A Supervised Machine-Learning Approach to Stratigraphic Surface Picking in Well Logs from the Mannville Group of Alberta, Canada*

Justin Gosses[1] and Licheng Zhang[2]

Search and Discovery Article #42403 (2019)**
Posted August 5, 2019

*Adapted from oral presentation given at AAPG 2019 Annual Convention & Exhibition, San Antonio, Texas, May 19-22, 2019
 DOI:10.1306/42403Gosses2019

[1]Gosses Data Science Consulting, Houston, TX, United States (jgosses82@gmail.com)
[2]CGG, Houston, TX, United States

Abstract

Variation in where stratigraphic surfaces are interpreted in well logs can significantly impact oil and gas resource calculations. Generating stratigraphic picks in hundreds of wells logs can take a geologist weeks to months. We propose a method for how supervised machine-learning can be used to extend human-generated well log picks to new wells, assuming a highly penetrated area where most depositional variance is captured in a training dataset. We use an open dataset of 2193 wells from the Mannville Group in Alberta Canada [Wynne et al., 1995] and our code is open source [Gosses and Zhang, 2018]. Curve matching approaches, while useful for lithologic correlation, are abandoned due to the difficulty of getting those methods to deal effectively with local controls on deposition or controls whose expression changes spatially. Instead, features are created to mimic the visual comparisons geologists make while correlating cross-sections but at a lower level of observation, letting the machine-learning algorithm, XGBoost, combine features and assign weights to reach higher-order conclusions. For each depth point, features are created using GR, ILD, DPHI, NPHI curve values at above, below, and around that depth point. Other features are generated from known picks in neighboring training wells, curve data features from neighboring wells, location, and summary descriptions of each full well. To address class imbalances that result from trying to predict one depth in a well of thousands of depth points, we create class labels for depth points at the pick, within 0.5 meters, above 5 meters, below 5 meters, and outside that range. Less than 15% of the depth points outside of the 5 meters range are kept for training, and two machine learning models are employed in sequence. The first for class prediction. The second to examine the depths labeled as either at the pick or within 0.5 meters and choose the best pick candidate in each well via regression. For the top McMurray pick, absolute mean errors are similar to that of a geologist new to the area mimicking the picking style of a geologist experienced with that formation. This type of approach may be useful in evaluation of nearby acreage, extending interpretation to infill wells, or quickly generating multiple probable picks in each well for Monte Carlo simulations.

Selected References

Gani, M.R., and J.P. Bhattacharya, 2005, Lithostratigraphy Versus Chronostratigraphy in Facies Correlations of Quaternary Deltas: Application of Bedding Correlation: River Deltas—Concepts, Models, and Examples, SEPM Special Publication No. 83, p. 31–48.

Gosses, J.C., 2019, JustinGOSSES/predictatops: v0.0.3: Zenodo, doi:10.5281/zenodo.3247092.

Gosses, J.C. and Licheng, Z., 2018, StratPickSupML; DOI: 10.5281/zenodo.1450597. https://github.com/JustinGOSSES/StratPickSupML/

Hein, F.J., and G. Dolby, 2001, Regional lithostratigraphy, biostratigraphy and facies models, Athabasca oil sands deposit, northeast Alberta: Ann. Conv. Proc. Rock the Foundation (Calgary), Can. Soc. Petroleum Geologists, 3 p.

Hein, F.J., G. Dolby, and B. Fairgrieve, 2013, A Regional Geologic Framework for the Athabasca Oil Sands, Northeastern Alberta, Canada: in Heavy-oil and Oil-sand Petroleum Systems in Alberta and Beyond: AAPG Studies in Geology 64, Chapter: 7.

Hall, B., 2016, Facies classification using machine learning: The Leading Edge, v. 35/10, p. 906-909.

Olea, R.A., and R.J. Sampson, 2003, User's Manual for Correlator, Version 5.2: Kansas Geological Survey Mathematical Geology Section, Lawrence, Kansas.

Wynne, D.A., M. Attalla, T. Berezniuk, H. Berhane, M. Brulotte, D.K. Cotterill, R.S. Strobl, and D.M. Wightman, 1994, Athabasca Oil Sands Database McMurray/Wabiskaw Deposit: Open-File-Report 1994-14, Alberta, Canada; Alberta Geological Survey, 51 p.

Wynne, D.A., Attalla, M., Berezniuk, T., Brulotte, M., Cotterill, D.K., Strobl, R. and Wightman, D., 1995, Athabasca Oil Sands data McMurray/Wabiskaw oil sands deposit - electronic data: Alberta Research Council, ARC/AGS Special Report 6. http://ags.aer.ca/publications/SPE_006.html

Ye, S-J., R.W. Wellner, and P.A. Dunn, 2017, Rapid and Consistent Identification of Stratigraphic Boundaries and Stacking Patterns in Well Logs – An Automated Process Utilizing Wavelet Transforms and Beta Distributions: SPE Annual Technical Conference and Exhibition, DOI: 10.2118/187264-MS

AAPG ACE 2019
ANNUAL CONVENTION & EXHIBITION
19-22 May • San Antonio, Texas

A Supervised Machine-Learning Approach to Stratigraphic Surface Picking in Well Logs From the Mannville Group of Alberta, Canada

