Application of Machine Learning Methods to Predict Well Productivity in Montney and Duvernay

Shengnan Chen, PhD, P.Eng.
Associate Professor

Department of Chemical and Petroleum Engineering
Schulich School of Engineering
University of Calgary

April 16, 2019

Outline

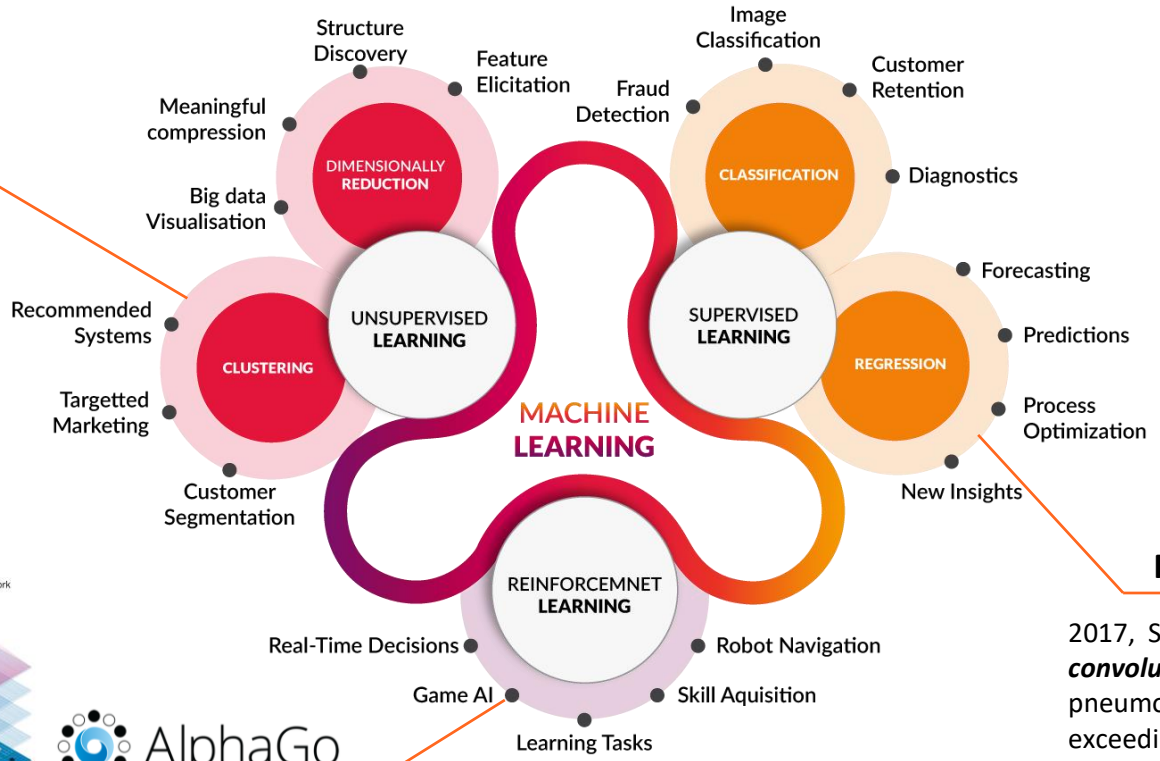
- 1. Introduction to Machine Learning
- 2. Unconventional Tight/Shale Reservoirs
- 3. General Data Visualization
- 4. Machine Learning Modeling
  - Case Study 1: Productivity Forecast in Montney Formation
  - Case Study 2: Productivity Forecast in Duvernay Formation
  - Case Study 3: Time Series Study in Montney
- 5. Conclusions

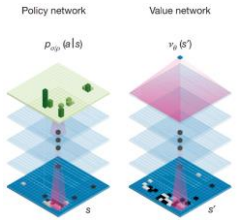
1. Introduction to Machine Learning

- ✓ Machine learning (ML) is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed.
- ✓ ML focuses on the development of computer programs that can access data and use it learn for themselves.

How to recognize cats?

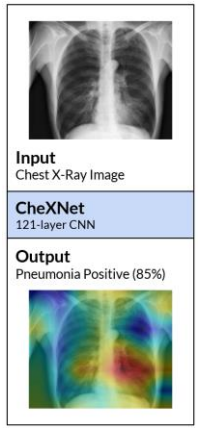
2012, Google AI built a neural network to recognize cats from ***unlabeled*** images using large scale unsupervised learning.





AlphaGo

2016, DeepMind developed a ***AlphaGo*** program to challenge professional Go players using reinforcement learning.

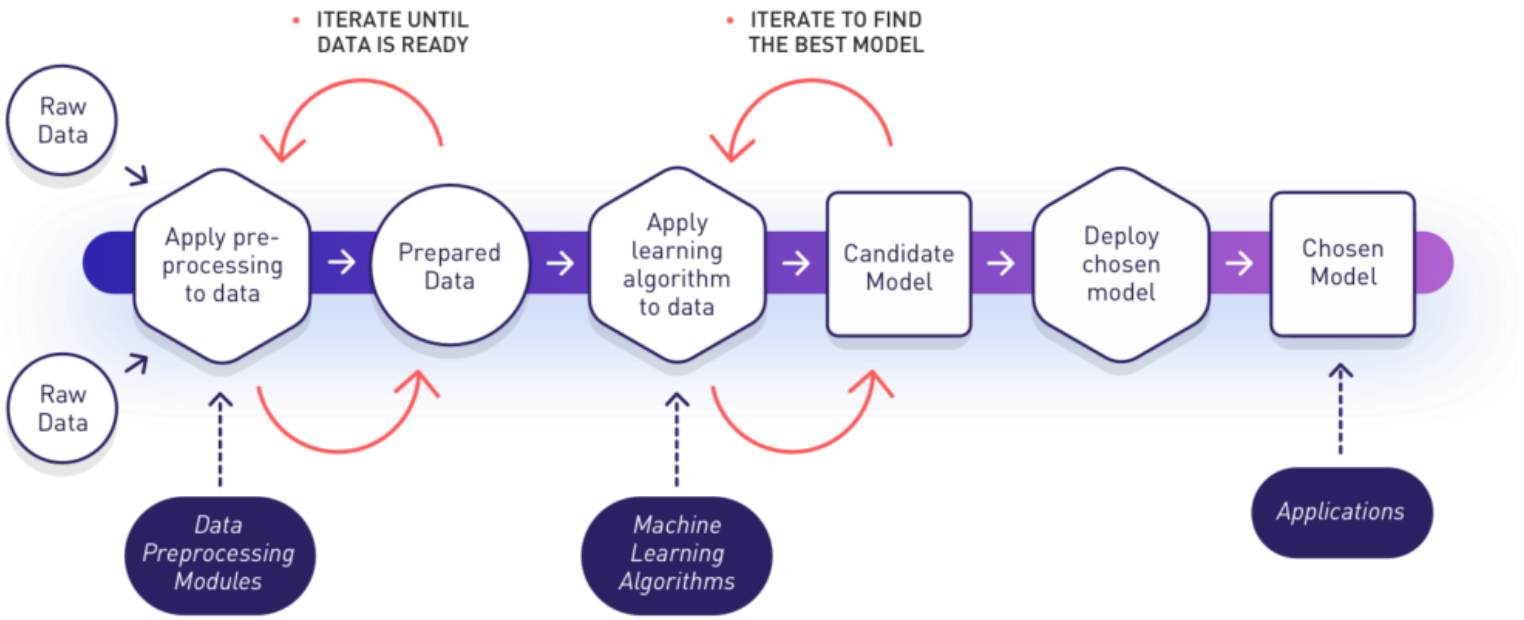


Pneumonia Detection

2017, Stanford University used 121-layer ***convolutional neural network*** to detect pneumonia from chest X-rays at a level exceeding practicing radiologists.

Image modified from: http://www.cognub.com/index.php/cognitive-platform/

1. Introduction to Machine Learning



A typical workflow for a machine learning project:

1) Integrate raw data sets from different data sources into a target database;

2) Clean target data by removing noise, duplicated, and inconsistent data;

3) Transform data into appropriate forms by dimension reduction or normalization;

4) Apply various machine learning algorithms to data set and select the best candidate model;

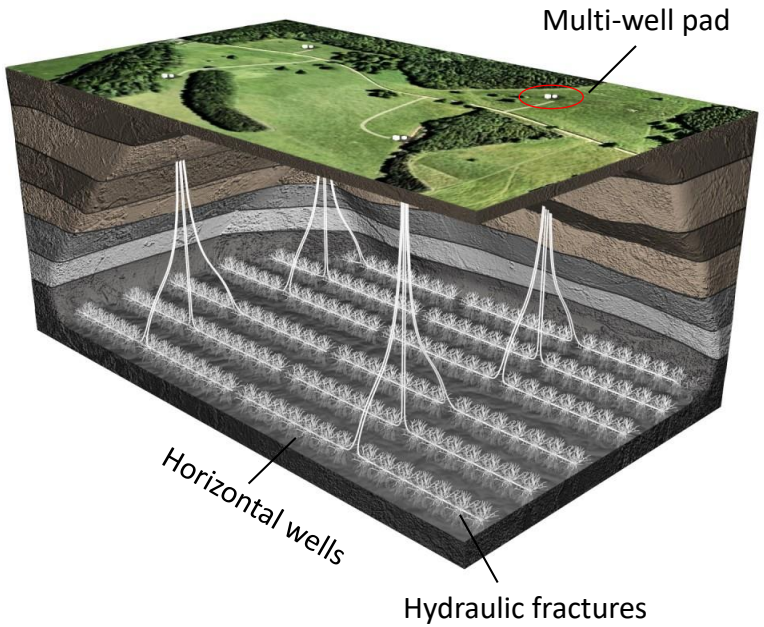
5) Use the machine learning models to help you make a decision.

Image source: https://www.pngfly.com/png-42dzkv/

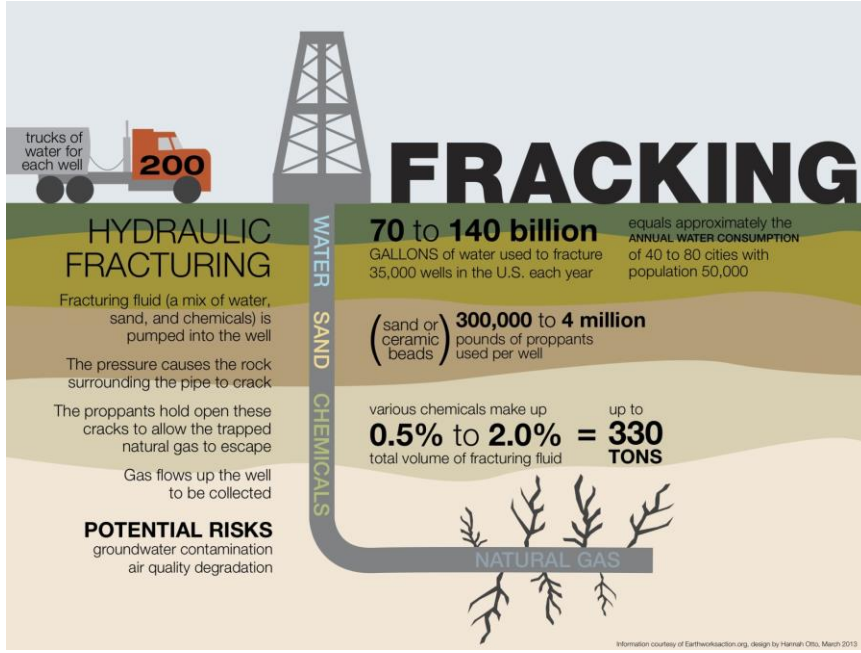
2. Unconventional Tight/Shale Reservoirs

- Very low permeability; Horizontal wells; Multi-stage hydraulic fracturing; Well pad with multiple wells,
- For well drilling + stimulation operations, engineers mainly concern about:
  - Well lateral length, well spacing, directions;
  - Fracturing fluid type and volume;
  - Proppants tonnage;
  - Fracture stages, spacing, clusters, etc.

Job size



Well pad with multiple wells



Injecting fluids to create fractures and proppants to keep them open

Image source 1: https://www.canadasnaturalgas.ca/en/explore-topics/hydraulic-fracturing
Image source 2: https://crudeoiltrader.blogspot.com/2012/09/pad-drilling-and-rig-mobility-lead-to.html
Image source 3: https://sites.psu.edu/fracking1/

2. Unconventional Tight/Shale Reservoirs

Project Objective: develop machine learning models to predict the well productivity and provide guidance on designing the hydraulic fracturing operations (e.g., fracturing fluid volume and proppant tonnage).

INPUT RAW DATA

↓

DATA PRE-PROCESSING → General Data Visualization

↓

MACHINE LEARNING

↓

Prediction

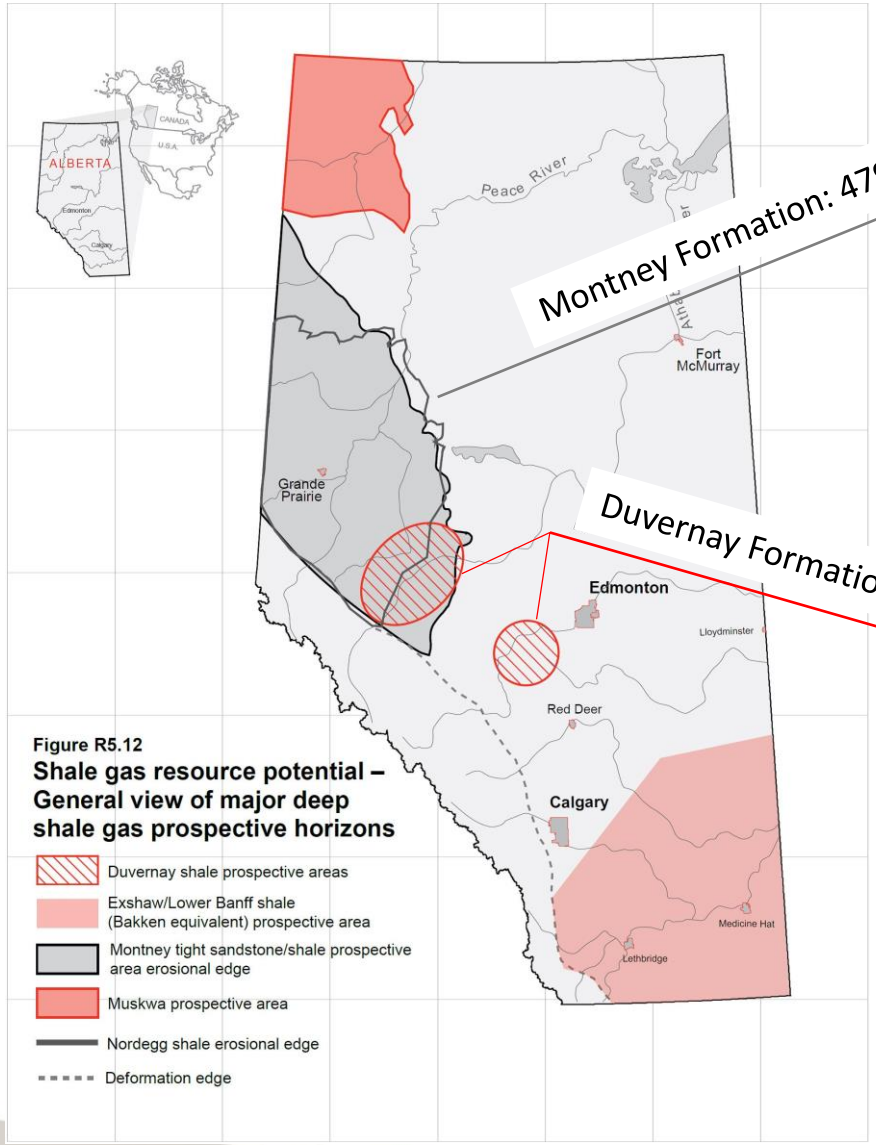
- Identify the importance of the features
- Clustering wells based on location or gas-oil-ratio (GOR)
- Select the most suitable machine learning algorithms
- Improve the performance of the chosen ML models
- Analyze the prediction uncertainty of ML models

