Webinar - Big Data Analytics for Petroleum Engineering: Hype or Panacea?

Presenters:

Dr. Michael Pyrcz, Petroleum and Geosystems Engineering (moderator)
Dr. Larry Lake, Petroleum and Geosystems Engineering
Dr. Joydeep Ghosh, Electrical and Computer Engineering
The Problem Setting

Swimming in Data and Information Starved!

- diverse data types: completions, production, geological, petrophysical, geophysical
- diverse data scale: pore imaging to drainage area
- only 1 trillionth of the subsurface is likely directly sampled
- indirect measures improve coverage, but at the cost of resolution and accuracy
The Problem Setting

Complicated Problems and Important Decisions

- heterogeneous rock system, multiphase fluids, coupled fluid flow and geomechanical response
- forecast production rates and protect environment, health and safety
- deliver reliable energy and optimal extraction of national resources

Fluvial reservoir modeling with connectivity approximated by a fast marching solution. Time of flight and fractional flow are shown.
The Problem Setting

Strategic:

• Getting more meaningful information from our data?
• Little Data + Simple Model = Big Data?
• Improve objectivity given the large number of experience-based decisions?
• What is the role of statistical vs. deterministic modeling?
• How to maximize profitability and support big data analytics?

Tactical:

• Modeling multivariate, multiscale, spatial phenomenon.
• Accounting for sampling bias and missing data.
• Improving accuracy and flexibility of proxy models.
• Modeling all important sources of uncertainty.
• Making optimum decisions and the presence of uncertainty.
Model Selection

Integrating Model Uncertainties in Probabilistic Decline Curve Analysis for Unconventional Hydrocarbon Production Forecast
Different types of models have been proposed for unconventional oil production forecast.

For example

**Arps model:**

\[ q_t = \begin{cases} 
q_0(1 + bD_i t)^{-1/b}, & b > 0 \\
q_0 e^{-D_i t}, & b = 0 
\end{cases} \]

- Empirical.
- May not be ideal for unconventional production.

**Stretched exponential model:**

\[ q_t = q_0 e^{-\left(\frac{t}{\tau}\right)^n} \]

- Empirical.
- Based on the analysis of the Barnett shale wells.

**Logistic growth model:**

\[ q_t = \frac{Knat^{n-1}}{(a + t^n)^2} \]

- Empirical.
- Used to forecast growth in many applications.

**Pan CRM:**

\[ q_t = \Delta P \left( \frac{\beta}{\sqrt{t}} + J_{\infty} \right) e^{-2\beta\sqrt{t} + J_{\infty} t} / (ctv_p) \]

- Analytical.
- CRM combined with analytical solution of linear flow into fractured wells.
Possible solution:
• Any model is regarded as a potentially good model whose goodness is described by a probability representation.
• The probability of a model is interpreted as a measure of the relative truthfulness of the model to the other models.
• The probability is further used to weight the model forecast.

How to assess the model probability?
• Maximum Likelihood Estimation
• Bayes’ Theorem
• Monte Carlo Simulation
Bayes Theorem
The model parameters are determined by matching production data. But, ...

- Which model should we use?
- What if they can match the data almost equally well?
Schematic Example 1

Estimated Cum. Oil Prod. [10^5 bbl]

Aggregated with equal weights

Aggregated with updated weights
Midland Well No. 29: match oil production rate
31 Midland Wells: match oil production rate
Simple Models for Isothermal EOR Displacements
Koval model:  

\[ f_{\text{solvent}} = \frac{1}{1 + \left(1 - \frac{C_{\text{solvent}}}{K_v C_{\text{solvent}}}ight)} \]

\[ K_v = H_k E \]

- \( H_k \): Heterogeneity factor
- \( E \): effective viscosity ratio

- \( H_k = 1 \) (homogeneous)
- \( E = 1 \) (tracer)
Flow Capacity Curves at Different Heterogeneity Factors

- $H_k=5$
- $H_k=10$
- $H_k=20$
Fractional Flow Solution (Two Fronts)

