Mahmood Shakiba, Larry Lake and Michael Pyrcz, funded and in collaboration with FRAC Consortium, UT BEG

Point Pattern Analysis

- Identify and describe spatial organization of fractures across multiple length scales
- Applicable to fracture measurements from outcrops, borehole image logs, core samples, etc.



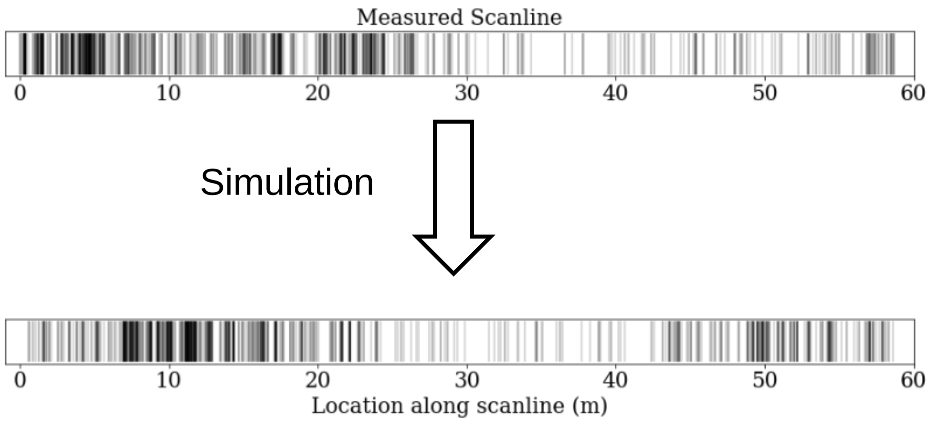
Multiscale spatial organization of fractures in Pedernales Falls Limestone

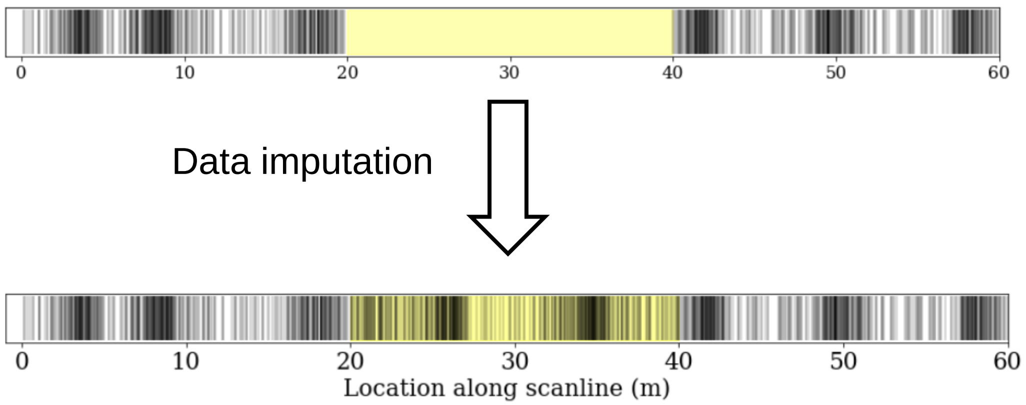


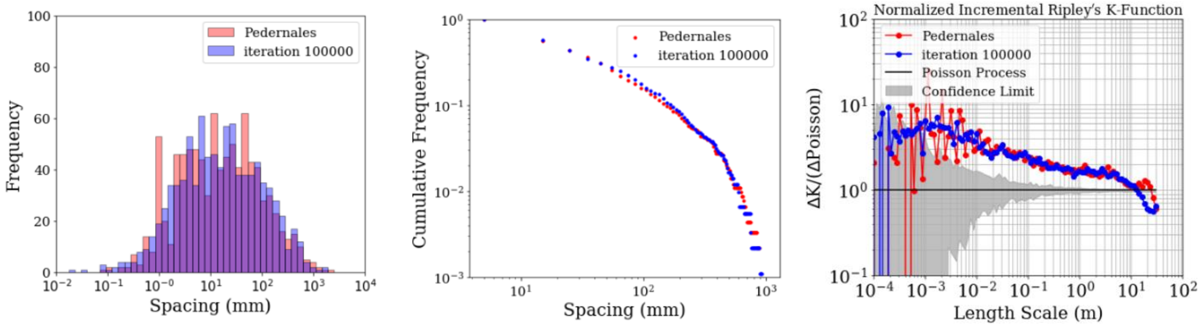
Demonstration of local structure of fractures (clustering/dispersion) using Ripley's K-function

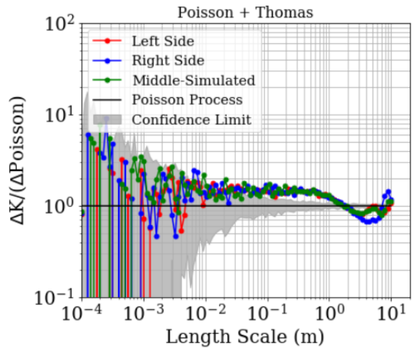
Fracture Pattern Reconstruction

Mahmood Shakiba, Larry Lake and Michael Pyrcz, funded and in collaboration with FRAC Consortium, UT BEG









- Fracture pattern reconstruction workflow based on Simulated Annealing algorithm
- Generating fracture realizations that statistically replicate actual fracture measurements

- Data imputation using Ripley's K-function and simulated annealing
- Honor multiscale spatial organization of fractures

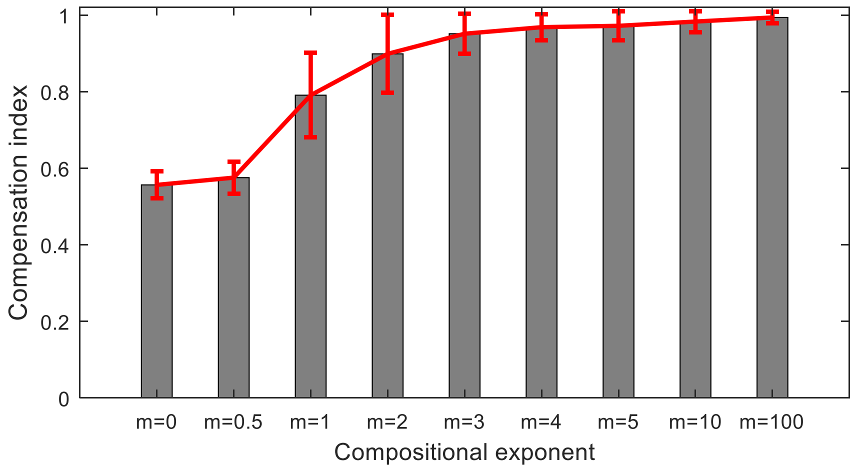
Modeling with Machine Learning Requires Good Training Models

Rule-based Subsurface Models and Flow Relevance

Honggeun Jo and Michael Pyrcz

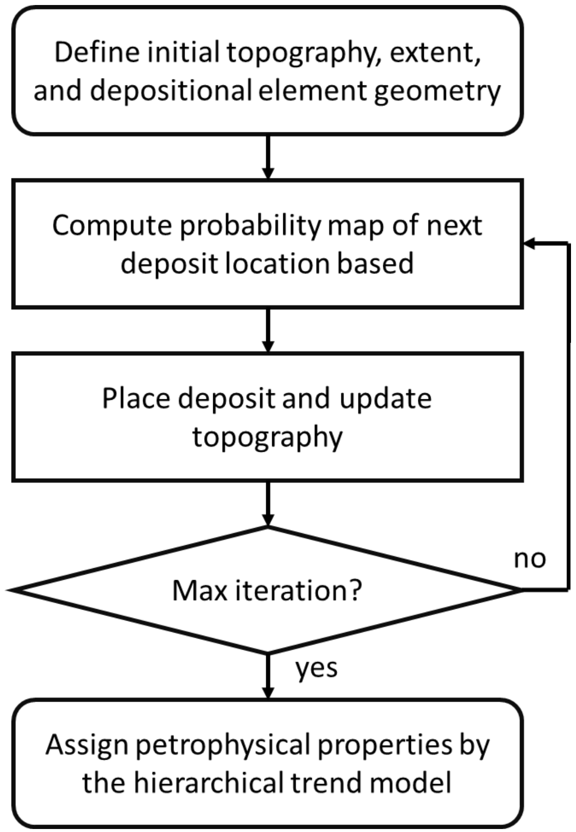
Rule-based Deepwater Lobe Model

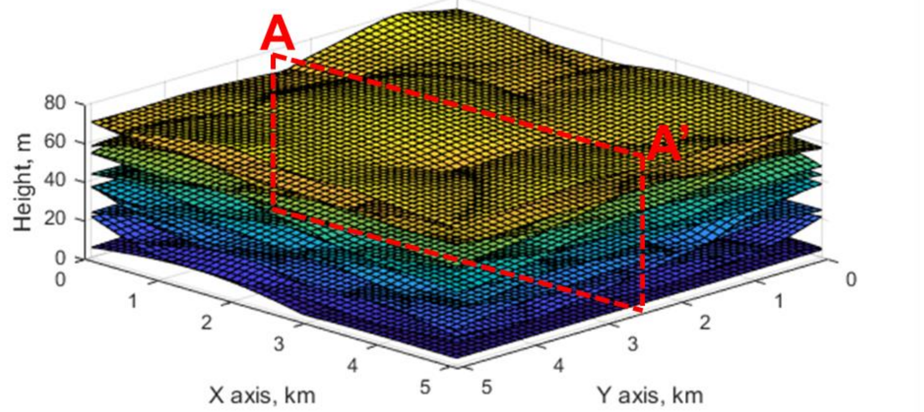
- Very efficient subsurface models calculated with rules tunned / calibrated to natural settings
- Training models for machine learning



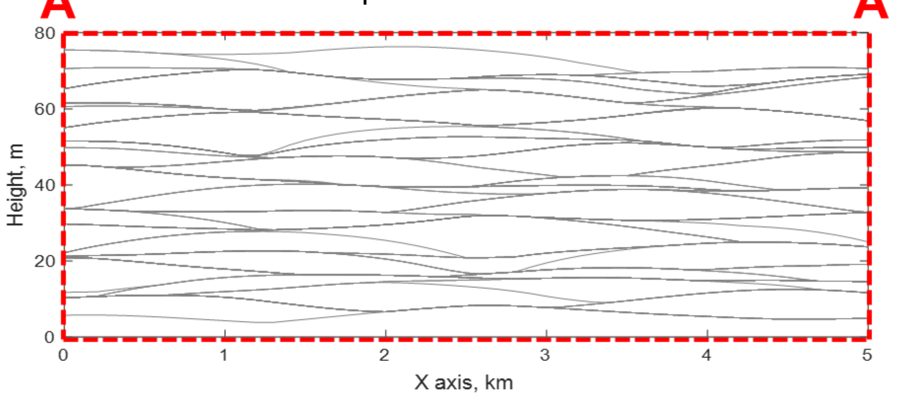
Compensational exponent calibrated to compensation index.