J. C. Gosses, L. Zhang

@JustinGosses
https://github.com/JustinGOSSES/predictatops

Predictatops

Talk Outline

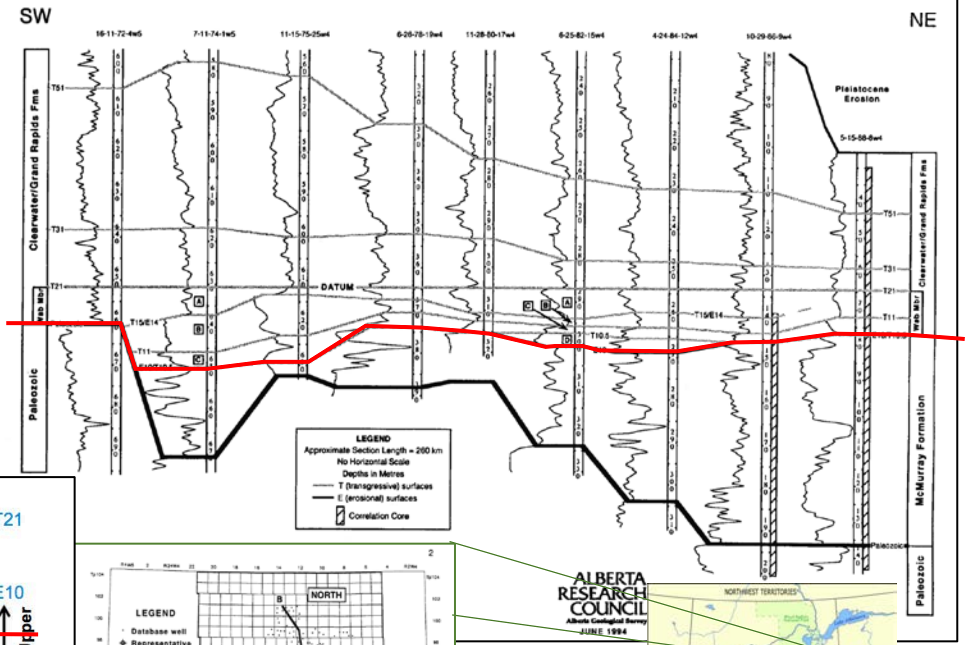
- Data: - Intro to an open-source dataset
- Theory - Human vs. machine-learning stratigraphy
- Methods - Introduction to Predictatops
- Application - How and when it might be useful

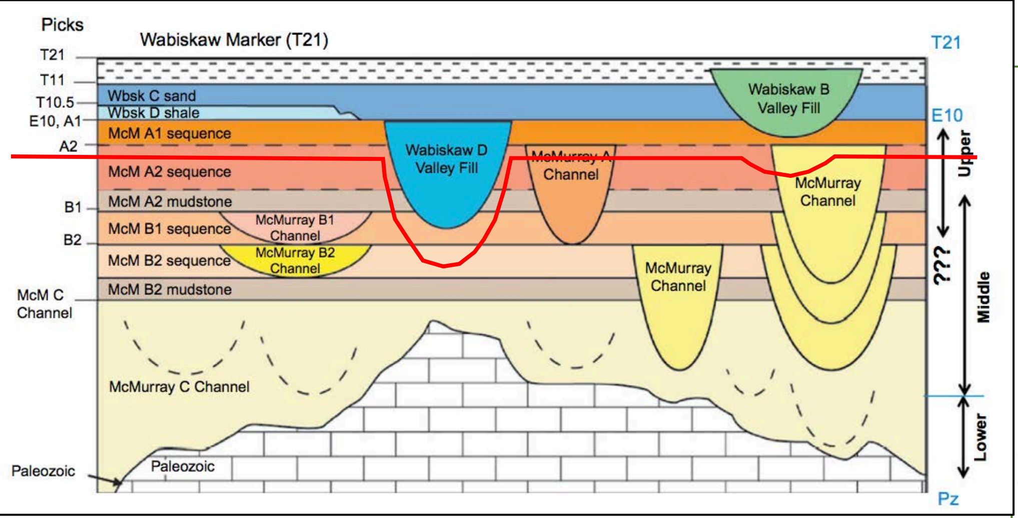
Predictatops https://github.com/JustinGOSSES/predictatops

AAPG ACE 2019 ANNUAL CONVENTION & EXHIBITION
19-22 May • San Antonio, Texas

Location: Mannville Group of Alberta, Canada
Goal: Predict Top McMurray
Dataset: 2193 wells, tops, & location data
Top McMurray is a regional transgressive, erosive surface. Dataset is public & described by Alberta Geological Survey Open File Report 1994-14

Wynne et al., 1994





Hein et al., 2001

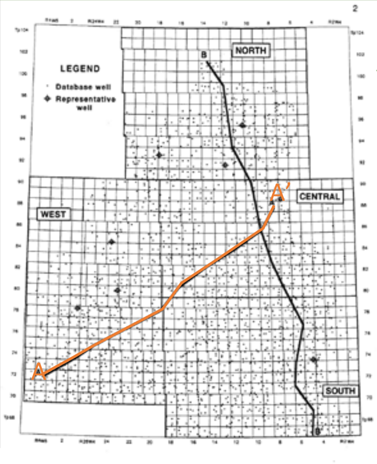


Figure 1. Study area with locations of wells, representative wells and cross sections.

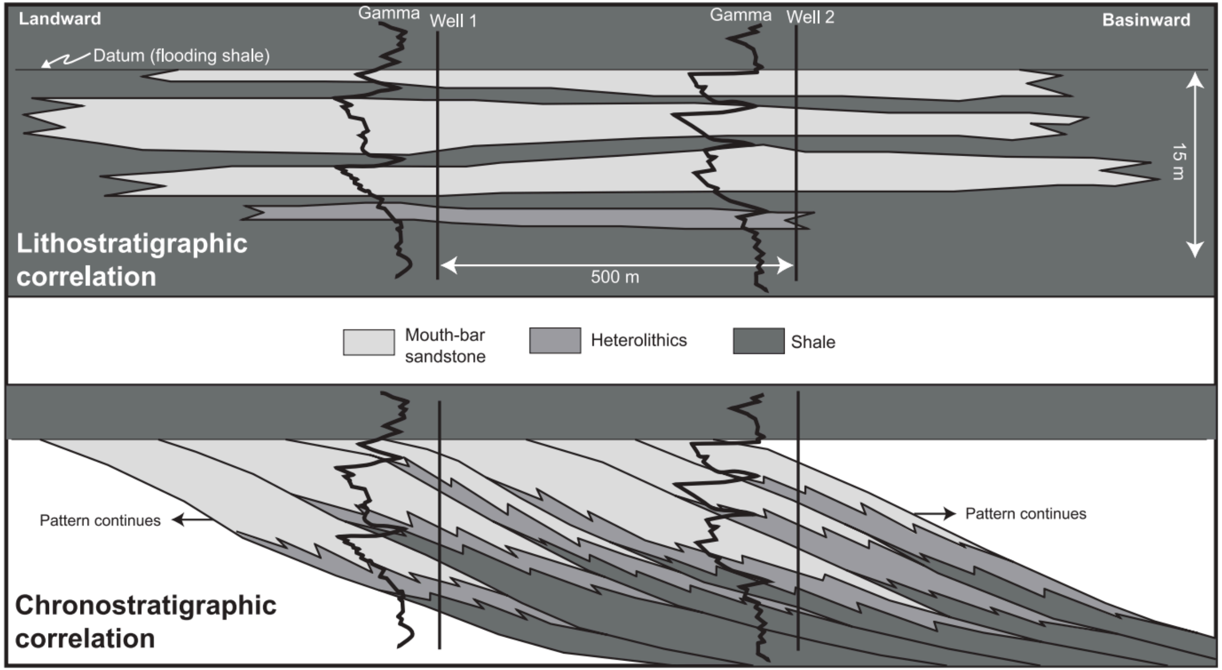
https://ags.aer.ca/document/OFR/OFR_1994_14.PDF

