SPE DISTINGUISHED LECTURER

Primary funding is provided by

The SPE Foundation through member donations and a contribution from Offshore Europe

The Society is grateful to those companies that allow their professionals to serve as lecturers

Additional support provided by AIME

SPE International

Society of Petroleum Engineers
Distinguished Lecturer Program
www.spe.org/dl

Big Data Analytics : What Can it do for Petroleum Engineers and Geoscientists?

Srikanta Mishra

BATTELLE

Society of Petroleum Engineers
Distinguished Lecturer Program
www.spe.org/dl

2018-19 DL Season

The Attraction ....

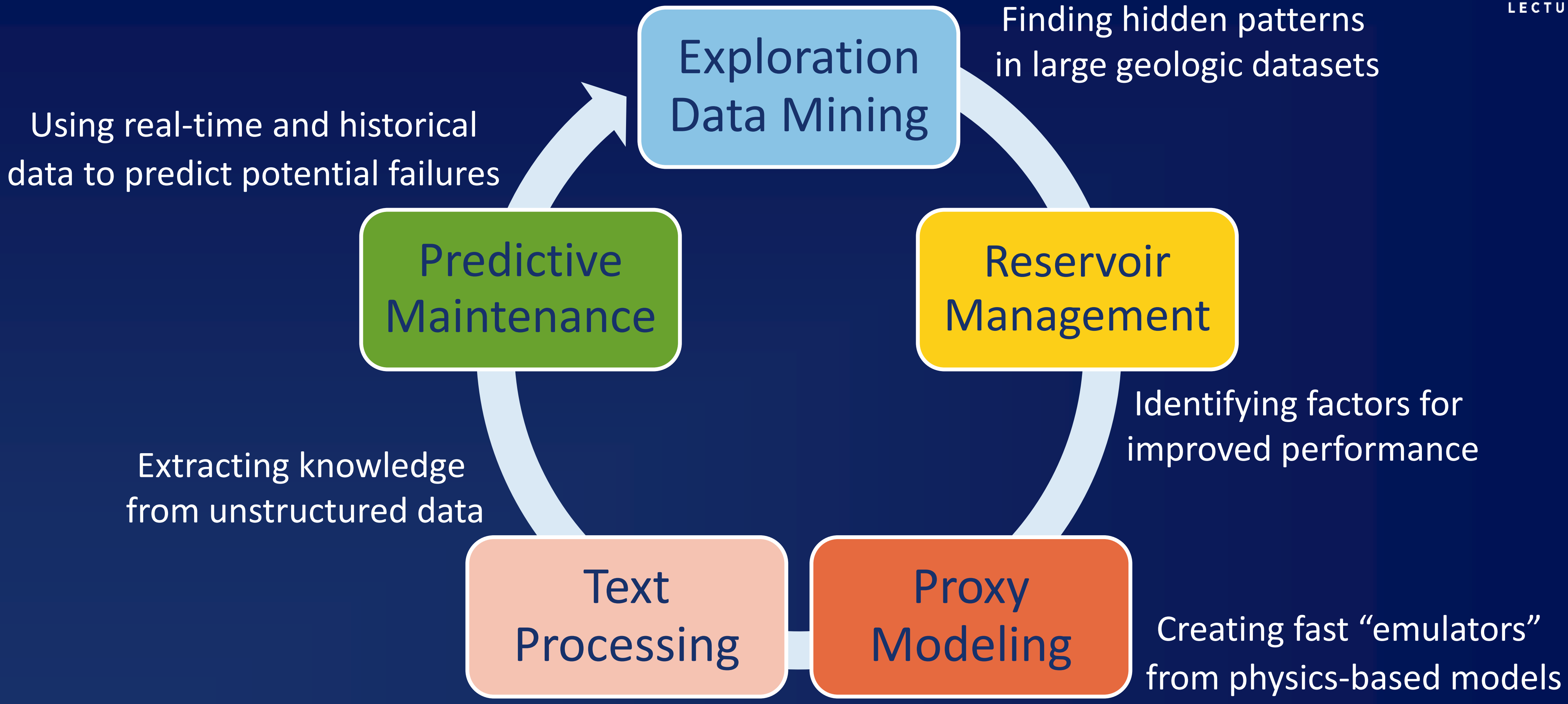
Big Data Analytics $\stackrel{?}{=}$ Game Changer

large volumes of data about subsurface, physical infrastructure and flows

New insights about reservoir from "data mining" can help increase operational efficiencies

Actionable information

The Possibilities ….



Reduce cost, improve productivity, increase efficiency

Outline of Talk

SPE DISTINGUISHED LECTURER

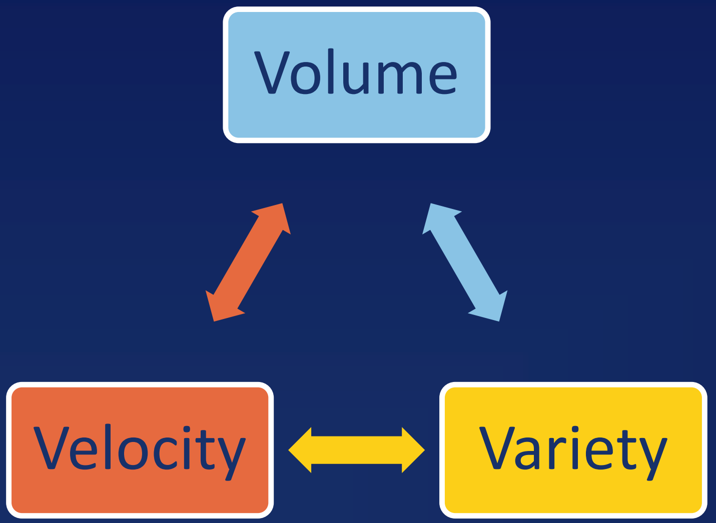
Basic Concepts

Case Studies

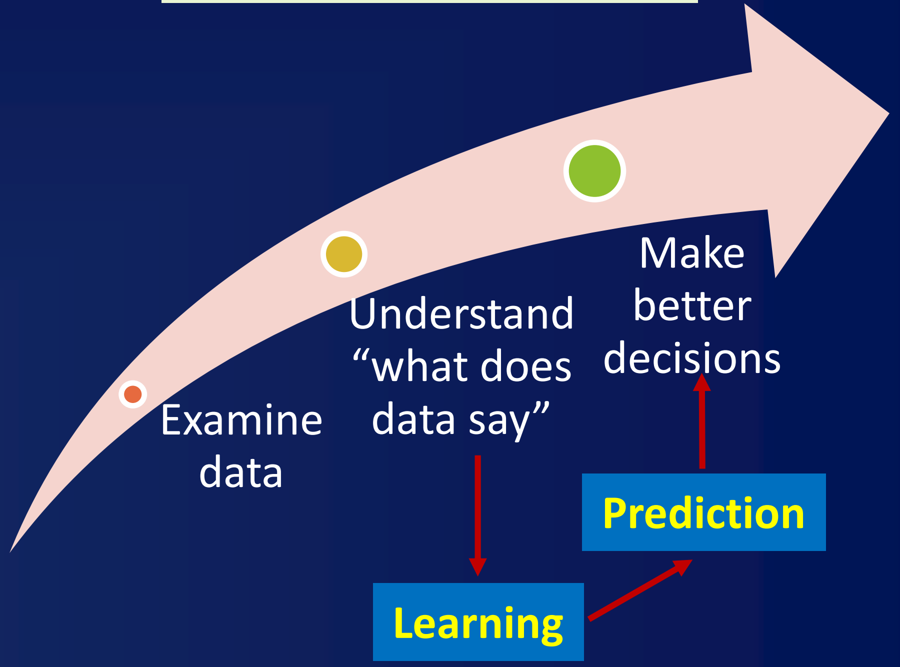
Lessons Learnt

Big Data Analytics – What & Why?

Big Data



Data Analytics



Data Analytics (*aka* Machine Learning, Data Mining) helps understand <u>hidden patterns and relationships</u> in large, complex datasets

Scope of Big Data and Analytics

Data Organization & Management

- data collection, warehousing, tagging, QA/QC, normalization, integration and extraction

Analytics & Knowledge Discovery

- software-driven analysis, predictive model building, and extraction of data-driven insights

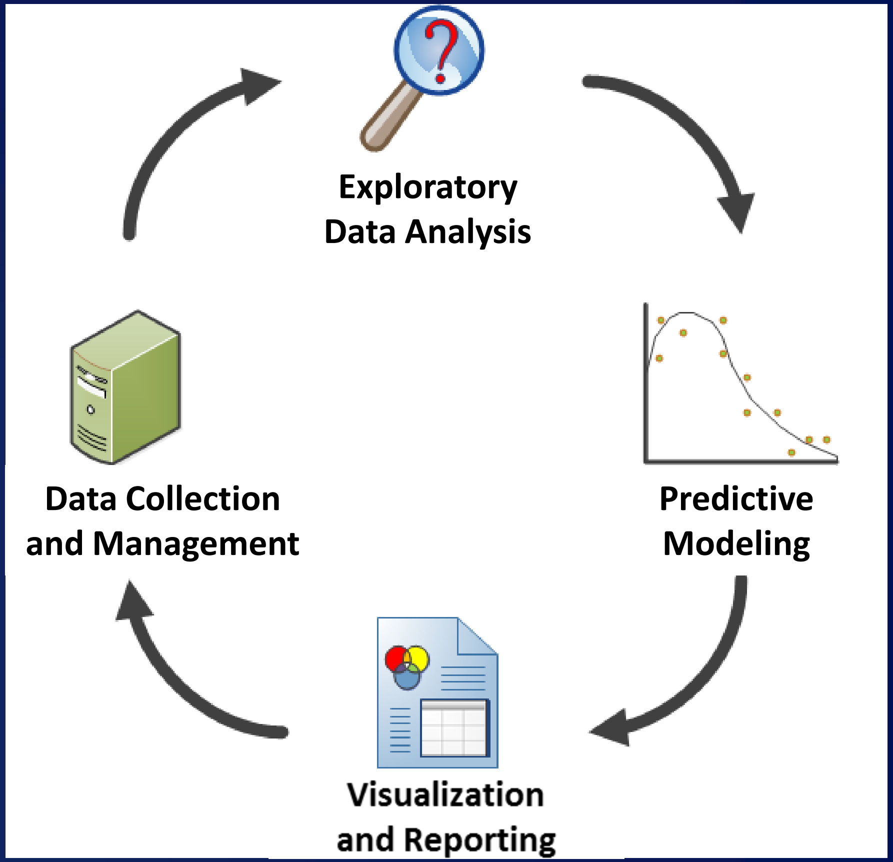
Decision Support & Automation

- rule-based systems with functionality to support collaboration and scenario / risk evaluation

