Application of Machine Learning for Production Data Analysis:

Premises, Promises & Perils

Dr. Srikanta Mishra

BATTELLE

2022 SPEE Conference, Napa, CA

June 13, 2022

Outline of Talk

Basic Concepts

Premises

Case Studies

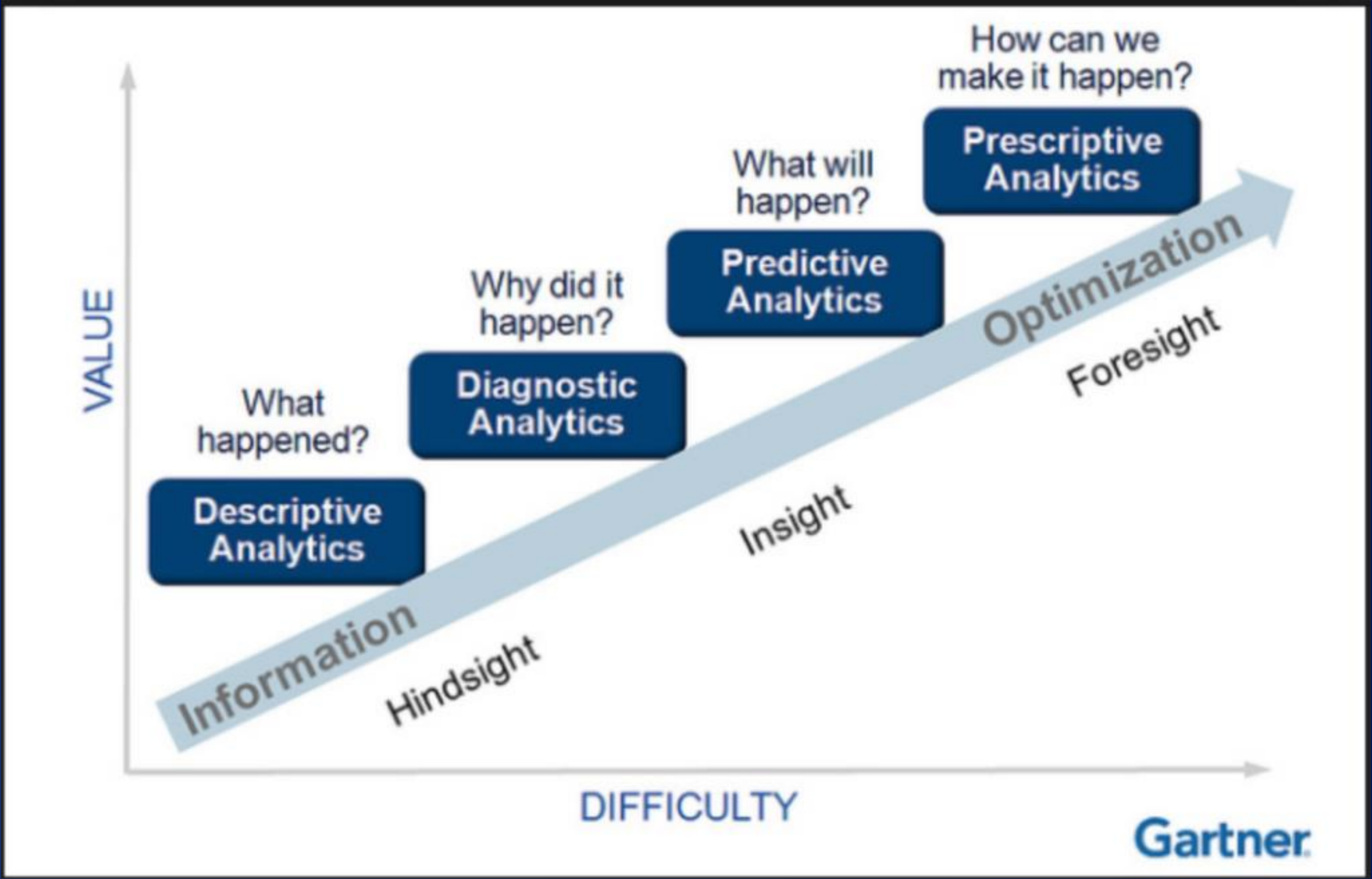
Lessons Learnt

A Few Definitions

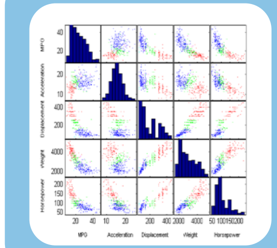
- ***Data analytics*** (DA) – sophisticated data collection + analysis to help understand hidden patterns and relationships
- ***Machine learning*** (ML) – building a model between predictors and response (often with a "black-box" algorithm)
- ***Artificial intelligence*** (AI) – applying predictive model with new data to make decisions without human intervention

***Mishra et al., 2021*, JPT (March), 25-30.**

Types of Analytics

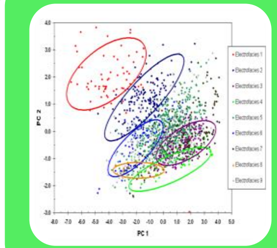


Predictive Analytics Process



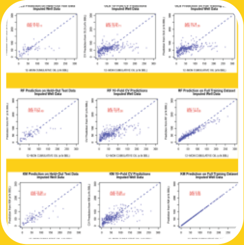
Exploratory Data Analysis

- Patterns, trends, outliers, imputation
- Scatter-plot matrix, trellis plots



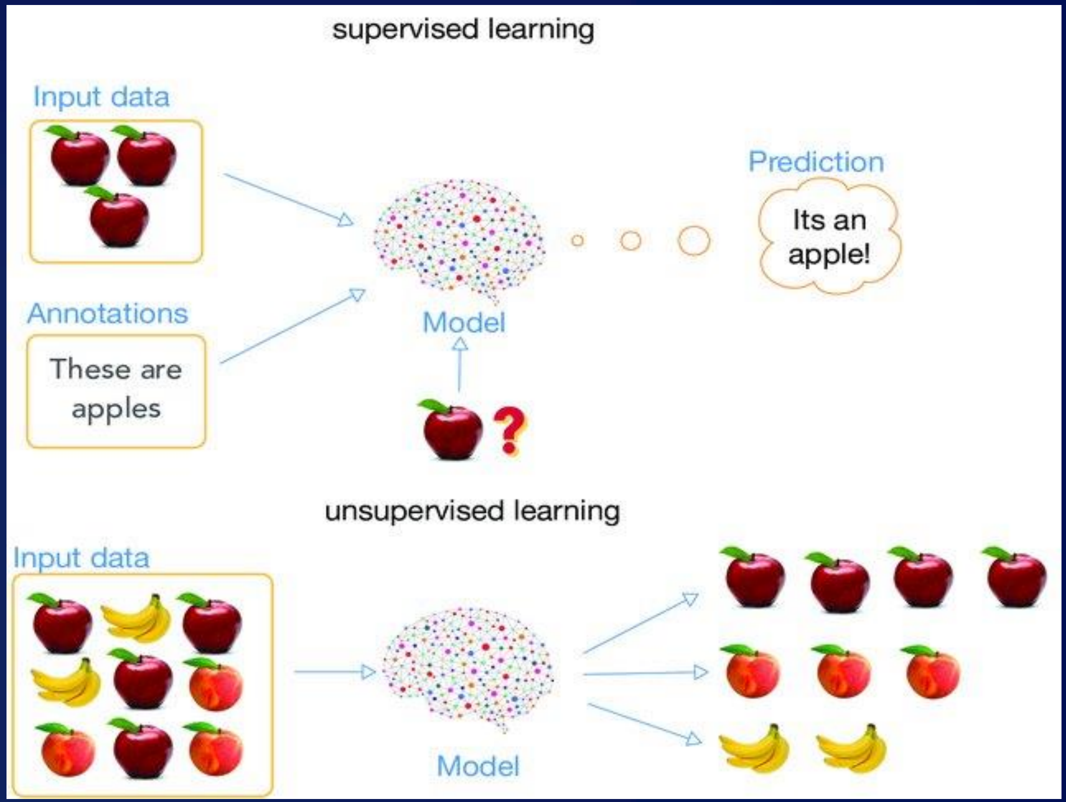
Unsupervised Learning

- Data reduction and clustering
- PCA, k-means, hierarchical methods



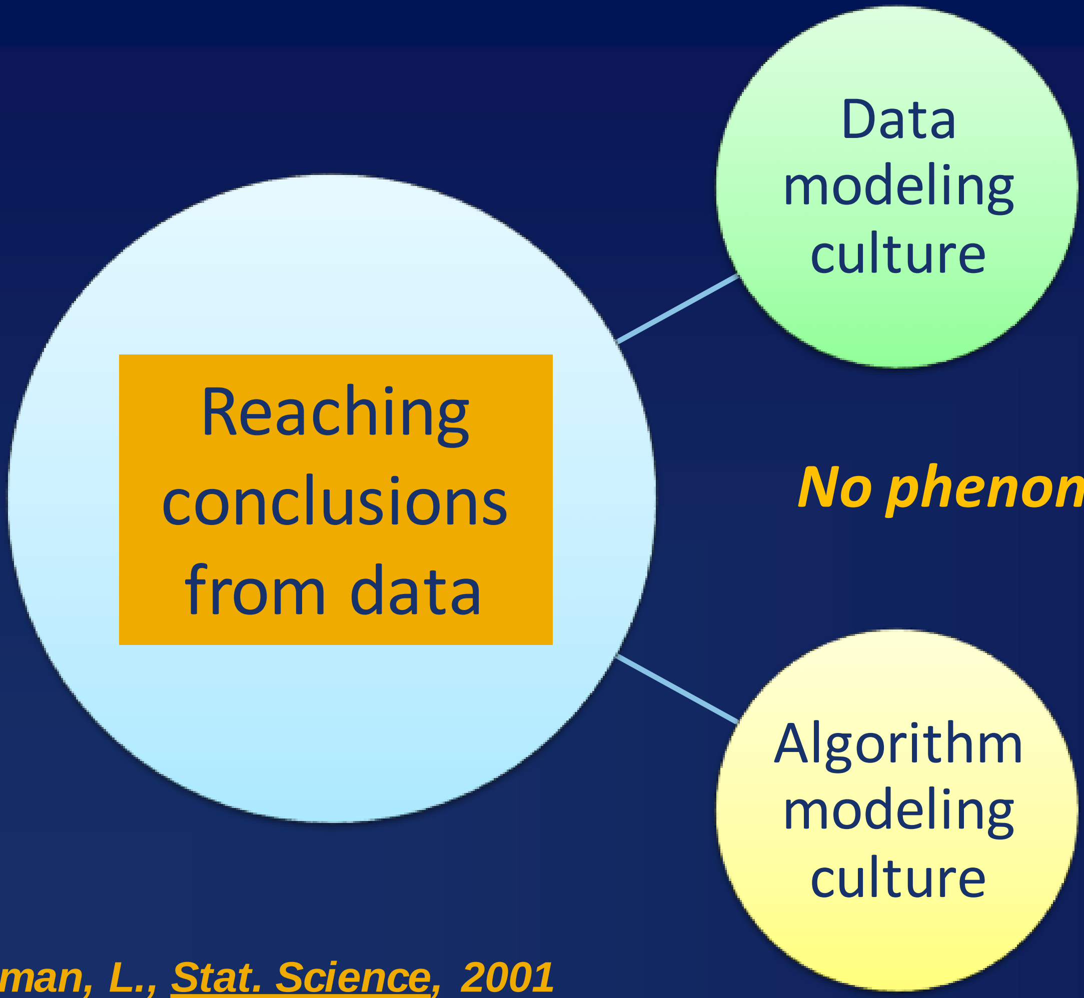
Supervised Learning

- Regression and classification
- Random Forest, ANN, kNN



Ma et al., 2018, Symmetry, 10, 734

Statistical Modeling v/s Machine Learning



- Assume some stochastic data model, estimate its parameters and apply model

No phenomenological models available

- Assume model form to be unknown, and generate fitting function from data

Breiman, L., <u>Stat. Science</u>, 2001

Observations on Where Things Stand

- Two tracks (state of practice) on Machine Learning
  - Significant self-learning and upskilling from technical staff
  - Fear, uncertainty and doubt from decision makers
- Some questions to ponder/discuss
  - Why ML models, and when
  - Mechanics of data-driven modeling
  - Predictive modeling approaches
  - ML-based workflow

Mishra et al., 2021, JPT (March), 25-30.

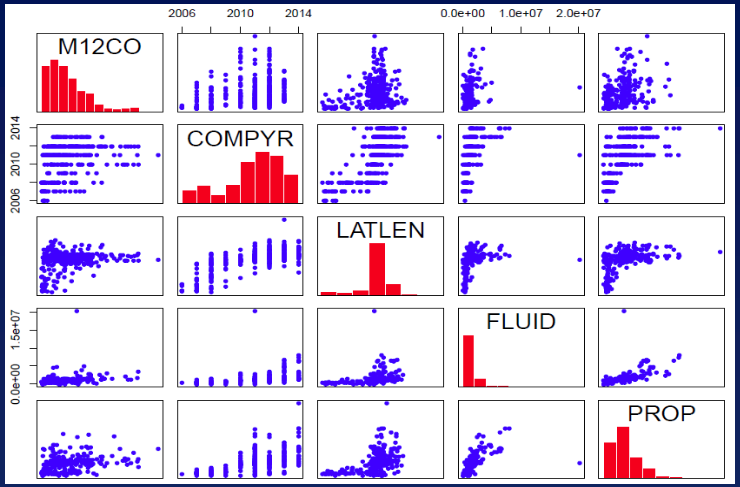
Why ML Models and When?