Rule-based model – Workflow





Compositional surfaces



Cross-sectional view

Rule-based modeling workflow, example model surfaces and cross section.

Rule-based Subsurface Models and Flow Relevance

Honggeun Jo and Michael Pyrcz

Rule-based Deepwater Lobe Model

- Very efficient subsurface models calculated with rules tunned / calibrated to natural settings
- Flow relevance studies to determine subsurface risk and prioritize patterns for integration into machine learning-based subsurface models.

Stage	Early	Early late	Middle	Late
Description	Before WBT	WBT just starts	$Q_{oil}>Q_{water}$	$Q_{water}>Q_{oil}$
PV water inj., %	~3%	~5%	~10%	~50% or more
WWCT, %	0%	~1%	~50%	~84% or more
Compositional Exponent	None	None	None	Positive
Radius of lobe	None	Negative	Negative	Negative
Thickness of lobe	Negative	Negative	Negative	None
STD of ln(k)	Positive	None	Negative	Negative

Rule based model flow relevance for model parameters over production time (Jo and Pyrcz, in press).

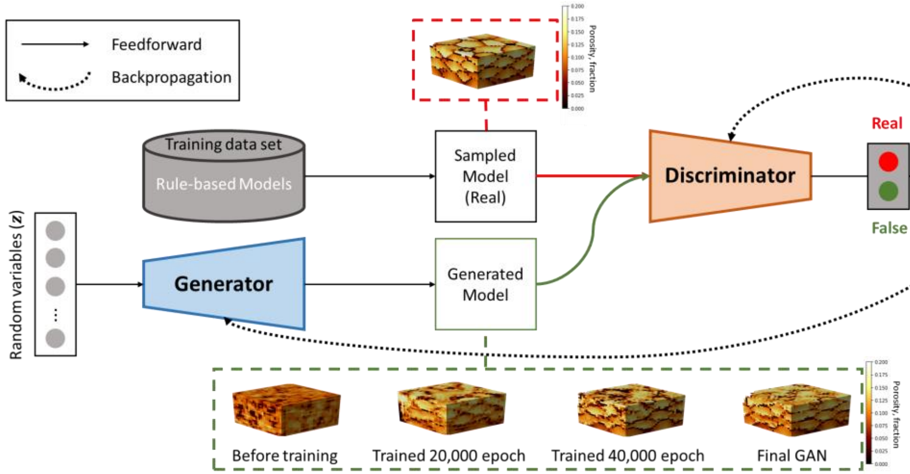
Subsurface Heterogeneity Models by Machine Learning?

ML-based Data Conditioning to Rule-based Model

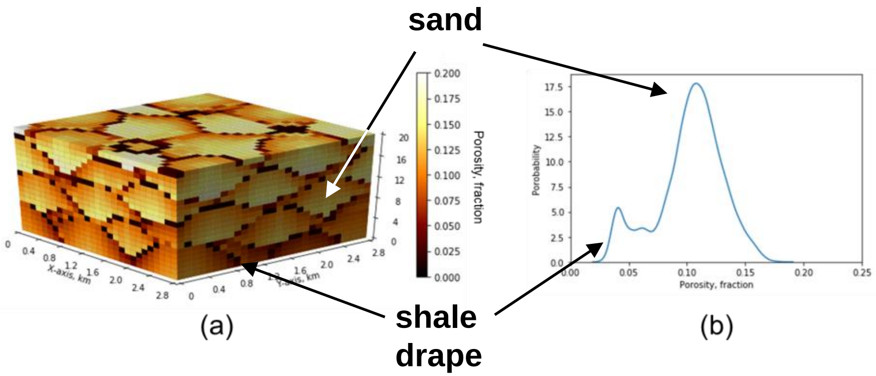
Honggeun Jo and Michael Pyrcz

Data Conditioning to Rule-based Models

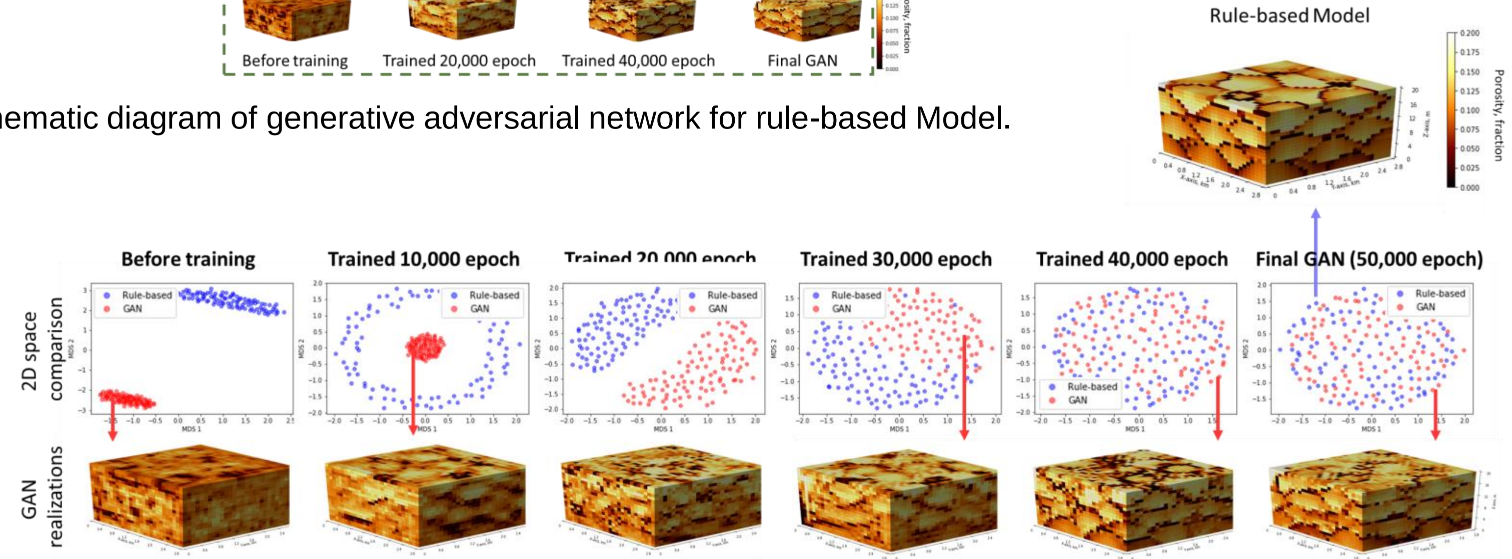
- Well data: GAN and image inpainting
- Stratigraphy profile: conditional GAN (i.e., image-to-image translation)
- Dynamic data: GAN and ensemble Kalman filter



Schematic diagram of generative adversarial network for rule-based Model.



A porosity realization of rule-based model: (a) 3D-view and (b) distribution of porosity (Jo, E. Santos, and Pyrcz, 2020).



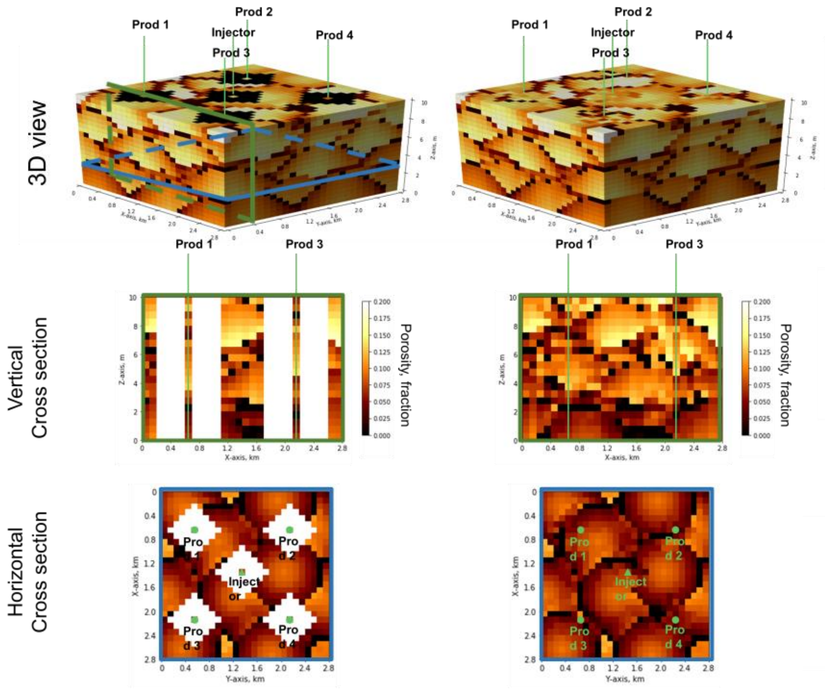
Comparison between rule-based models and GAN realizations in 2D space.