Source: IDC Energy Insights

Data Analysis Cycle

- **Data Collection and Management**
  - Combine data from multiple sources
  - Clean and prepare data
  - Make data easily available for analysis
- **Exploratory Data Analysis**
  - Better understand relationships
  - Formulate questions
- **Predictive Modeling**
  - Explicitly model relationships
  - Use models to answer the questions
- **Visualization and Reporting**
  - Summarize what has been learned
  - Transfer information to decision makers
  - Identify new data to collect



Why Machine Learning?

- Benefits of machine learning:
  - Identify hidden patterns in data
  - Capture non-linear relationships between variables
  - Avoid explicitly defining variable transformations
  - Automatically handle correlation between predictors
  - Guided/automated tuning of model
- Some degree of interpretability lost due to model complexity

Fitting models to data

Assessing quality of fit

Identifying key variables

How to Fit Models?

Regression & Classification Tree

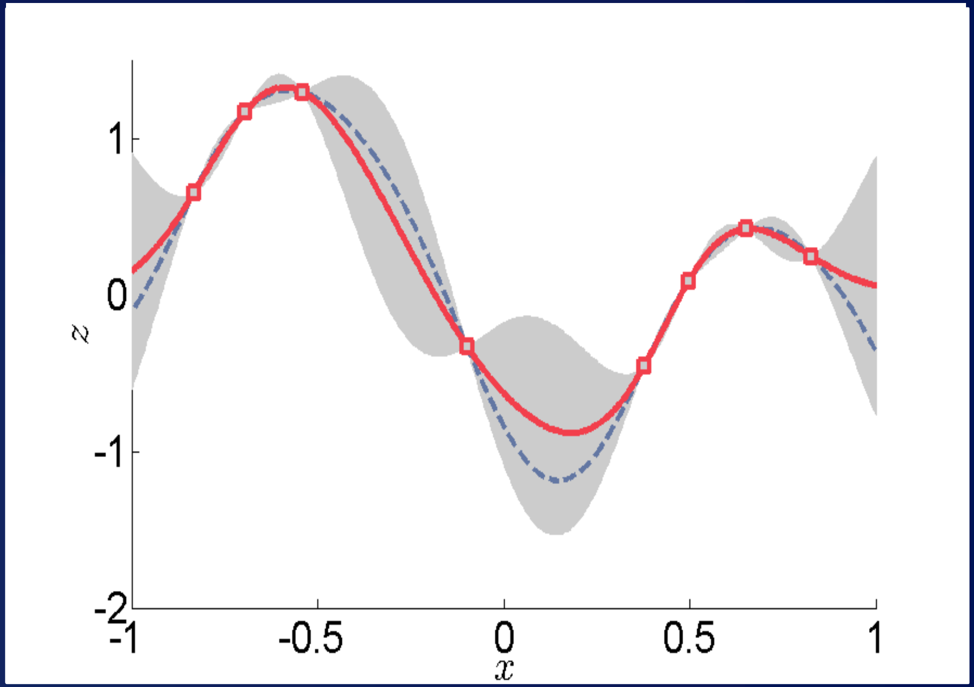
Random Forest

Gradient Boosting Machine

Support Vector Machine

Artificial Neural Network

Gaussian Process (Kriging)



Multidimensional interpolation considering trend and autocorrelation structure of data

How to Assess Quality of Fit?

SPE DISTINGUISHED LECTURER

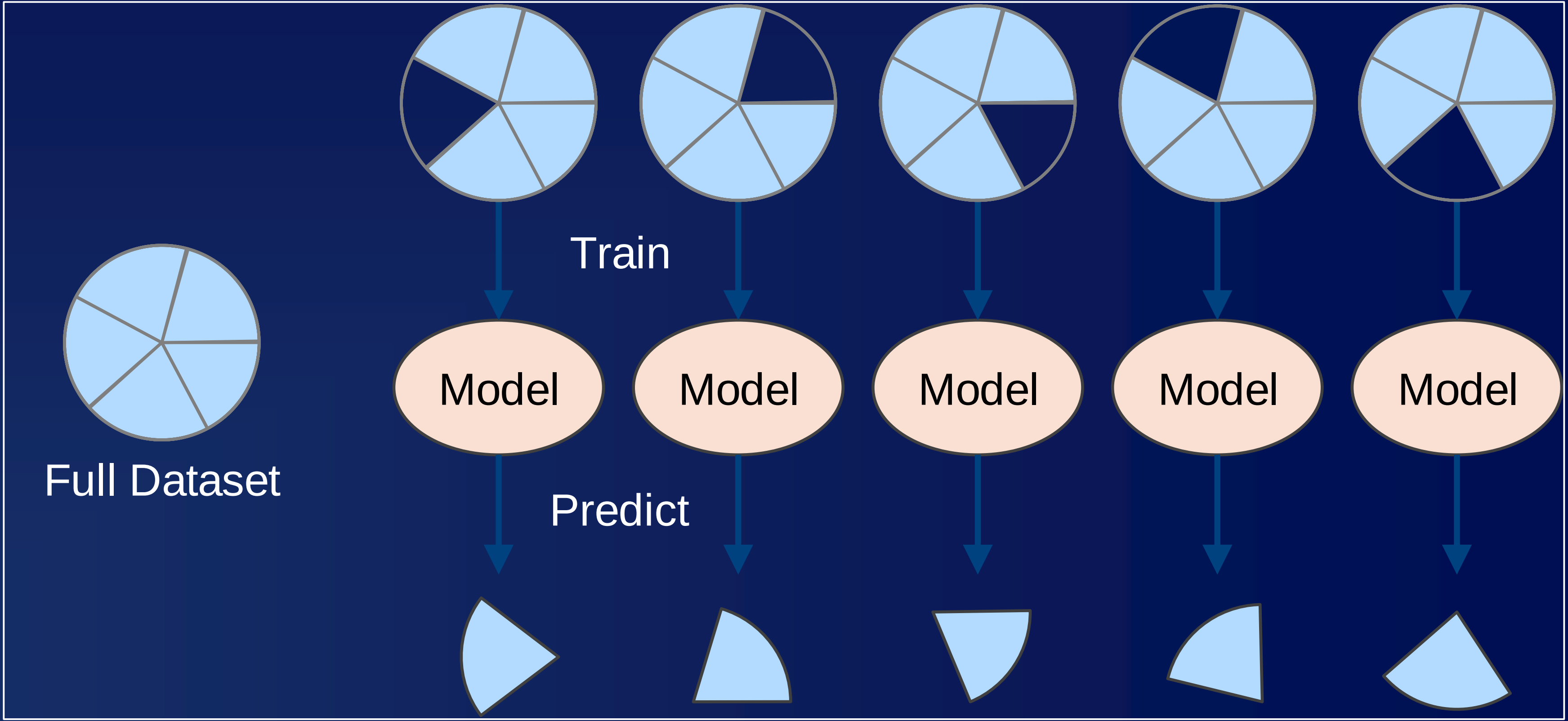
Metrics

$$AAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

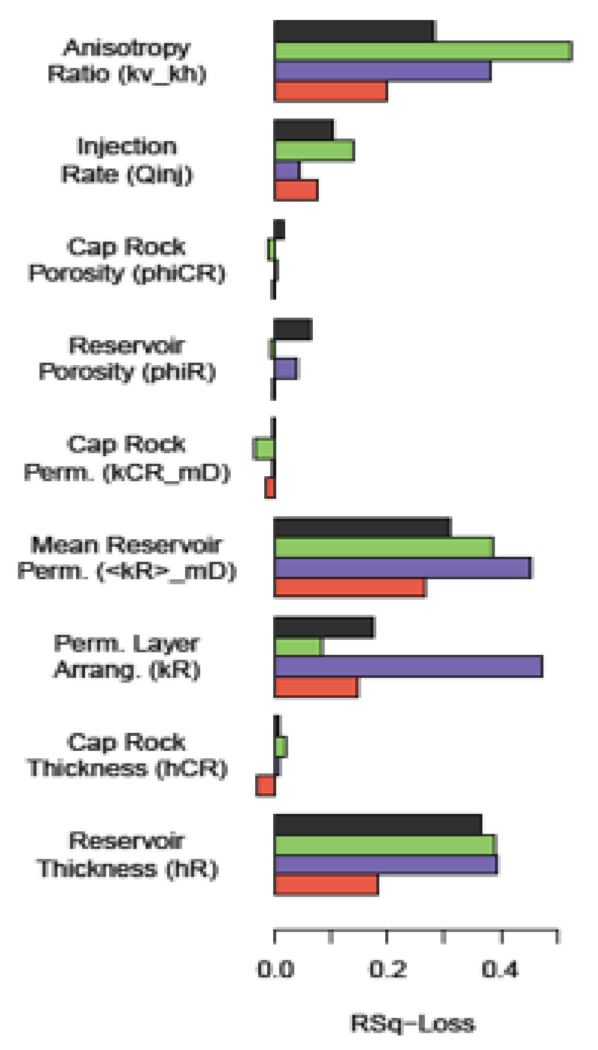
$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

k-fold Cross Validation



How to Identify Key Variables?