Flow

Final $S_{OF}$
Bank $S_{OB}$
Initial $S_{ol}$

Fractional Flow Solution (Two Fronts)

Cumulative Residual Oil Recovery

97% oil recovery
0.009 final oil saturation

Pope, et al., 2007
Field Oil Bank Formation

Flow

Final $S_{oF}$

Bank $S_{oB}$

Initial $S_{ol}$

Flow

Final

$S_{oF}$

Bank

Initial $S_{ol}$

Flow

Inaccessible $S_{ol}$

Final $S_{oF}$

Bank $S_{oB}

Flow

Rate

Time

Rate

EL

Time

$C=1$

$C=0$
CO₂ Project Results

Lost Soldier Field

Slaughter Pilot

Rangely Field

Twofreds Field
Estimated Volumetric Sweep

Volumetric Sweep at One Pore Volume
Capacitance Resistance Modeling
CRM Development

• Bruce (1943) - Analogy to resistance and capacitance
• SPE 414 (1962) - Saudi reservoir simulator
• CPARM
  • UT PGE dozens of researchers over a decade
  • Technical and peer-reviewed papers
• Case studies (25+)
• CRPOT
  • Consulting projects
    • Licensed to BP
      • Currently used
      • BP developed work-flow patents around the technology
Hypothesis

Characteristics of a reservoir can be inferred from analyzing production and injection data only.
Capacitance-Resistance Modeling

How it works...
- Analogous to a resistance capacitor—electric potential (input) is converted to current or voltage (output)
- Converts input signals (injection rates) to output signals (production rates) after a time lag and attenuation
- Uses the basic continuity equation
- Characterizes relationships between wells
- Optimizes based on the economics

Reservoir Simulation in 1962 (SPE-414)
Continuity:
\[ c_t V_p \frac{d\bar{p}}{dt} = i(t) - q(t) \]

Production Rate:
\[ q(t) = J(\bar{p} - p_{wf}) \]

Ordinary Differential Equation:
\[ \frac{dq(t)}{dt} + \frac{1}{\tau} q(t) = \frac{1}{\tau} i(t) - J \frac{dp_{wf}}{dt} \]

\[ \tau = \frac{c_t V_p}{J} \quad \text{Time constant} \]

Solution:
Production
\[ q(t_k) = q(t_{k-1}) e^{-\frac{\Delta t}{\tau}} + \left( 1 - e^{-\frac{\Delta t}{\tau}} \right) I(t_k) - J (p_{wf}(t_k) - p_{wf}(t_{k-1})) e^{-\frac{\Delta t}{\tau}} \]
CRM for Multiple Injectors

Production

Injection

Pressure

\[ q(t_k) = q(t_{k-1}) e^{-\frac{\Delta t}{\tau}} + \left( 1 - e^{-\frac{\Delta t}{\tau}} \right) \left( I(t_k) - J(p_{wf}(t_k) - p_{wf}(t_{k-1})) \right) e^{-\frac{\Delta t}{\tau}} \]

For multiple injectors and neglecting pressure:

\[ q_j(t_k) = q_j(t_{k-1}) e^{-\frac{\Delta t}{\tau_j}} + \left( 1 - e^{-\frac{\Delta t}{\tau_j}} \right) \sum_{j=1}^{n_p} f_{ij} I_j(t_k) \quad \text{where} \quad \sum_{j=1}^{n_p} f_{ij} \leq 1 \]

Gain:

\[ \text{Gain:} \]

\[ \text{Time Constant:} \quad \tau = \frac{c_t V_p}{J} \]

Drainage volume around a producer
Ventura Field—History Match

Average R-Squared = 0.71
Ventura Field, CA—Connectivity Maps

Connectivity > 10%

Connectivity > 50%
Optimization—Maximize Net Present Value

\[
\text{NPV} = \sum_{j=1}^{n_p} \sum_{k=1}^{n_t} \frac{p_0 q_{0j}(t_k, f_{ij}, \tau_j)}{(1 + D_f)^k} \Delta t - \sum_{i=1}^{n_i} \sum_{k=1}^{n_t} \frac{p_w}{(1 + D_f)^k} I_i(t_k) \Delta t
\]