ML-based Data Conditioning to Rule-based Model

Honggeun Jo and Michael Pyrcz

Well Data (hard data)

- GAN and image inpainting

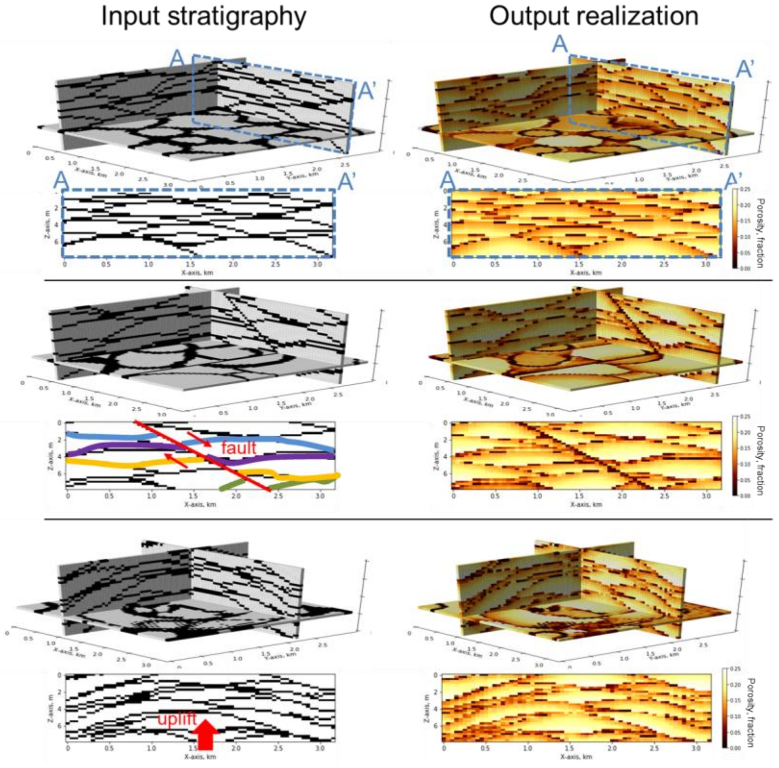


(a) (b)

Demonstration of well data conditioning: (a) Rule-based model with voids and well data in each center and (b) restored rule-based model (Jo, E. Santos, and Pyrcz, 2020).

Stratigraphy Profile

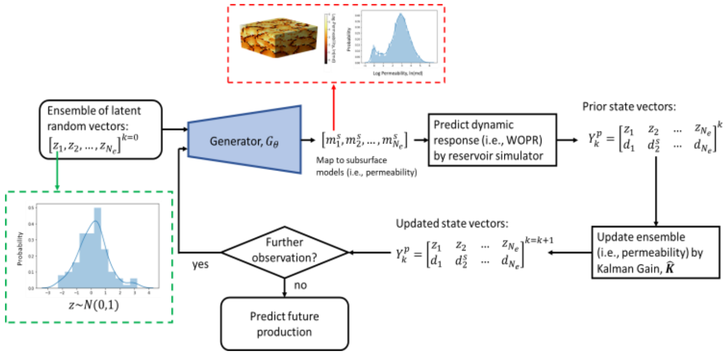
- Conditional GAN



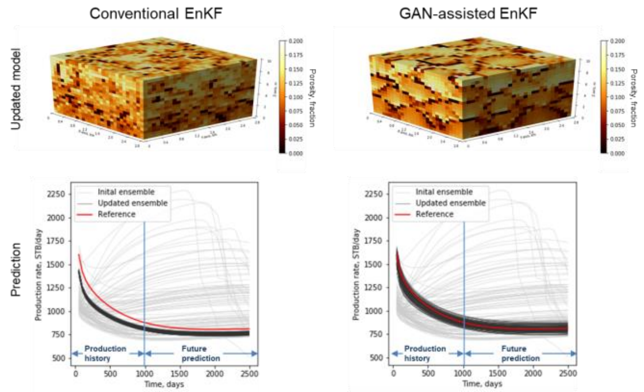
Examples of GAN realizations (right) that are conditioned to the given stratigraphy profiles (left) (Jo and Pyrcz, 2020).

History Matching

- Update input parameters of GAN, not porosity/permeability



Workflow of GAN-assisted EnKF



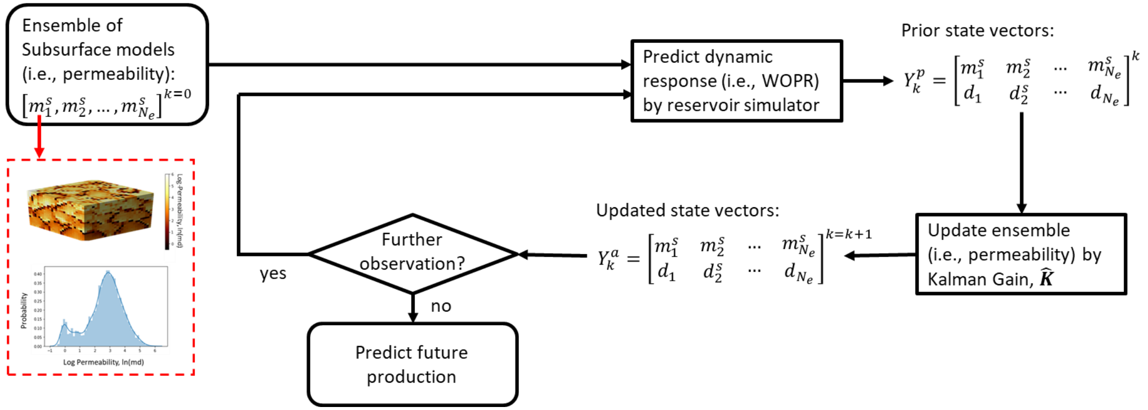
History matching results after assimilating 500 days production observation

ML-based Data Conditioning to Rule-based Model

Honggeun Jo and Michael Pyrcz

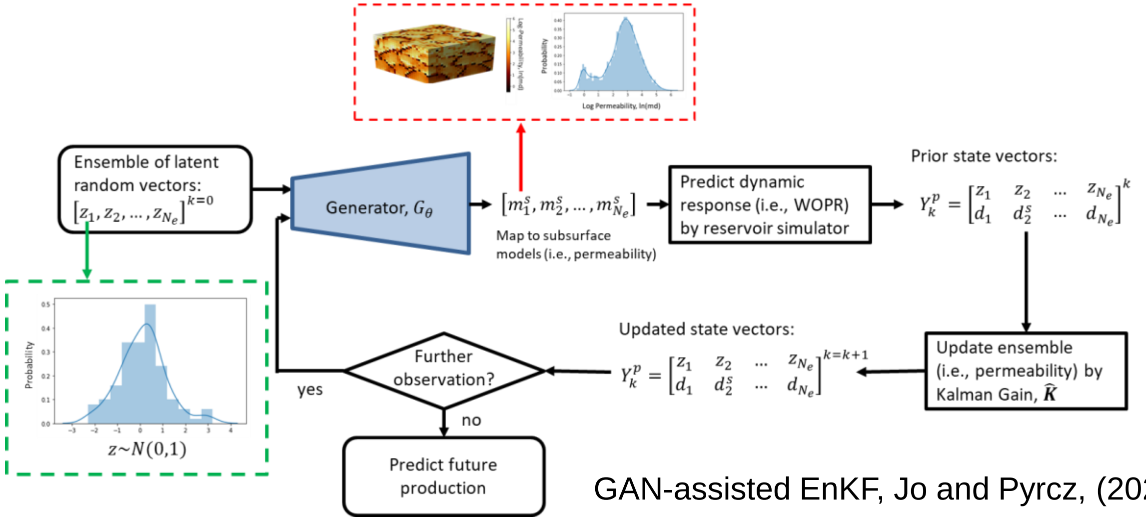
Comparison between conventional EnKF and GAN-assisted EnKF

Conventional EnKF



- Update subsurface model parameters (i.e., porosity and permeability) directly
- No guarantee that initial geological features are conserved in the updated models

GAN-assisted EnKF



GAN-assisted EnKF, Jo and Pyrcz, (2021)

- Update latent random vectors (i.e., inputs for GAN)
- Latent random vector follows Gaussian, which is the underlying assumption in EnKF
- GAN always generate realizations with realistic heterogeneity and continuity

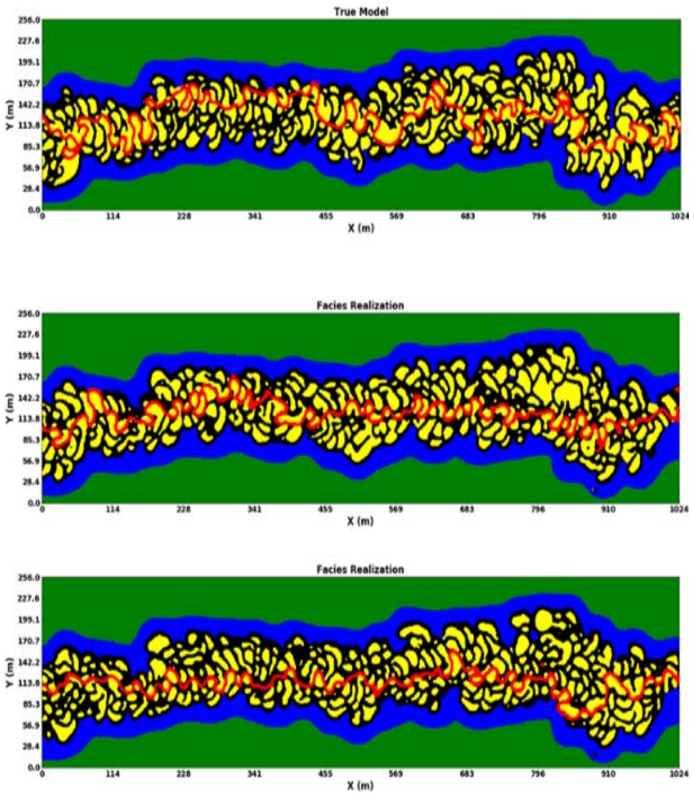
Stochastic pix2pix for Subsurface Modeling

Wen Pan, Michael Pyrcz and Carlos Torres-Verdin

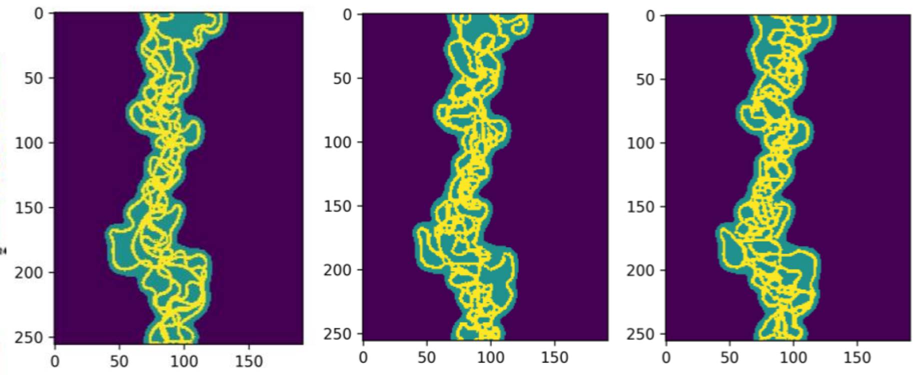
Highlights of This Method

- Models are generated “on the fly”, it is over 90% faster than conventional MPS method.
- A good tool for model reduction.
- Introducing data conditioning as a loss function avoids mode collapse in other conditional generative methods.
- It is able to capture large-scale patterns with less memory requirement compared to MPS method.
- Field data and geological concept conditionings are good.