- Identification of variable importance can be model specific (e.g., for RF, GBM)
- Model independent metric based on R^2-loss
  - [R^2 for full model] minus [R^2 for model without predictor of interest]
  - larger R^2–loss ⇨ greater influence



Outline of Talk

Basic Concepts

Case Studies

Lessons Learnt

Example Applications

SPE DISTINGUISHED LECTURER

- **Regression** ⇨ Explaining production from shale oil wells in terms of completion and well attributes
- **Classification** ⇨ Identifying advanced log outputs (e.g., vug v/s no vug zones) using basic well log attributes
- **Proxy modeling** ⇨ Fitting statistical response surface to mimic output of full-physics model (reservoir simulator)

Example [1] – Key Factors Affecting Hydraulically Fractured Well Performance

SPE DISTINGUISHED LECTURER

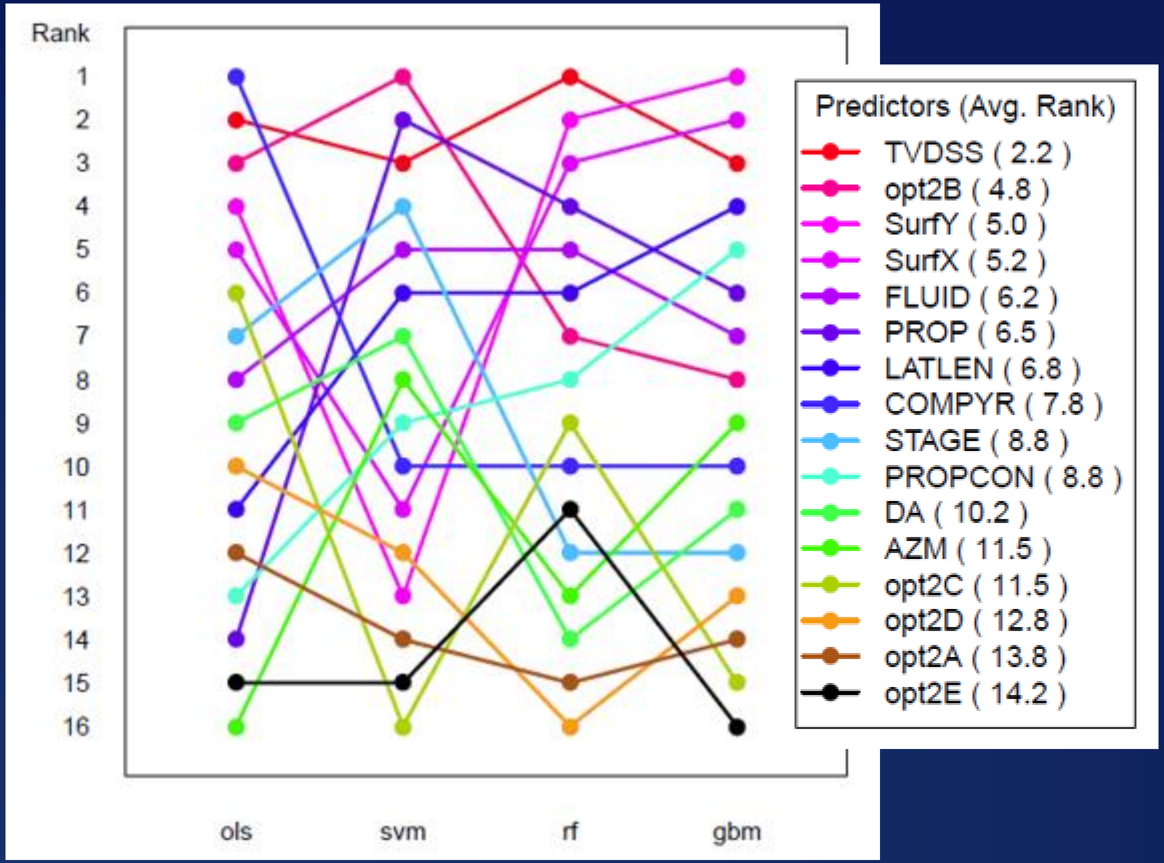
- Wolfcamp Shale horizontal wells
  - Data from 476 Wells
  - **Goal** ⇨ Fit M12CO ~ ***f*** (12 predictors)
  - Multiple machine learning methods
  - Model validation + variable importance

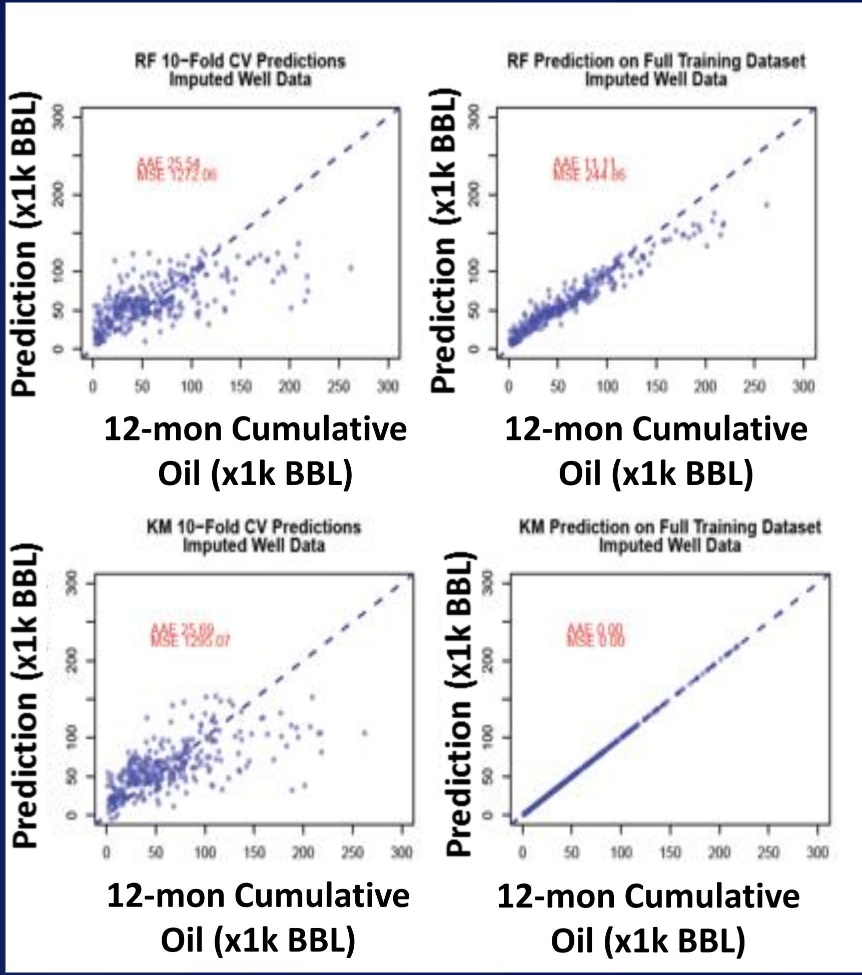
Field	Description
M12CO	Cum. production of 1st 12 producing months (BBL)
Opt2	Categorized operator code
COMPYR	Well completion year
SurfX, SurfY	Geographic location
AZM	Azimuth angle
TVDSS	True vertical depth (ft)
DA	Drift angle
LATLEN	Total horizontal lateral length (ft)
STAGE	Frac stages
FLUID	Total frac fluid amount (gal)
PROP	Total proppant amount (lb)
PROPCON	Proppant concentration (lb/gal)

Schuetter, Mishra, Zhong, LaFolette, 2018, SPEJ, SPE-189969-PA

Variable Importance Using R^2-Loss Metric

Multiple Models Fitted and Validated





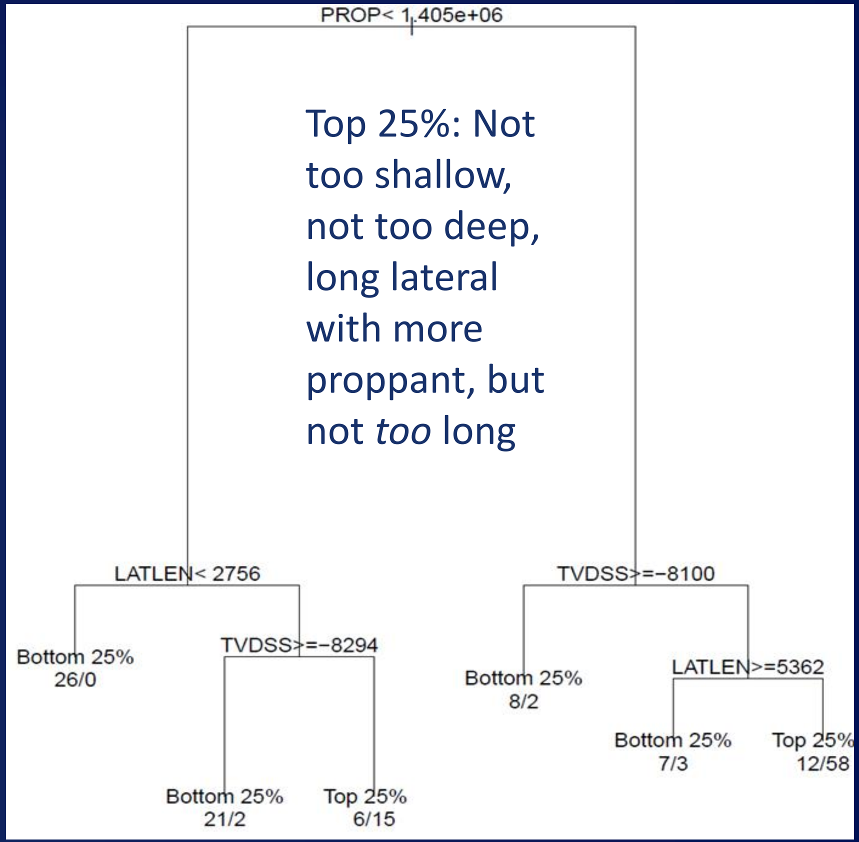
Classification Tree Analysis to Identify Factors Driving Extreme Outcomes

SPE DISTINGUISHED LECTURER

- [Q] What separates top 25% from bottom 25% of producing wells in terms of well productivity?