↓

Optimization

- Identify hot zones that tend to yield highly productive wells
- Sensitivity analysis of each factor
- Optimize the hydraulic fracturing parameters

3. General Data Visualization



Montney Formation: 4790 wells

Duvernay Formation: 766 wells

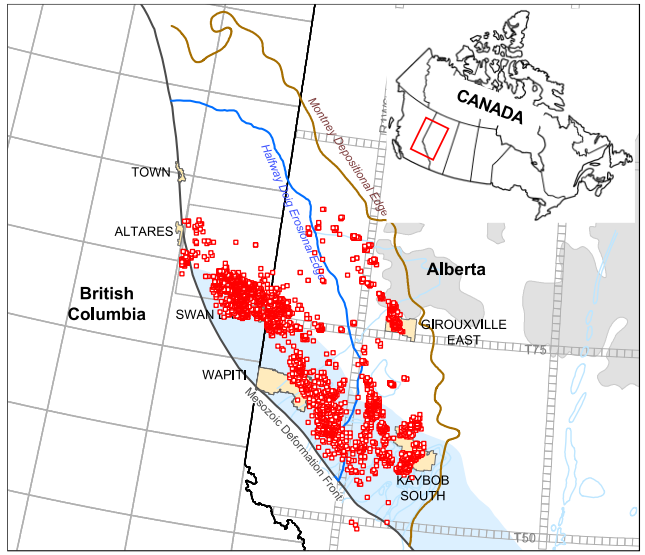


Image modified from: AER, 2014, ST98-2014, ISSN 1910-4235.

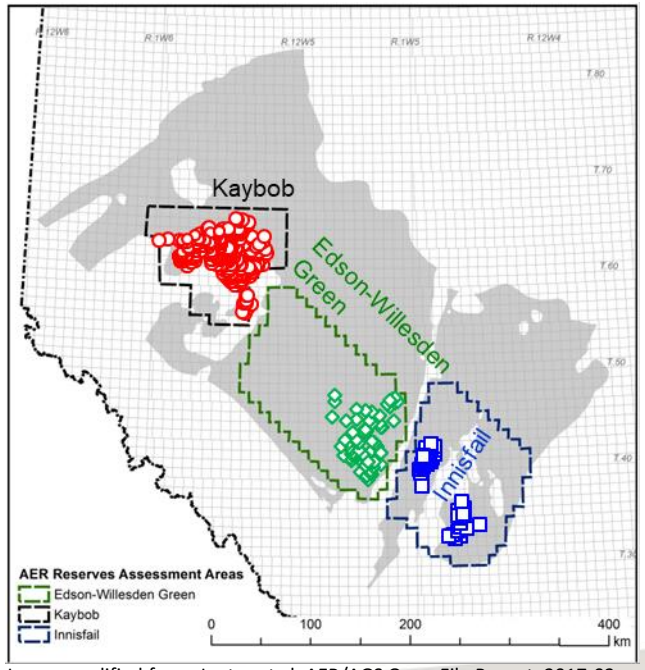
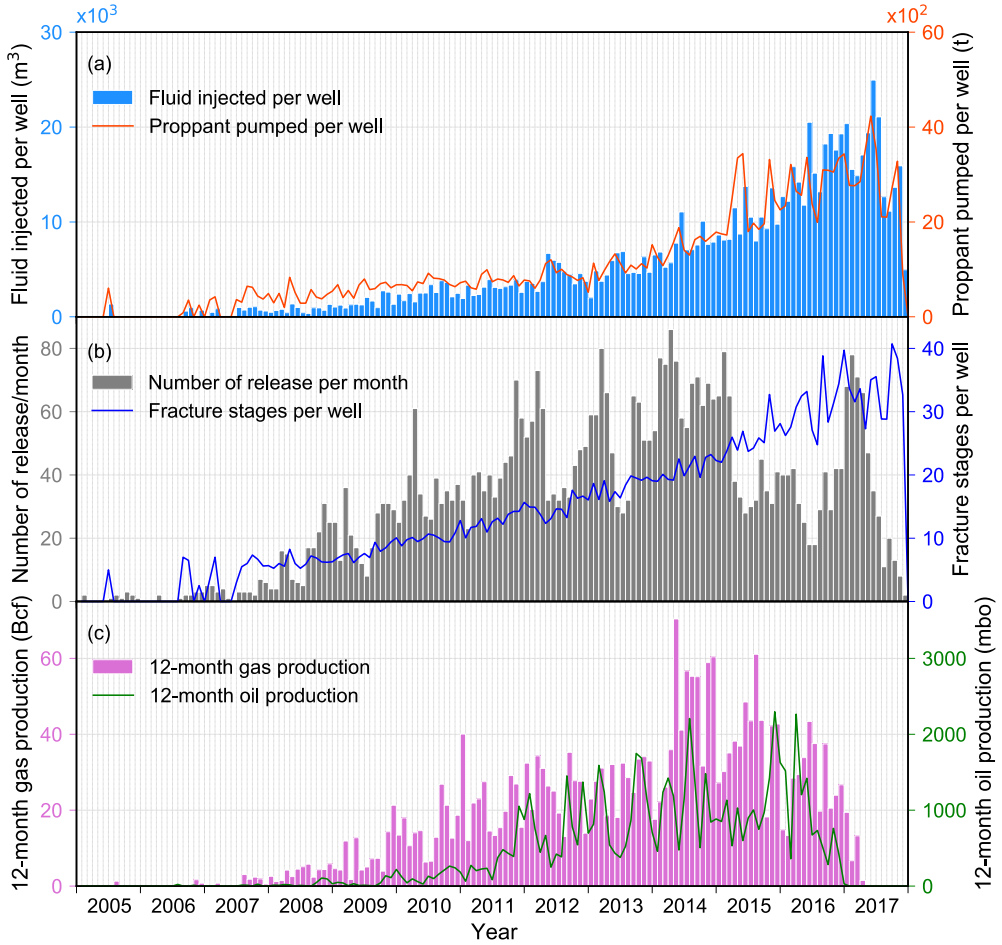
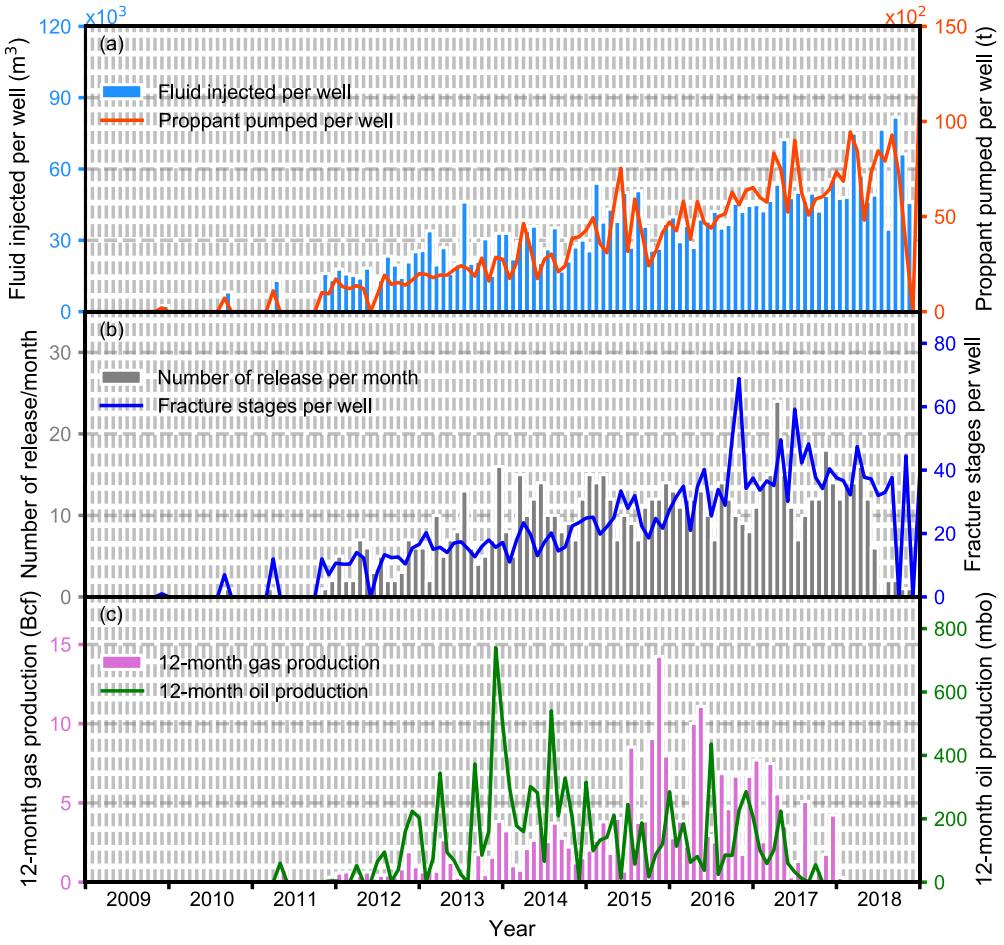


Image modified from: Lyster et al. AER/AGS Open File Report, 2017-02.

3. General Data Visualization



Development history of **Montney** Formations

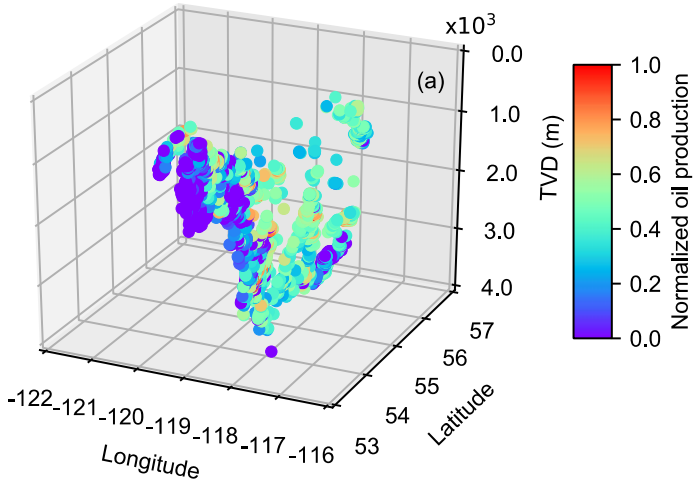


Development history of **Duvernay** Formations

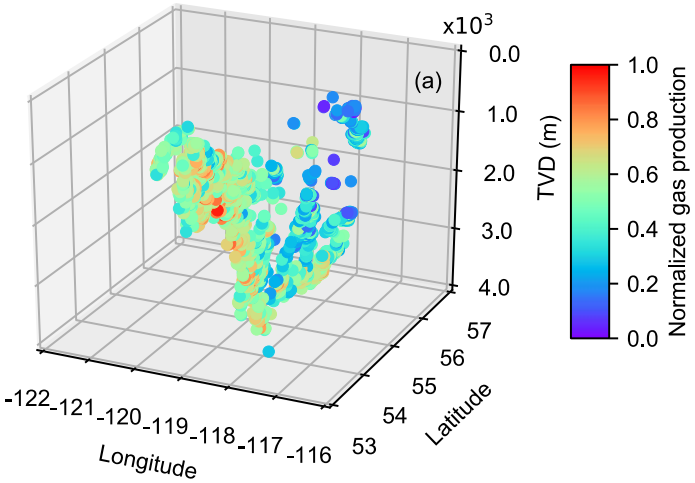
- Fluid/proppant injected in a Duvernay well is 2-3 times of that in a Montney well
- Similar fracture stages
- Gas and oil production for a Duvernay well is roughly half of that in a Montney well.