Different Types of Stratigraphic Labeling

Facies

Lithostratigraphy

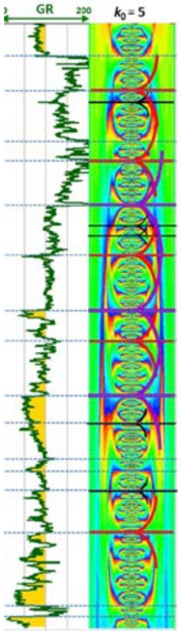
Chronostratigraphy



Gani and Bhattacharya., 2005

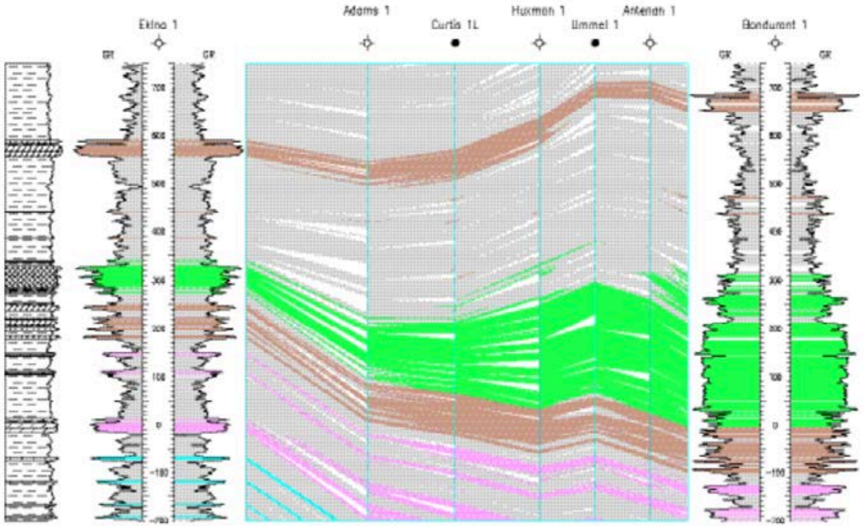
Machine-Learning in Stratigraphy

1D stacking pattern break identification via wavelet transforms



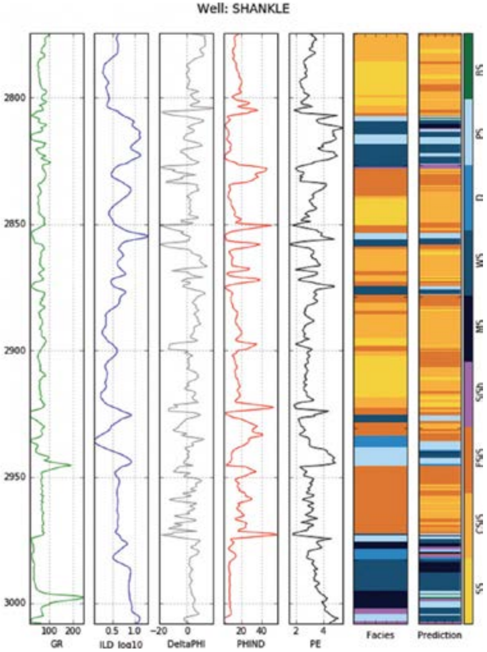
Ye et al., 2017

Correlator: Fortan program for well-to-well lithostratigraphy



Olea and Sampson, 2002, Kansas Geological Survey

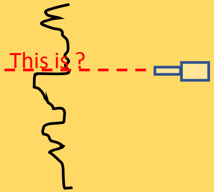
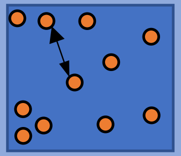
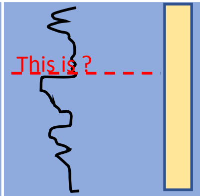
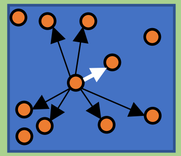
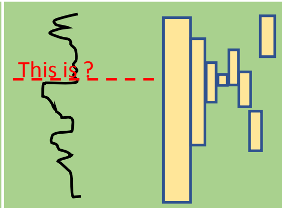
SEG facies prediction contest



Hall., 2016

https://wiki.seg.org/wiki/Facies_classification_using_machine_learning
Brendon Hall, Enthought
THE LEADING EDGE October 2016
https://github.com/seg/2016-ml-contest

Comparing Different Types of Stratigraphic Labeling to Find Key Parts

	Min # of wells	All training wells used in prediction	Wells compared to one another?	Information used from above or below a depth point?	What features & how are they used?	What is the prediction?
Facies	1	Probably	No		<u>Classification</u>	Facies labels for each depth point
Litho-stratigraphy	2	No			<u>Curve matching</u>: often dynamic time warping	Lines connecting 2 wells that may or may not overlay with tops
Chrono-stratigraphy	100s to 1000s (enough for models to be discovered)	Yes			<u>Classification</u>: Features similar to low-level human observations generated across different windows.	A Top Scored by distance between predicted & actual

Repackaging Chronostratigraphy as a Machine-learning Problem

Human Chrono-stratigraphy

Outcrop & analogue studies → Conceptual Chronostratigraphic Model + High-level human observations about wells relative to other wells & models → Geologist labeled Tops

Supervised machine-learning Chrono-stratigraphy

Machine-learning model that can mimic human chronostratigraphic interpretation ← Machine-learning algorithm good at clustering, finding threshold, etc. to classify + Rule-based features programmatically created to mimic low-level human observations + Geologist labeled Tops in training wells

How to Code Low-Level Geologic Observations as Features?