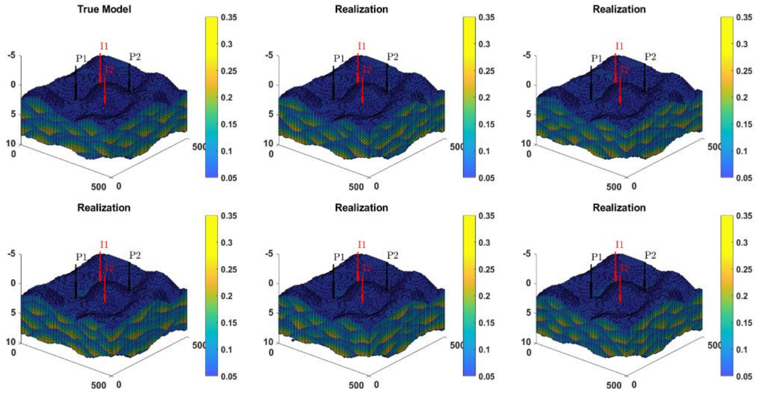
Fluvial System Realizations



Turbidite Channel System Realizations



Lobe System Realizations



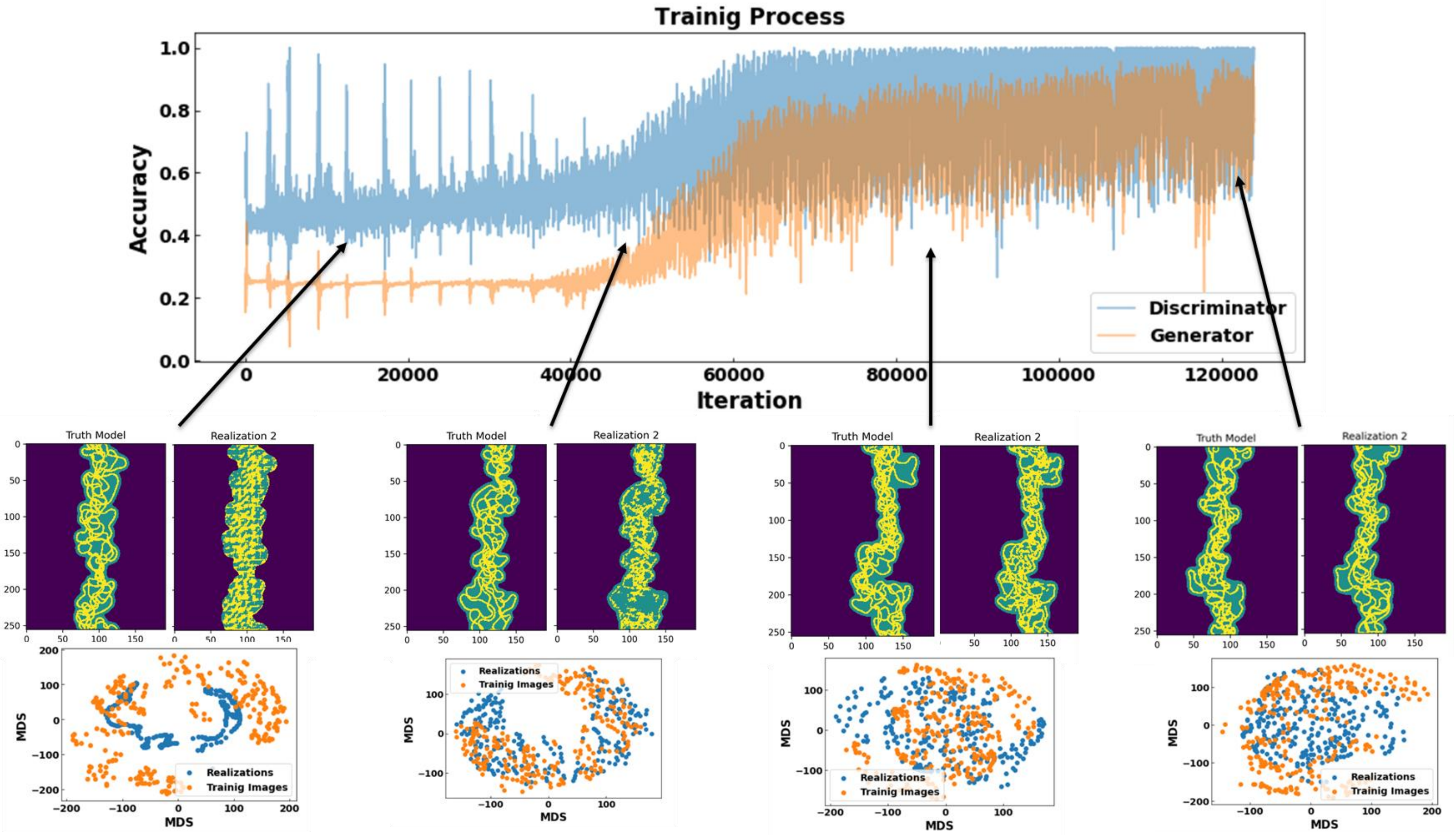
Truth model and realizations with good horizon/outline conditioning and feature reproduction Pan, Pyrcz & Torres-Verdin, (in press)

Stochastic pix2pix for Subsurface Modeling

Wen Pan, Michael Pyrcz and Carlos Torres-Verdin

Training Process

- MDS is calculated to show the convergence of this method during training.
- A more reliable but more computationally expensive method to check convergence is to calculate the MPS difference between training images and realizations.



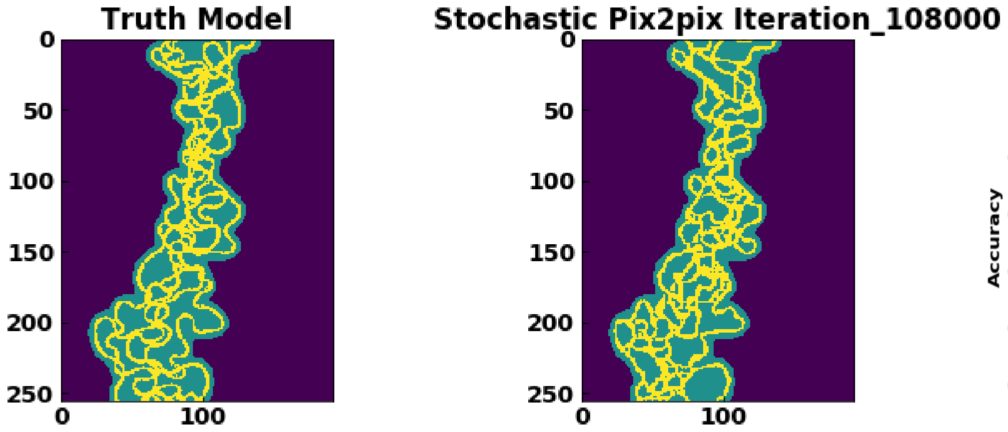
Model evolution over training epochs and coverage of the subsurface space of uncertainty with resulting realizations.

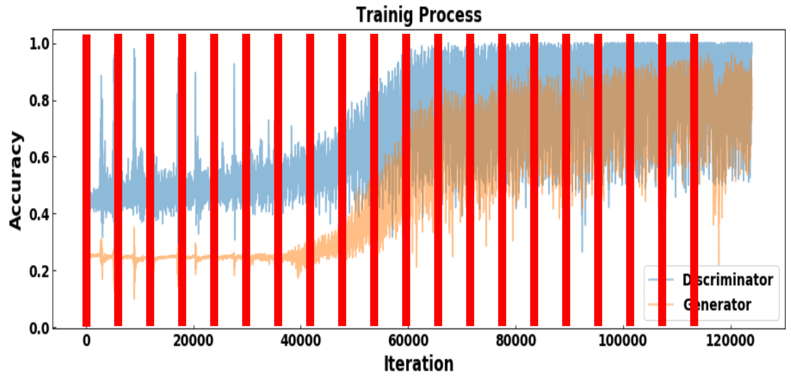
Stochastic pix2pix for Subsurface Modeling

Wen Pan, Michael Pyrcz and Carlos Torres-Verdin

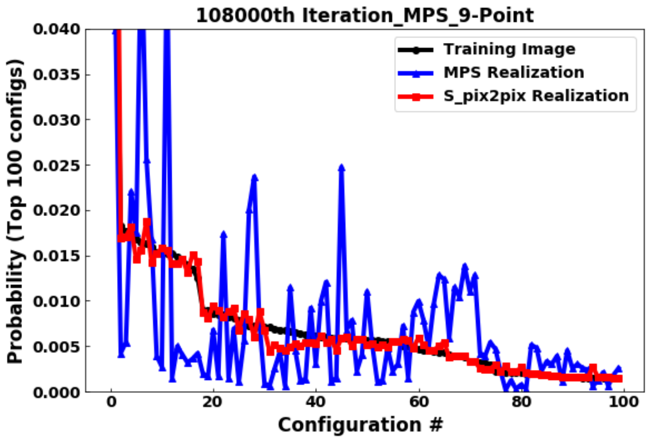
Training Process

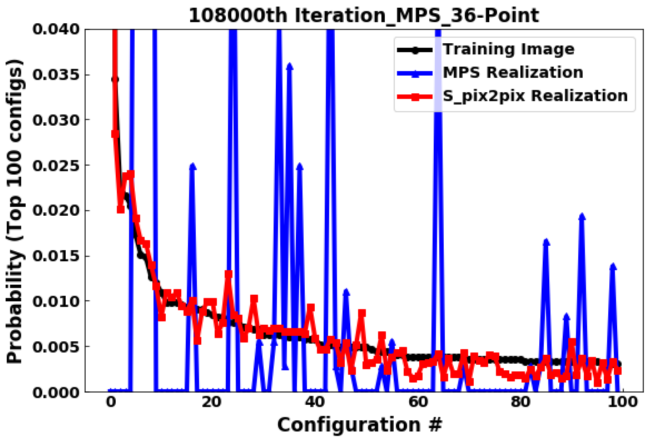
- A more reliable but more computationally expensive method to check convergence is to calculate the MPS difference between training images and realizations.