- Historically, subsurface science and engineering analyses have relied on mechanistic (physics-based) models
- Incorporation of causal input-output relationship
- Experienced professionals are wary of purely data-driven "black-box" ML models that lack such understanding
- Nevertheless, the use of ML models is easy to justify - if
  - relevant physics-based model is computation intensive and/or immature
  - suitable mechanistic modeling paradigm does not exist

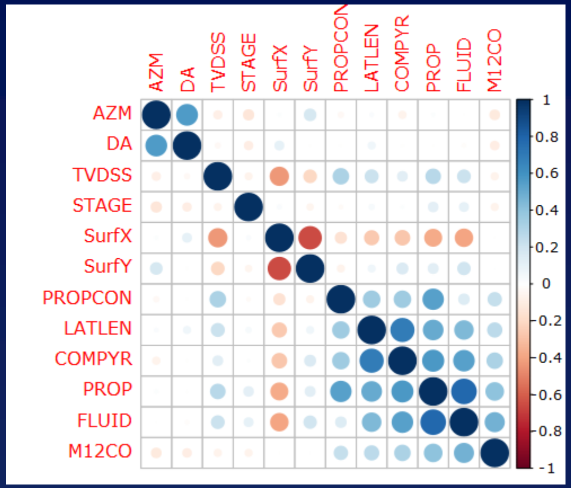
Rationale for Data-Driven Models

- Mechanistic modeling in unconventional reservoirs complex
  - fluid flow in a network of induced and natural fractures
  - coupled processes such as geomechanical effects, water blocking, non-Darcy flow in nano-scale pores, adsorption/desorption etc.
  - robust and computationally-efficient physics-based modeling frameworks and software tools under continued development
- Empirical models (e.g., decline curves) popular alternative but have many limitations (model form, parameterization)
- Data-driven models are emerging as alternative approach (*let the "machine" learn about the system from the data*)

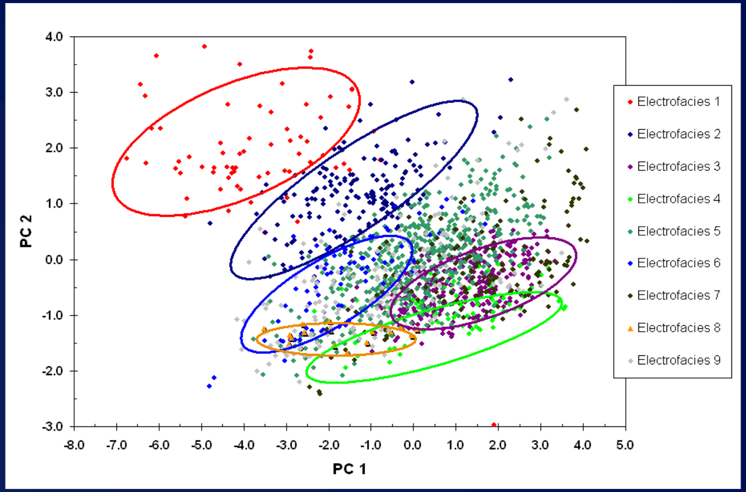
Mechanics of Data-Driven Modeling



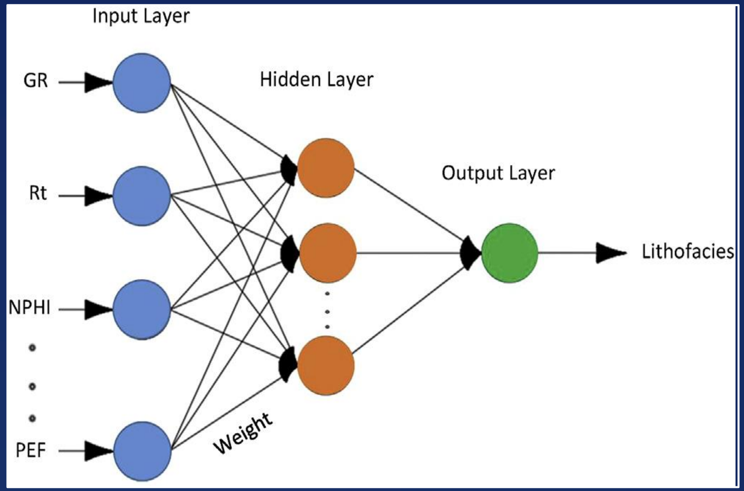
Exploratory Data Analysis



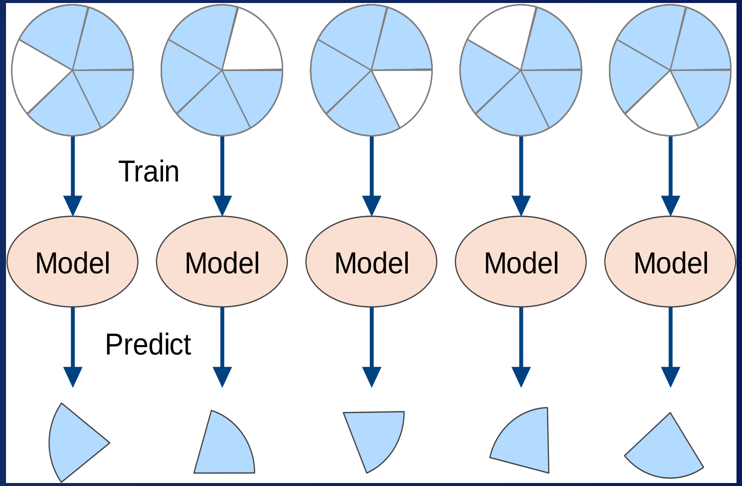
Feature Selection



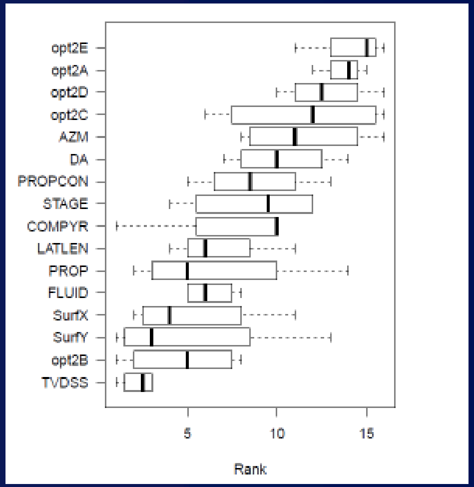
Multivariate Analysis



Model Building



Cross-Validation



Variable Importance

Mishra 2022 SPEE

Predictive Modeling Approaches

Regression & Classification Tree

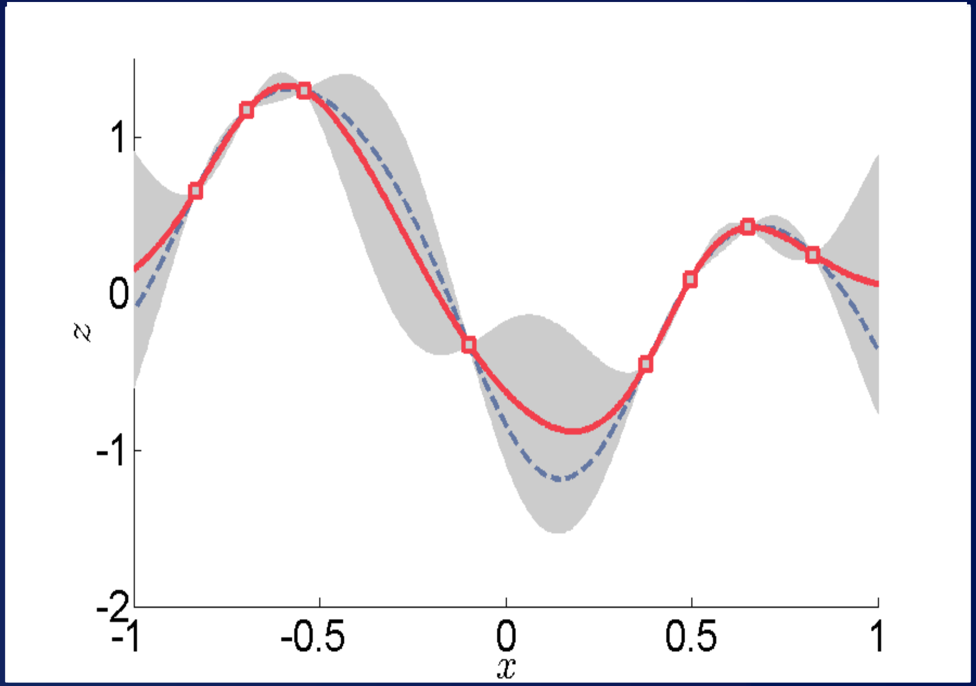
Random Forest

Gradient Boosting Machine

Support Vector Machine

Artificial Neural Network

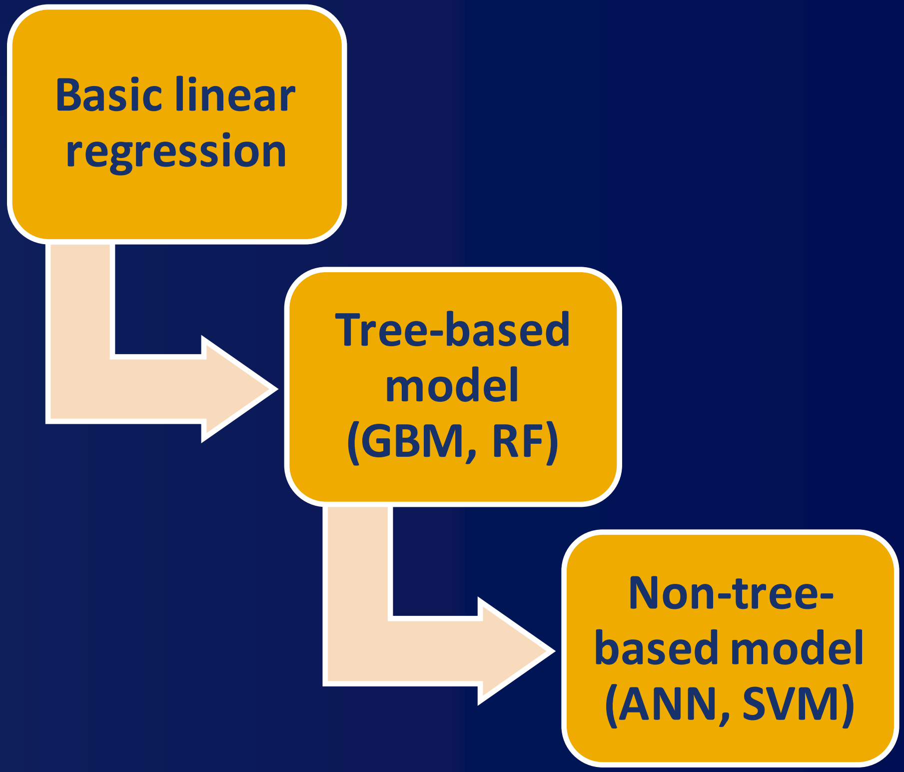
Gaussian Process (Kriging)



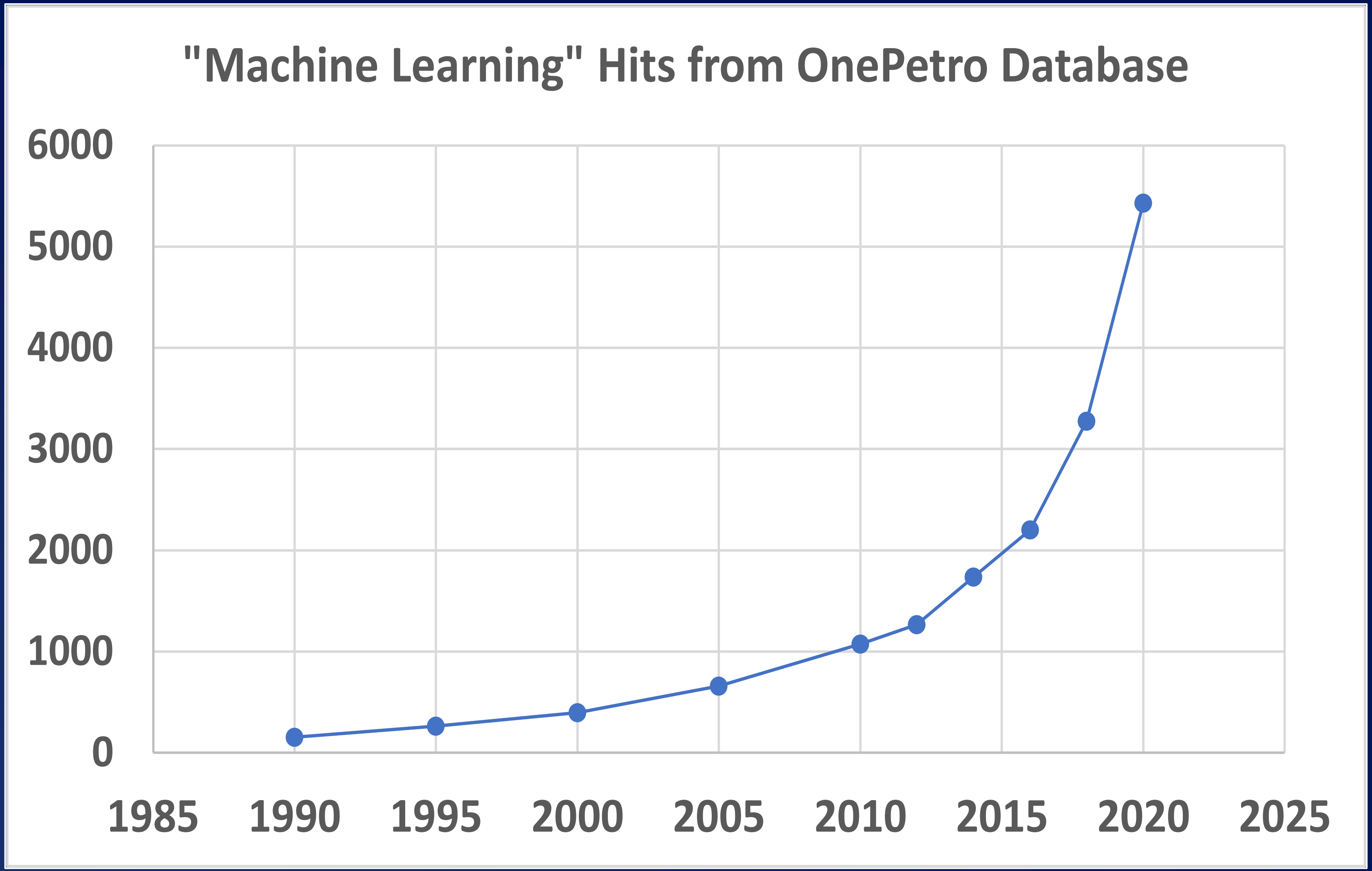
Multidimensional interpolation considering trend and autocorrelation structure of data