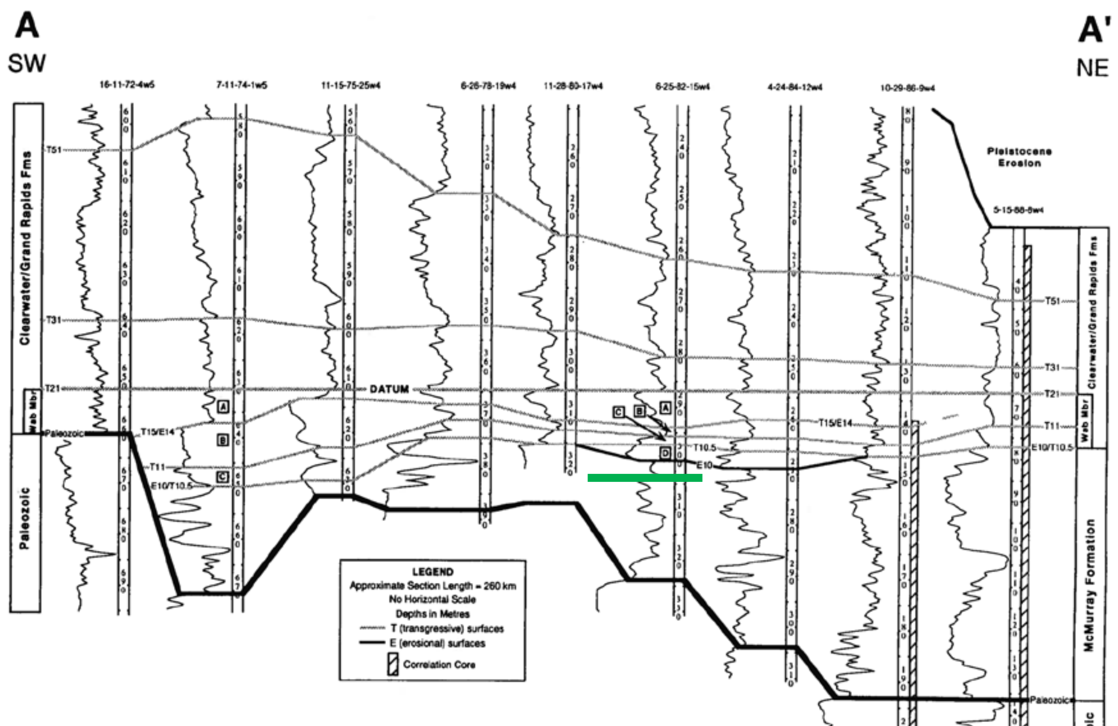


Figure 2. Stratigraphic cross section A-A' of Lower Cretaceous McMurray/Wabiskaw interval, Athabasca area, northeast Alberta.

ALBERTA RESEARCH COUNCIL
Alberta Geological Survey
JUNE 1994

Hein et al., 2001

We want to create features to determine if each depth point is the top.

https://ags.aer.ca/document/OFR/OFR_1994_14.PDF

How to Code Low-Level Geologic Observations as Features?

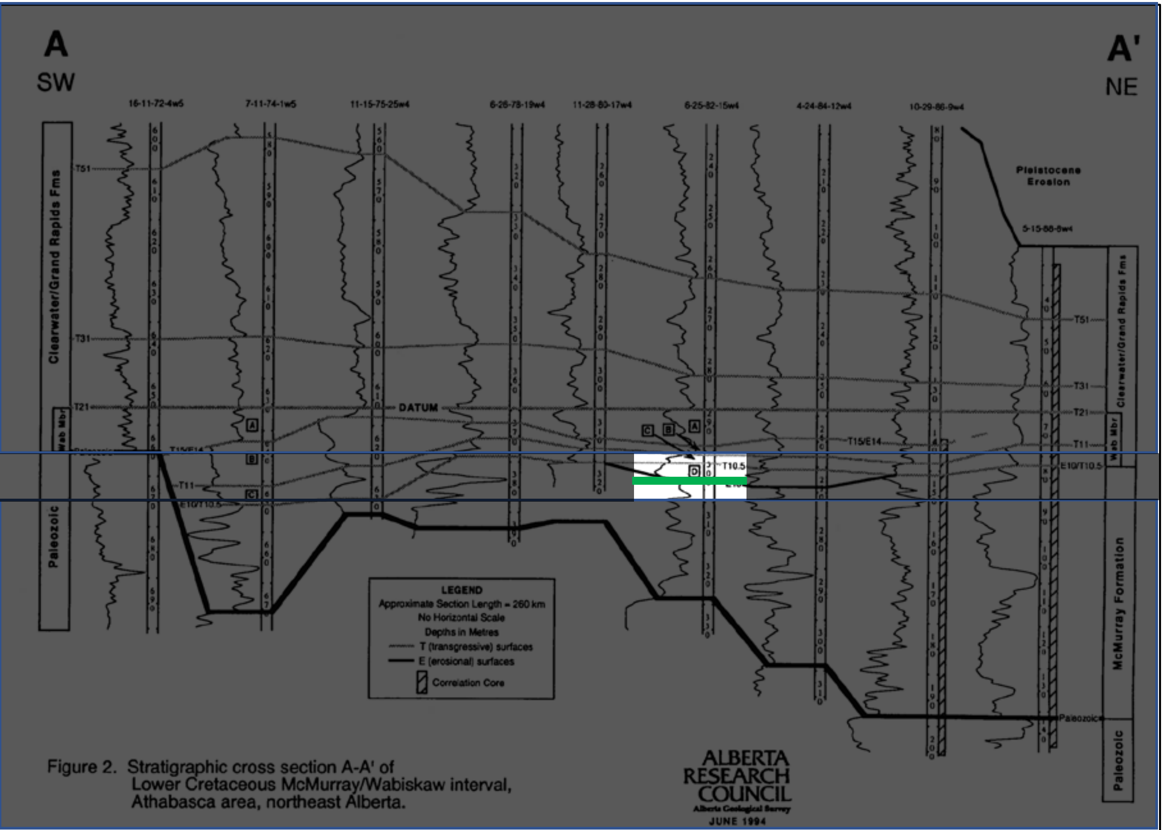


Figure 2. Stratigraphic cross section A-A' of Lower Cretaceous McMurray/Wabiskaw interval, Athabasca area, northeast Alberta.