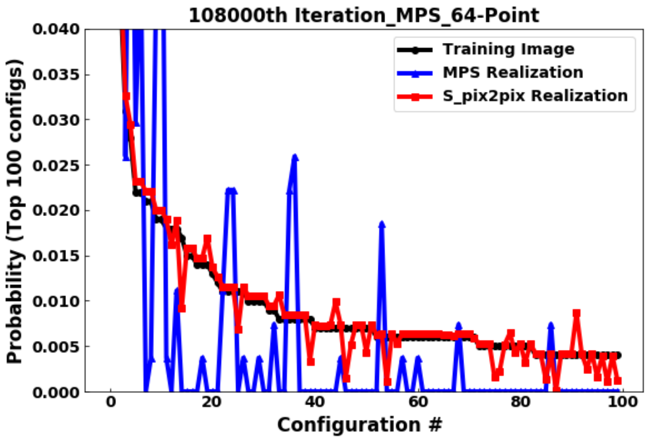




Truth model and realizations with good horizon/outline conditioning and feature reproduction.





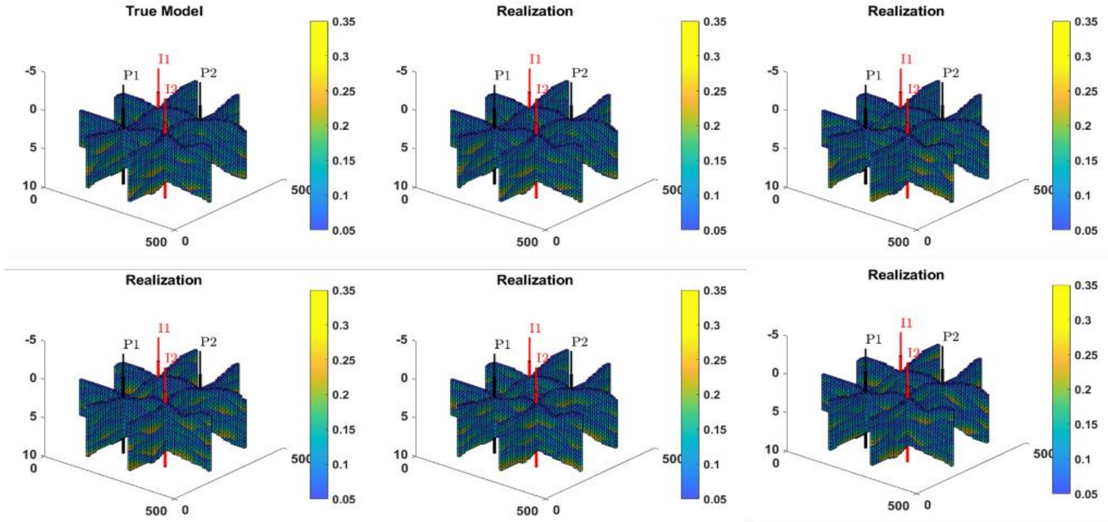


Stochastic pix2pix for Subsurface Modeling

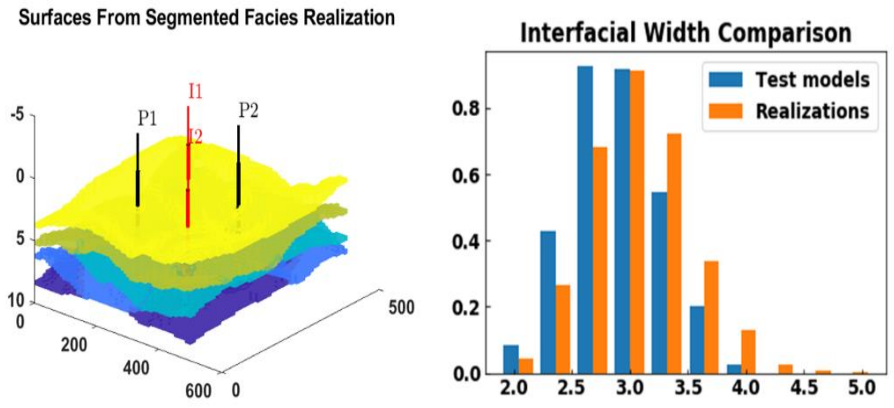
Wen Pan, Michael Pyrcz and Carlos Torres-Verdin

Model Validation I

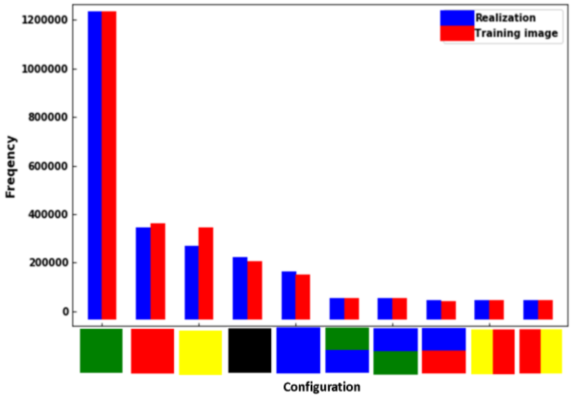
- It reproduces single-,two-point and multi-point statistics.
- It reproduces depositional rules used in the rule-based simulator



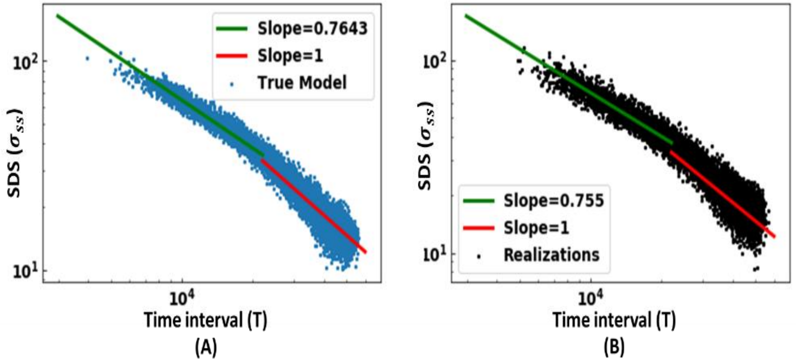
Reproduction of well conditioning.



Reproduction of bounding surfaces, bed boundary check with surface statistics.



Reproduction of multiple point statistics.



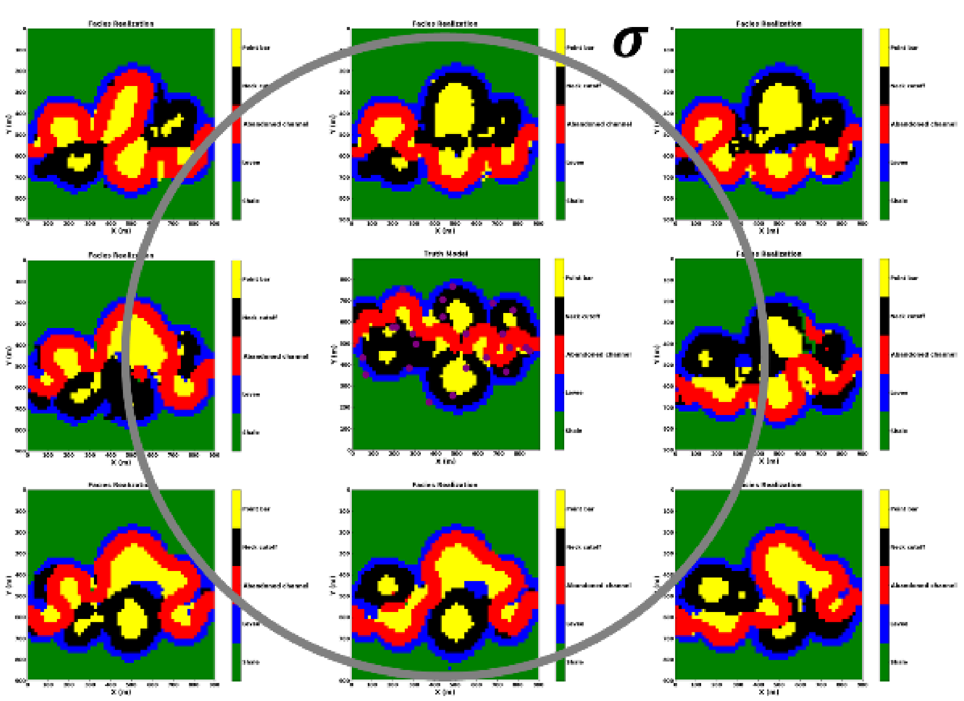
Reproduction of stacking patterns, compensational index.