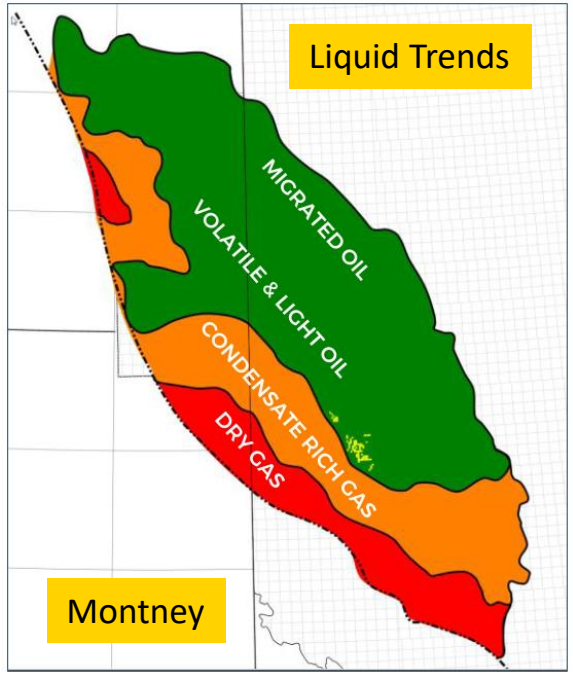
3. General Data Visualization

Oil Production Distribution

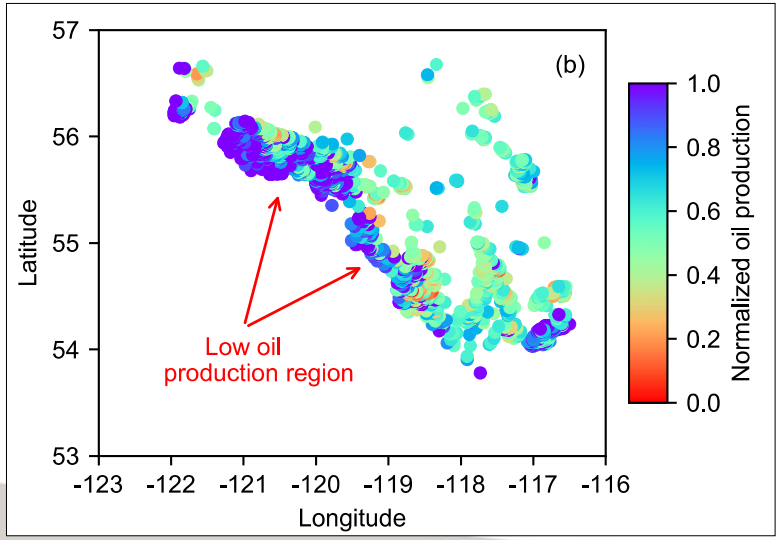


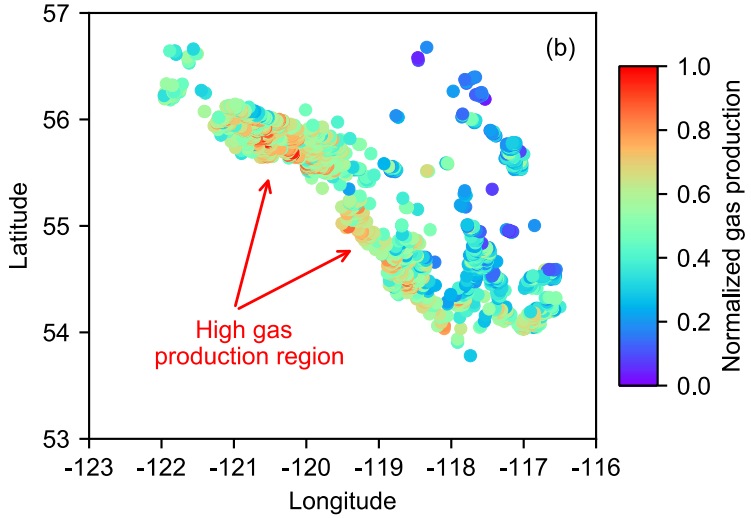
Gas Production Distribution

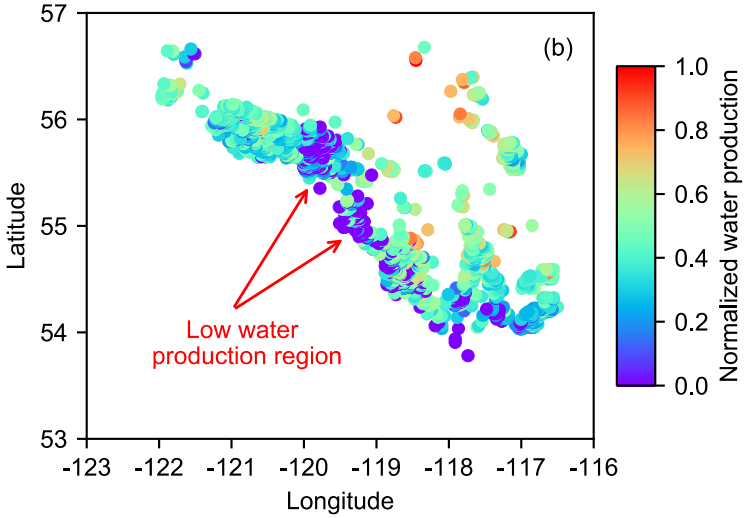




Figures modified from Presentation: http://ironbridgeres.com/wp-content/uploads/2018/05/CAPP-Q2-2018-Investor-Presentation-FINAL.pdf

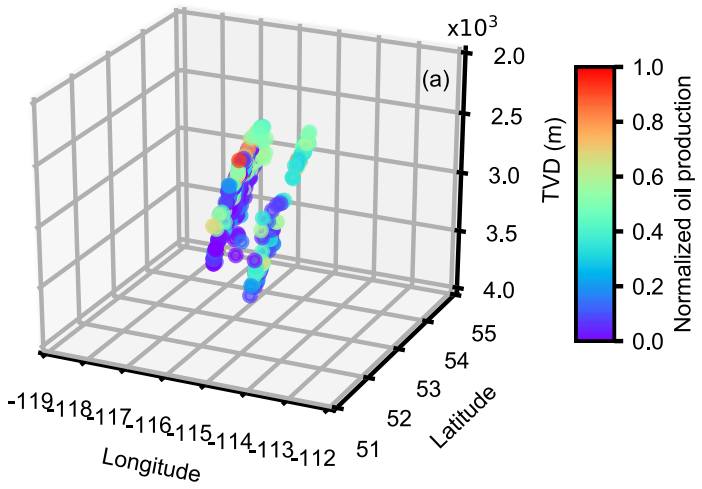




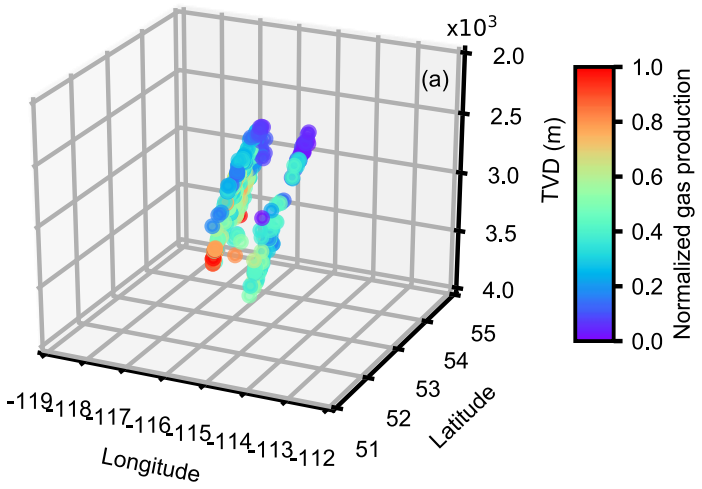


3. General Data Visualization

Oil Production Distribution



Gas Production Distribution



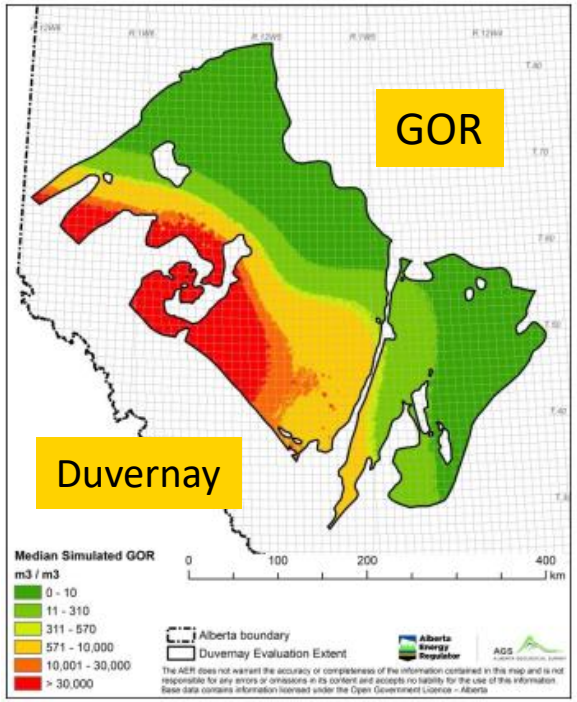
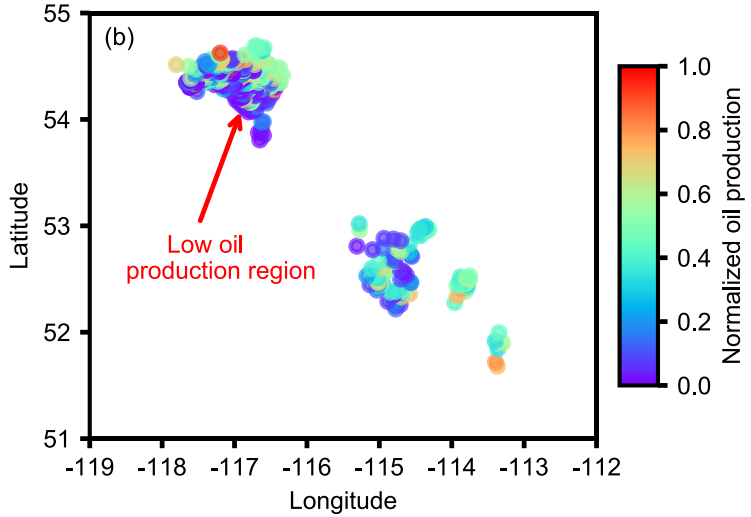
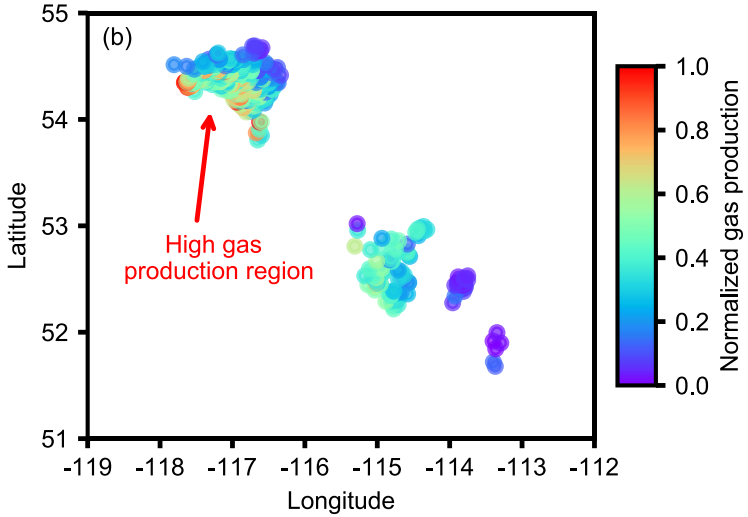
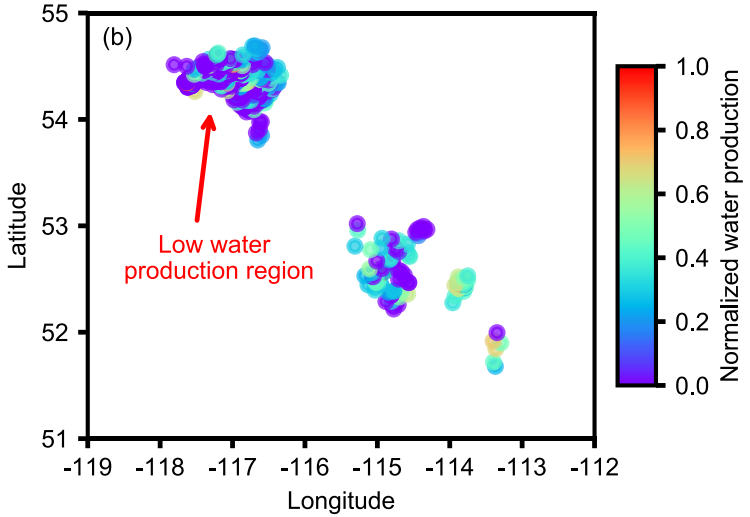


Image modified from: Lyster et al. AER/AGS Open File Report, 2017-02.







Machine Learning Modeling

- Case Study 1: Montney Formation

Objective: Develop a machine learning model to provide guidance on designing the hydraulic fracturing process and predict the well productivity in the Montney Formations.

UNIVERSITY OF CALGARY

INPUT RAW DATA → MACHINE LEARNING → Prediction / Optimization

Type	Parameters
Well information	Latitude
	Longitude
	Well TVD
	Lateral length
	Well completion type
	Wellbore direction
Fracture stimulation design	Total proppant
	Total fracturing fluid
	Fracturing fluid type
	Fracture stages
	Use of interval clusters
Production	12-months BOE

BOE: barrels of oi equivalent
TVD: true vertical depth

Input features

$$y = f(a, b, c, \cdots)$$

Output

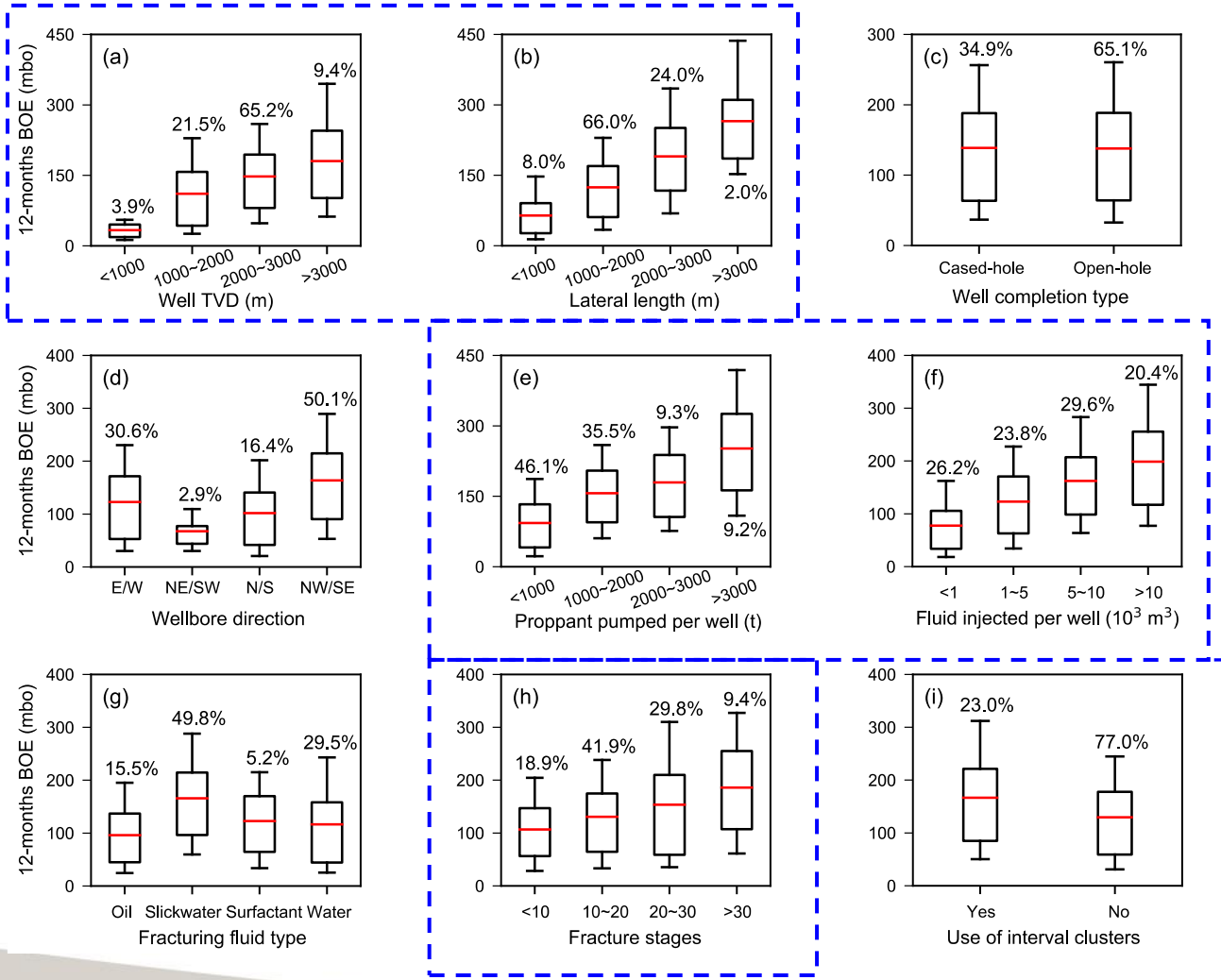
Well completions and frac database by geoLOGIC systems