A game: Pay attention to what you can't observe when aspects of the cross-section are taken away.

For each depth point, need to create features that gather information around it.

AAPG ACE 2019 ANNUAL CONVENTION & EXHIBITION
19-22 May • San Antonio, Texas

How to Code Low-Level Geologic Observations as Features?

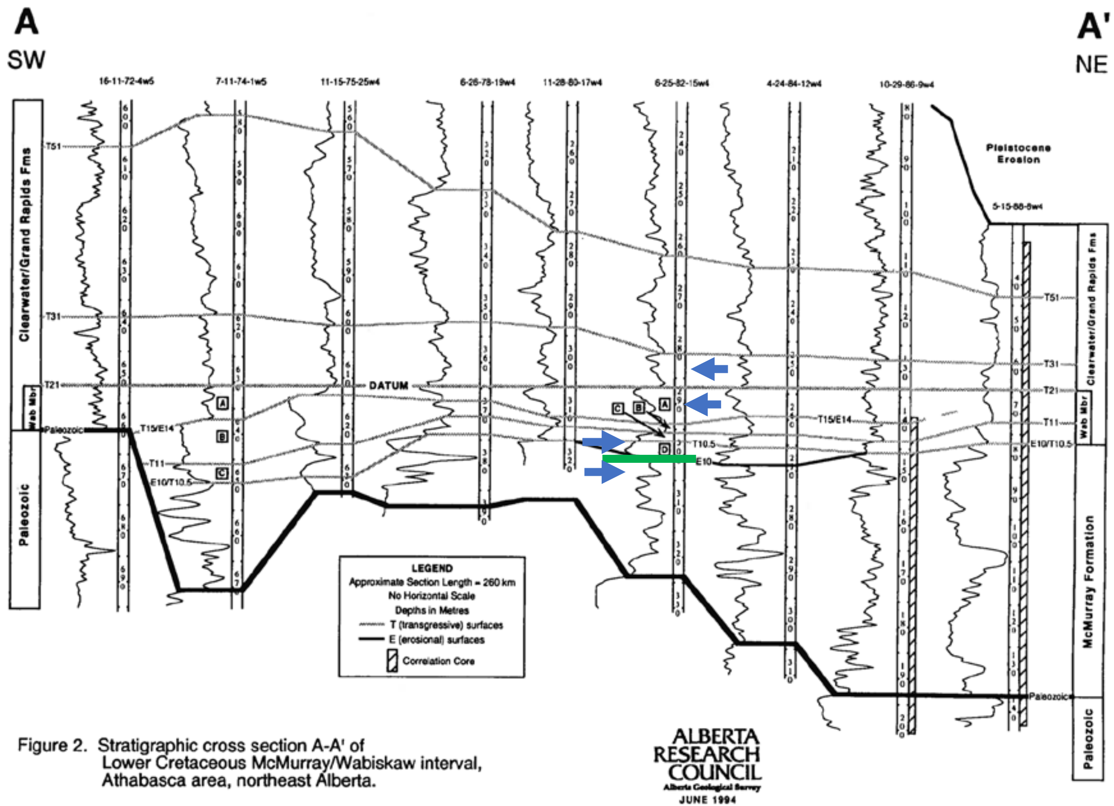
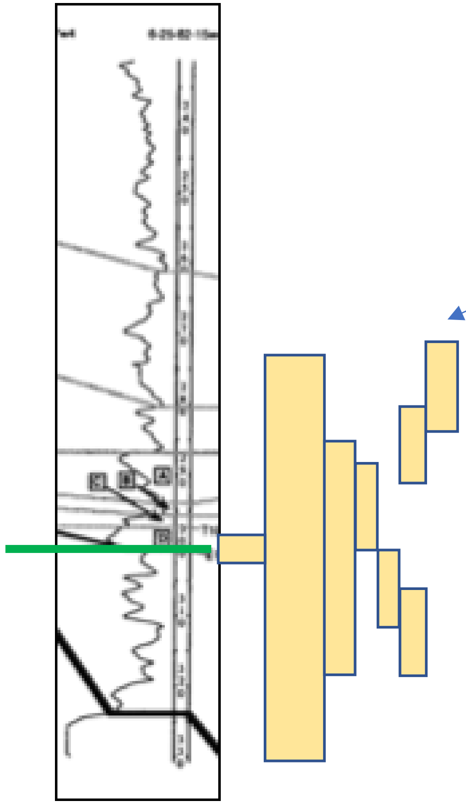


Figure 2. Stratigraphic cross section A-A' of Lower Cretaceous McMurray/Wabiskaw interval, Athabasca area, northeast Alberta.

Neighboring training wells can be used for features: here unit thickness of neighbors represented by blue arrows.

How to Code Low-Level Geologic Observations as Features?