- Accuracy:

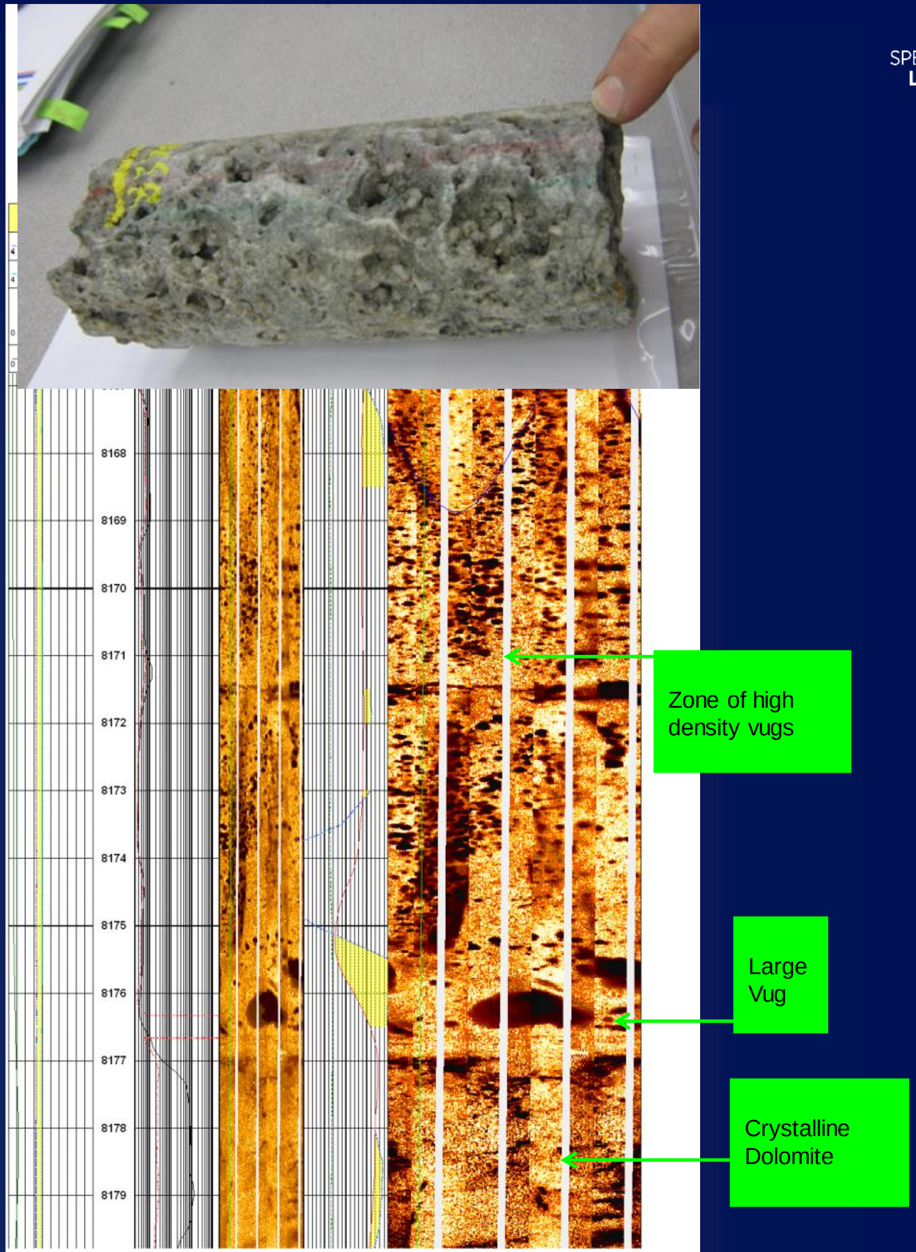
	Bottom 25%	Top 25%	Correct ID
Bottom 25%	62	18	**78%**
Top 25%	7	73	**91%**
Total	69	91	**70%**



Example [2] – Vug Detection from Proxies

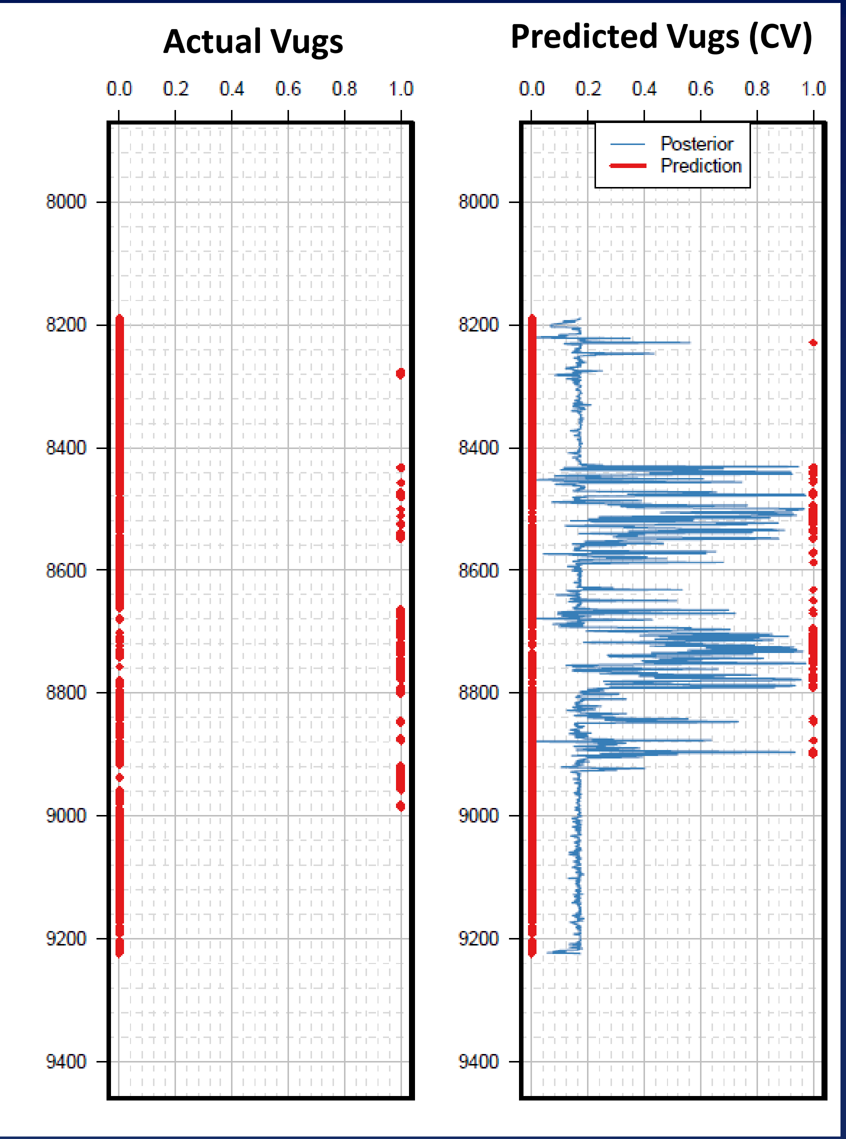
- Vuggy zones create high-permeability pathways in carbonate rocks
- Generally identified from cores and image logs
- **Challenge**: Identify vuggy zones from well-log response (PEF, GR, NPHI, RHOB)
- **Approach**: Use machine learning for classification