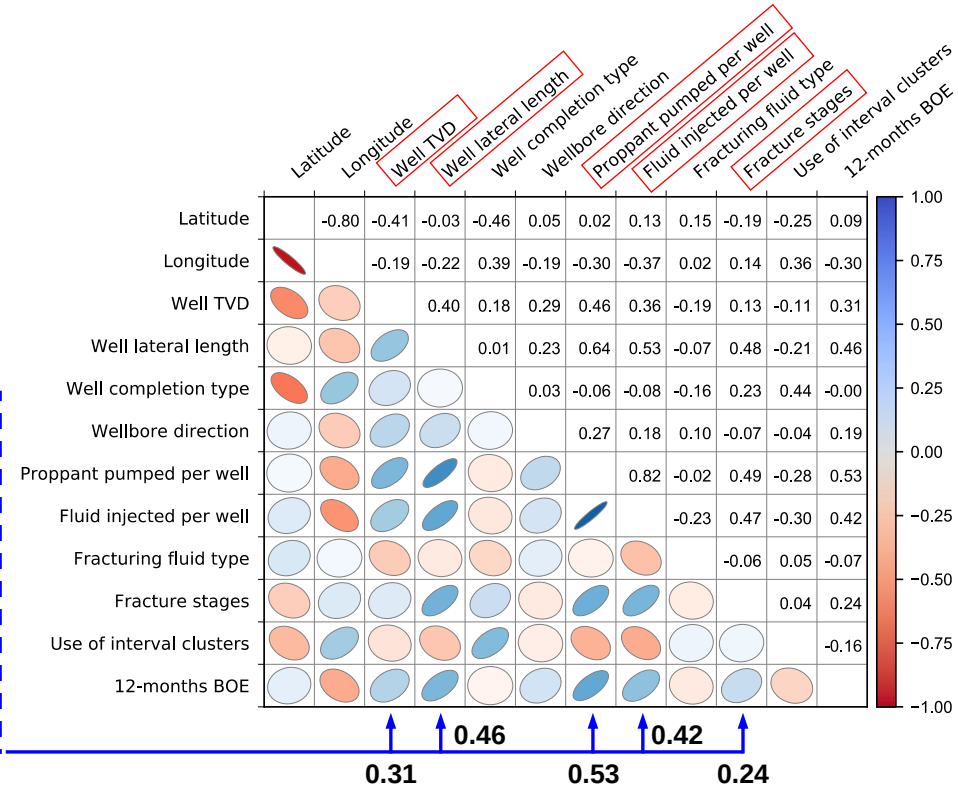
(1) Feature Importance

- 0 – no linear relationship between two variables
- 1 – strong positive linear relationship
- -1 – strong negative linear relationship

UNIVERSITY OF CALGARY

- Relationship between stimulation parameters and 12-month BOE





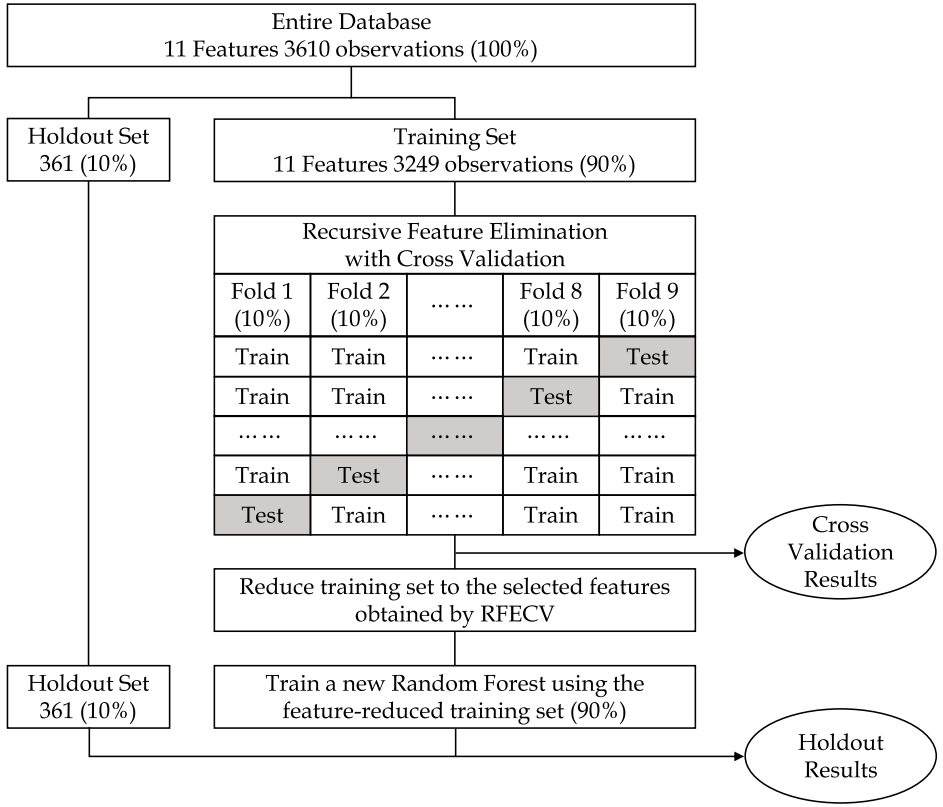
- **Strong positive factors:**
  - Proppant pumped per well (R=0.53)
  - well lateral length (R=0.46)
  - fluid injected per well (R=0.42)
- **Proppant tonnage has the greatest positive impact on 12-month BOE**
- **Proppant tonnage and fluid volume is highly correlated (R=0.82)**

(2) Feature Selection

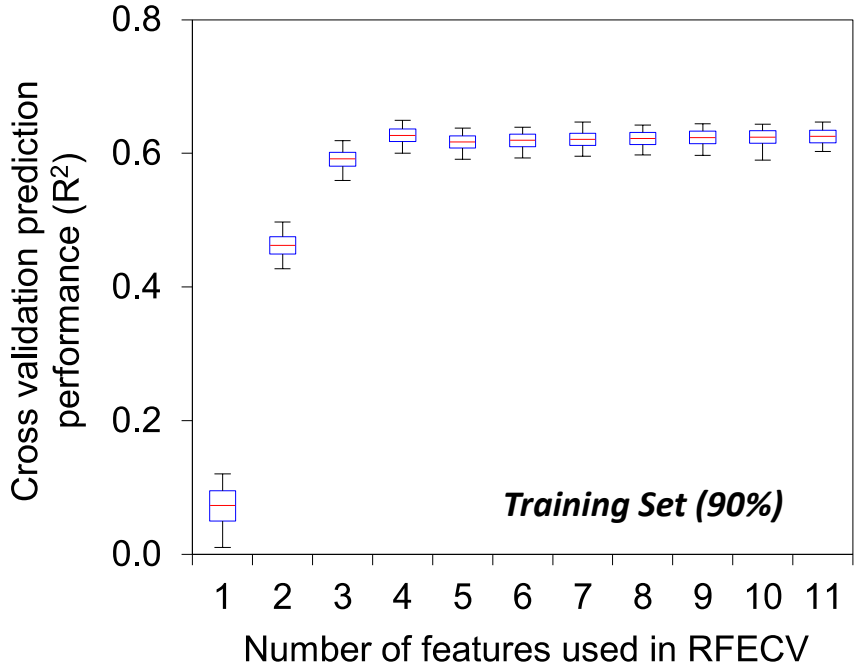
- Select features that contribute most to the prediction variable or target;
- Reduce the training time, improve the accuracy of prediction models and avoid the overfitting problems;
- ***Recursive Feature Elimination with cross validation:*** recursively removes features, then builds models using the remaining features and calculates model's accuracy.

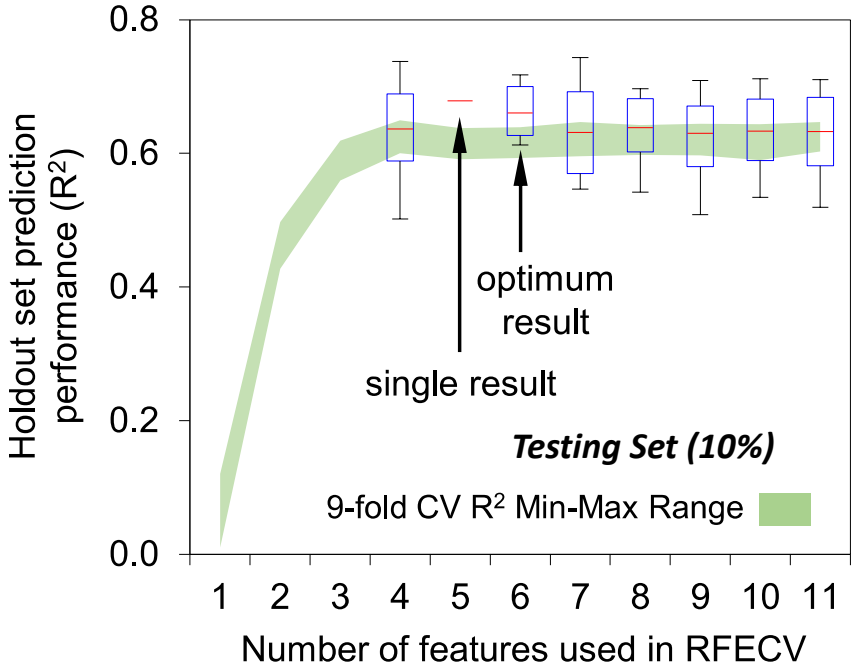
Type	Parameters
Well information	Latitude
	Longitude
	Well TVD
	Lateral length
	Well completion type
	Wellbore direction
Fracture stimulation design	Total proppant
	Total fracturing fluid
	Fracturing fluid type
	Fracture stages
	Use of interval clusters
Production	12-months BOE

Well completions and frac database by geoLOGIC systems



(2) Feature Selection

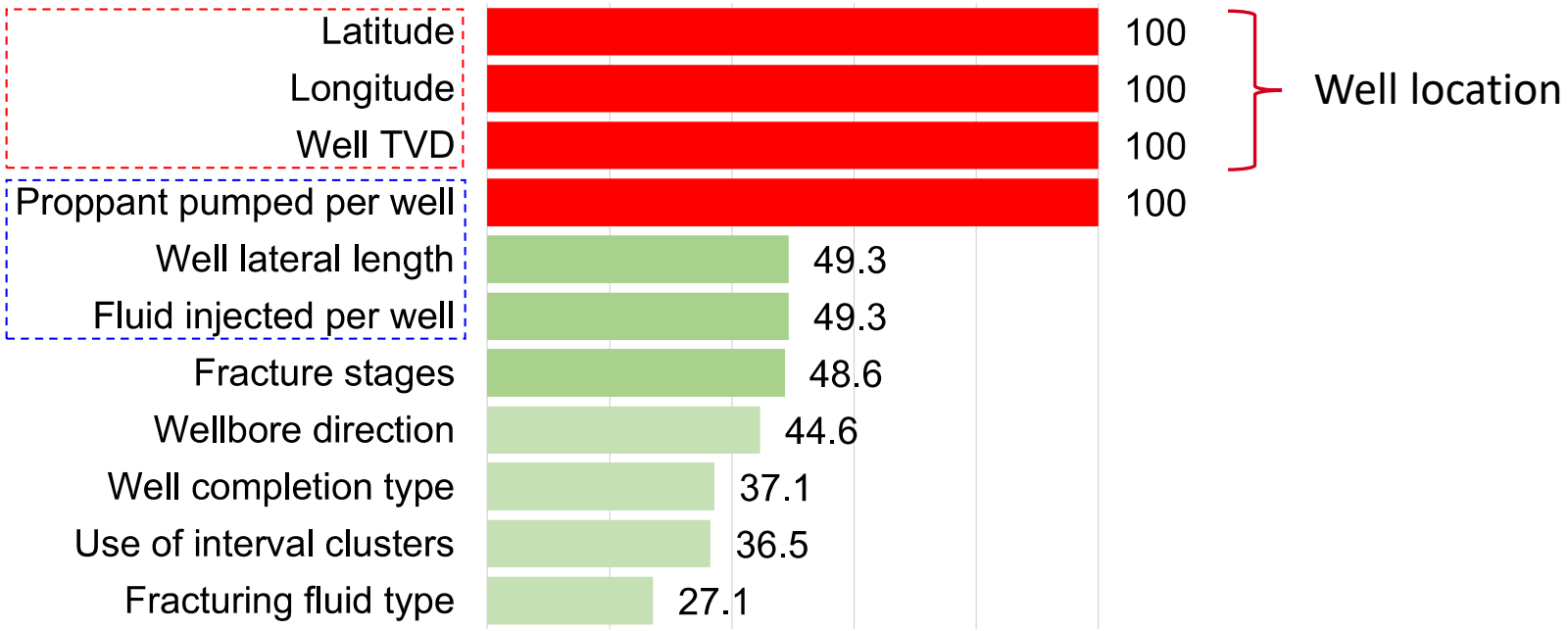




- If the number of features is less than 4, the prediction accuracy of the model is significantly reduced;
- The model with 6 features has the largest mean value and a relatively lower standard deviation;
- Random forest is used as the prediction model in the RFECV process.

(2) Feature Selection

The frequency of features included through RFECV in percent of number of model runs (e.g., 100 model runs, then 50% means this feature was selected by 50 models).



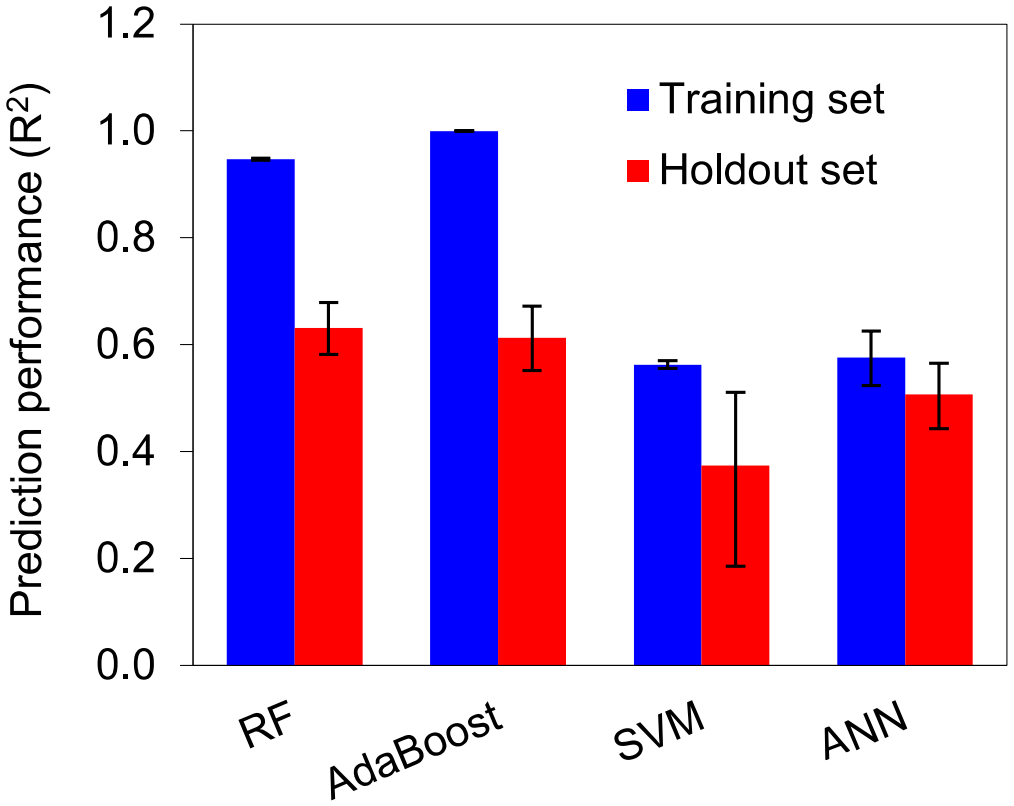
- 6 features are selected by using RFE;
- Well locations/geological properties: well latitude, longitude, true vertical depth;
- Well completion and stimulation strategy: proppant pumped per well, well lateral length, fluid injected per well.