ML-Based Workflow (Analysis, Review)

- Framing the problem
- Selecting causal variables
- Checking data quality
- Fitting model(s) and aggregating
- Validating model(s)
- Identifying key variables
- Communicating results



Exponential Growth in O&G ML Applications



Outline of Talk

Basic Concepts

Case Studies

Promises

Lessons Learnt

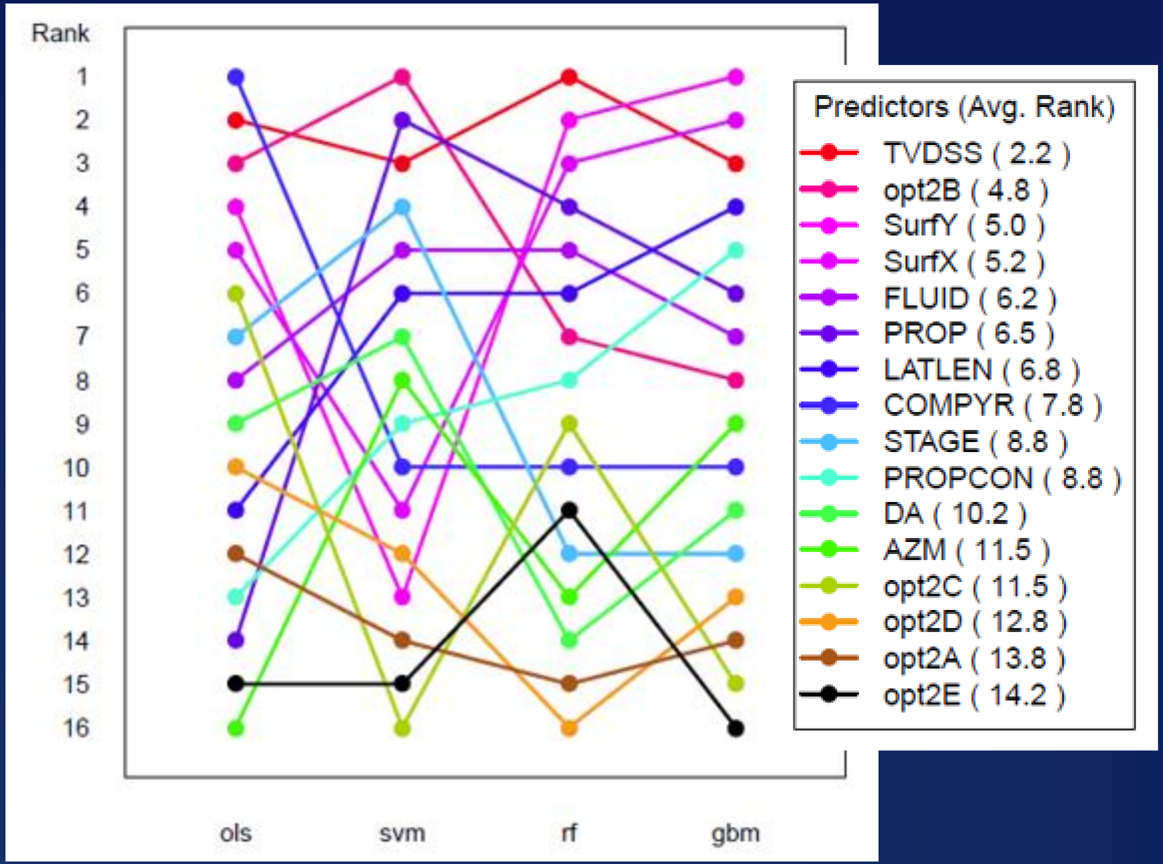
Case Study [1] – Key Factors Affecting Hydraulically Fractured Well Performance

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

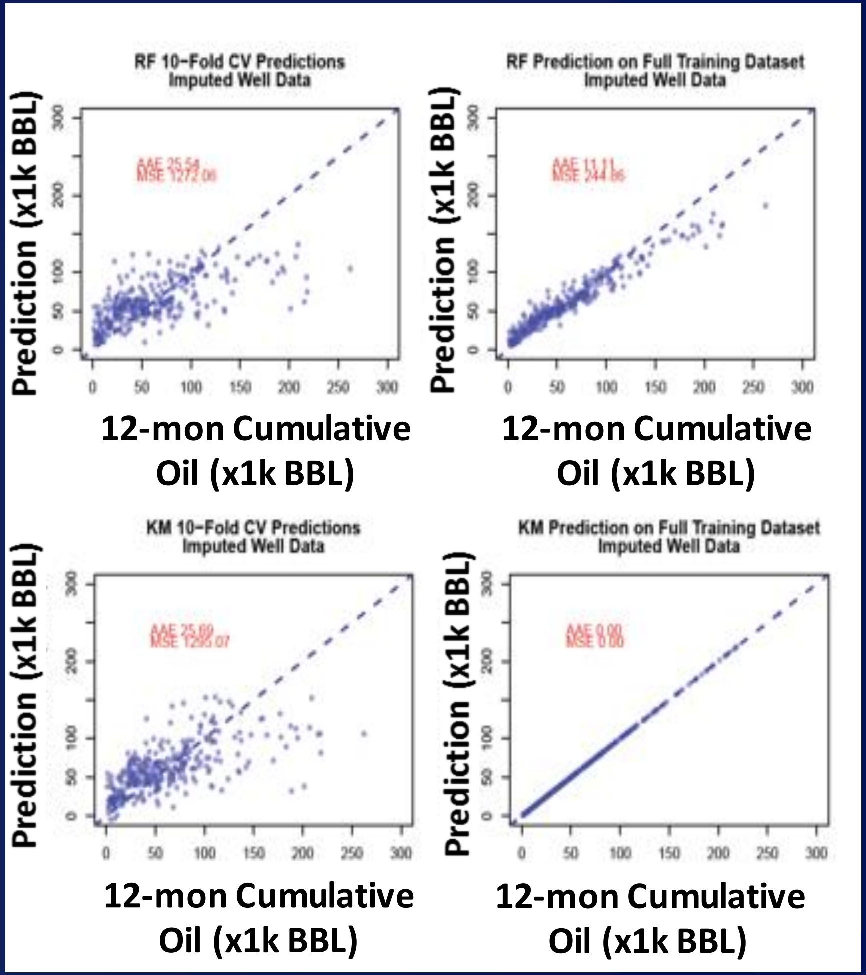
Field	Description
M12CO	Cum. production of 1st 12 producing months (BBL)
Opt2	Categorized operator code
COMPYR	Well completion year
SurfX, SurfY	Geographic location
AZM	Azimuth angle
TVDSS	True vertical depth (ft)
DA	Drift angle
LATLEN	Total horizontal lateral length (ft)
STAGE	Frac stages
FLUID	Total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

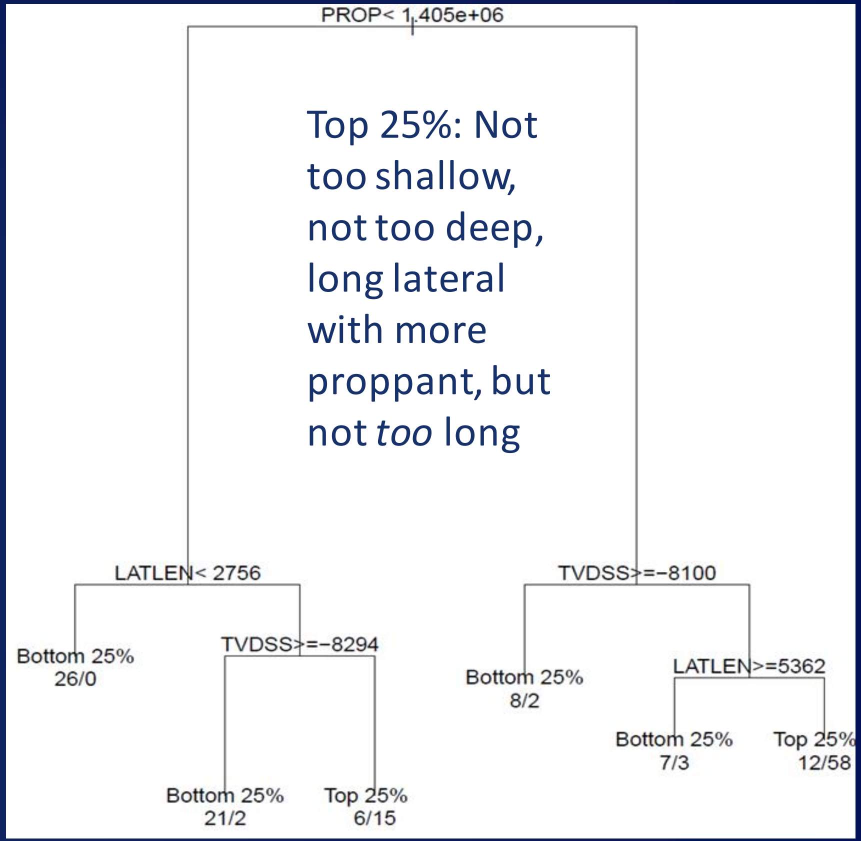


Classification Tree Analysis to Identify Factors Driving Extreme Outcomes

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

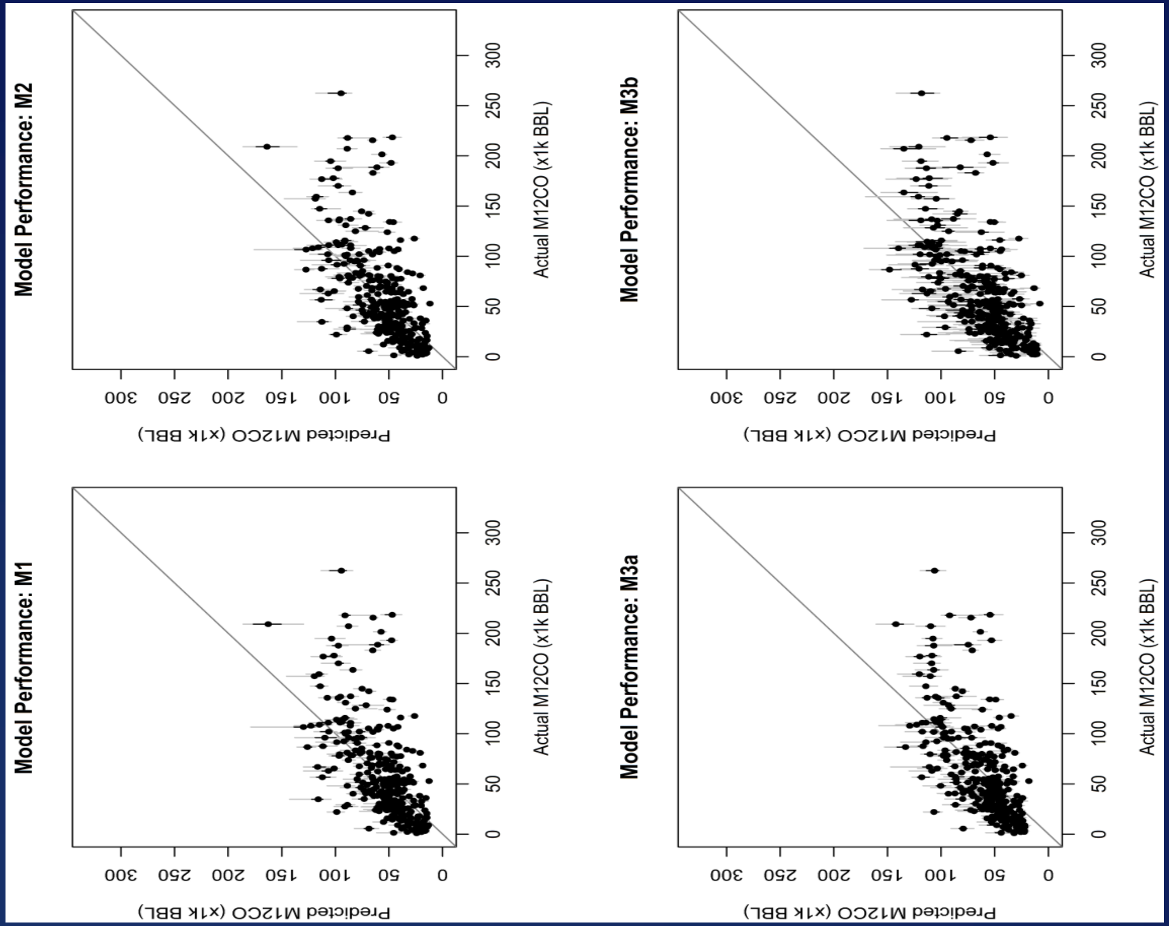
- Accuracy:

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



Ensemble Modeling

M1 — direct averaging; M2 — weighted averaging;
M3a — stacking with NN; M3b — stacking with RF