Mishra, Howat, Schuetter, Haagsma, 2018, in preparation



SPE DISTINGUISHED LECTURER

Synthetic Vug Log from Triple Combo Data

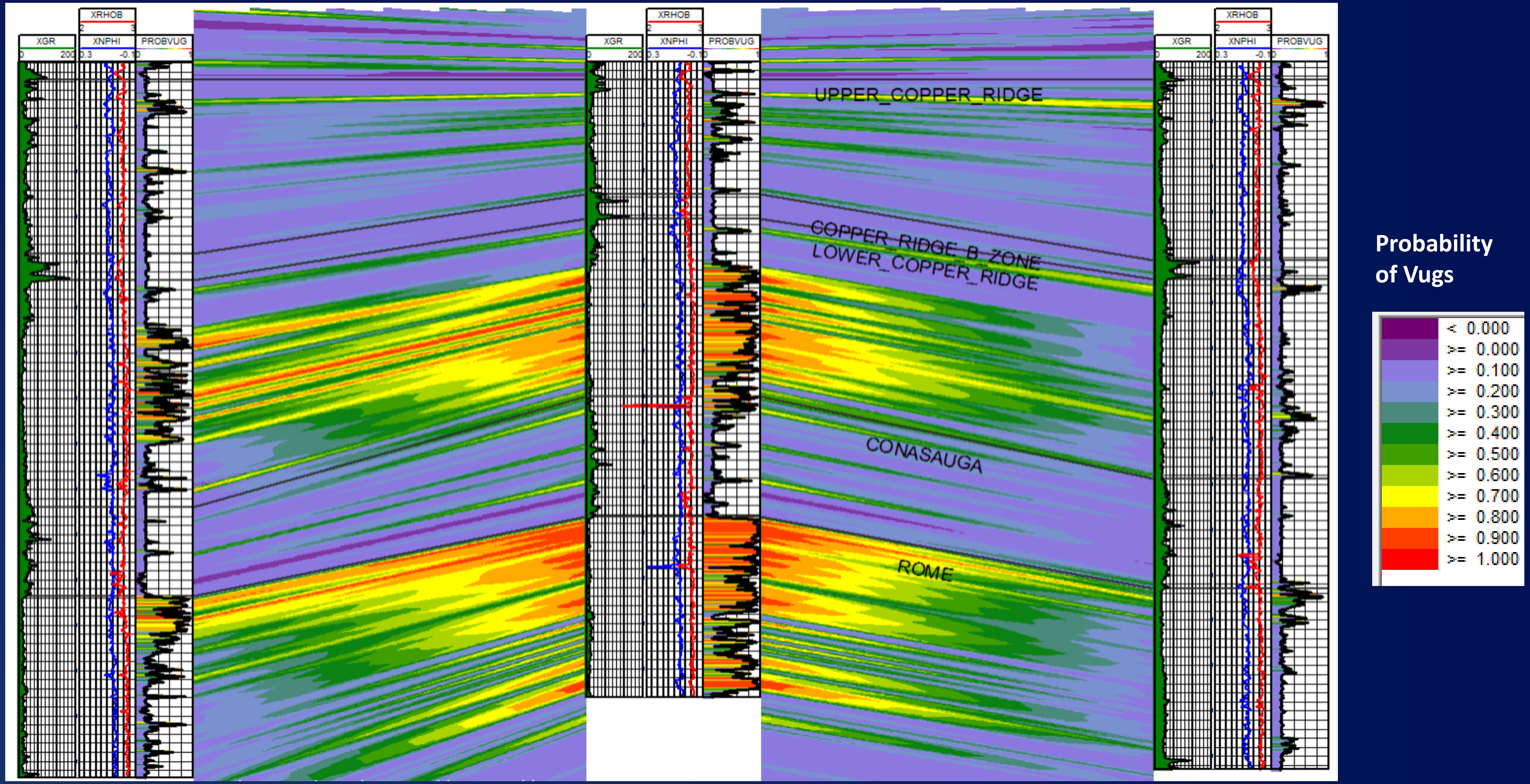


Model Fitting and Validation Process

Held Out Well	Correct ID Rate
Well #1	0.721
Well #2	0.675
Well #3	0.748
Well #4	0.820
Well #5	0.767
Well #6	0.885
Well #7	0.733
Well #8	0.604
Well #9	0.810
Well #10	0.820

SPE DISTINGUISHED LECTURER

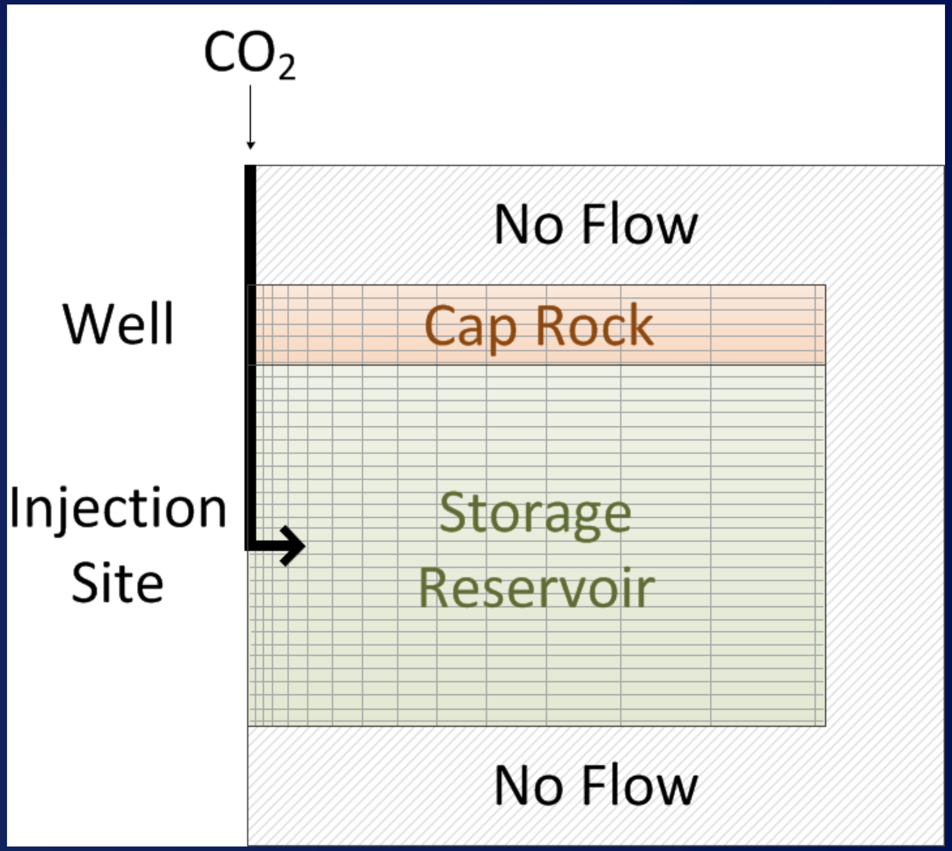
Mapping Vugs in Multiple Wells and Correlating to Well Injectivity



q = 5 bbl/min　　q = 5 bbl/min　　q = 1 bbl/min

Example [3] – Statistical Proxy Modeling for Reservoir Simulation

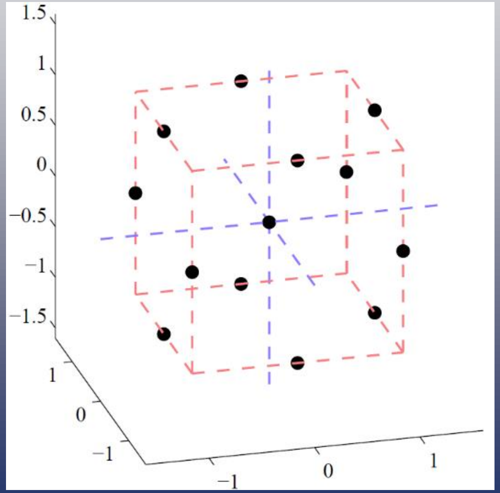
- **Goal** ⇨ Fit fast/accurate response surface to output of full-physics model
- **Problem** ⇨ CO_2 injection into deep saline aquifer
- 9 uncertain inputs
  - Reservoir and caprock k, h, ϕ
  - q, k_h/k_v, k-layering
- 3 responses (E_s, R_{CO2}, P_{avg})



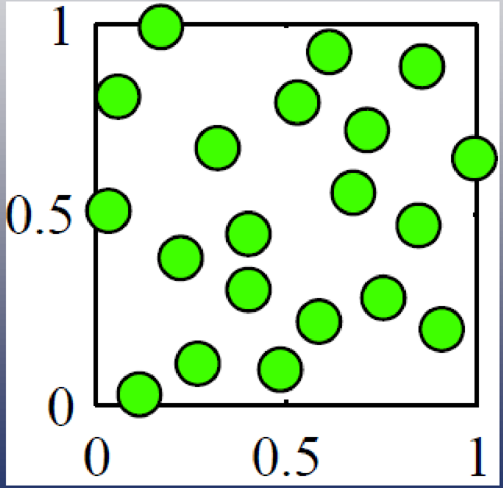
Schuetter and Mishra, **2015, SPE 174905**

Comparing Designs (Discrete Model Run Points)