(3) ML Algorithms Comparison

- Machine Learning algorithms
  - Random Forest (RF)
  - Adaptive Boosting (AdaBoost)
  - Support Vector Machine (SVM)
  - Artificial Neural Network (ANN)

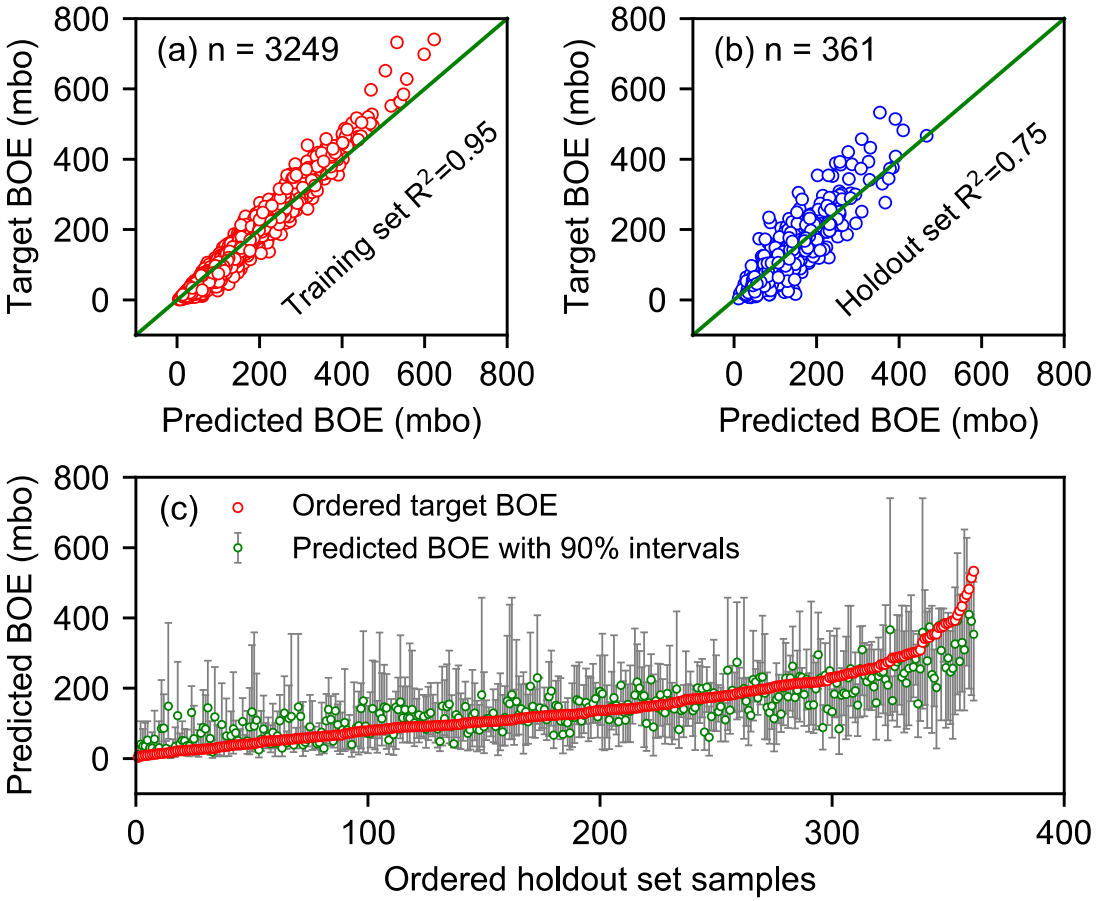
Machine learning algorithms	Training set (90%) Average R^2	Holdout set (10%) Average R^2
RF	0.9468	0.6310
AdaBoost	0.9998	0.6131
SVM	0.5621	0.3736
NN	0.5758	0.5064

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$



- For each ML algorithm, a total of 1000 model runs are performed and average result is shown;
- For each model run, 90% data is used to train the model, and 10% data is used to test the model;
- Random Forest performs best and overcoming the overfitting problems compared to other methods.

(4) Prediction Uncertainty of the Random Forest



- The values of R^2 for training and holdout set are calculated to be 0.95 and 0.75, respectively.
- Developed RF-based production forecasting model is reasonable and robust for fracture stimulation design.

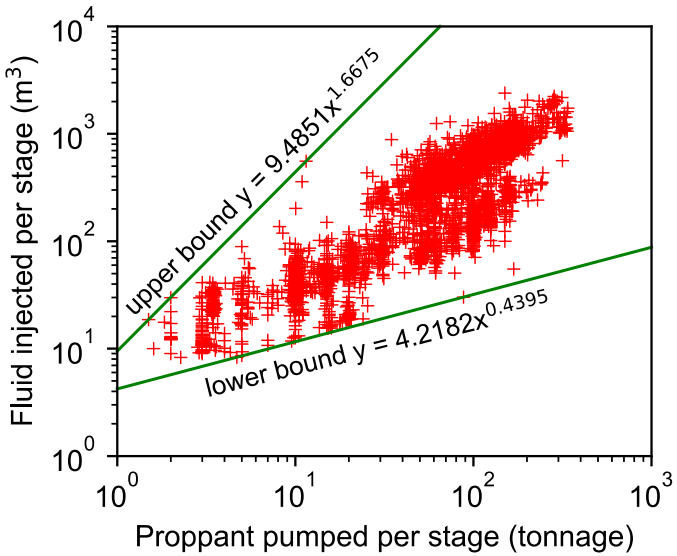
(5) Fracture Design Optimization

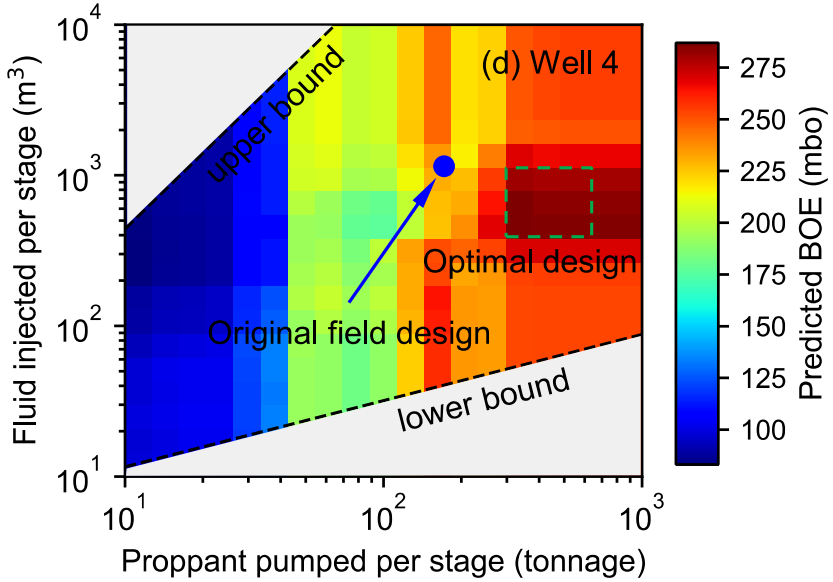
- Case study: optimal proppant and fracturing fluid
  - Applying the production forecasting model to optimize the fracture stimulation design in the field;
  - Well was drilled along the northwest direction;
  - Open-hole completion without interval clusters.

Parameters	Well
Well TVD (m)	3196.1
Lateral length (m)	2548.82
Well completion type	open-hole
Wellbore direction	NW
Total proppant (tonnage)	4117.3
Total fluid (m^3)	27550.1
Fracturing fluid type	Slickwater
Number of stages	24
Use of interval clusters	No
12-months BOE (mbo)	229.6

(5) Fracture Design Optimization

- RF model is applied to predict the 12-month BOE;
- Field cases are applied to set the upper/lower boundary;
- Effect of proppant tonnage on the predicted BOE is more significant compared to that of the fluid volume;
- Proppant tonnage per stage is better to be higher than 100 ton/stage; the fluid volume per stage to be higher than 300 m^3 per stage;
- Need to consider the cost of the fluid and proppant;
- Too much fluid injected per stage could lead to the fluid leakage, water blockage around the hydraulic fractures, and reduction in the well productivity;
- On the contrary, when the fracturing fluid is less than 300 m^3 per stage, proppant suspension in the fracturing fluid could be a problem.