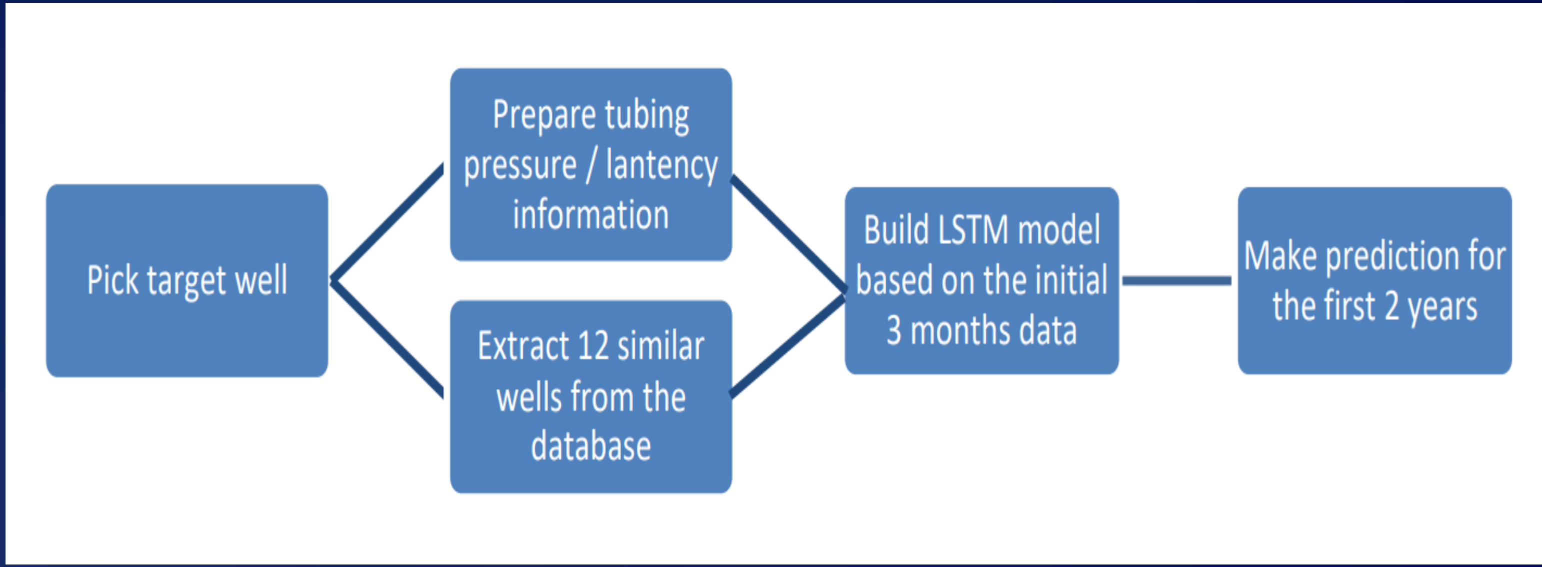


Model Name	RMSE (x1k BBL)
M1	37.57
M2	37.45
M3a	36.21
M3b	36.15
LPM	47.12
QPM	40.03
SVR	39.00
RF	38.33
GBM	40.40

Mishra 2022 SPEE

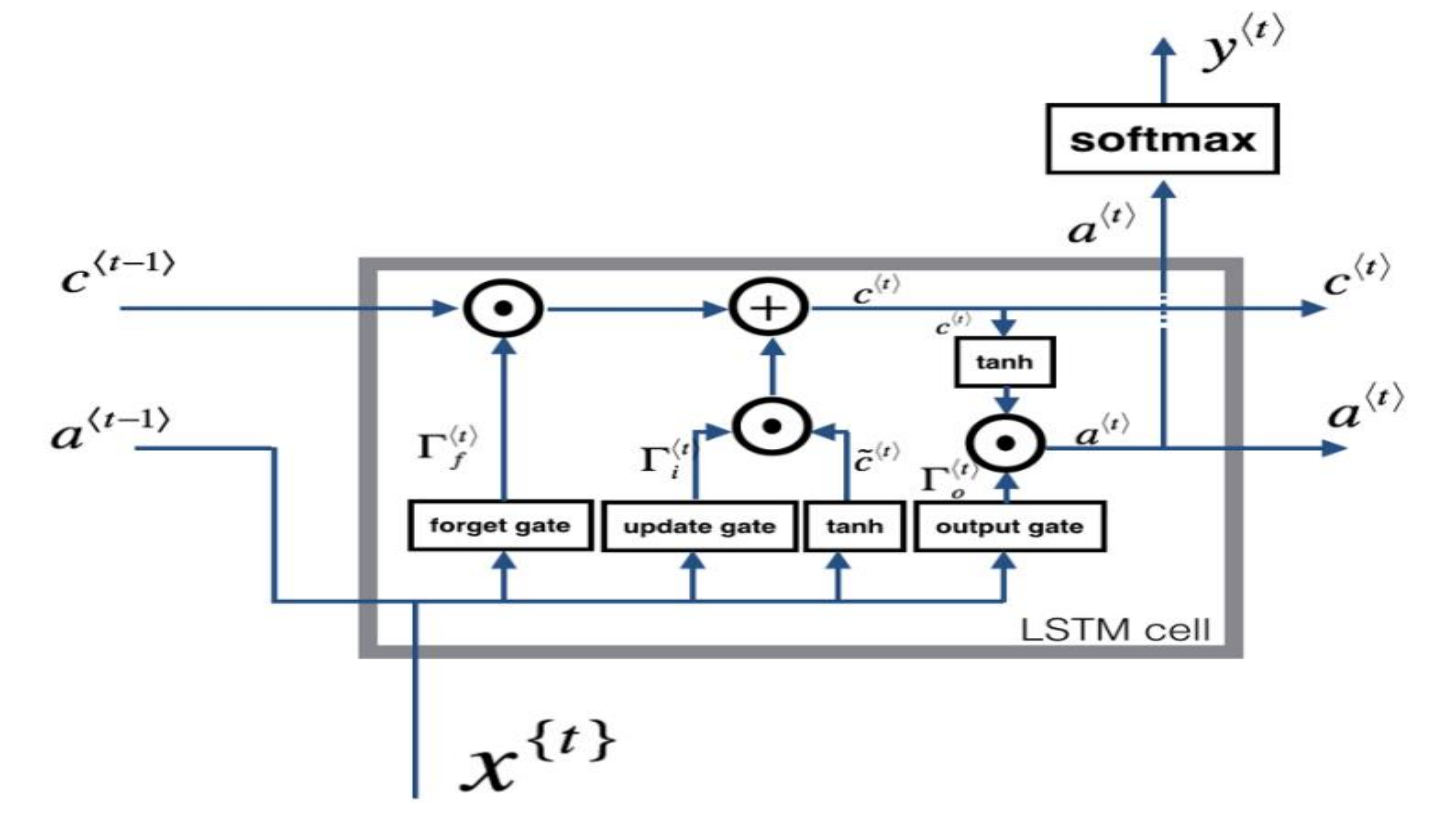
Case Study [2] – Application of Machine Learning for Production Forecasting

- Time-series approach to forecasting (v/s DCA)
  - Training on early-time data (~ 3 months); forecast for > 2 yrs
  - Long short-term memory (LSTM) method
  - 300+ wells analyzed in hindcasting of strategy



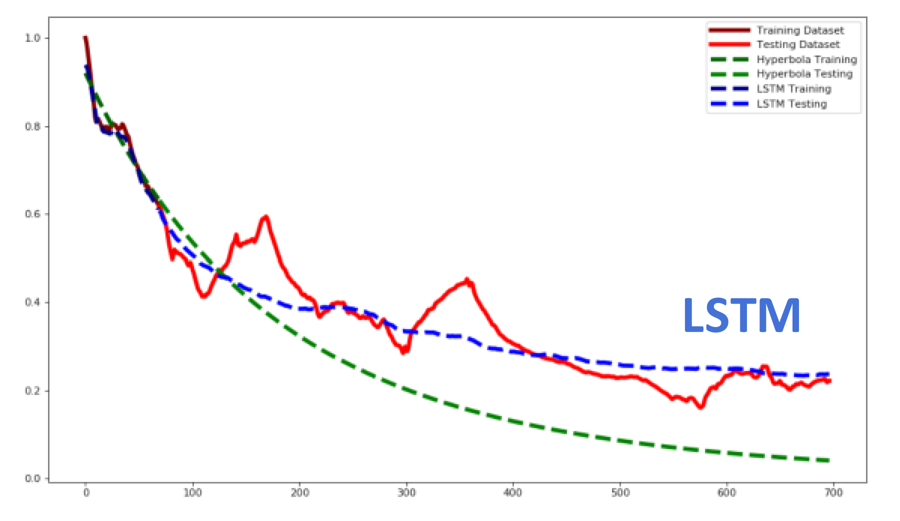
***Zhan, Sankaran, LeMoine, Graybill, Mey,* 2019, URTeC-47**

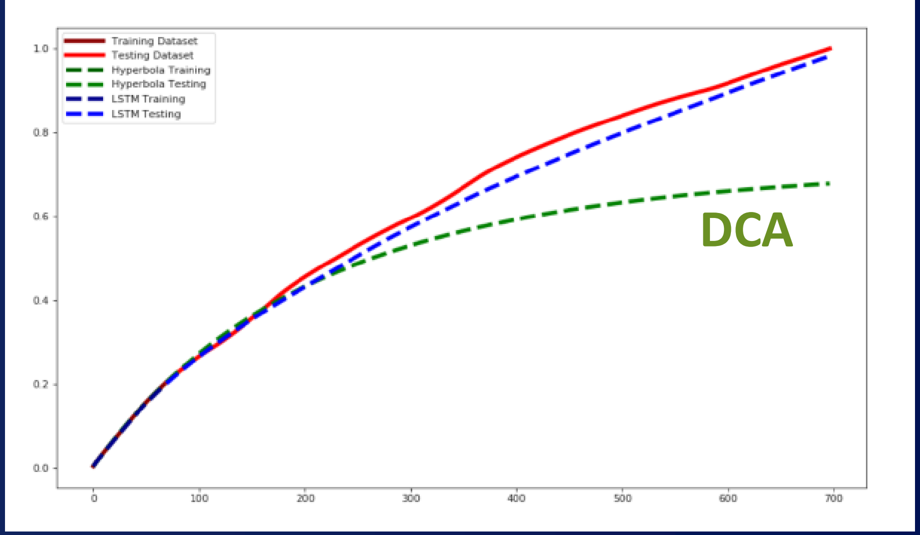
LSTM Architecture



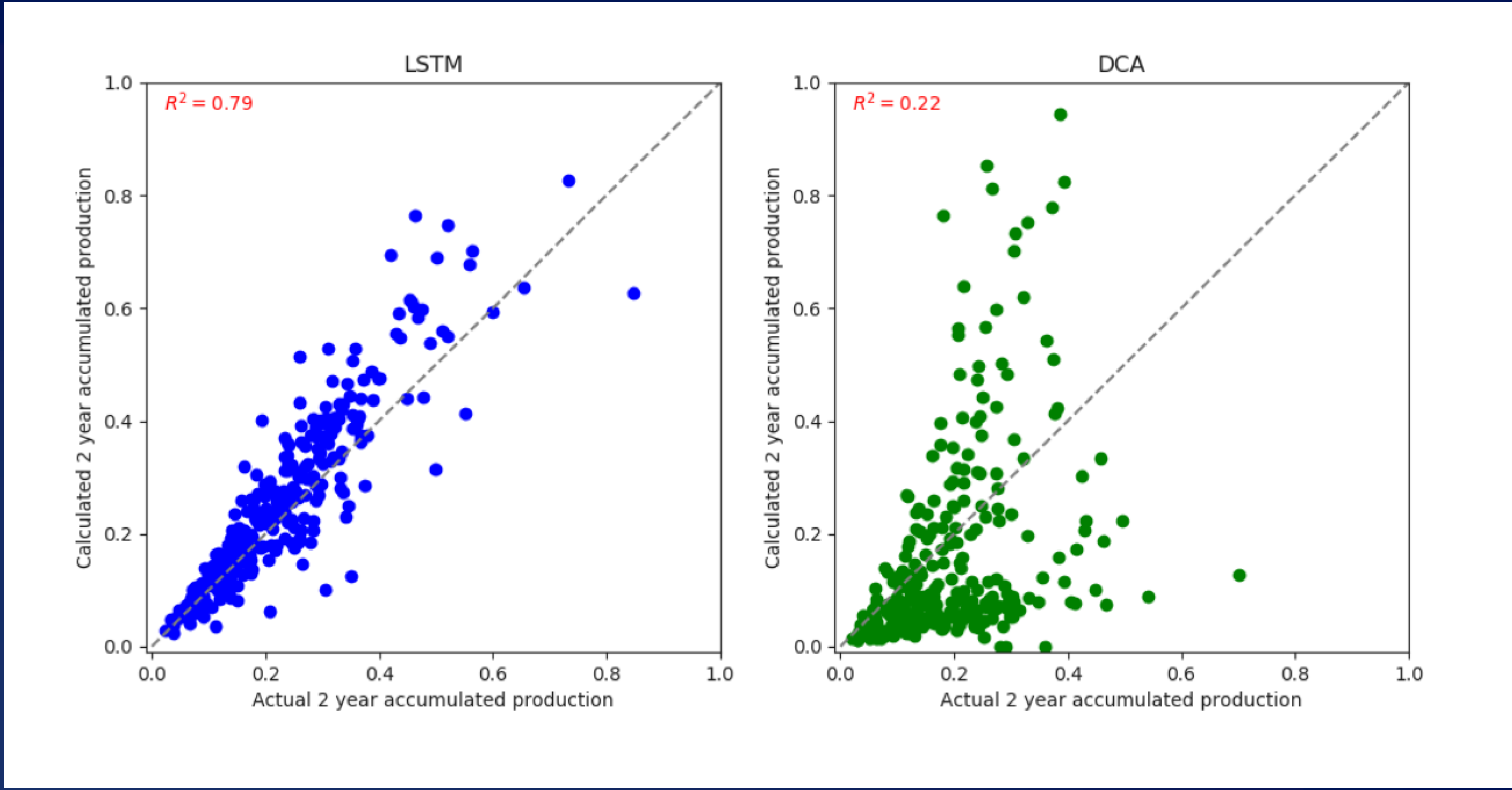
Key Aspects of Approach

- Purely data-driven – (tubing pressure, oil production from past 3 days, select nearby wells
- Leverages historical data from other wells (similarity measure)
- Separate models built for rate and cumulative production – aggregated with data-driven weights
- No static or completion parameters (similar to DCA)