Box-Behnken (BB)
inputs sampled using -1, 0, +1



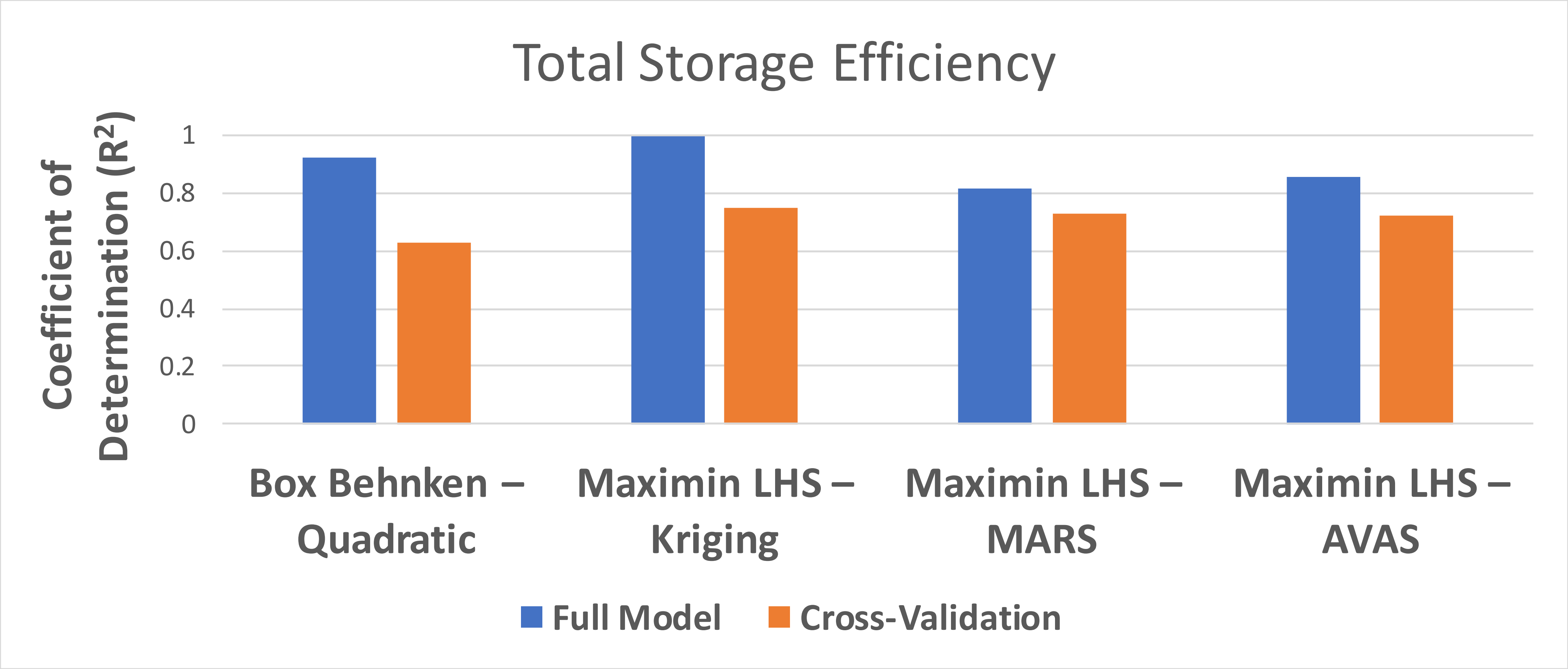
Maximin LHS (MM)
sampling using equi-probable bins



- BB common statistical design - number of runs ↑↑ for n>10
- Higher granularity and space-filling properties for MM design
- More flexibility for model fitting with MM (beyond *quadratic*)
  - Kriging – MARS – AVAS
  - Also RF, GBM, SVM, ANN etc.

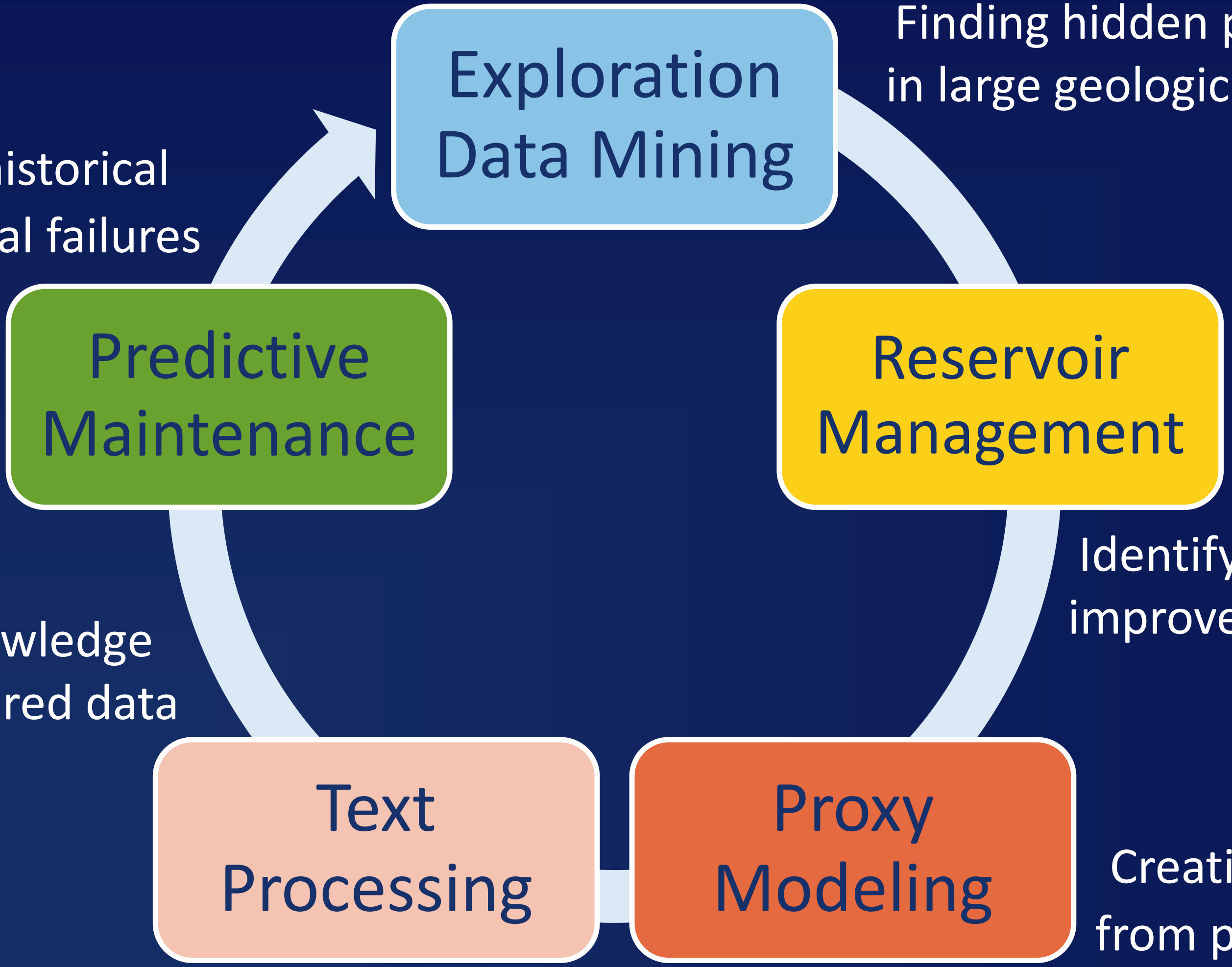
97 sample BB and MM designs for 9 factors

Comparing Model Performance



Better model fits with Maxmin LHS designs (more flexibility)

Other Recent Examples



Finding hidden patterns in large geologic datasets

Identifying factors for improved performance

Creating fast “emulators” from physics-based models

Extracting knowledge from unstructured data

Using real-time and historical data to predict potential failures

Example [1]

Perez et al.

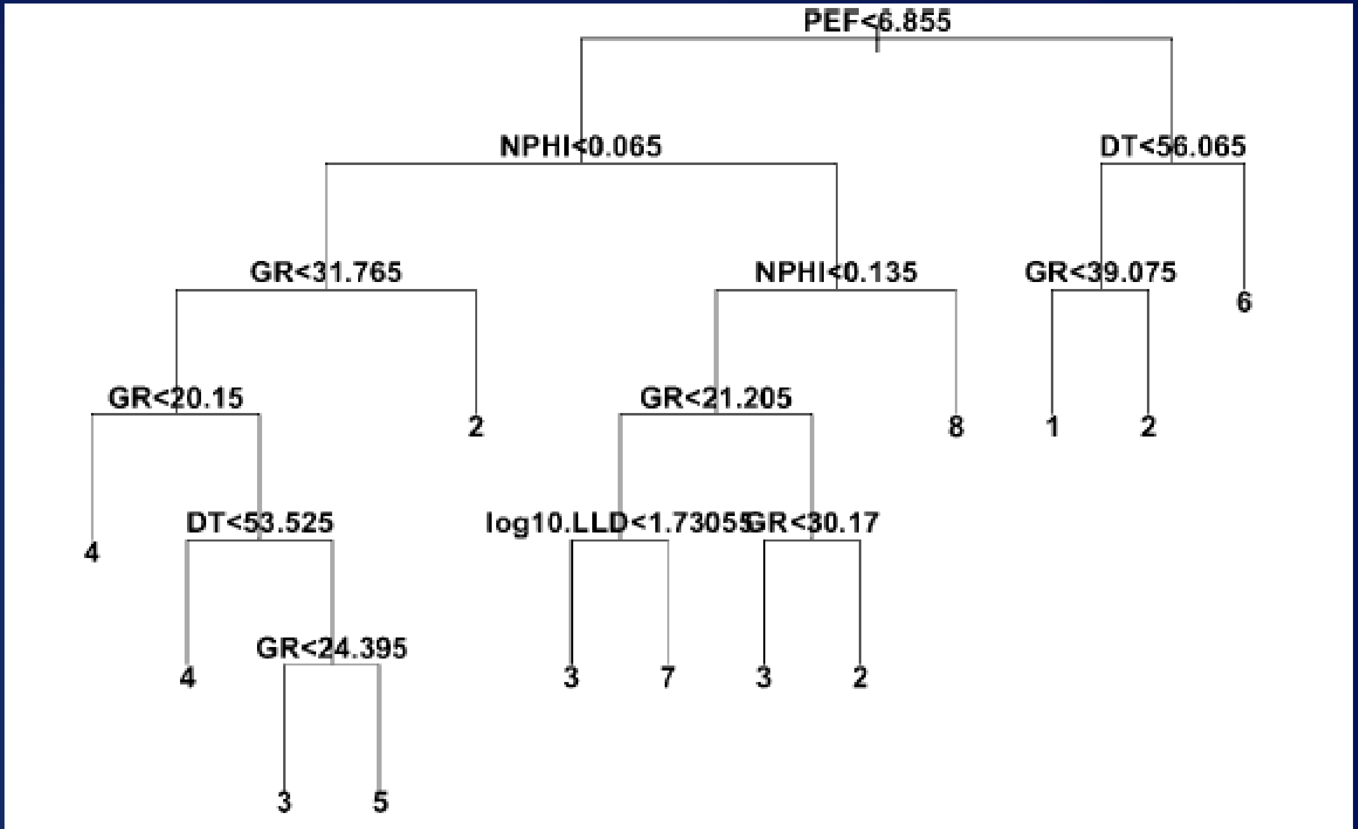
SPERE April 2005

The Role of Electrofacies, Lithofacies, and Hydraulic Flow Units in Permeability Prediction From Well Logs: A Comparative Analysis Using Classification Trees

Hector H. Perez,* SPE, and **Akhil Datta-Gupta,** SPE, Texas A&M U., and **S. Mishra,** SPE, Intera Inc.