Subject to

- CRM
- Fractional flow model
- Upper limit on total injection rate
- Upper limits on rate of each injector
- Price of oil and cost of water
- Discount rate
On Data-Intensive Approaches to Complex Engineering and Business Problems

Prof. Joydeep Ghosh

Schlumberger Centennial Chaired Professor
Dept of ECE
Director, IDEAL
(Intelligent Data Exploration and Analysis Lab)
University of Texas at Austin
The Fourth Paradigm (2009)

Major Paradigms for Scientific Exploration & Discovery

- Empirical
- Theoretical
- Computational
- Data-Intensive Scientific Discovery
  (but not in Kuhn’s sense)

- 2017: AI Rules
Going Deep

From Facebook's Deepface paper

Our method reaches an accuracy of 97.35% on the Labeled Faces in the Wild (LFW) dataset,
reducing the error of the current state of the art by more than 27%, closely approaching human-level performance.
Big Data Allows

• Complex models with many variables
  • “Automated” feature engineering, (somewhat) compensating for lack of domain knowledge
  • Reduce “noise component”
• Flexible, General Models
  • Adaptable Complexity
  • Reduce uncertainty
• New Insights
Big data: are we making a big mistake?

By Tim Harford

- Digital exhaust $\rightarrow$ retrospective analysis
- Theory-free analysis of correlations is fragile
- Bias in data
  - Alfred Landon vs. FDR, 1936
  - If the term “doctor” is more associated with men than women, then an algorithm might prioritize male job applicants over female job applicants for open physician positions.
- Difficult to Comprehend
  - New emphasis on Explainable AI
- The Parable of Google Flu Trends
5 Tribes of ML (From Domingos)

My Speciality: Multi-learner systems

Use multiple, complementary approaches for more robust modeling of complex engineering problems

Custom models, where “canned solutions” are inadequate.
Multi-view Learning
Multi-Task & Transfer Learning

Multitask

Transfer

Active & Semi-supervised Scenarios also possible
Active + Multi-task Learning (+semi-supervised + knowledge transfer)

Active: get to optimality quicker with minimal human involvement
Field-wide Distributed Monitoring and Prediction using DXS Technology (DTS, DAS, DPS,..)

- TB of data/day
- Large number of (potential) applications
  - 2 page list by Dria (2012) of just DTS.
    - gas lift monitoring/optimization, injection profiling, well integrity and monitoring, real-time stimulation monitoring, etc.

- Some visual analysis or descriptive statistics; but little predictive modeling so far
Revisiting existing data-driven E&P applications

- "Predicting equipment failures"
- rock permeability and its spatial distribution
- hydraulic fracture design and post-fracture well performance prediction.
- virtual sensors, e.g. synthetic MR logs (Mohaghegh)
- Surrogate Reservoir Modeling (SRM), to model and analyze an enhanced coalbed methane project.
- group of horizontal well placement attributes are defined to capture the location of horizontal wells in a heterogeneous reservoir.
- Seismic inversion
- Reservoir Surveillance
- Well Stimulation, e.g. predict job pause time
- monitoring of subsea structures, pipelines, NG pipes (size, location of leaks)
- Shale exploration and production

Current Practice: Largely BI (dashboarding), and use of off-the-shelf prediction models)

Possible: Specialized Regression/Classification:
  - 1000+ variables, sparse data, structured outputs, low rank, networked info,......
  - Integrate heterogeneous data types
  - ....
Big Data in Climate

- Satellite Data
  - Spectral Reflectance
  - Elevation Models
  - Nighttime Lights
  - Aerosols

- Oceanographic Data
  - Temperature
  - Salinity
  - Circulation

- Climate Models
- Reanalysis Data
- River Discharge
- Agricultural Statistics
- Population Data
- Air Quality
- ...