A single depth point with curve data

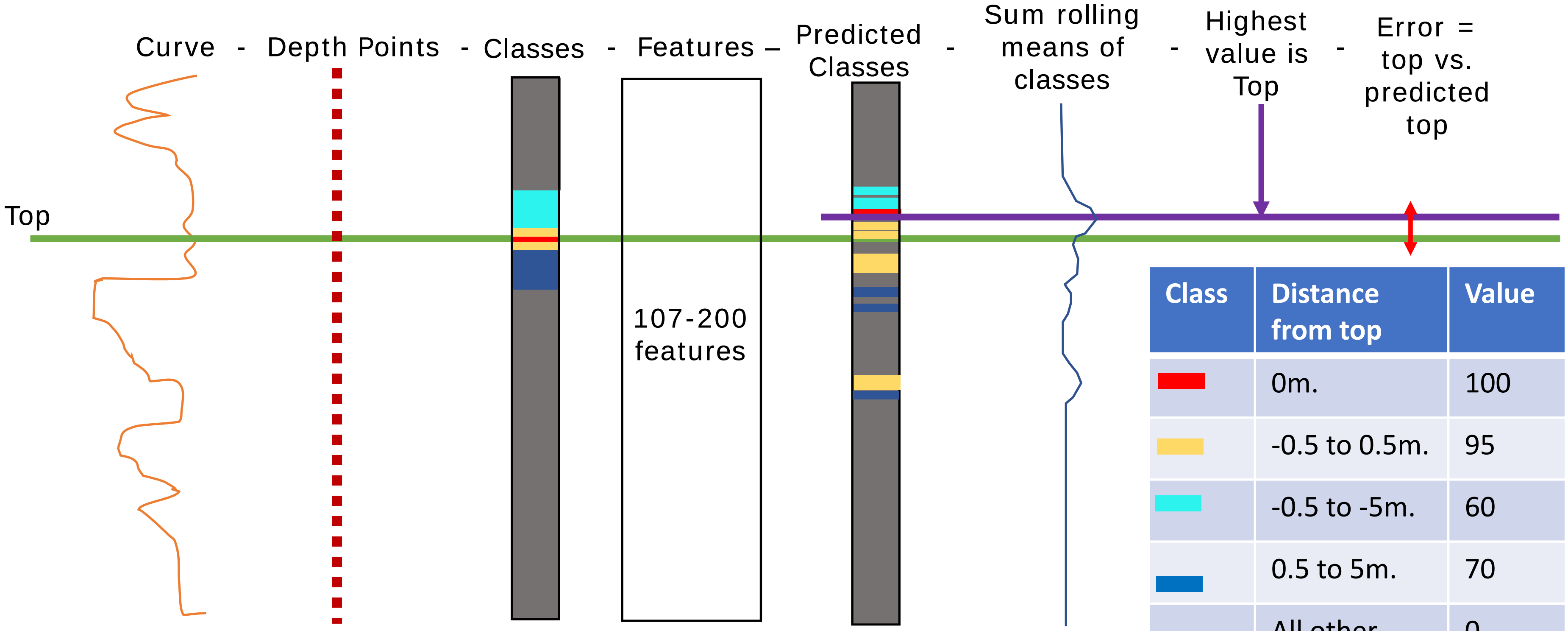


Within each window:

- Max, min
- Rate of change
- Variance
- Etc.

Information from above & below a point in question turned into features

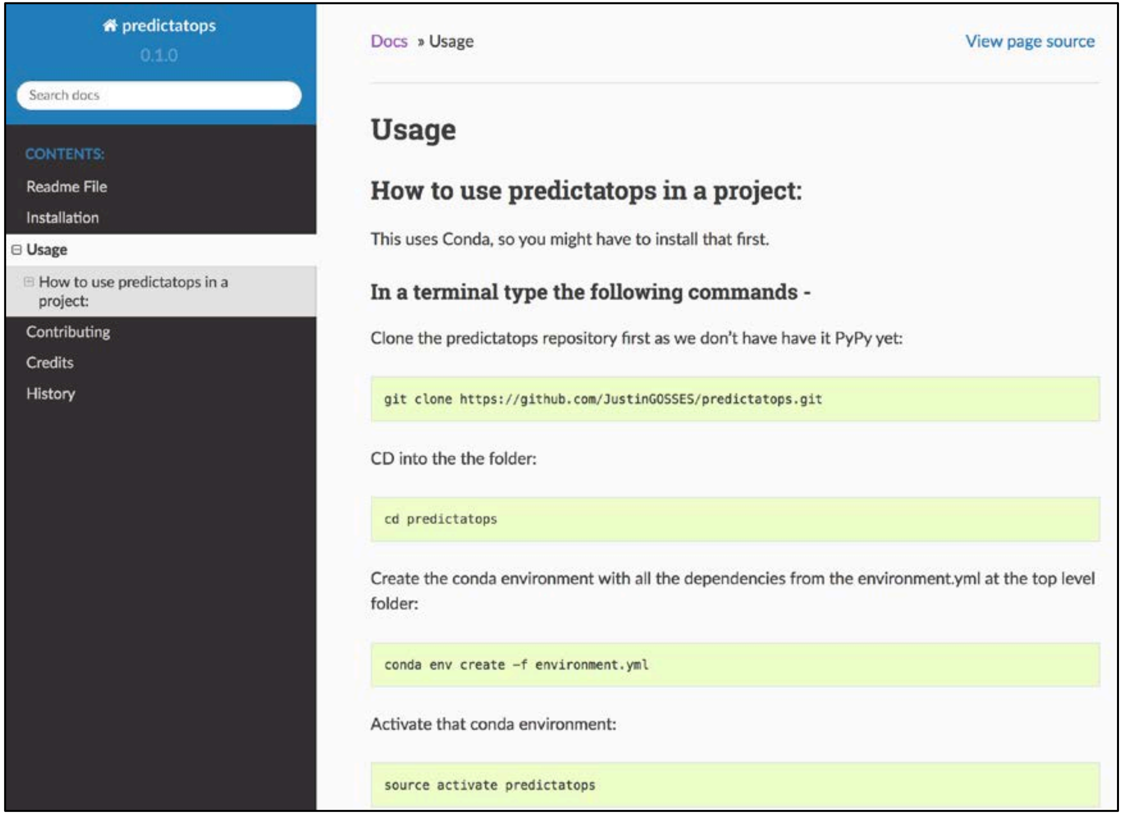
Predicting a single top by creating classes based on distance from top



Class	Distance from top	Value
(red)	0m.	100
(yellow)	-0.5 to 0.5m.	95
(cyan)	-0.5 to -5m.	60
(blue)	0.5 to 5m.	70
(grey)	All other	0

Predictatops

https://github.com/JustinGOSSES/predictatops


predictatops
0.1.0
Search docs
CONTENTS:
Readme File
Installation
Usage
How to use predictatops in a project:
Contributing
Credits
History

Docs » Usage   View page source

Usage

How to use predictatops in a project:

This uses Conda, so you might have to install that first.