SPE DISTINGUISHED LECTURER

- Classification tree analysis for identifying rock types from basic well log attributes
- Accounting for missing well logs
- Application for permeability prediction in Salt Creek field



Example [2]

Shelley et al.

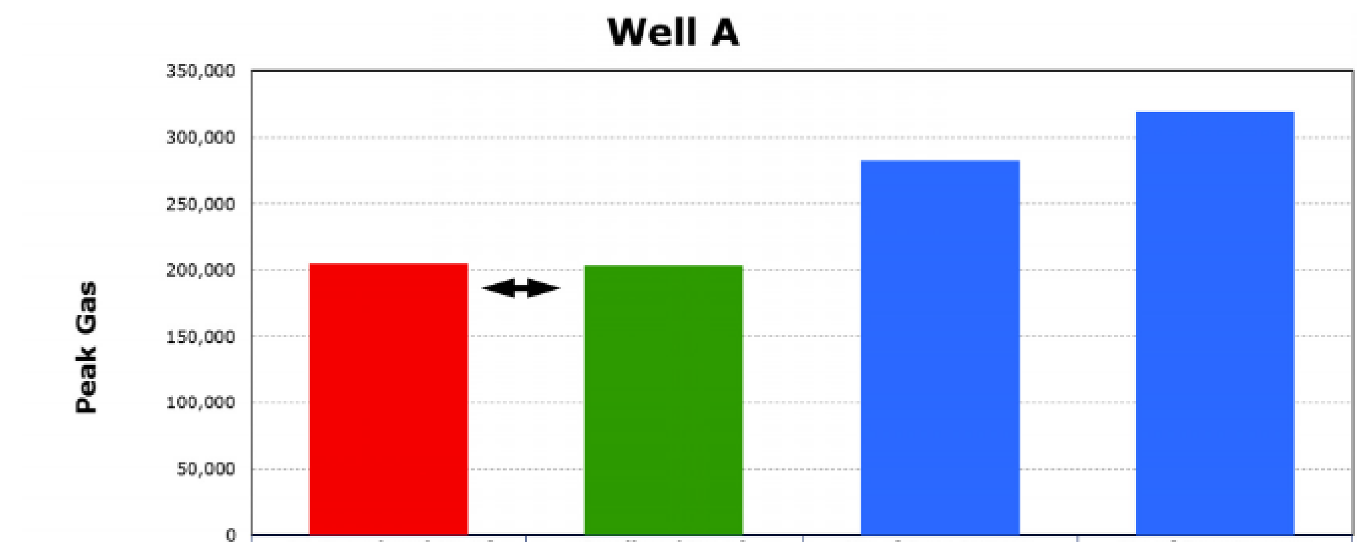
SPE-171003, 2014

SPE-171003-MS

Understanding Multi-Fractured Horizontal Marcellus Completions

Robert Shelley, Amir Nejad, and Nijat Guliyev, StrataGen; Michael Raleigh, and David Matz, Epsilon Energy USA, Inc.

- Identifying performance drivers and completion effectiveness for Marcellus shale wells
- Predictive model using ANN (Artificial Neural Networks)
- Role of different variables evaluated



	As Completed No of Frac Stages: 10	Predicted No of Frac Stages: 10	No of Frac Stages: 15	No of Frac Stages: 15
Peak Gas	204,557	203,061	282,190	318,832
%Change			39%	57%
Proppant Mass (lb)	6,117,532	6,117,532	6,117,532	9,176,298
Fluid Volume (bbl)	116,832	116,832	116,832	175,248

Example [3]

Guerrilot et al.

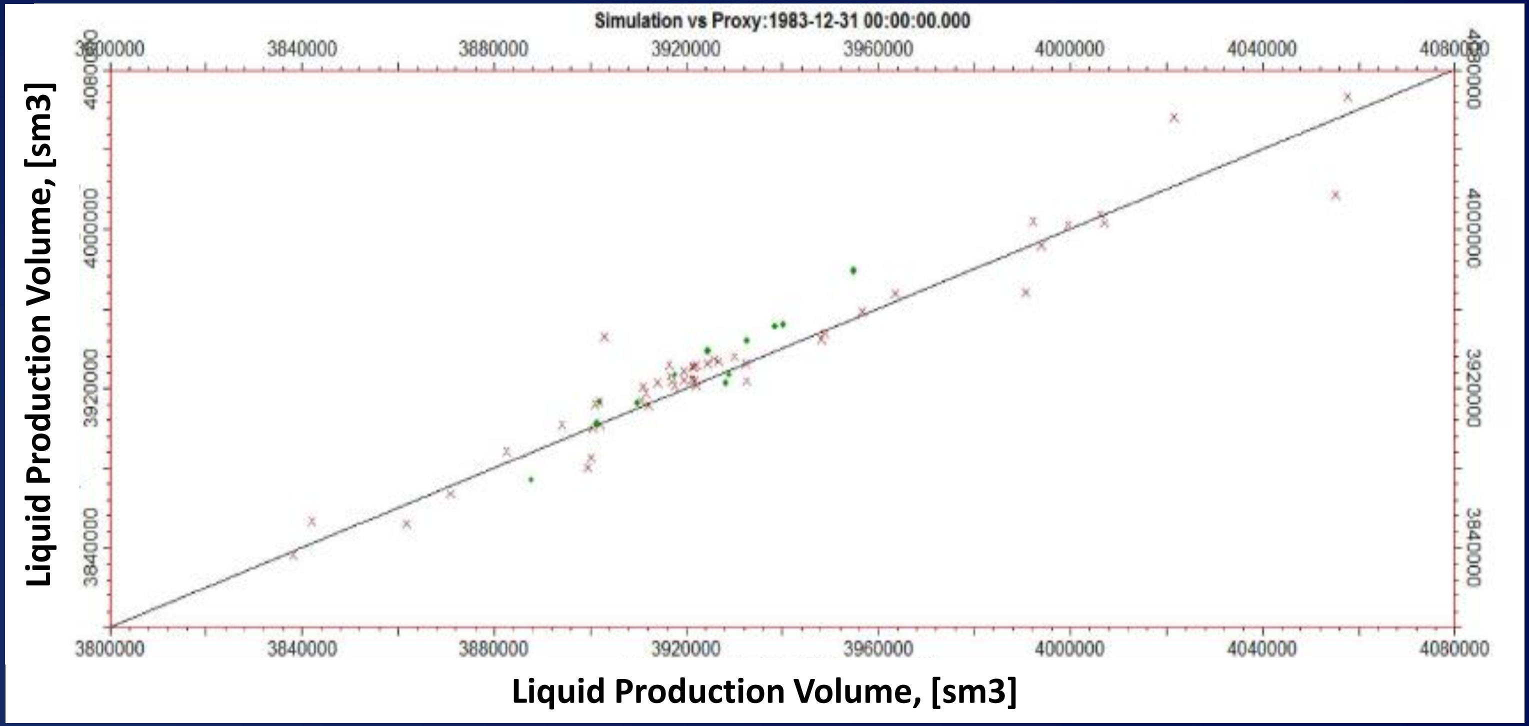
SPE-183921, 2017

SPE-183921-MS

Uncertainty Assessment in Production Forecast with an Optimal Artificial Neural Network

D. R. Guérillot, Texas A&M University; J. Bruyelle, Terra 3E

- Building proxy model for synthetic reservoir using simulator output
- 6 facies each with 3 fitted parameters (ϕ, k_h, k_v)
- ANN proxy model better than kriging and quadratic versions for history match
- Probabilistic forecasts



Example [4]

Arumugam et al.