Model Performance



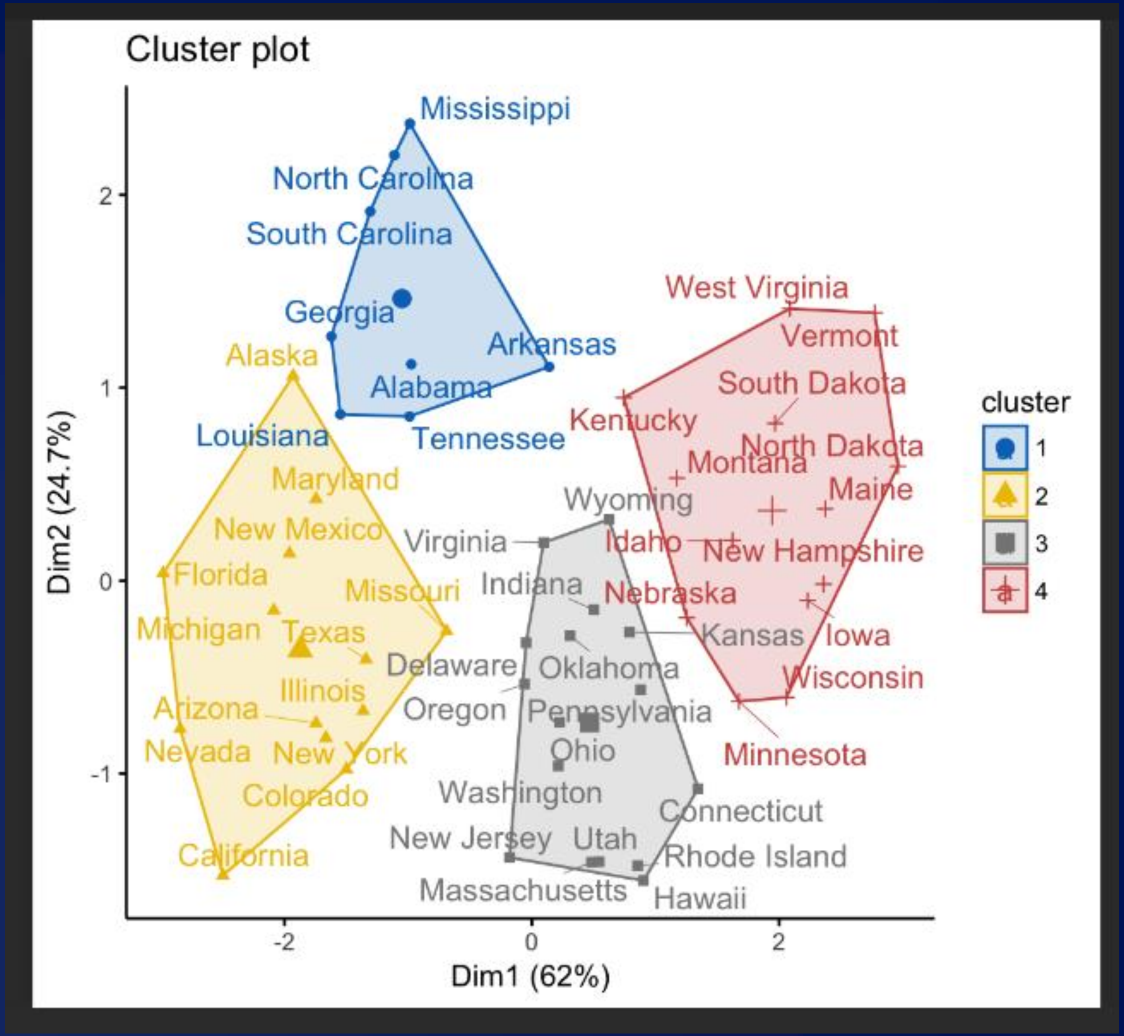
Case Study [3] – Generation of Type Wells via Cluster Analysis

- Conventional approach
  - Select analogous wells (reservoir type, well length, completion, age)
  - Calculate average performance of group
  - Use this "type well" for long-term forecasting
- Statistical approach
  - Create clusters based solely on shape of production decline
  - Create type well for each of the clusters
  - For a target well, find cluster that is most similar and use its type well for forecasting

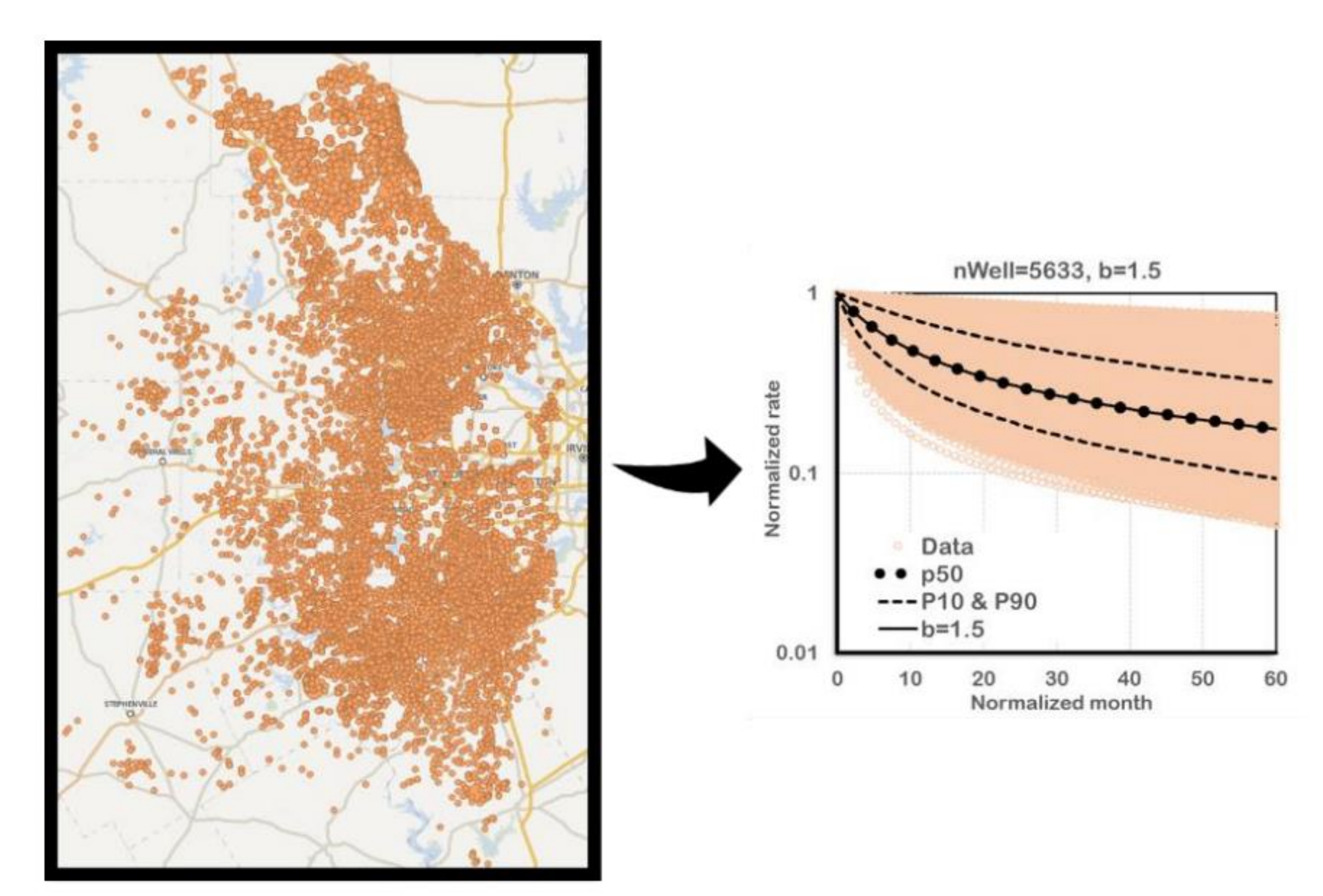
***Khaksarfard, Tabatabaie, Mattar,* 2019, URTeC-992**

Clustering and Barnett Dataset

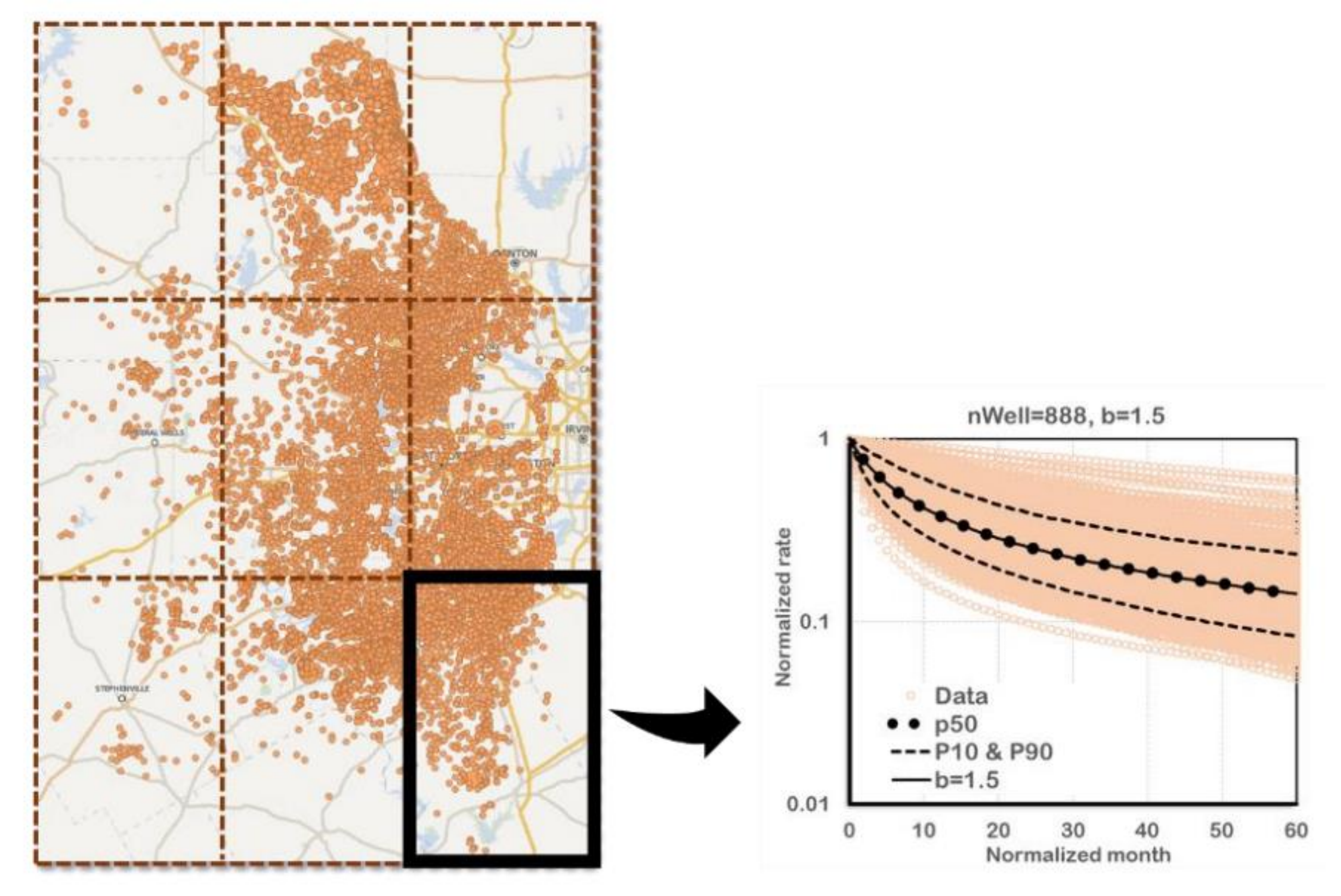
- Clustering with k-means
  - Minimizes average squared distance between data points and their corresponding cluster centers
- 7000 multi-fraced horizontal wells producing gas for 5+ years with non-anomalous + continuous production
- 80-20 split for training and testing
- Data normalized to peak rate