In a terminal type the following commands -

Clone the predictatops repository first as we don't have have it PyPy yet:

```
git clone https://github.com/JustinGOSSES/predictatops.git
```

CD into the the folder:

```
cd predictatops
```

Create the conda environment with all the dependencies from the environment.yml at the top level folder:

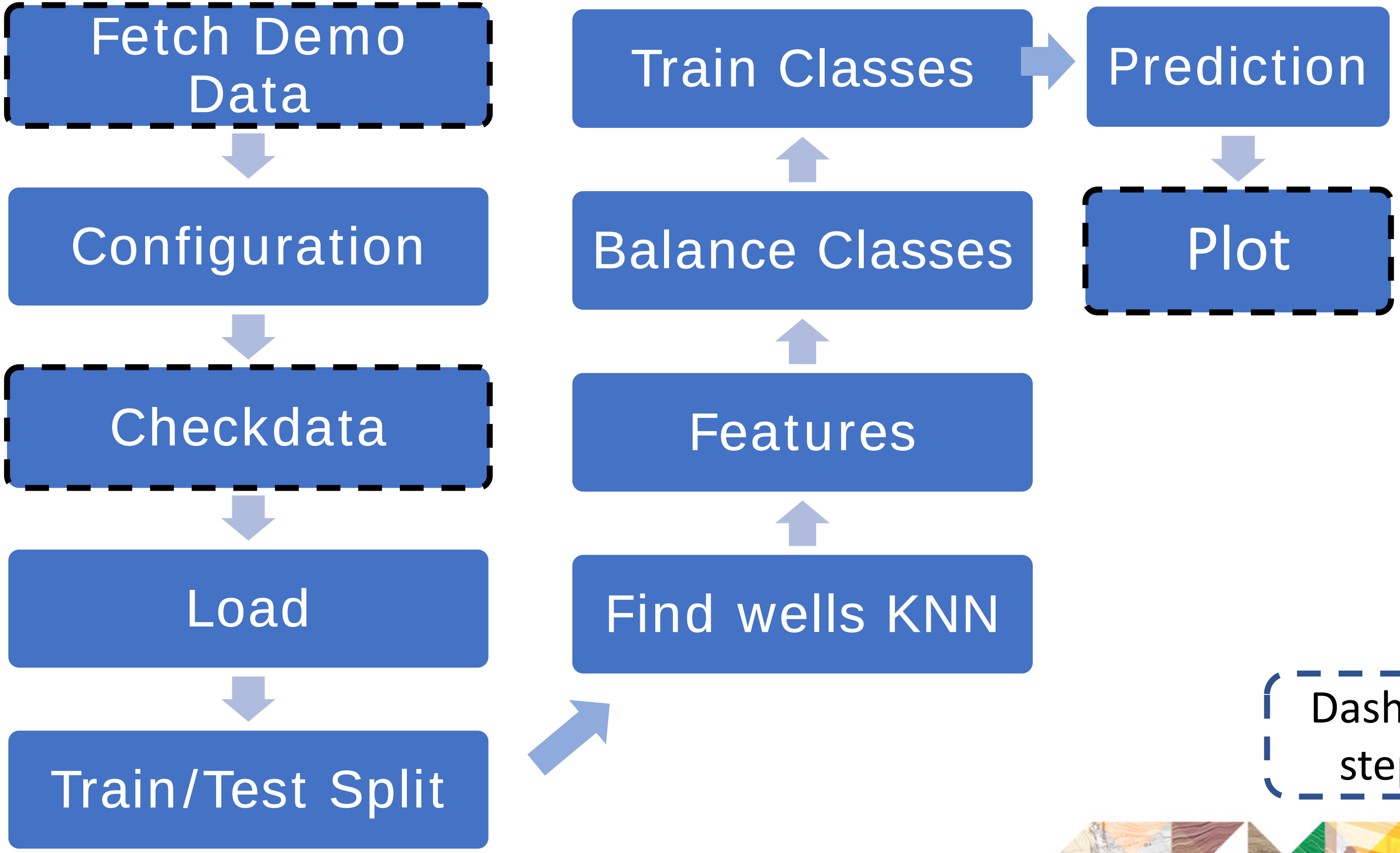
```
conda env create -f environment.yml
```

Activate that conda environment:

```
source activate predictatops
```

- Python code for top prediction
- M.I.T. License
- Run interactive in Jupyter or all at once via config file
- Alpha state

Predictatops ML pipeline



Dashed Black outline steps are optional

Parts of Machine-learning Code Worth Mention

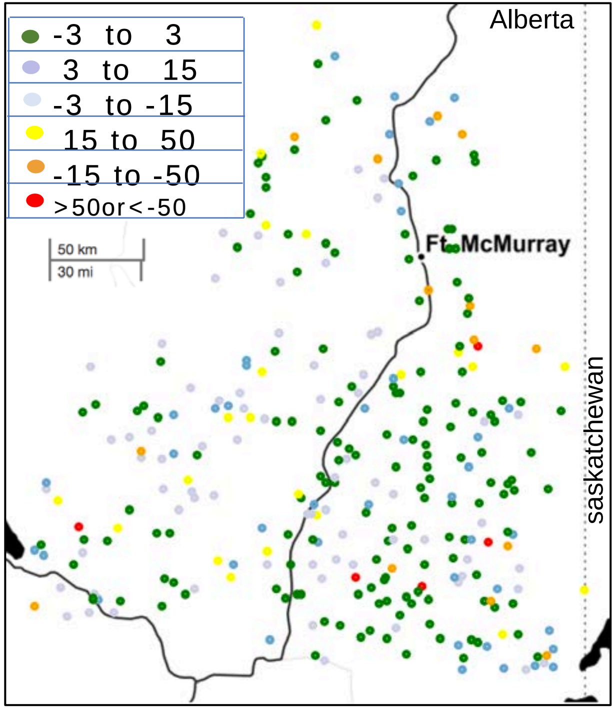
Create Train/Test **split before creating features**, so you don't cheat when you create features using spatial knowledge.