Source: NASA
**Pattern Mining:**

*Monitoring Ocean Eddies*
- Spatio-temporal pattern mining using novel multiple object tracking algorithms
- Created an open source data base of 20+ years of eddies and eddy tracks

**Extremes and Uncertainty:**

*Heat waves, heavy rainfall*
- Extreme value theory in space-time and dependence of extremes on covariates
- Spatiotemporal trends in extremes and physics-guided uncertainty quantification

**Network Analysis:**

*Climate Teleconnections*
- Scalable method for discovering related graph regions
- Discovery of novel climate teleconnections
- Also applicable in analyzing brain fMRI data

**Change Detection:**

*Monitoring Ecosystem Disturbances*
- Robust scoring techniques for identifying diverse changes in spatio-temporal data
- Created a comprehensive catalogue of global changes in surface water and vegetation, e.g. fires and deforestation.

**Sparse Predictive Modeling:**

*Precipitation Downscaling*
- Hierarchical sparse regression and multi-task learning with spatial smoothing
- Regional climate predictions from global observations

**Relationship mining:**

*Seasonal hurricane activity*
- Statistical method for automatic inference of modulating networks
- Discovery of key factors and mechanisms modulating hurricane variability

http://climatechange.cs.umn.edu/
**Challenges**

- Multi-resolution, multi-scale data
- High temporal variability
- Spatio-temporal auto-correlation
- Spatial and temporal heterogeneity
- Large amount of noise and missing values
- Lack of representative ground truth
- Class imbalance (changes are rare events)
Machine Intelligence and Decision Systems

UT-MINDS

Prof. Joydeep Ghosh, UT-MINDS director

http://data.ece.utexas.edu
Overview

- UT-MINDS faculty perform R&D in **data science** and **machine learning**
  - Theory and algorithms
  - Deployable Applications based on real, complex data

- **Robust, Scalable, Well-Engineered solutions** for design of reliable **full-stack systems**
Data/sensors

- Images/video
- Signals (wireless, sensors,..)
- Text and Speech
- Networked data sources
- Databases/streams
  ...

ML core

- Deep Learning and GANs
- Reinforcement Learning
- Explainable ML/AI
- Optimization and Robust ML
- Parallel/Distributed Implementations
- Lifelong/continual learning
- Model Lifecycle Management
  ...

Applications

- Human-Machine interaction
- Context-Aware Personalization
- ML for Healthcare
- Security
- Hardware/Software Verification
- Infrastructure: Transportation, Energy
- Network Models
  ...

...
Thank You!
What we do

• Data-Driven Modeling & Knowledge Discovery
  “Big Data Predictive and Prescriptive Analytics”

• Data Types:
  • relational databases, distributed sensors, signals, images, web-logs, key-value....
  • data (continuous + symbolic) + domain knowledge

• Tools:
  • Data mining/stats; web mining; machine learning, Neural nets, signal/image processing....

• Large Scale System issues

• Speciality: Multi-learner systems
  • Use multiple, complementary approaches for more robust modeling of complex engineering problems
  • Custom models, where “canned solutions” are inadequate.
Neuro-Symbolic Hybrids for Knowledge Refinement

• Decision Support for LCRA Dams near Austin:
  • initial domain knowledge + follow-on data
  • Can extract refined domain knowledge!
Sample Project on DAS Analysis

• ML+ SP for
  • semi-automating the process of detecting and characterizing various events of interest
  • determining the (spatio-temporal) resolution of data that is necessary and sufficient
  • predicting field-wide developments and potential problem-spots (with alerting mechanism) based on detected signals/events and their spatio-temporal correlations.

• Potential Benefits
  • more accurate monitoring with less personnel, reduce data requirements, and lead to (near)-real time alerting systems and post-job diagnostics. May suggest timely interventions.
Knowledge Pipeline:

http://www.cpge.utexas.edu/

This webinar will be posted to our website.