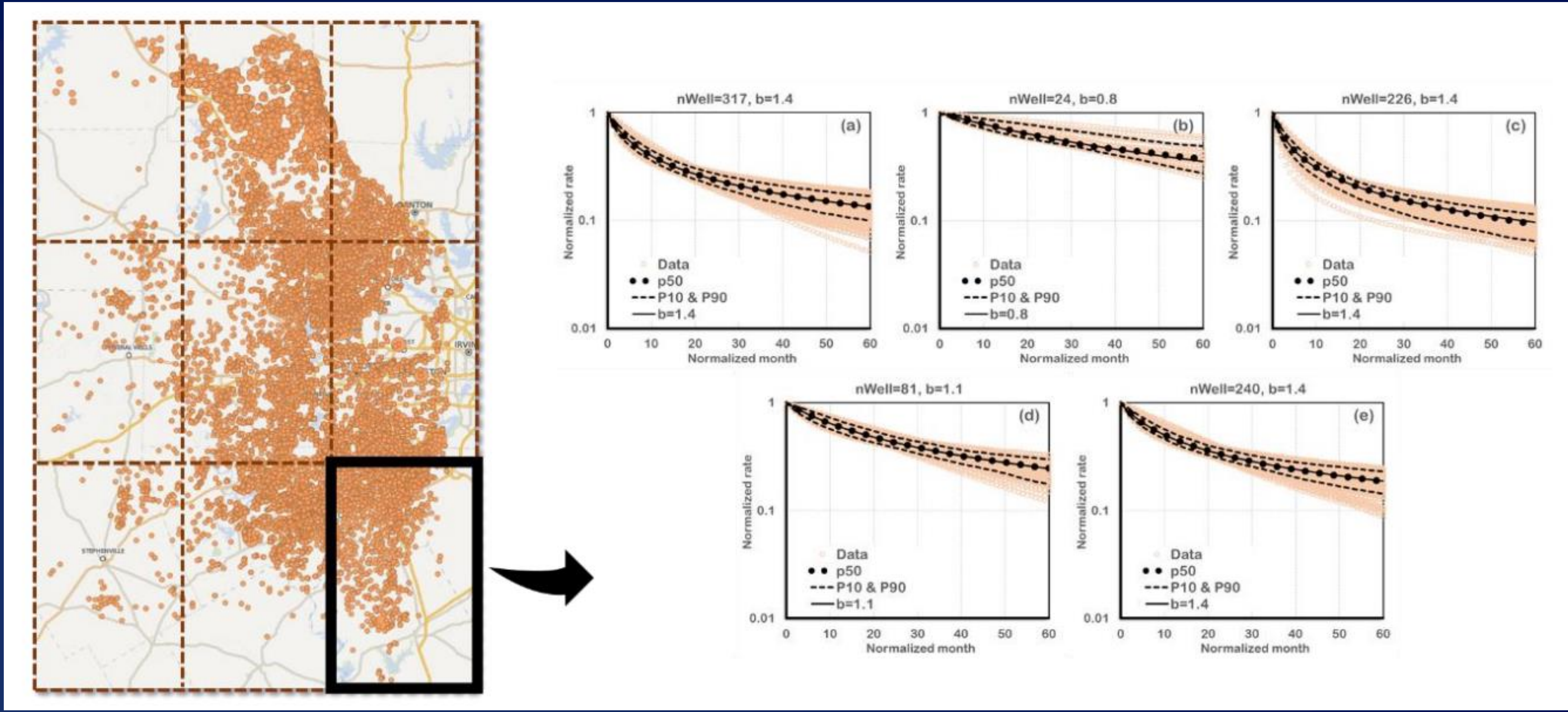
Single Area – Single Cluster (SA-SC)



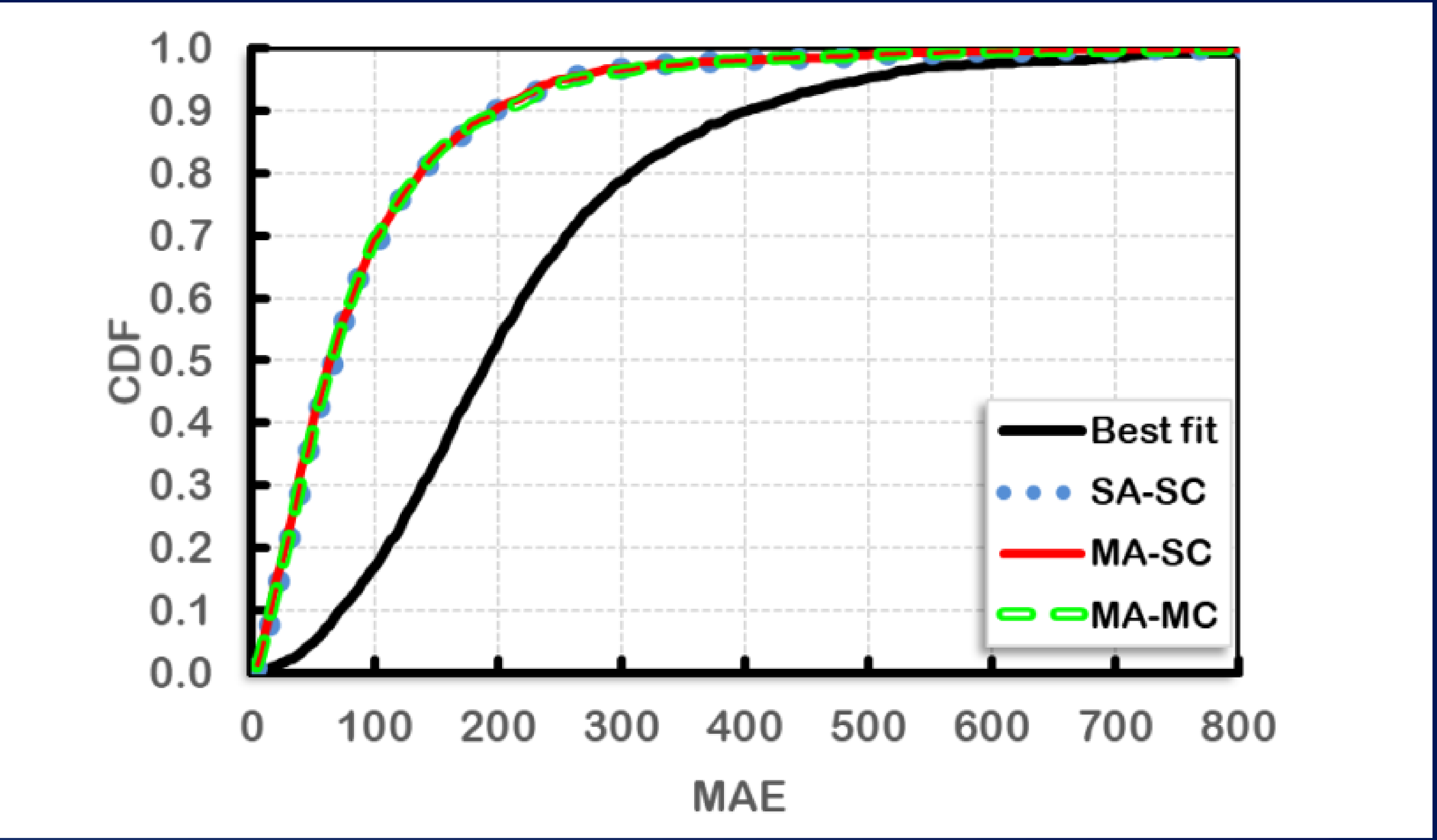
Multiple Area – Single Cluster (MA-SC)



Multiple Area – Multiple Clusters (MA-MC)

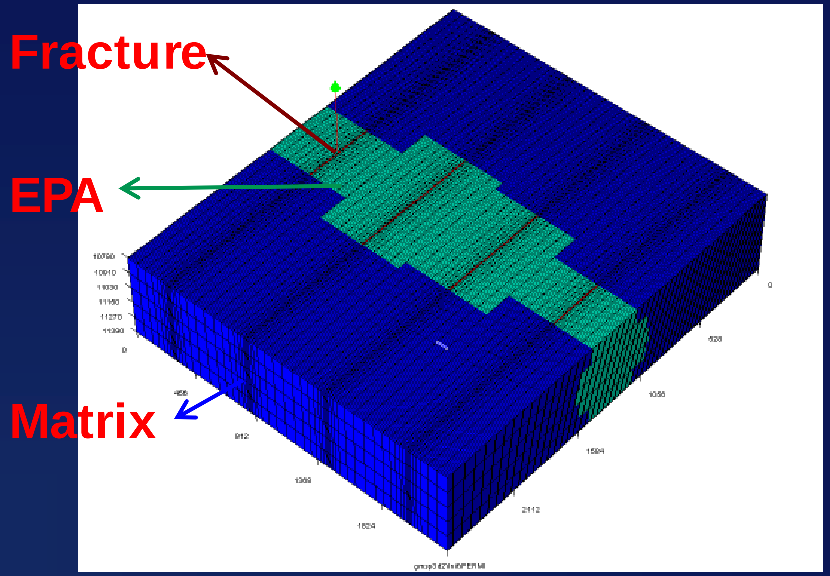


Testing with 6 Months Data to 60 Months



Case Study [4] – History Matching of Production Data with ML-Based Proxies

- History match flowing BHP data + SRV estimate
  - Proxy for dynamic reservoir model with ED/RS
  - Approximation of SRV using time-of-flight drainage volume
  - Sensitivity analysis, global optimization, uncertainty analysis



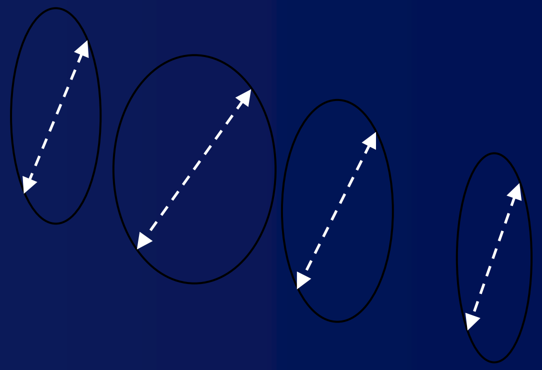
Objectives:

1) Match flowing BHP, SRV for 0-295 days

2) Predict BHP and gas rate for 295-730 days

Uncertainty Variables:

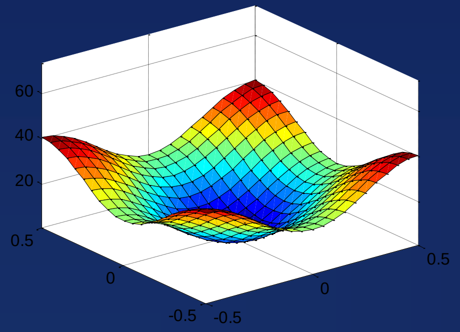
- ✓ Fracture/EPA/Matrix perm
- ✓ Elliptical Fracture half axis

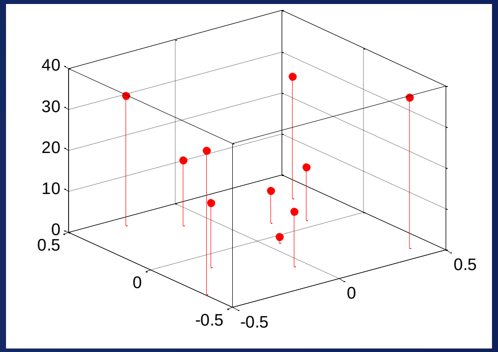


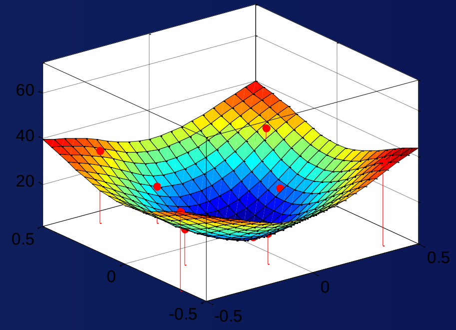
Yin, Xie, Datta-Gupta and Hill, **2011,JPSE, 127, pp. 124-136.**

Why and How to Build Proxy?

- Typical model run times too long (~multiple hours) – unsuitable for HM
- Solution ⇨ build surrogate (proxy) model (~seconds)
  - Create experimental design (incomplete factorial, space-filling LHS)
  - Run full-physics model at these parameter combinations
  - Fit response surface to observed results (quadratic fit, kriging, other ML models)



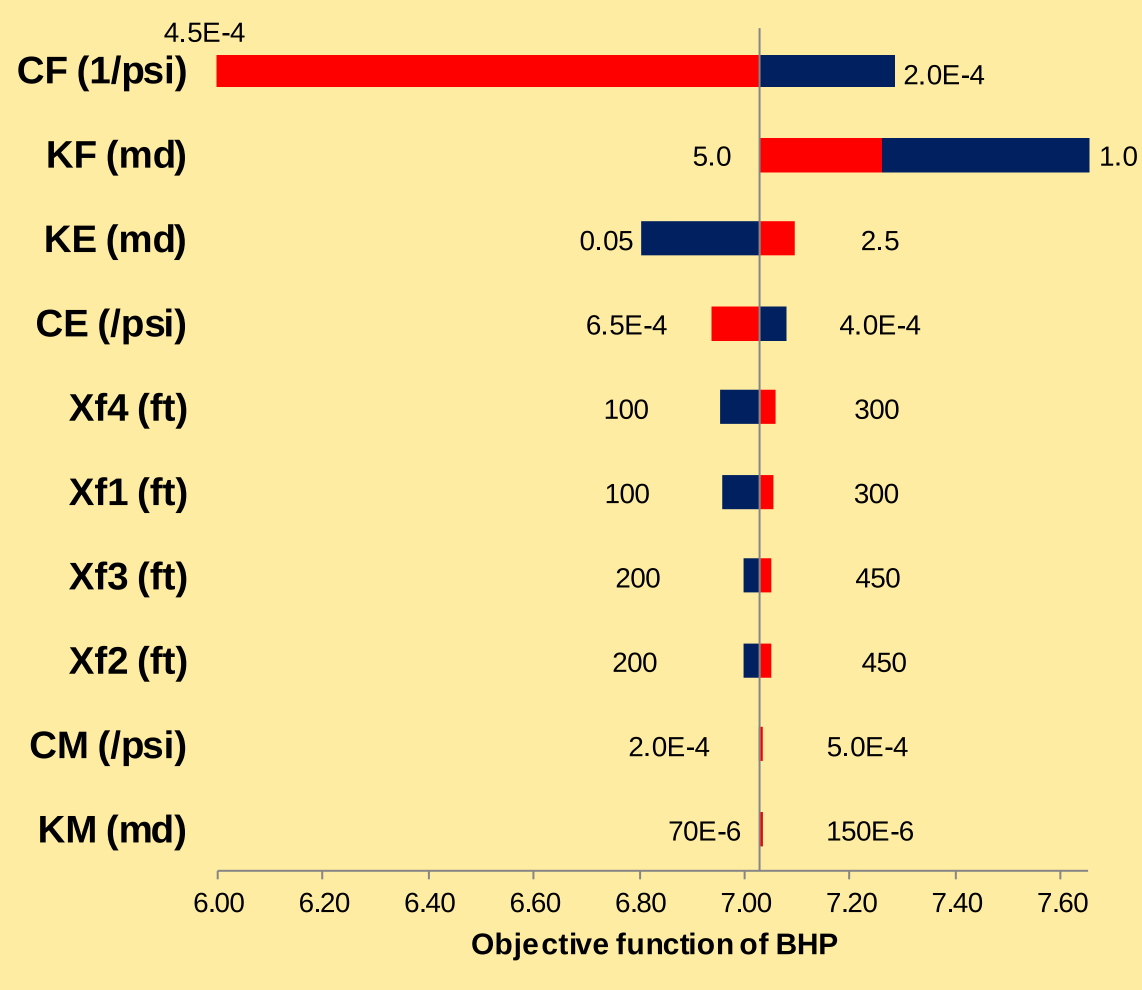




Sensitivity analysis identify key parameters → Proxy construction (Response Surface) → Parameter updating by GA with proxy → Uncertainty analysis