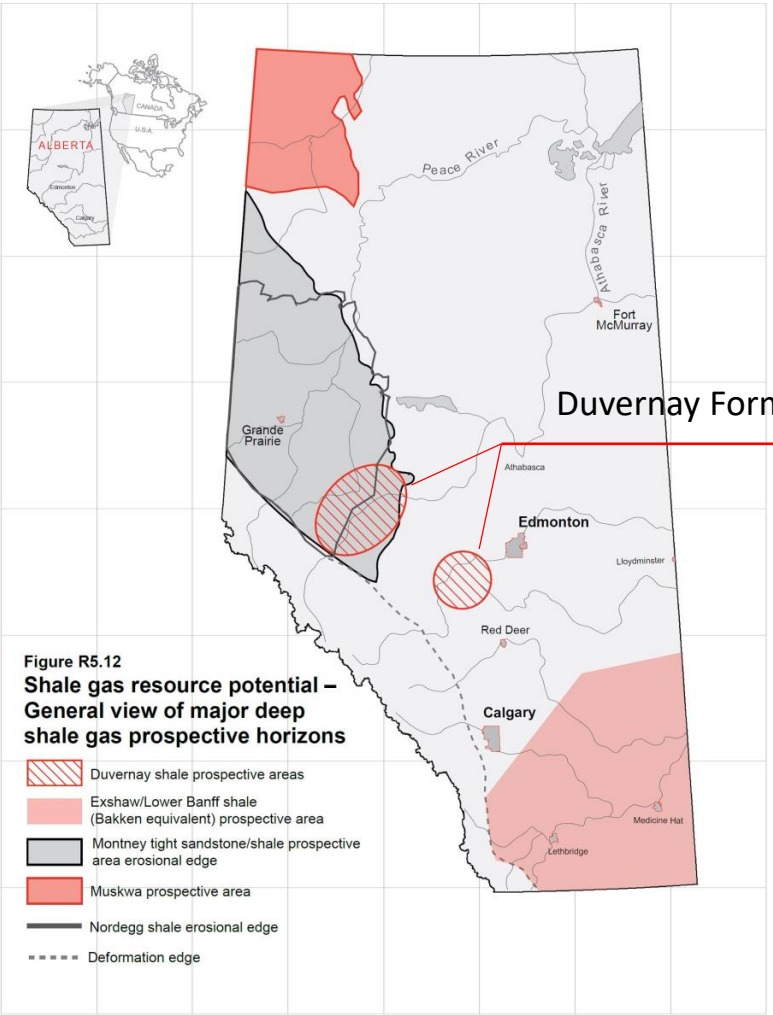




Machine Learning Modeling

- Case Study 2: Duvernay Formation

(1) Well Location in Duvernay Formation



Duvernay Formation: 766 wells

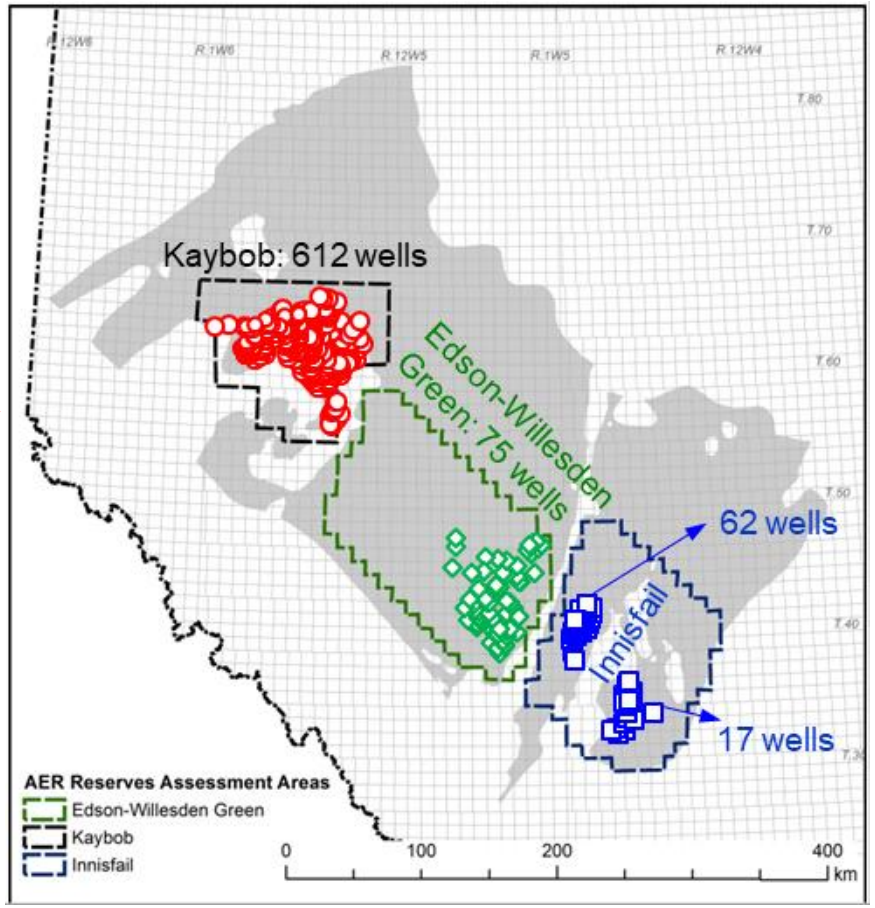


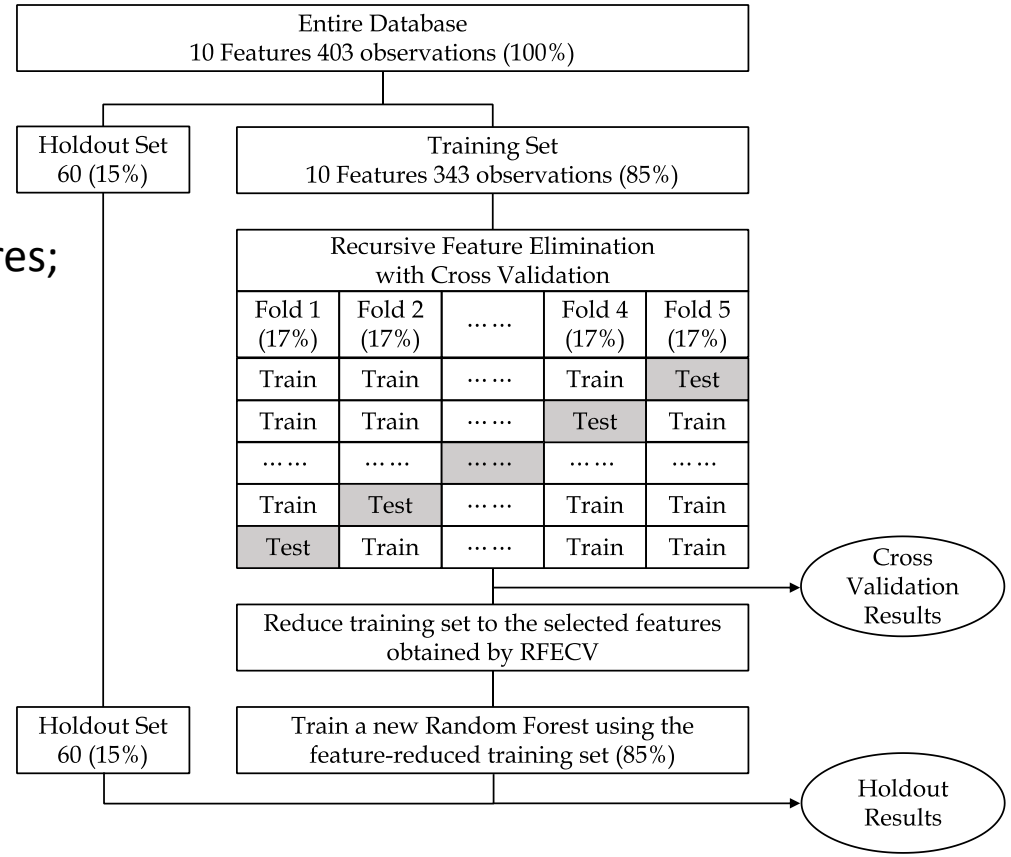
Image modified from: AER, 2014, ST98-2014, ISSN 1910-4235.

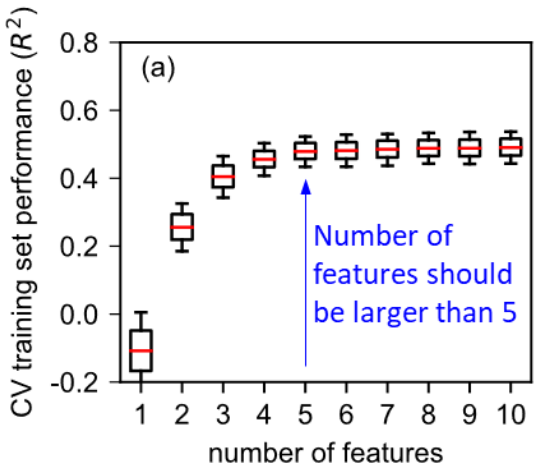
Image modified from: Lyster et al. AER/AGS Open File Report, 2017-02.

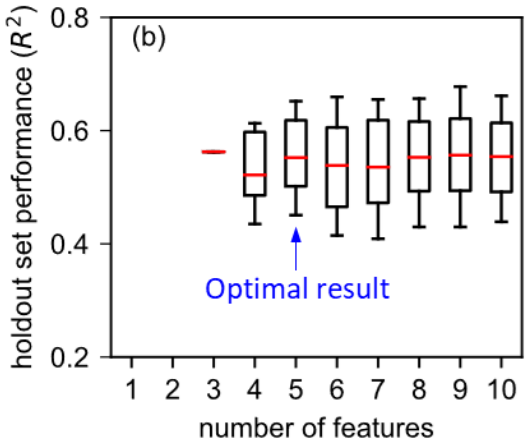
(2) Feature Selection

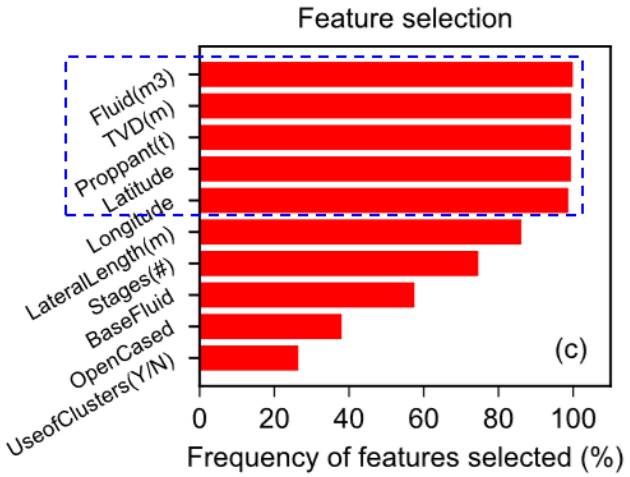
- Wells in the Kaybob Field are used (403 wells);
- Recursive feature elimination is used to select important features;
- Natural gas equivalent as target.

No.	Parameter type	Parameters
1	Well information	Latitude
2		Longitude
3		Well TVD (m)
4		Lateral length (m)
5		Well completion type (Open-hole or cased)
6	Fracture stimulation design	Total proppant placed per well (tonnage)
7		Total fracturing fluid pumped per well (m^3)
8		Fracturing fluid type (Water/Slickwater)
9		Number of fracture stages
10		Use of interval clusters (Yes/No)
11	Production data	12-month cumulative gas production (mmcf)

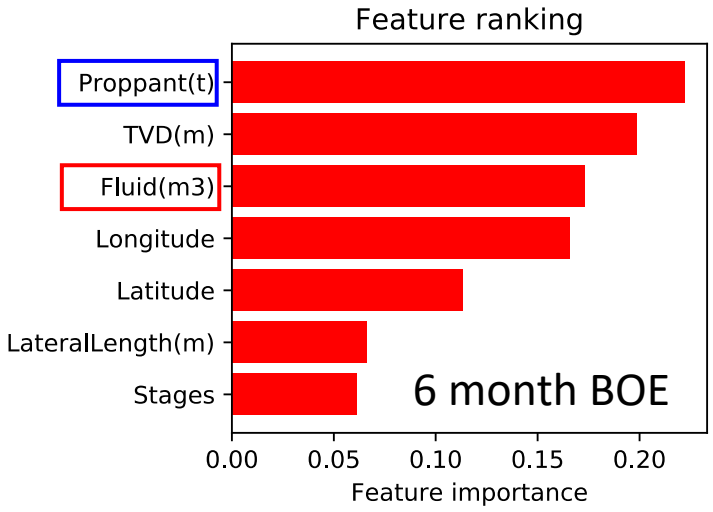


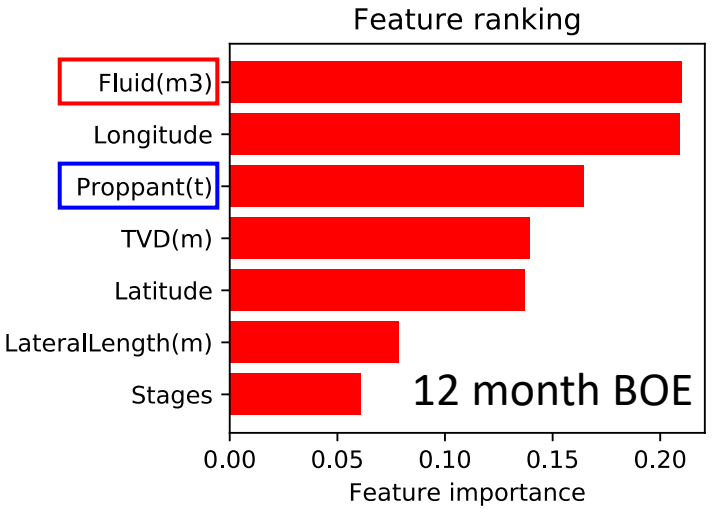


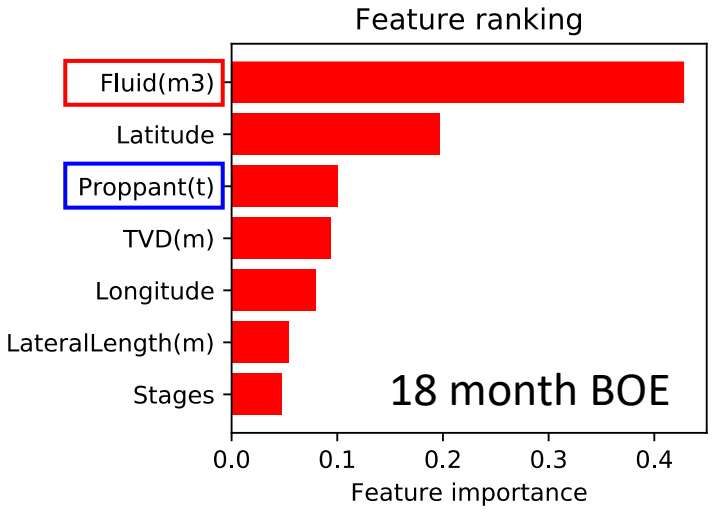




(3) Feature importance





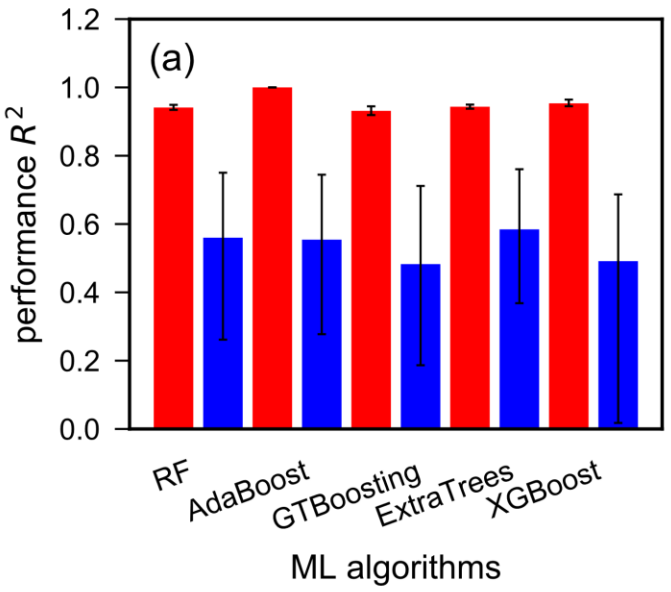


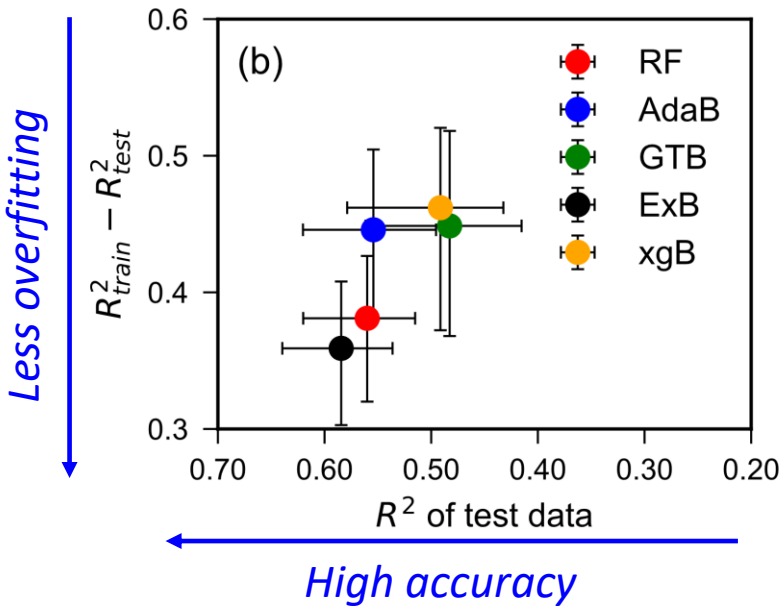
- Propped fracture has great contribution to the gas and condensate productions in short-term production (6 months);
- The un-propped fracture is the most important factor for the long-term production (12-18 months);
- While in Montney, well location and proppant are always selected as the most important factors.

(4) ML Algorithms Comparison

- Machine Learning ***Ensemble Models***
  - Random Forest (RF)
  - Adaptive Boosting (AdaBoost)
  - Gradient Boosting (GTBoost)
  - Extra Trees Boosting (ExBoost)
  - XGBoost

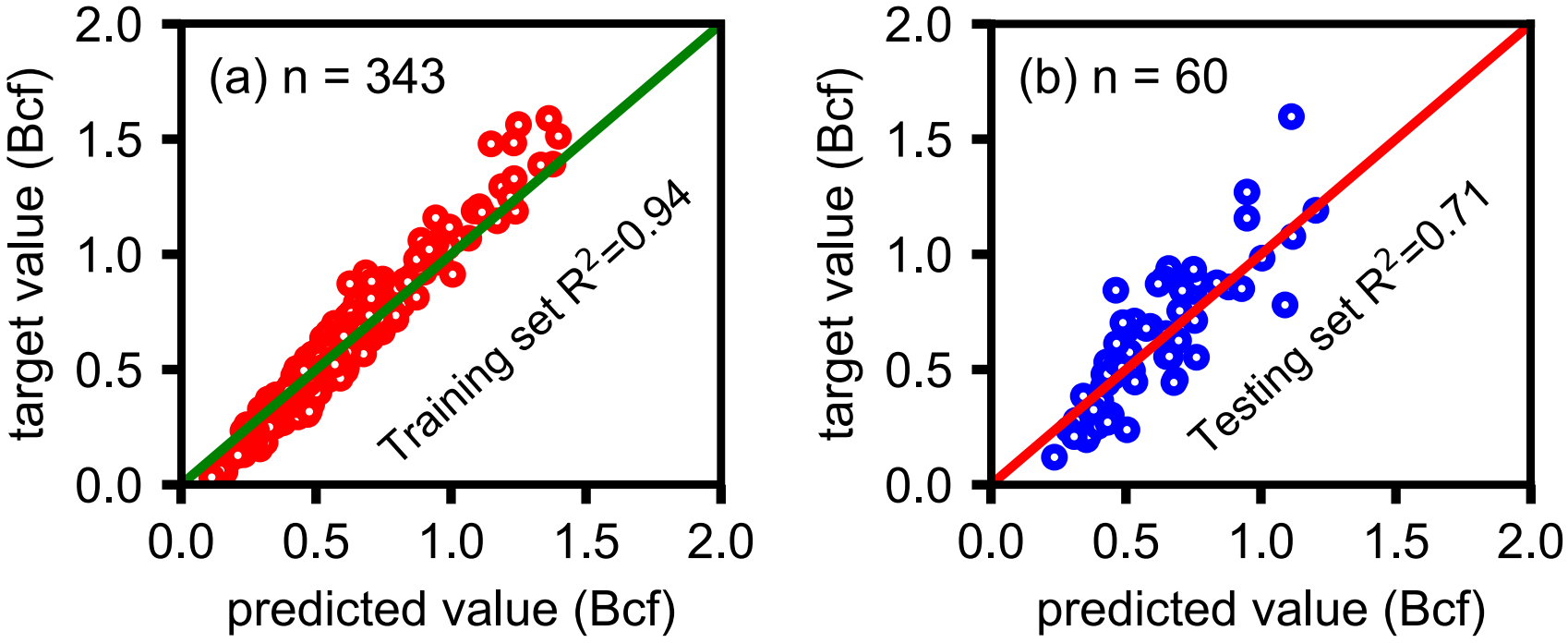
$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$





- For each ML algorithm, a total of 100 model runs are performed and average result is reported;
- For each model run, 85% data is used to train the model, and 15% data is used to test the model;
- Extra-Trees Boosting method is found to be the best prediction model with the lowest overfitting problems and highest prediction accuracy.