Stochastic pix2pix for Subsurface Modeling

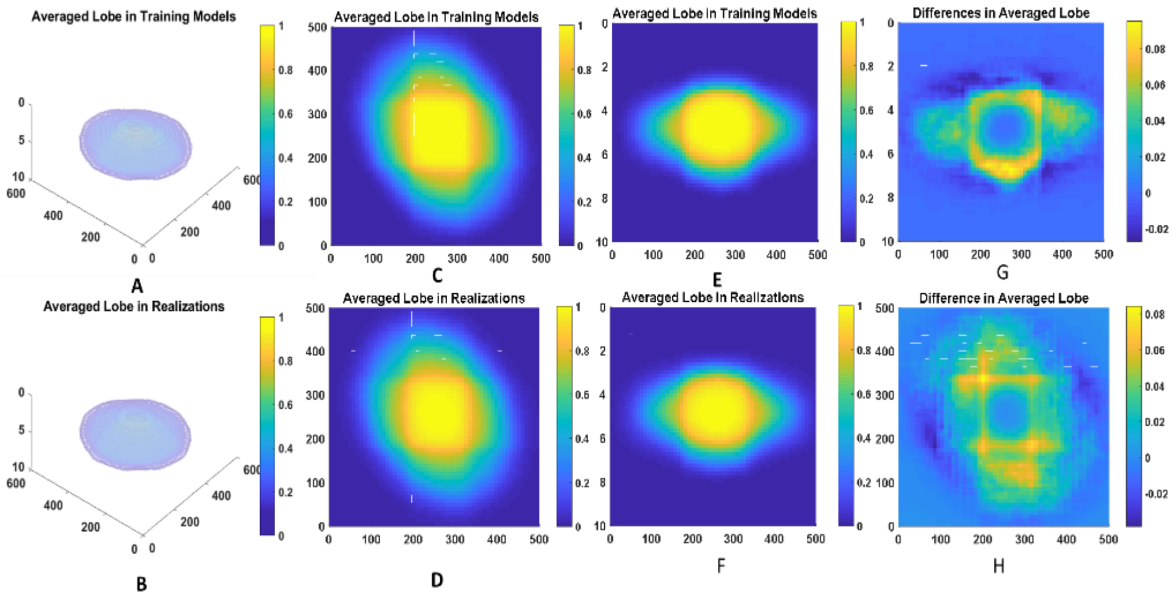
Wen Pan, Michael Pyrcz and Carlos Torres-Verdin

Model Validation II

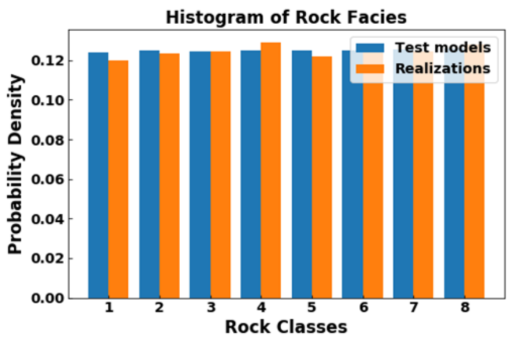
- Continuous change in latent space corresponds to continuous change in model space.

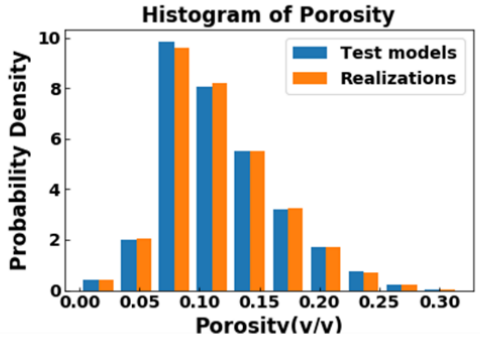


Continuous change in both latent and model space.



Good geobody geometry stats reproduction.





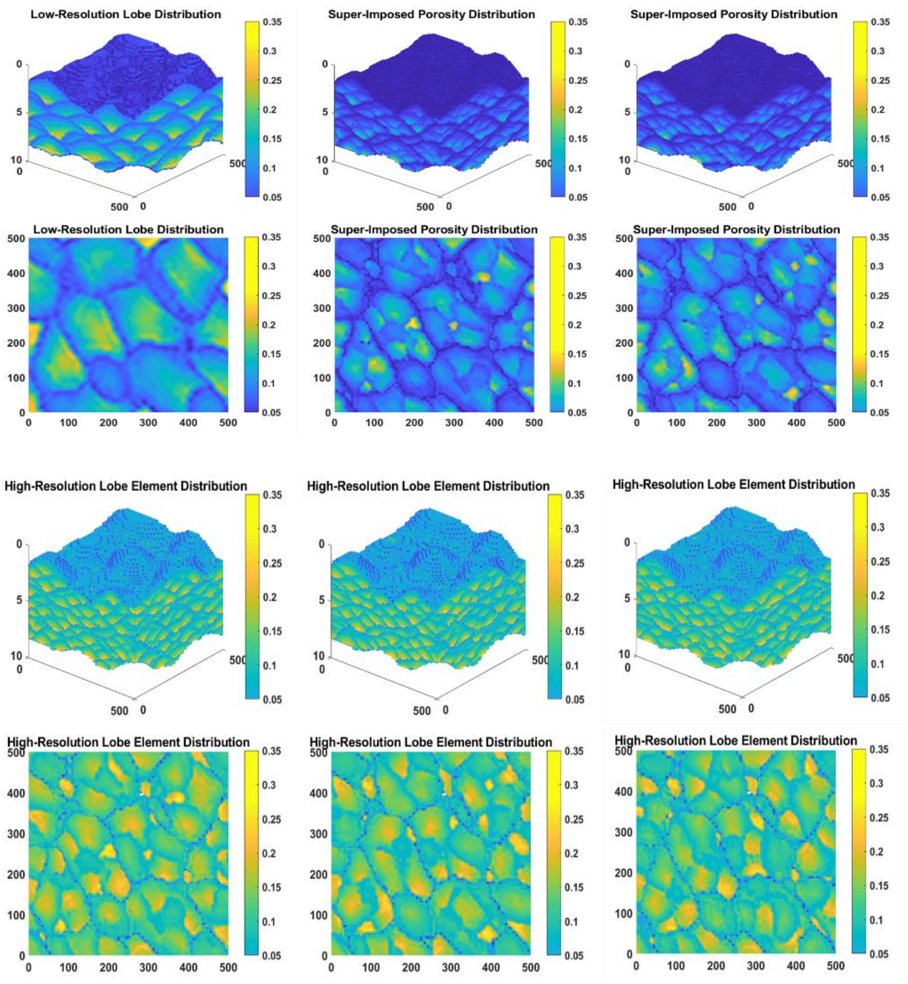
Reproduction of univariate statistics.

Stochastic pix2pix for Hierarchical Modeling

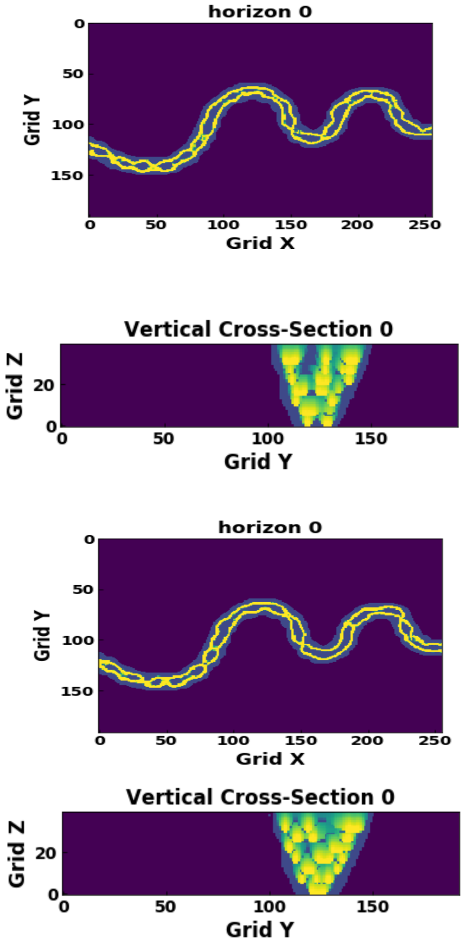
Wen Pan, Michael Pyrcz and Carlos Torres-Verdin

Highlights of This Method

- Consistent hierarchical modeling. Avoid the cookie-cutter approach!
- Enabling top-down modeling to avoid over-complicated models.
- Different frequency components of production and seismic data can be attributed to the geomodels at different scales, faster convergence and better uncertainty quantification can be achieved.
- Pan, Pyrcz and Torres-Verdin (in prep)



Hierarchical lobe modeling example.



Hierarchical channel modeling example.

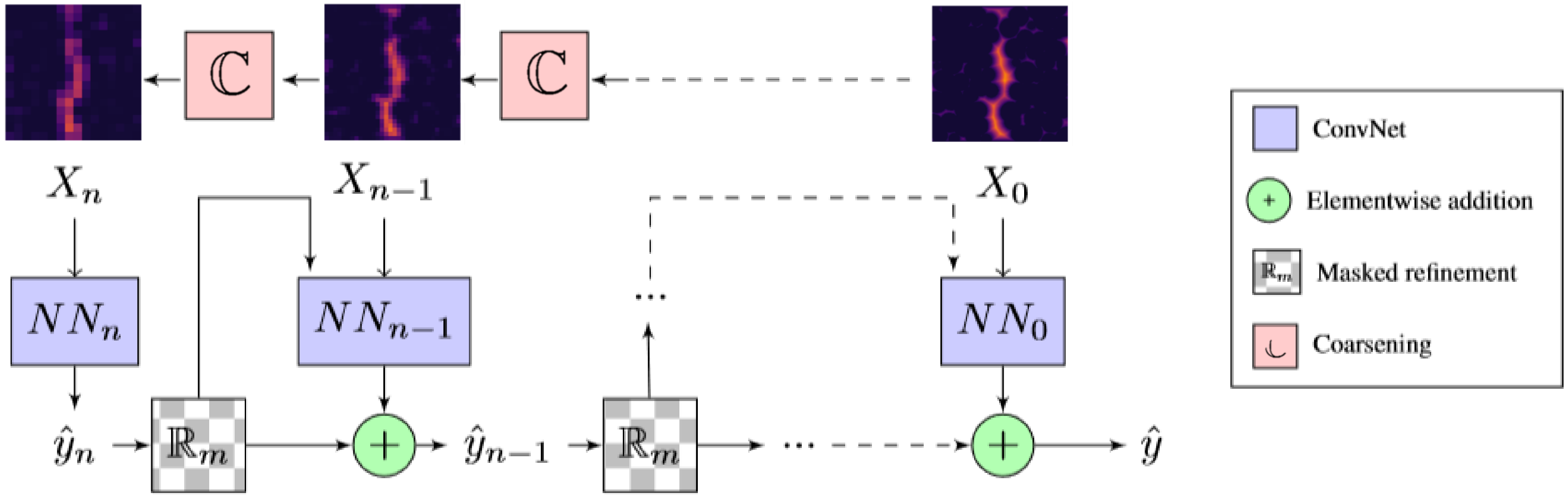
Subsurface Forecasting by Machine Learning?

MS-Net: Multiscale Flow Surrogate Model

Javier E. Santos, Masa Prodanovic and Michael Pyrcz

MS-NET Architecture:

- Multiscale flow in porous media prediction.
- A system of interacting convolutional neural networks that work at different scales.



MS-Net architecture (Santos et al., in press).