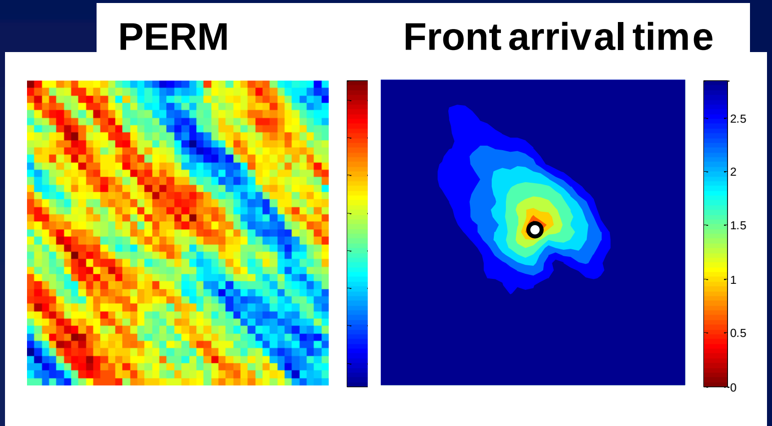
History Matching Steps

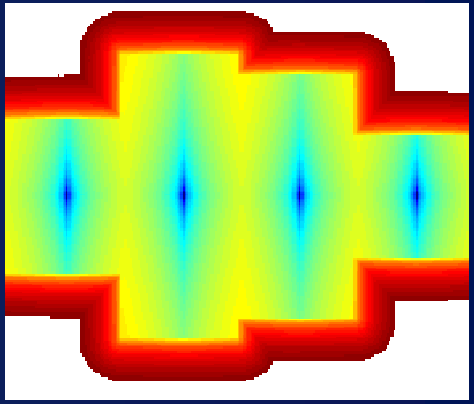
- **Sensitivity analysis with heavy hitters**
- Proxy construction using LHS+ kriging
- Drainage volume estimation from TOF
- Screening for DV vis-à-vis SRV (from microseismic or RT/PTA)
- Model calibration with GA
- Representative models from clustering
- Uncertainty estimation



History Matching Steps

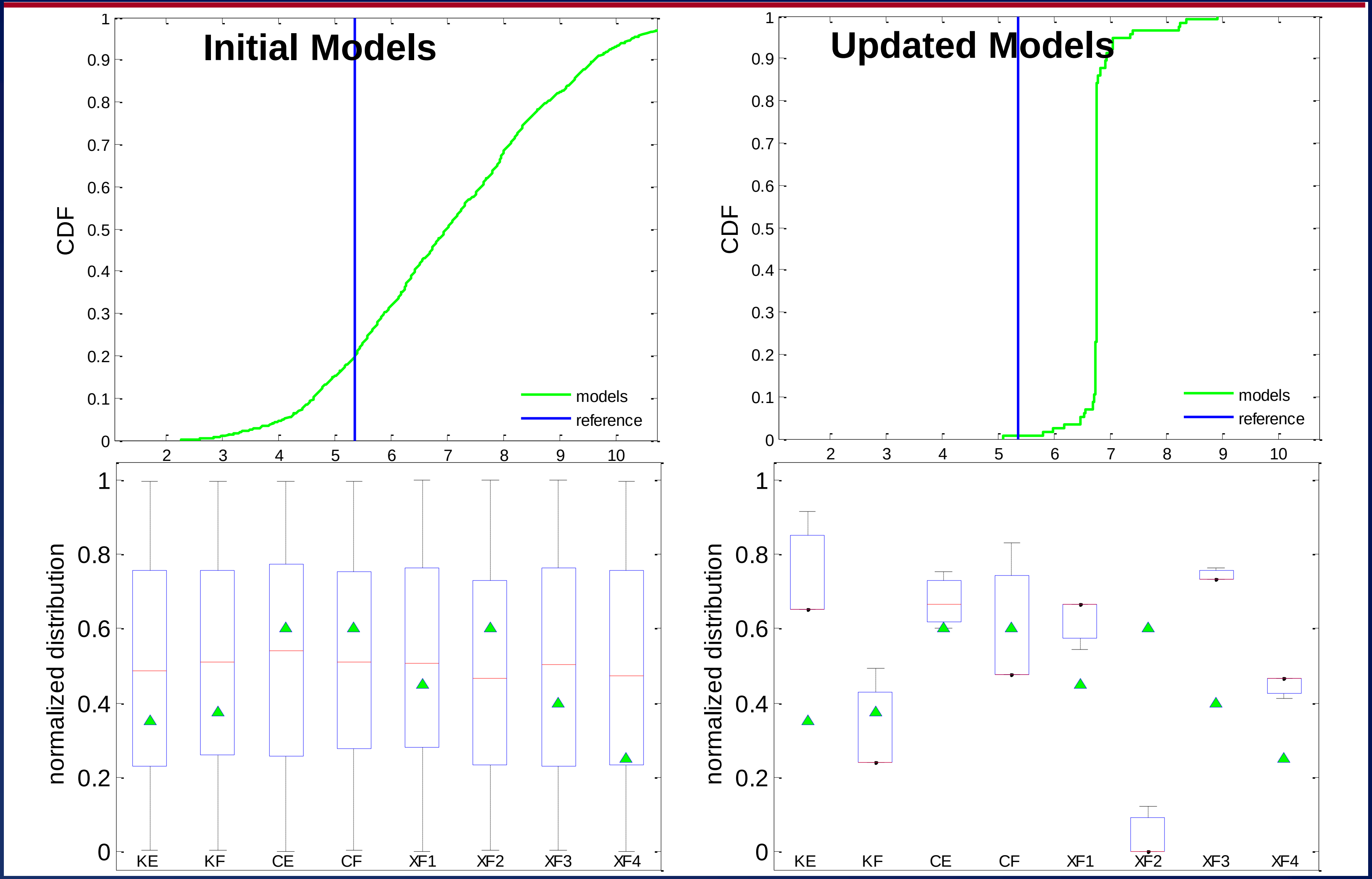
- Sensitivity analysis with heavy hitters
- Proxy construction using LHS+ kriging
- **Drainage volume estimation from TOF**
- Screening for DV vis-à-vis SRV (from microseismic or RT/PTA)
- Model calibration with GA
- Representative models from clustering
- Uncertainty estimation



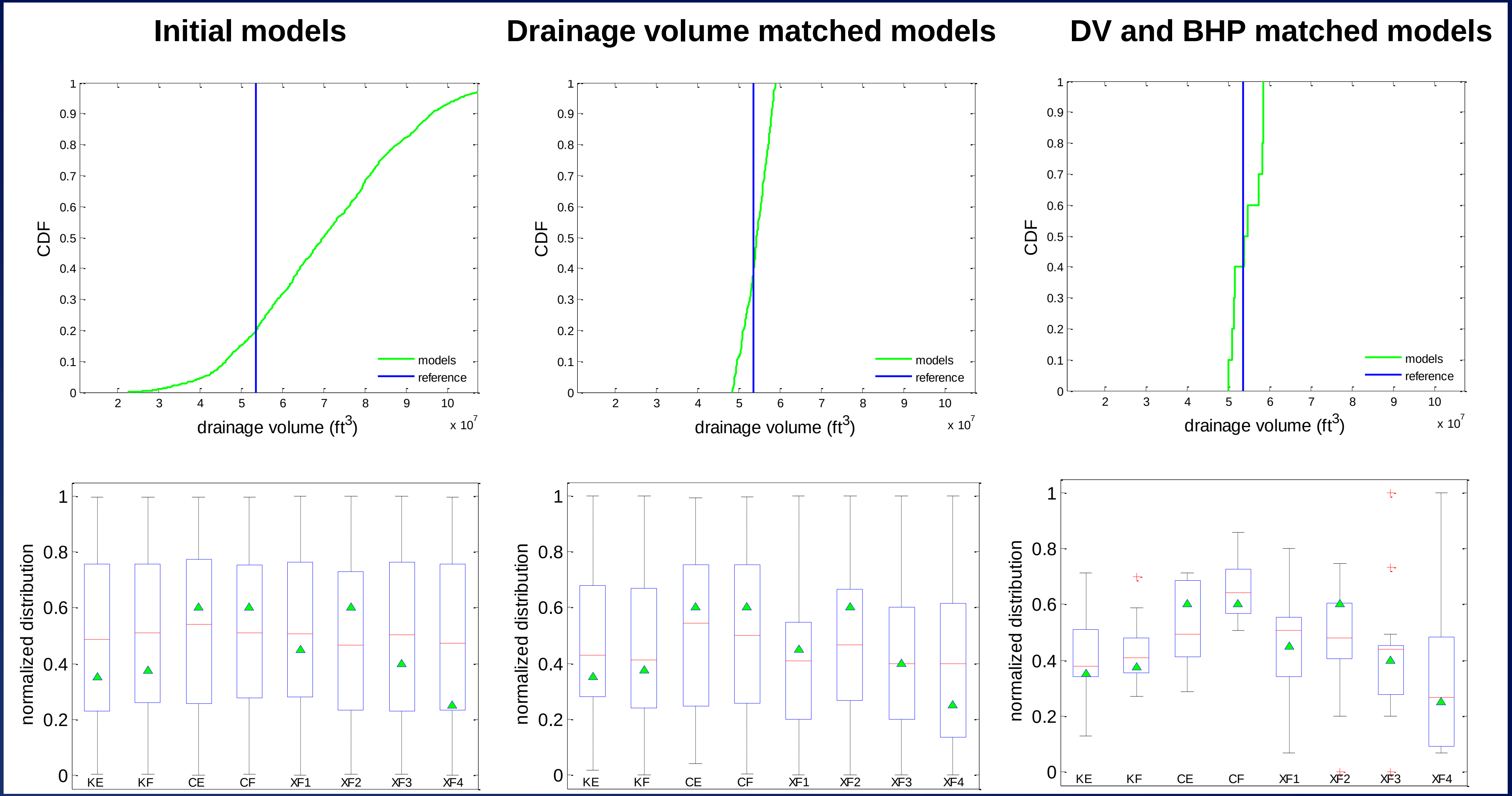


1.00E4 days
(very long time)

History Matching with Proxy using BHP



History Matching with Proxy using BHP and DV



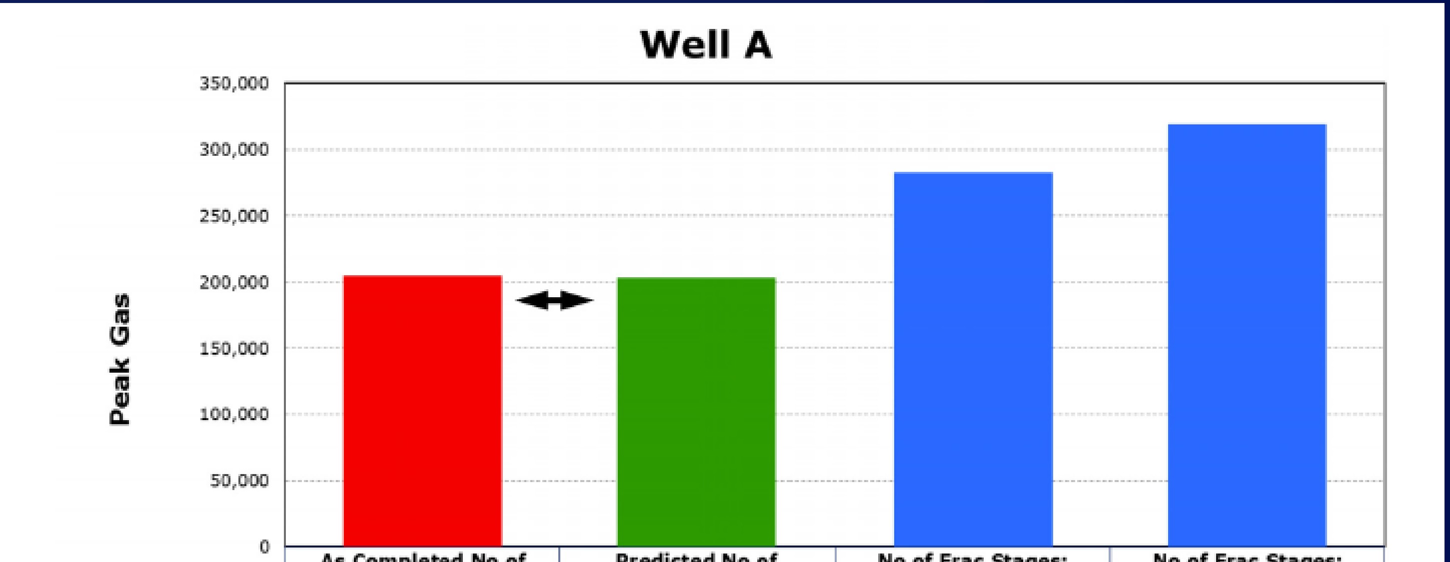
Example [1]

Shelley et al.