Class **rebalancing is critical** as the class you care most about (those nearest the top pick) will be the more sparsely populated in your original dataset.

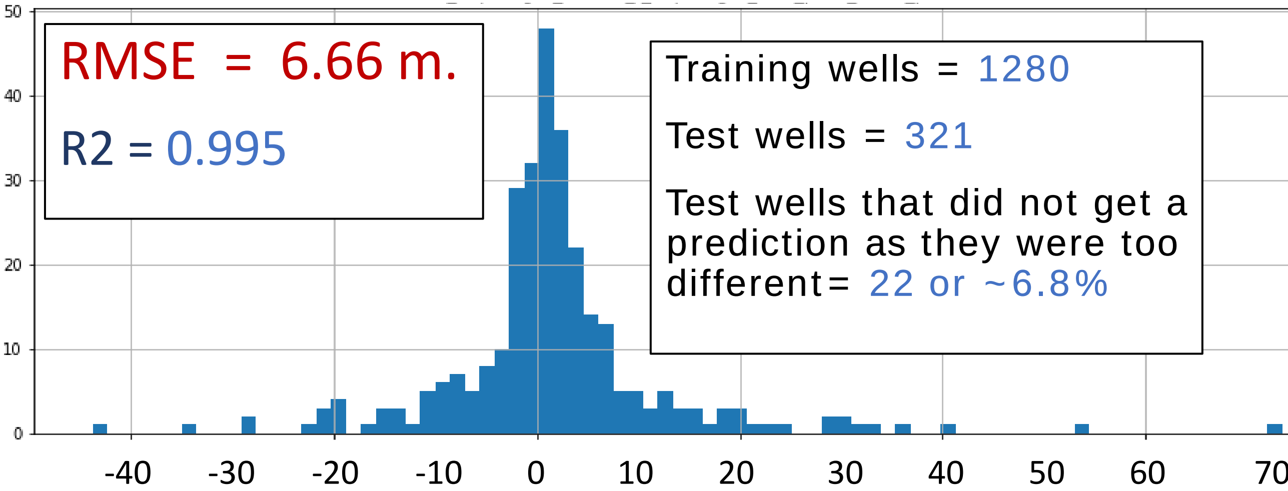
Sometimes a well doesn't have any depths predicted as remotely close to the top. Which is great! **Lets you know that well is different than training wells and needs a human touch**!

Results

Map test well prediction errors



Error [Predicted vs. Actual Top Depth meters]



Requirements for a well to be used:

Curves = ['ILD', 'NPHI', 'GR', 'DPHI']

Tops = Top McMurray, Base McMurray

When to use? How to use?

Constraints on when to use?

- Need a large number of wells
- Need a large number of tops you trust
- Need tight enough well spacing to capture variance in order to produce model

Possible Applications

- **Time Reduction:** Interpret 1200 wells, and automate the other 1200
- **Compare Interpretations:** train two models in two areas, then predict on each other to see where differences in interpretation happen.
- **Better Represent Uncertainty:** easy to generate and track multiple top predictions & flag the wells with highest uncertainty

Acknowledgements

- Coauthor: Licheng Zhang
- Additional Hackathon participants: Zhenshen Zong, Jonathon Parker
- Hackathon organizer Agile Geoscientific
- Open-Source Geoscience Libraries: Welly & Lasio!
- Dataset Suppliers!!!
  - Alberta Research Council & Alberta Energy Regulator.
  - Many authors of Alberta Geological Survey Open File Report 1994-14
    - Recently, AGS has made public 35,000 more tops, but the logs need to purchased.

Conclusions

Philosophy:

- Instead of trying to encode a geologic model in code directly or find mathematical patterns in the raw data, create features that map to low-level geologic observations & then let the program figure out the relationships that human would describe with a model.

Requirements for use: 1000s of wells & acceptable to have slightly worse than human performance

Possible Application: Time reduction on regional scale work & new uncertainty management options

Future Work: Different algorithms + More features + Different Datasets + Better Visualizations + Better Docs

Predictatops

https://github.com/JustinGOSSES/predictatops

References

- Gani and Bhattacharya (2005) Lithostratigraphy Versus Chronostratigraphy In Facies Correlations of Quaternary Deltas: Application of Bedding Correlation, River Deltas—Concepts, Models, and Examples SEPM Special Publication No. 83, SEPM (Society for Sedimentary Geology), ISBN 1-56576-113-8, p. 31–48
- Gosses, J.C., 2019, JustinGOSSES/predictatops: v0.0.3: Zenodo, doi:10.5281/zenodo.3247092.Hein (2013) A Regional Geologic Framework for the Athabasca Oil Sands, Northeastern Alberta, Canada,Heavy-oil and Oil-sand Petroleum Systems in Alberta and Beyond: AAPG Studies in Geology 64, Chapter: 7, American Association of Petroelum Geologists.
- Hall, B. (2016) Facies classification using machine learning, The Leading Edge, 35 (10): 906-909.
- Hein, F. J., and Dolby, G., 2001, Regional lithostratigraphy, biostratigraphy and facies models, Athabasca oil sands deposit, northeast Alberta: Ann. Conv. Proc. Rock the Foundation (Calgary), Can. Soc. Petroleum Geologists, 3 p.
- Olea and Sampson (2003) User's Manual For Correlator, Version 5.2, Lawrence, Kansas; Kansas Geological Survey Mathematical Geology Section
- Wynne et al., (1994) Athabasca Oil Sands Database McMurray/Wabiskaw Deposit, Open-File-Report 1994-14, Alberta, Canada; Alberta Geological Survey. Links to report & dataset.
- Ye et al. (2017) Rapid and Consistent Identification of Stratigraphic Boundaries and Stacking Patterns in Well Logs – An Automated Process Utilizing Wavelet Transforms and Beta Distributions, SPE Annual Technical Conference and Exhibition, DOI: 10.2118/187264-MS