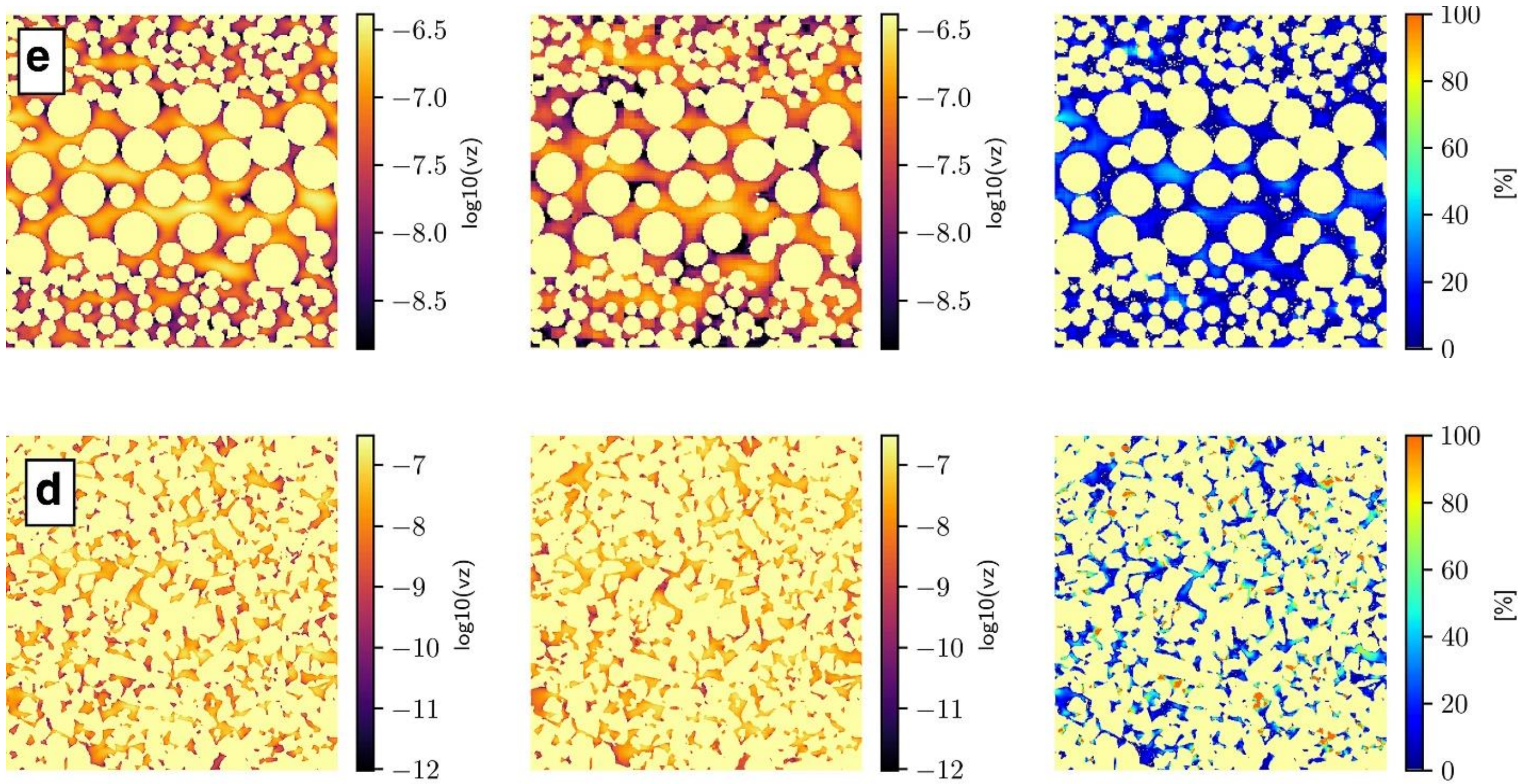
MS-Net: Multiscale Flow Surrogate Model

Javier E. Santos, Masa Prodanovic and Michael Pyrcz

MS-NET Architecture:

Propped Fracture

Castlegate Sandstone



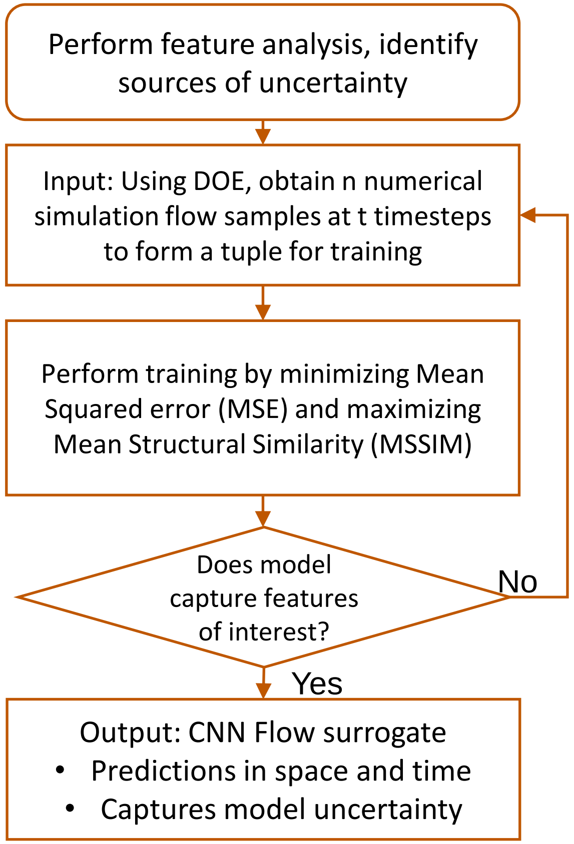
“Hard” samples from the test set. The MS-Net produces excellent pixel-wise results in a natural samples with a model trained with a sphere-pack.

ML Deep Convolutional Network for Flow Surrogates

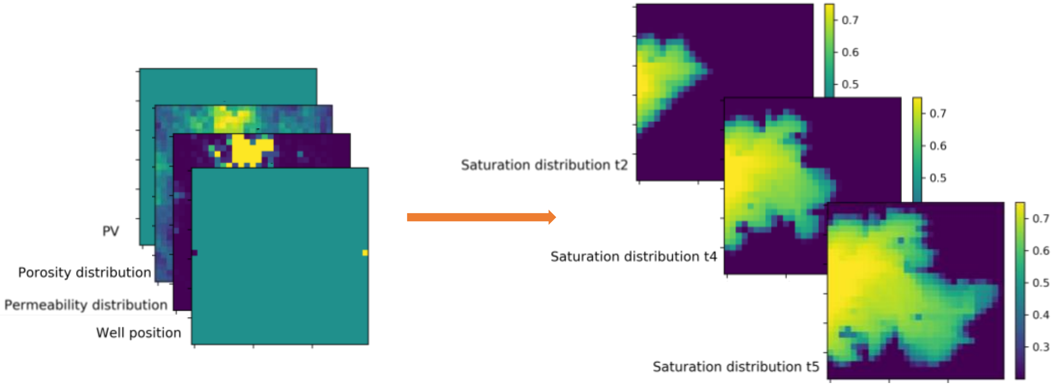
Eduardo Maldonado and Michael Pyrcz

Deep Convolutional Network for Surrogate Modeling

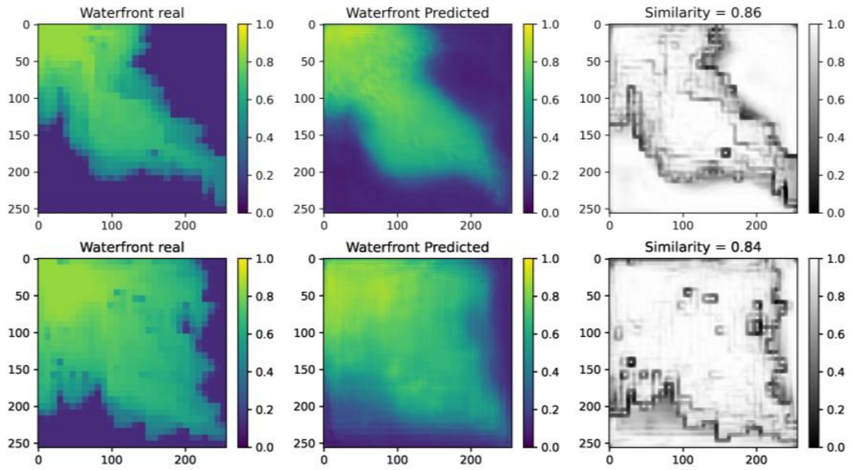
- New and general workflow for a machine learning-based flow surrogate model
- The model output shows saturation changes and estimates recoverable volumes along discrete timesteps.
- Generalizable to a variety of practical flow forecasting problems for real-time modeling decision significance feedback.



Workflow for surrogate modeling.



Tuple used for training: Multi-channel integrates static and dynamic flow properties.



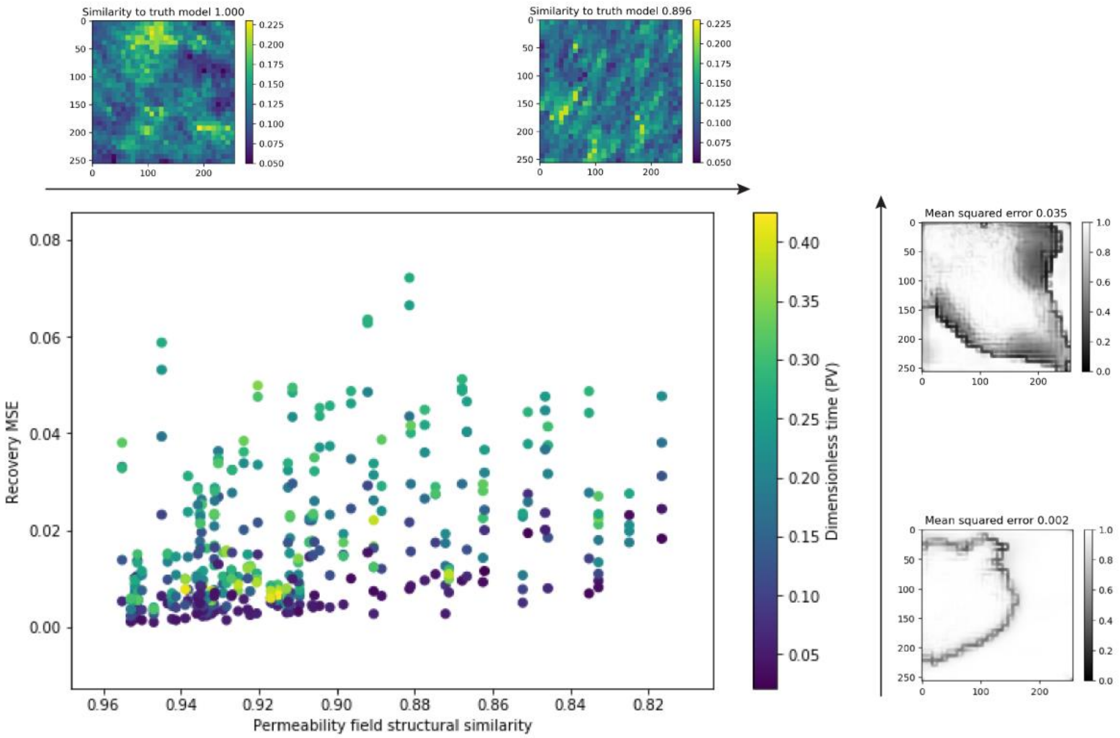
Saturation change from simulator, b) Saturation estimated by Convolutional Neural Network and c) Similarity between estimates.