(5) Prediction Results for Duvernay (Kaybob Field)



- The values of R^2 for training and holdout set are calculated to be 0.94 and 0.71, respectively.
- Developed Extra Tree Boosting-based production forecasting model is reasonable and robust for fracture stimulation design.

(6) ML Models for Other Fields

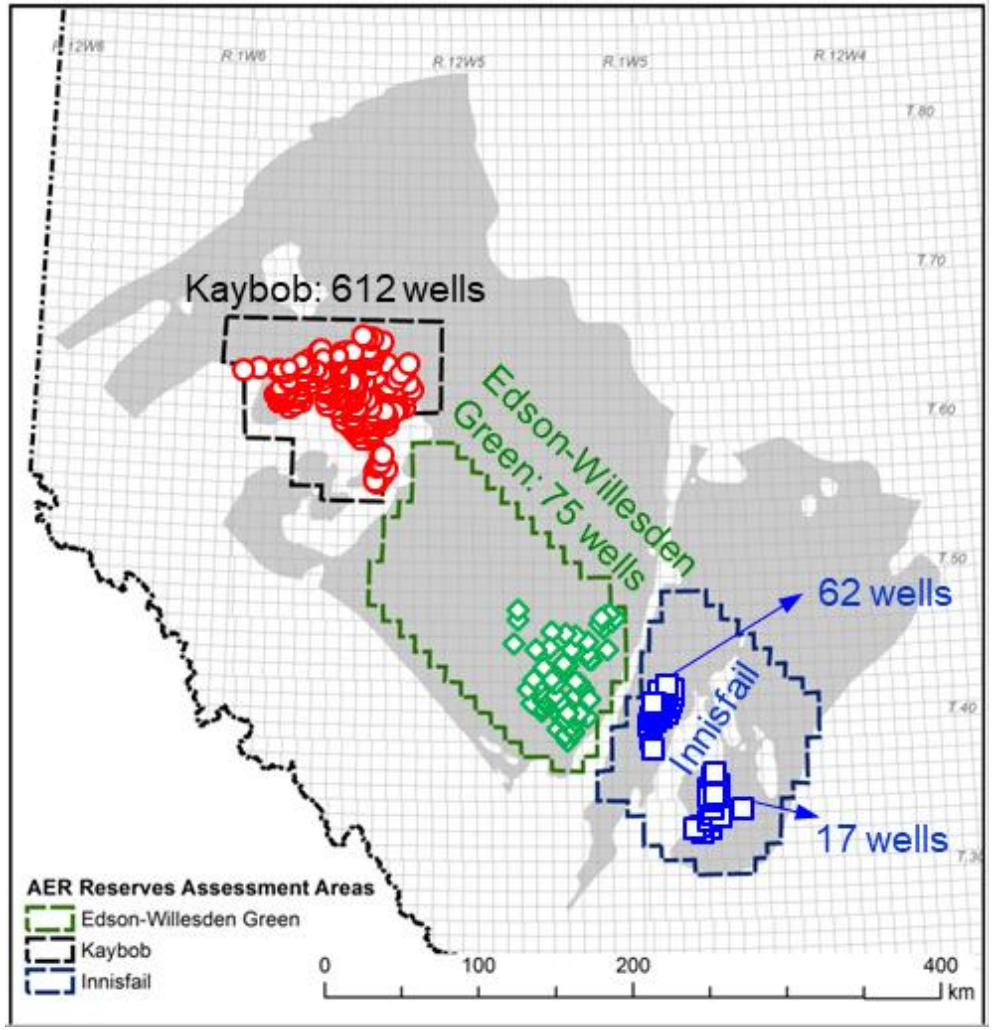
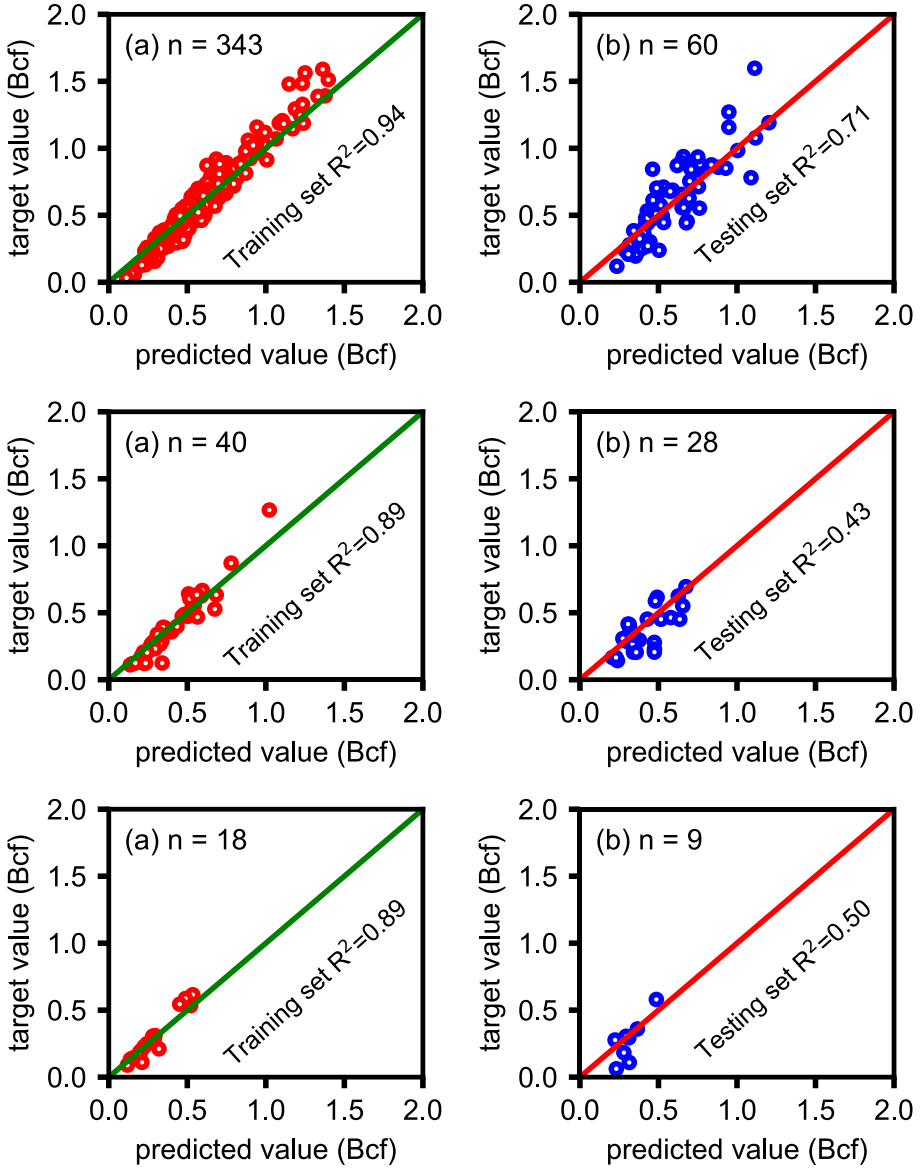


Image modified from: Lyster et al. AER/AGS Open File Report, 2017-02.

Kaybob Area

Edson-Willesden Green

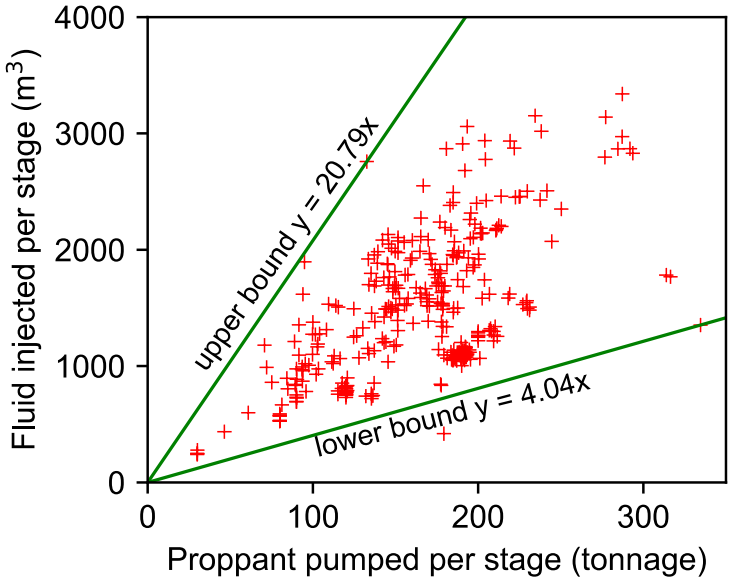
Innisfail Area

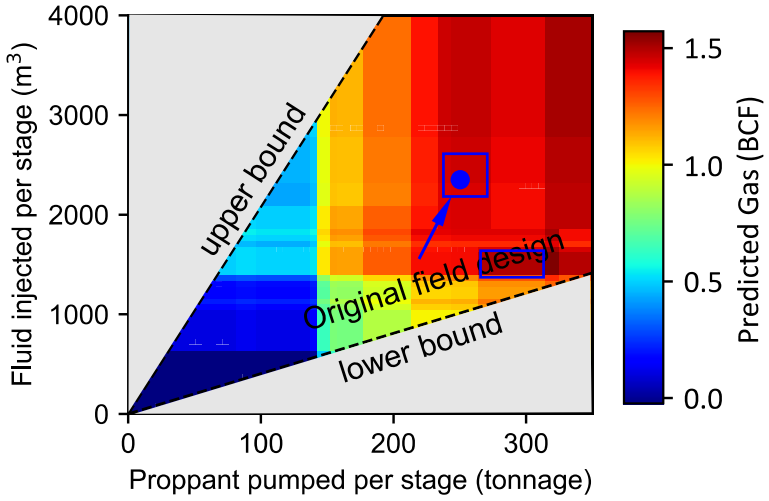


(7) Fracture Design Optimization

- Case study: optimal proppant and fracturing fluid
  - Applying the production forecasting model to optimize the fracture stimulation design in the field.

No.	Parameters	Well
1	Well TVD (m)	3749
2	Lateral length (m)	2230
3	Well completion type (Open-hole or cased)	Cased
4	Total proppant placed per well (tonnage)	5506.64
5	Total fracturing fluid pumped per well (m^3)	51728.37
6	Fracturing fluid type (Water/Slickwater)	Slickwater
7	Number of fracture stages	22
8	Use of interval clusters (Yes/No)	Yes
9	12-month cumulative gas production (Bcf)	1.2

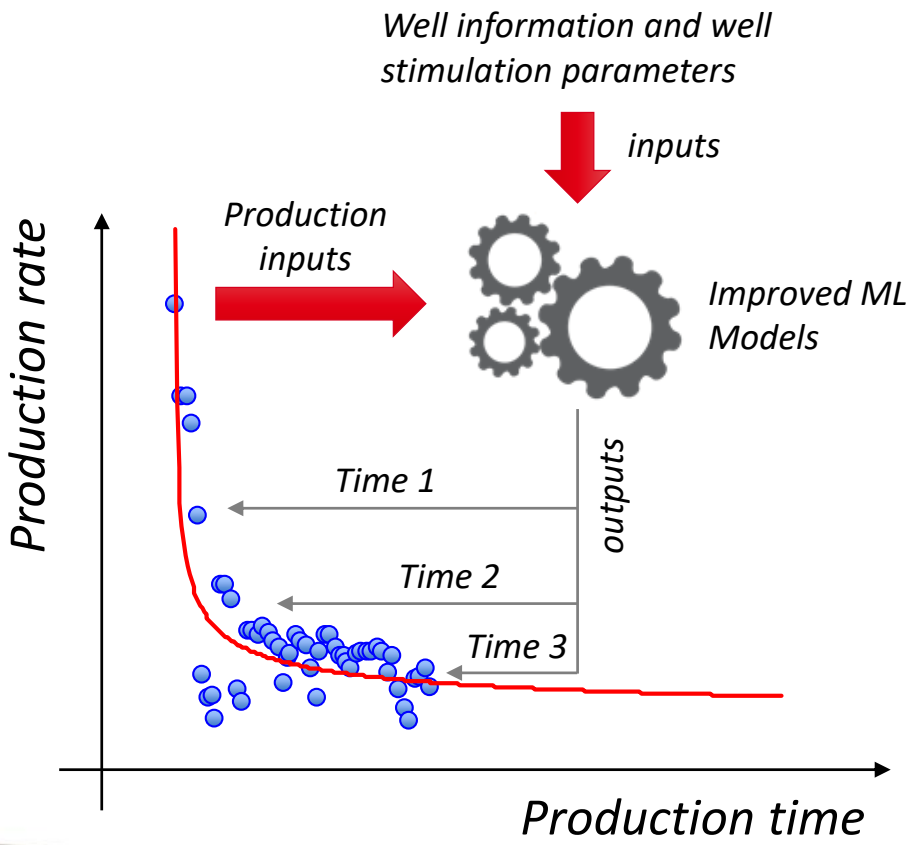




Time Series Study in Montney

Time Series Study in Montney

- Integrating production data into ML training process.
  - Production data is a ***time-series*** data set. The later-time production data is closely related to the previous data.
  - Incorporate the previous production data into input variables.