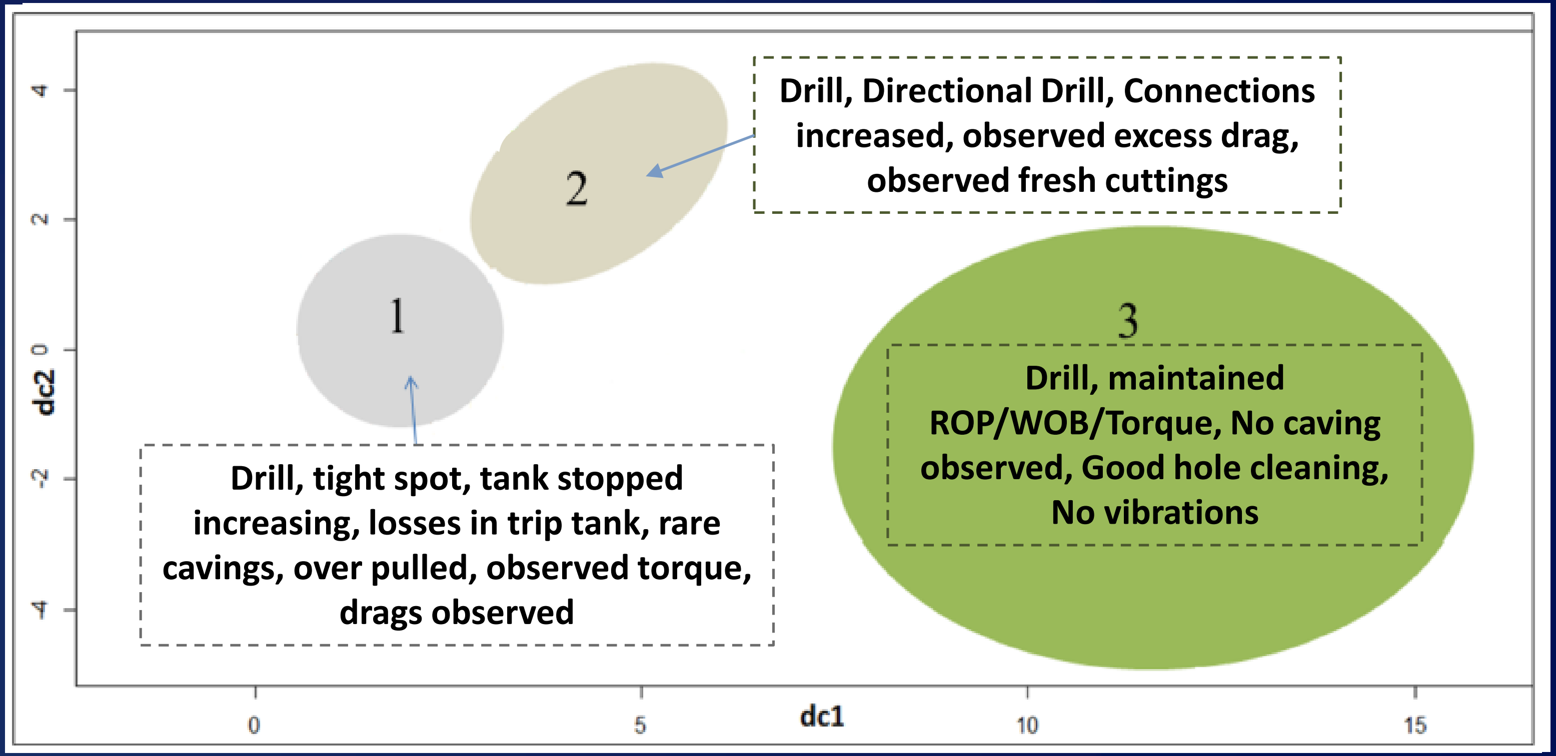
SPE-184062, 2016

SPE-184062-MS

Revealing Patterns within the Drilling Reports Using Text Mining Techniques for Efficient Knowledge Management

Sethupathi Arumugam, Sanjay Gupta, Biswaranjan Patra, Shebi Rajan, and Satyam Agarwal, Infosys Limited

- Processing of daily drilling data to identify drilling anomalies / best practices
  - Information retrieval
  - Conversion to structured data
  - Clustering
  - Pattern identification
  - Knowledge management



Example [5]

Santos et al.

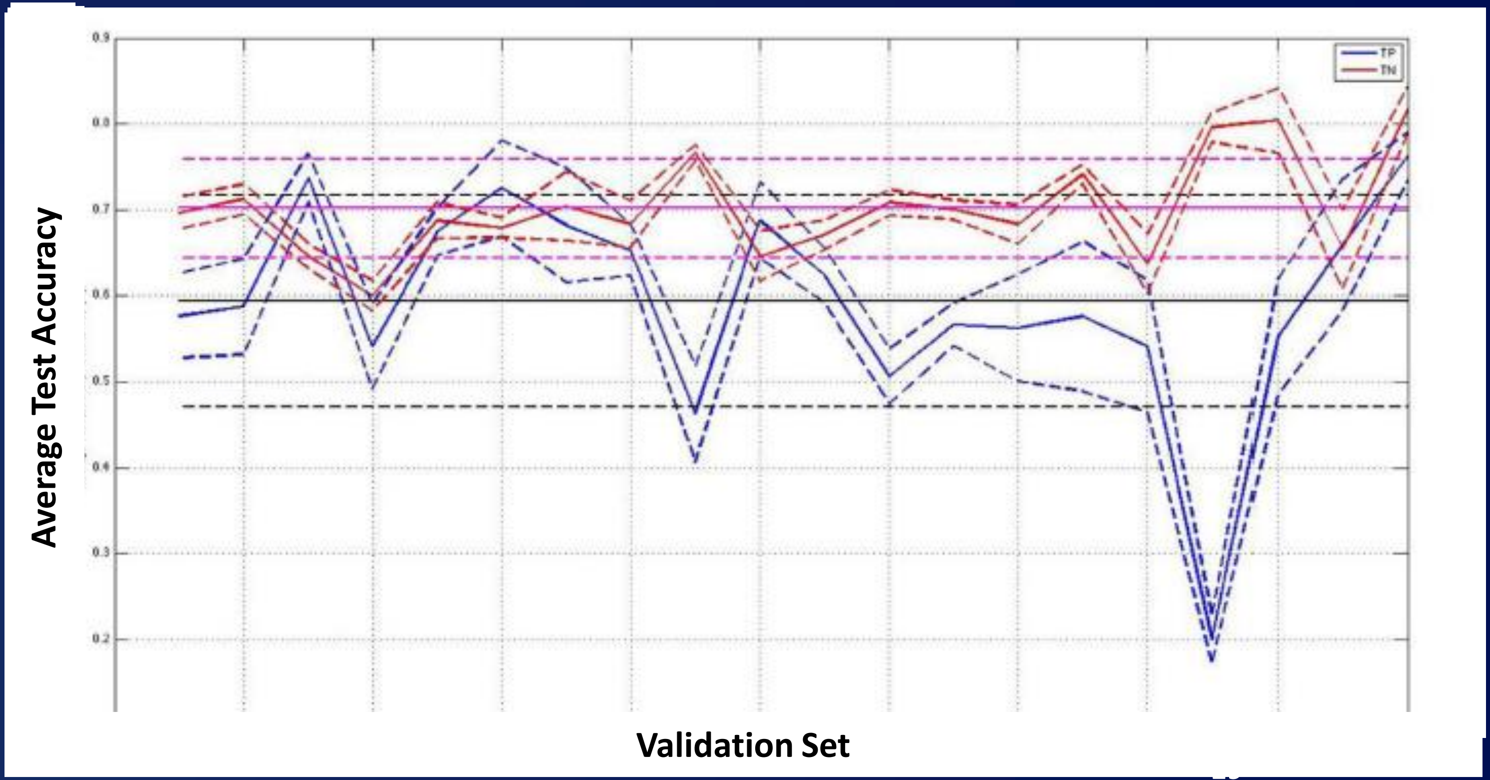
OTC-26275, 2014

OTC-26275-MS

Big Data Analytics for Predictive Maintenance Modeling: Challenges and Opportunities

I. H. F. Santos, M. M. Machado, E. E. Russo, D. M. Manguinho, V. T. Almeida, R. C. Wo, M. Bahia, and D. J. S. Constantino, Petrobras; D. Salomone, M. L. Pesce, C. Souza, and A. C. Oliveira, EMC - Brazil Research Center; A. Lima, J. Gois, L. G. Tavares, T. Prego, S. Netto, and E. Silva, PEE-COPPE / UFRJ.

- Building prognostic classifier for specific turbogenerator failures during startup
- Data from offshore facility – extraction of fuel burning related features
- RUSBoost and RF models
- Multi-fold validation approach for evaluation



Outline of Talk

Basic Concepts

Case Studies

Lessons Learnt

Key Takeaways

- Ensure availability of good and appropriate (causal) data
- Predictive modeling can be nuanced – avoid overfitting!
- Multiple models may provide comparable fits ⇨ aggregate!
- Limited ability of data-driven models to project the “unseen”
- Many resources available
  - Software, training, courseware (R/RATTLE; Python)
  - Rapidly expanding literature (see OnePetro)

Recommended Workflow

- Framing the problem
- Checking the data
- Selecting the causal variables
- Picking the software
- Choosing the modeling technique(s)
- Validating the model
- Understanding and communicating the results

Reduce cost
Improve productivity
Increase efficiency

Looking Ahead

- Machine learning applications in oil & gas rapidly growing
  - exploration and production
  - digital oil field management
  - predictive maintenance
  - natural language processing

- Significant potential for data analytics to provide useful insights (data ⇨ information ⇨ knowledge ⇨ wisdom)

- Petroleum engineers and geoscientists need better understanding of data science fundamentals + applicability + limitations

Reference

- Applied Statistical Modeling and Data Analytics: A Prcatical Guide for the Petroleum Geoscience
  - Srikanta Mishra and Akhil Datta-Gupta (Texas A&M U.)
  - Published by Elsevier, October 2017
- Also, SPE short course (and customized courses for companies) offered on same topic

ACKNOWLEDGMENTS

Battelle Memorial Institute

US DOE – NETL

Jared Schuetter

Thank you for your attention

Your Feedback is Important

Enter your section in the DL Evaluation Contest by completing the evaluation form for this presentation
Visit SPE.org/dl

Society of Petroleum Engineers
Distinguished Lecturer Program
www.spe.org/dl