ML Deep Convolutional Network for Flow Surrogates

Eduardo Maldonado and Michael Pyrcz

Results

- The workflow presented allows the generation of spatially aware flow surrogate models.
- The flow surrogate model correctly identifies:
  - Well positions.
  - Position and maximum saturation changes.
- Addresses an unsolved task with the inclusion of dimensionless time.



Mean squared error between predicted recovery factor vs permeability field similarity

We Need a Good Uncertainty Model?

ML Hyperparameter Tuning for Fair Uncertainty

Eduardo Maldonado and Michael Pyrcz

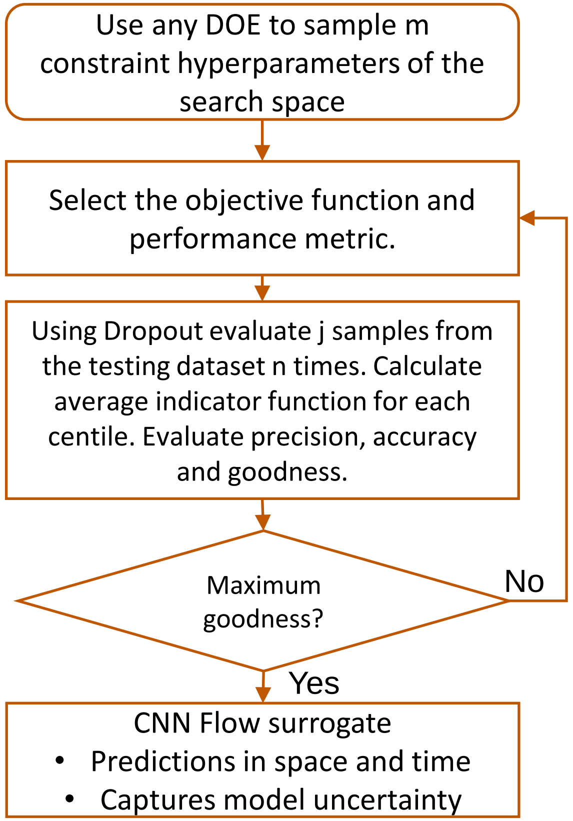
Uncertainty Model Goodness

- We integrate model goodness (Deutsch, 1996) with deep learning models to tune dropout frequency for improved uncertainty models

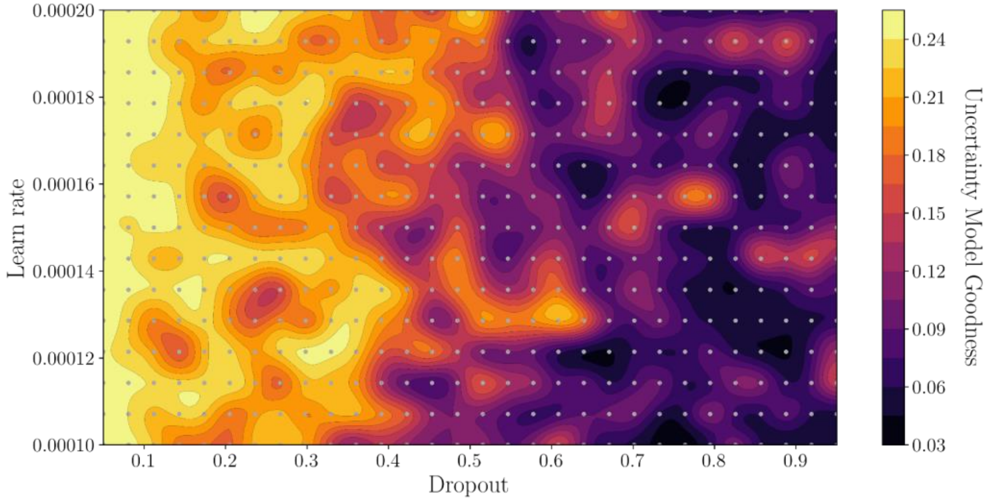
- Use of a metric to summarizes model uncertainty through a comparison of the entire uncertainty model based on the prediction realization.

- This function can be used for hyperparameter tuning beyond estimation accuracy.

Workflow for hyperparameter search for uncertainty models



Proposed workflow for fair ML uncertainty models.



Hyperparameter search based on uncertainty model goodness (Maldonado-Cruz and Pyrcz, 2021).

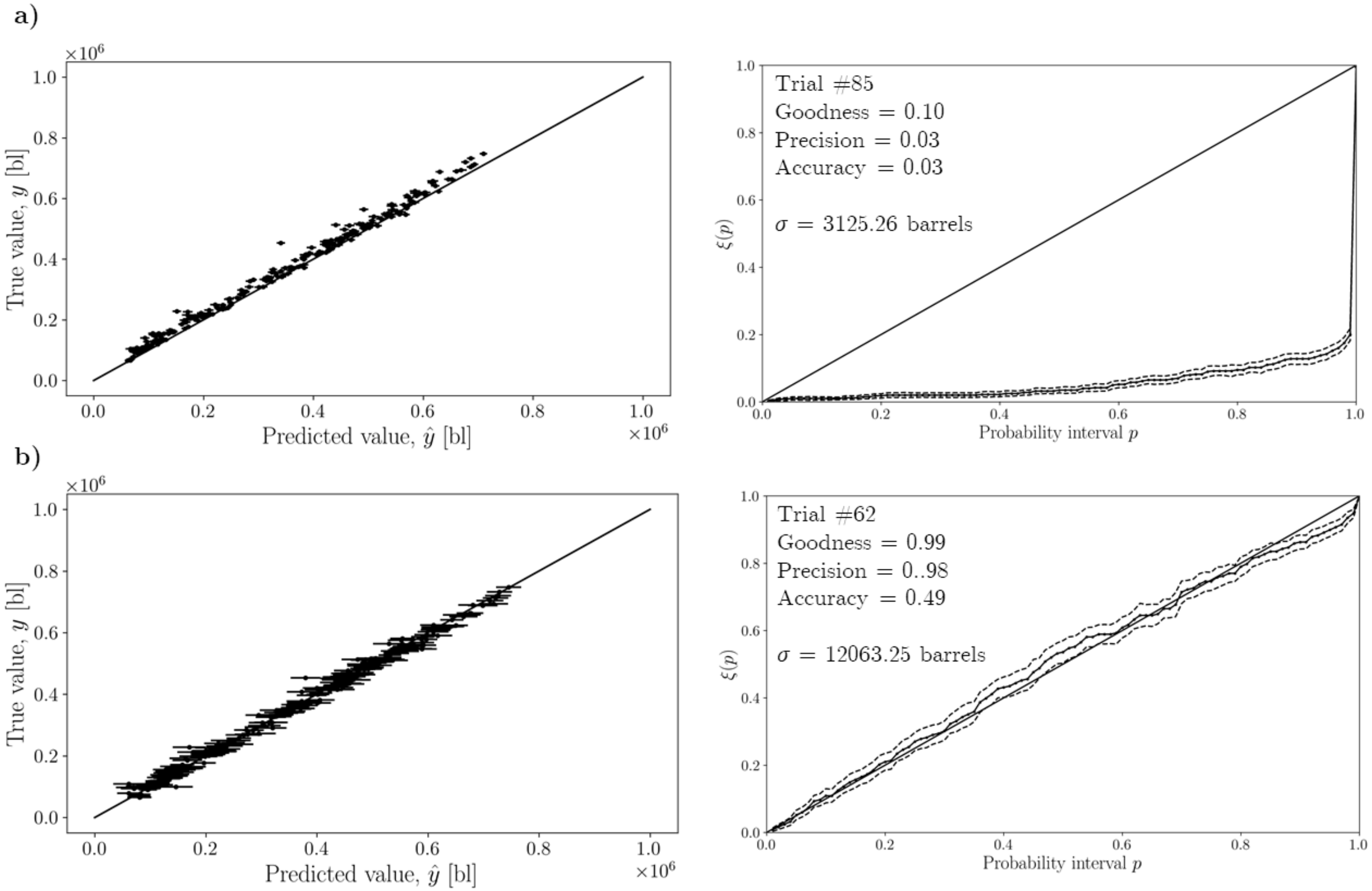
ML Hyperparameter Tuning for Fair Uncertainty

Eduardo Maldonado and Michael Pyrcz

Hyperparameter Selection Tuning for Fair Uncertainty

- The inclusion of uncertainty model goodness allows the construction of accurate and precise uncertainty models.

- Our metric addresses uncertainty quantification and optimization of network hyperparameters for surrogate modeling.



Both models have high cross-validation scores in testing data, but the proposed method with the new objective function results in an improved uncertainty model. (Maldonado and Pyrcz, 2021)

Concluding Remarks

There are new digital solutions for subsurface data analytics and machine learning, adding **new capabilities and value!**

- efficiency and automation – **geoscientists do more geoscience, engineers to do more engineering!**
- maximizing the value of data – **find new patterns, focus on what matters!**
- assisted modeling – **real-time feedback for learning while modeling!**
- new improved models – **better integration of expert knowledge!**

New digital skills for new opportunities for engineers!