- Using 6 month BOE to better predict 12 month BOE:

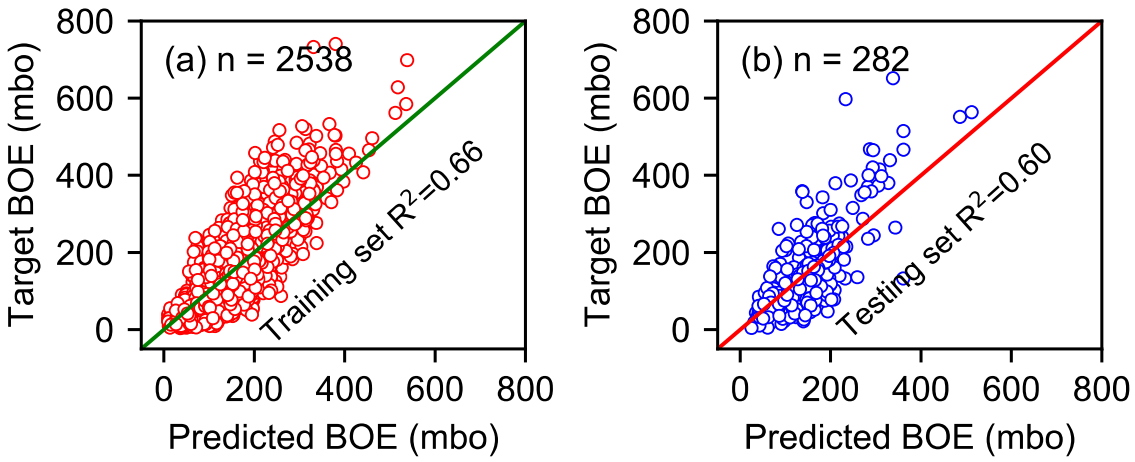
No.	Parameter type	Parameters
1	Well information	Latitude
2		Longitude
3		Well TVD
4		Lateral length
5		~~Well completion type~~
6		~~Wellbore direction~~
7	Fracture stimulation design	Total proppant
8		Total fracturing fluid
9		~~Fracturing fluid type~~
10		~~Fracture stage~~
11		~~Use of interval clusters~~
12	**Production data**	**6-month BOE**
13	Production data	12-month BOE

Input features

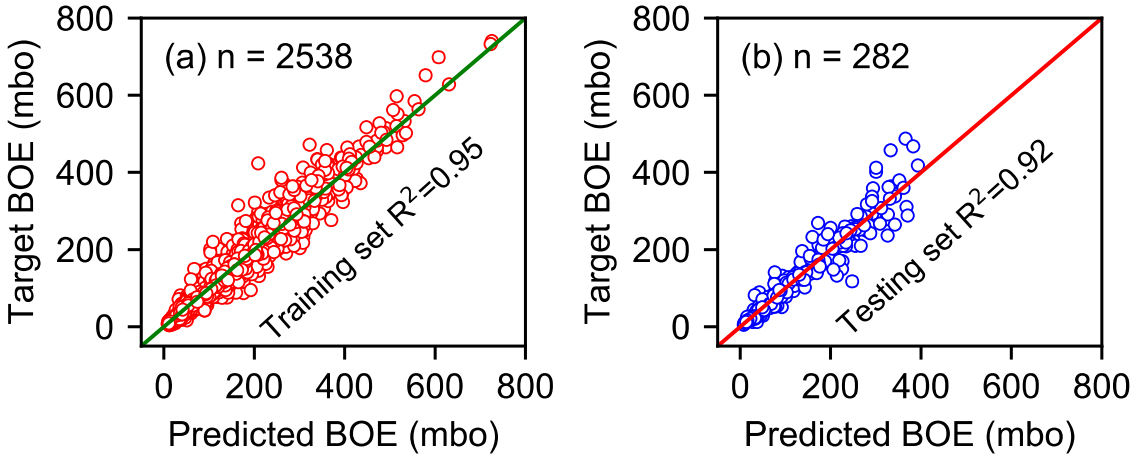
$$y = f(a, b, c, \cdots)$$

Output

- 12-month BOE prediction without 6-month input variable



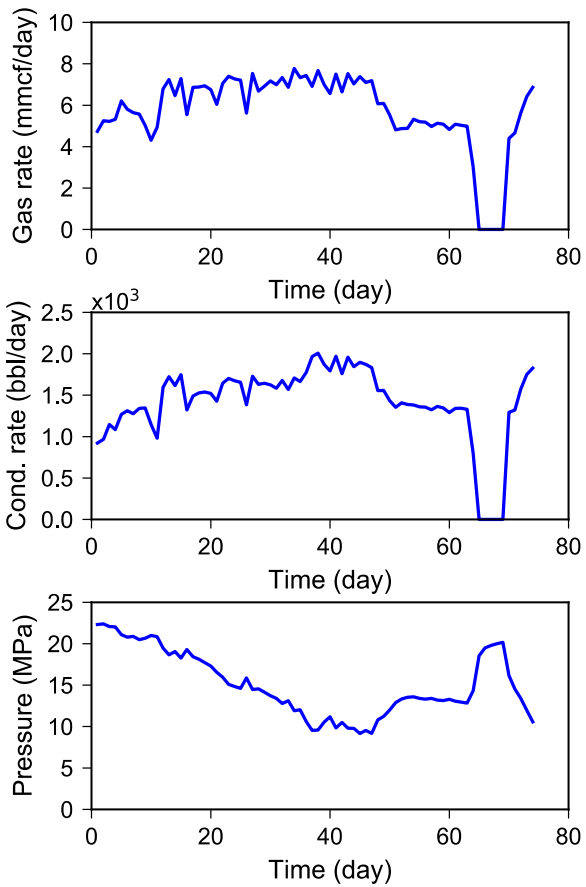
- 12-month BOE prediction with 6-month input variable



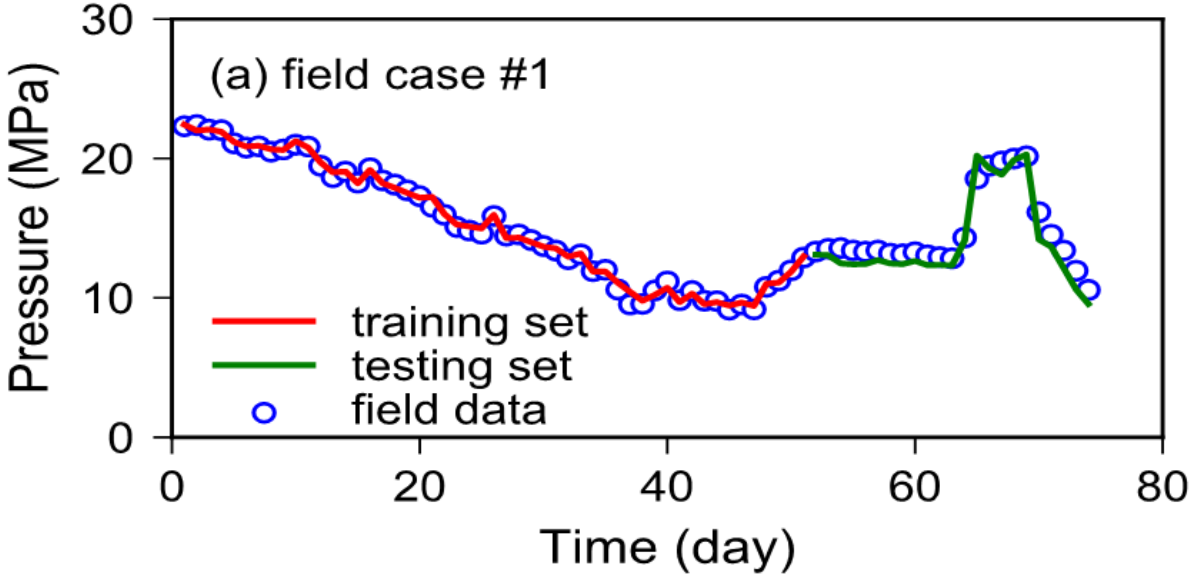
Prediction accuracy is significantly improved!!

Well-Testing Data Interpretation - Long Short-Term Memory Network

- ✓ The interpretation model for the horizontal well in liquid-rich tight formations is difficult to construct due to the complex flow mechanisms due to matrix/fractures.
- ✓ The data-driven models developed by LSTMs may provide an alternative approach to learn the pressure transient behavior from the real field data.



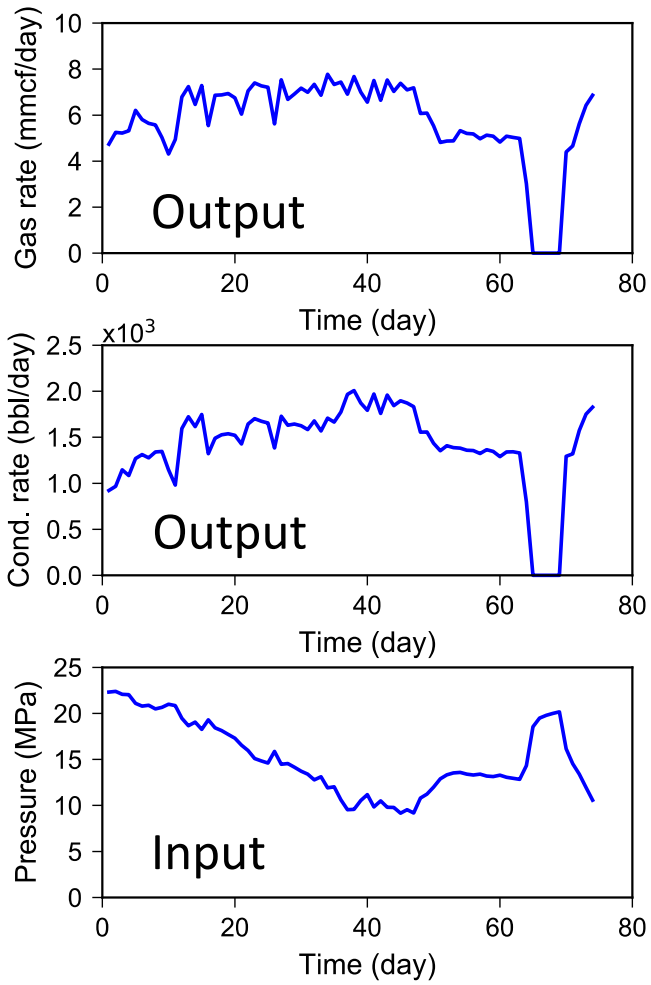
(a) field case #1



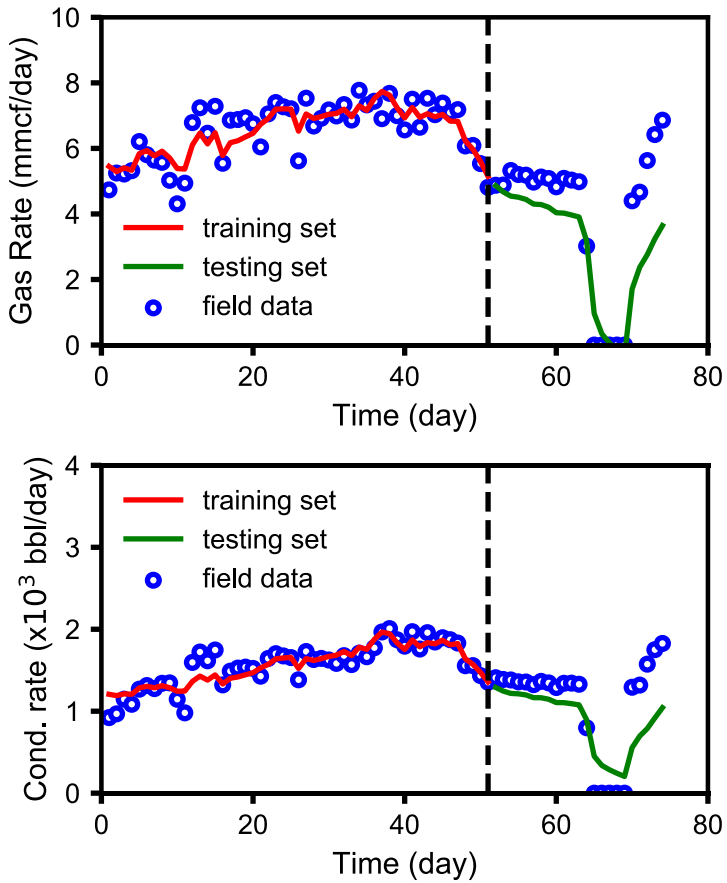
- Good prediction results are achieved;
- Well is put on production and then shut-in at the 65th day;
- Although the training data set does not contain any information on the pressure response of the well shut-in, the LSTM model successfully captures and accurately forecasts the pressure response during the shut-in period.

Time Series Production Prediction

$x_t = [p_t, t]$ → **LSTM Model** → $y_t = [q_{gas,t}, q_{cond,t}]$



(a) field case #1



Conclusions

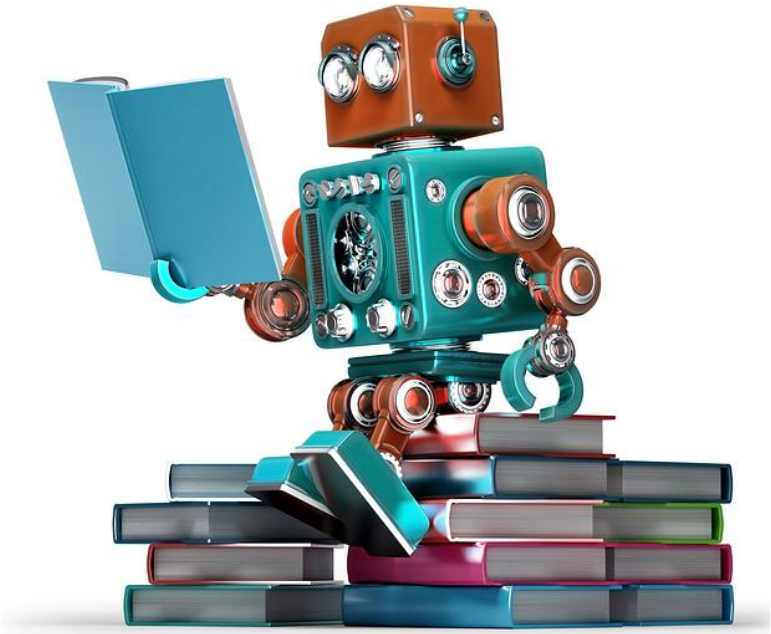
UNIVERSITY OF CALGARY

- Machine learning methods were successfully applied to predict well productivity and design hydraulic fracturing parameters in Montney and Duvernay Formations;
- Well location is found to be the most important parameter to affect the well after-stimulation productivity in Montney formation, while proppant tonnage and fluid volume are the main influencing factors for short term (6 month) and long term (12-18 month) production;
- Appropriate proppant tonnage and fracturing fluid volume can be suggested using the machine learning models to provide references for the hydraulic fracturing design;
- Extra-Tree Boosting method is found to be the best machine learning algorithms for well productivity prediction in Duvernay Formations, which has less overfitting problems and highest prediction accuracy;
- A higher prediction accuracy can be achieved by clustering wells based on GOR in Montney Formations.
- Prediction accuracy of machine learning models can be improved by integrating the production data into inputs.

Acknowledgement

geoLOGIC SYSTEMS
Canadian Discovery Ltd.
NSERC CRSNG
CANADA FIRST RESEARCH EXCELLENCE FUND
APOGÉE CANADA FONDS D'EXCELLENCE EN RECHERCHE
CMG COMPUTER MODELLING GROUP LTD.
IHS
UNIVERSITY OF CALGARY
Alberta Innovates Technology Futures

- ***Thank you!***
- ***Questions?***



Shengnan Chen
Associate Professor
University of Calgary

(403)-220-2040
snchen@ucalgary.ca

Check out website:
http://wcm.ucalgary.ca/nancychen/