SPE-171003, 2014

Understanding Multi Fractured Horizontal Marcellus Completions

- 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

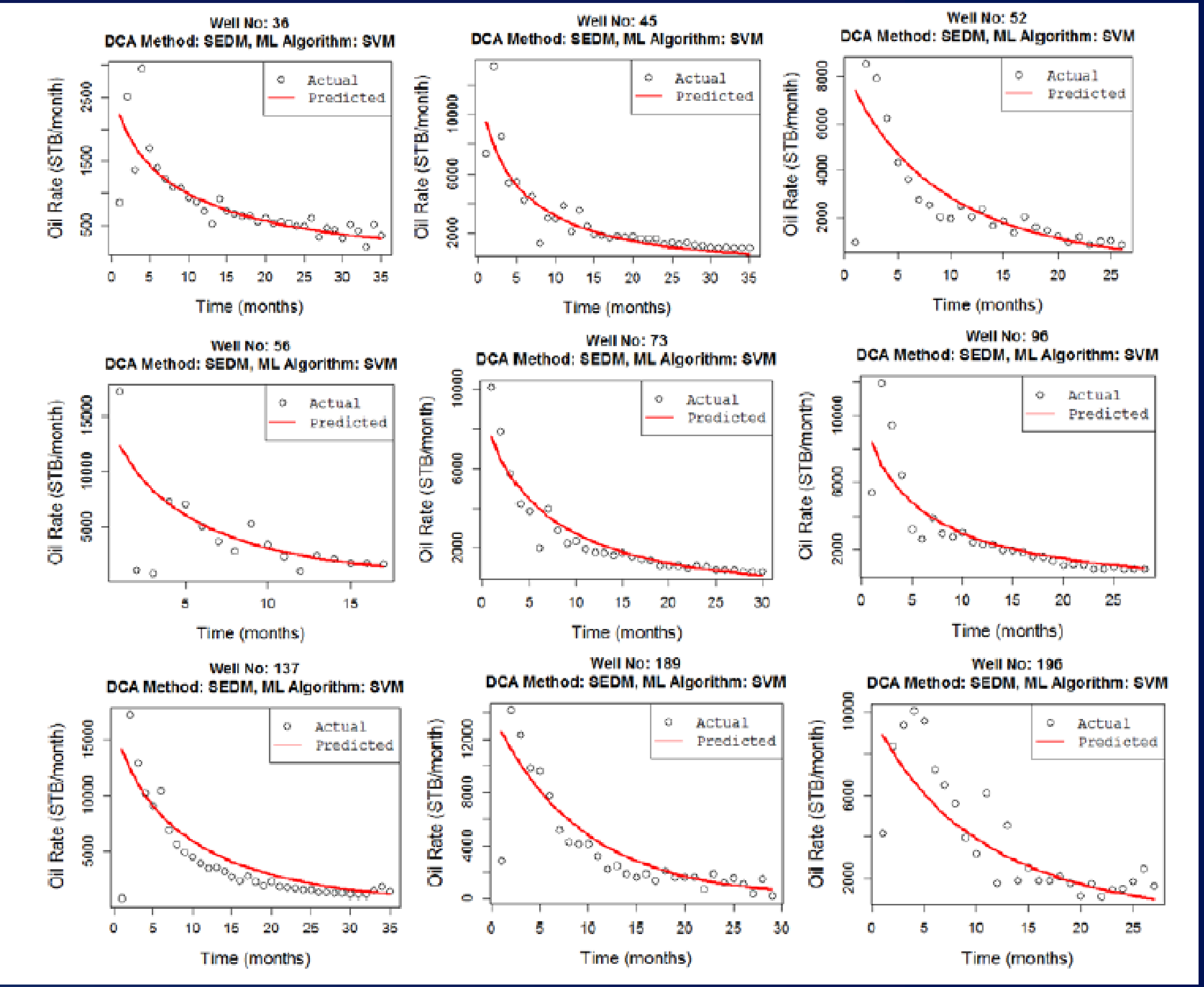
Example [2]

Vyas et al.

SPE-188231, 2017

Modeling Early-Time Rate Decline Using Machine Learning

- Decline curve model parameters linked to well completion related variables
- DCA Methods – Arps, Duong, SEDM, Weibull
- ML methods – RF, SVM, MARS
- Applied to Eagleford wells
- SEDM + SVM most suitable for forecasting



Outline of Talk

Basic Concepts

Case Studies

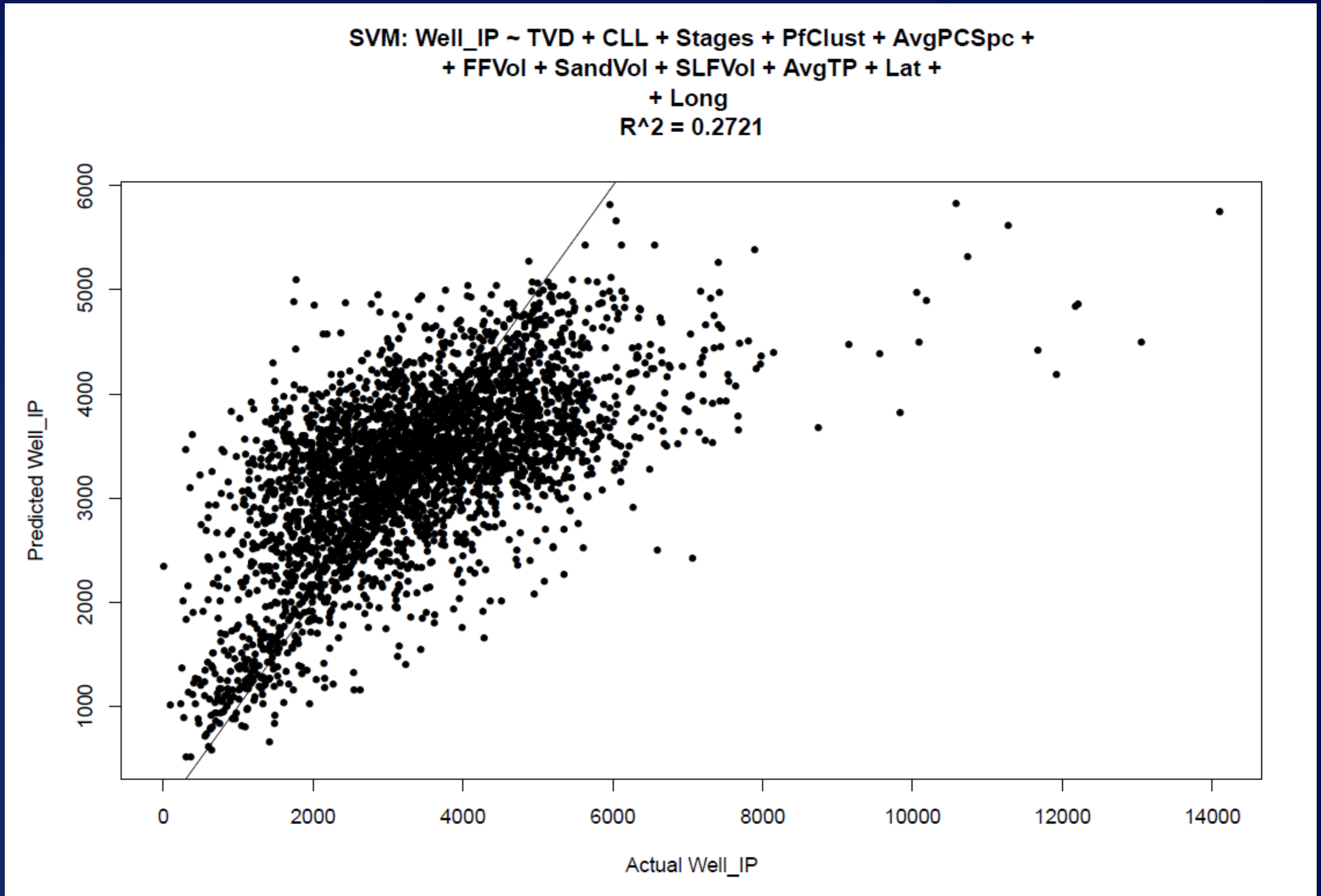
Lessons Learnt

Perils

An Unsuccessful (☹) Example

Production Data Analysis in Shale Gas Wells

- Some shale formation
- Data from ~3000 wells
- **Goal** ⇨ Fit well_IP ~ ***f*** (11 predictors related to well completion + location)
- **Regression** ⇨ SVM (also similar results with other techniques)
- *Issue ⇨ missing key causal variables in modeling!*



Recap of Lessons Learned

- Proper problem formulation is crucial
- Data quality/quantity can compromise results
- Predictive modeling is nuanced (many options)
- Multiple competing models may exist
- Unwrapping black-box models is difficult
- Communicating results can be challenging

Challenges for Acceptance of ML

- Our ML models are not very good.
- If I don't understand the model, how can I believe it?
- We are still waiting for the "Aha" moment!
- My staff need to learn data science, but how?

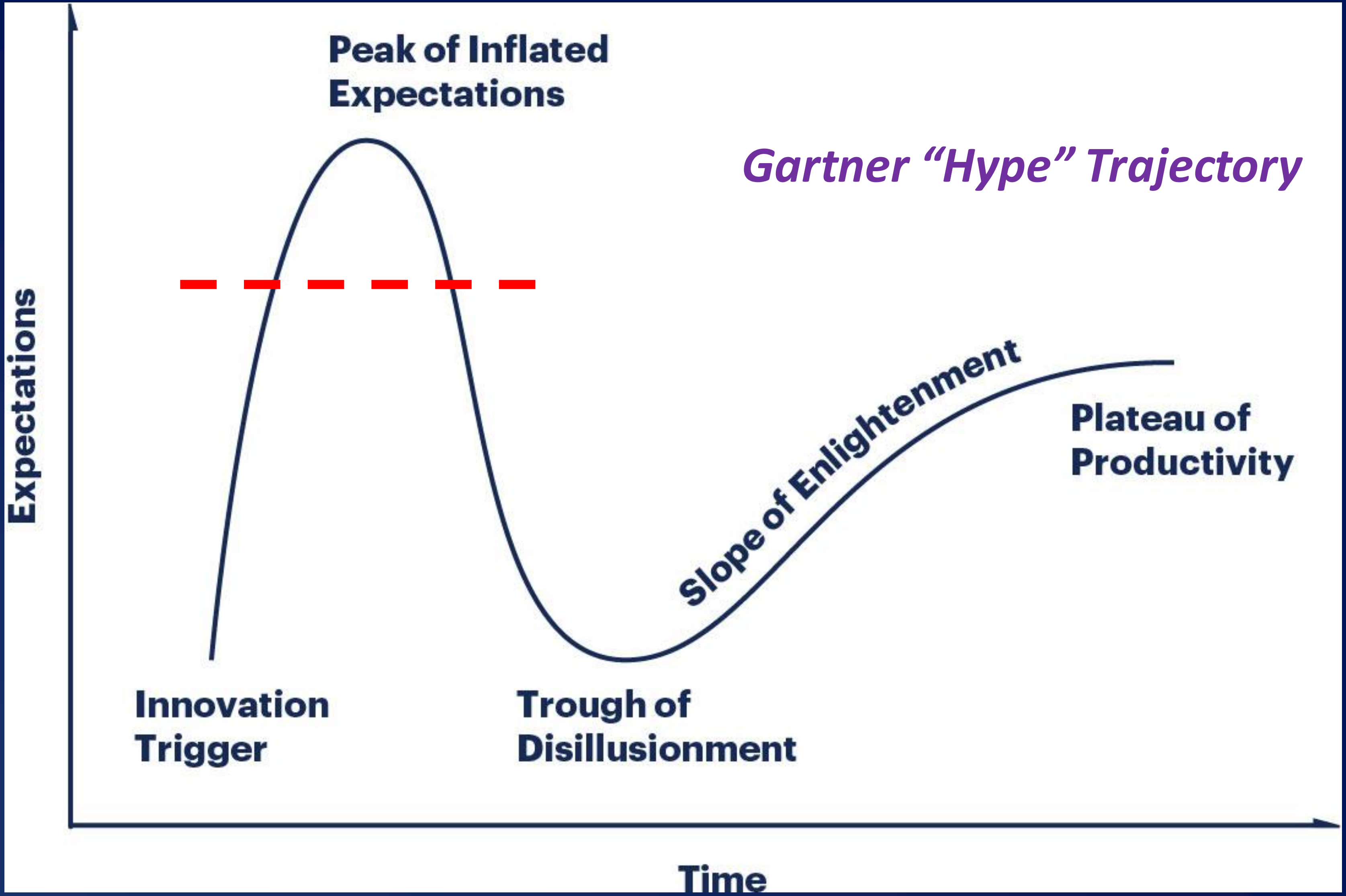
- Manage expectations
- Focus on added value
- Adequate/robust model?
- Key variables ID-ed?
- A new input-output tool
- Mechanistic model alternative
- Formal knowledge of statistics, programming (R/Python), ML

***Mishra et al., 2021*, JPT (March), 25-30.**

Closing Thoughts – Future

- Focus on issues for making data-driven models more robust (i.e., accurate, efficient, understandable, and useful)
- Promote foundational understanding of ML-related technologies among subsurface engineers and geoscientists
- Appropriate mindset
  - NOT curve-fitting exercises using very flexible and powerful algorithms
  - BUT extraction of insights consistent with mechanistic understanding

So, Where Are We?



ACKNOWLEDGMENTS

Battelle Memorial Institute

US DOE-NETL (SMART)

Thank you for your attention

mishras@battelle.org