Lithology Prediction of Slabbed Core Photos Using Machine Learning Models

Thomas Martin, Ross Meyer, Zane Jobe

*Colorado School of Mines*

Thanks to:

thomasmartin@mines.edu

: ThomasM_geo
The Problem

- Legacy data of variable interpretation quality
- Core is stored around the world in public and private repositories
- In the age of big data, core photos are underutilized
- Logging core is a manual, time consuming, subjective process

Probably a core repository, courtesy of Indiana Jones

Photo: Walt Disney Company

Clark Gilbert and Wylie Walker logging at USGS Core research Center
The Solution

Image-based ML model

- BGS Tray
- Core Column
- Labeled Dataset
- Supervised ML Model
- Predictions
Test Case: Quadrant 204, UKCS

- 11 Wells chosen from Schiehallion Area, West of Shetlands
  - Ten wells from reservoir intervals
  - One non-reservoir well

- All data freely available with unencumbered licensing

- Core images downloaded from the British Geological Survey

- Wireline information from UK Oil and Gas Authority

Geologic Map from Freeman et al. 2008
Quadrant 204 Geology

- Conventional hydrocarbon system, produced ~400 Million BOE, projected end of life in 2035
- Reservoir targets are T25 to T35 sands in the Vaila Formation
- 25-30% Porosity sands, 500-1500 mD
- Interpreted to be a confined submarine channel system (Ward 2017)
- Turbidites, hybrid event beds are present in the cores
• BGS stores all geologic material from offshore hydrocarbon wells

• Entire inventory was imaged under the same conditions

• Automated workflow to go from core tray to stacked core column

• All pixels are depth registered

• Manual QC of depths and some tray editing
• Each 32 (high) by 600 (wide) pixel wide image is compared to training data for texture, color, and patterns. This is done on a sliding window, and predictions are row-averaged where the windows overlap.

• Differences in lighting minimized due to consistent image acquisition techniques

• Affected by shadows, dirt and dust on core

Example of one image subset, each chunk is ~0.5cm
“Pseudo Gamma” Data

- For each pixel row, the mean and variance is calculated for red, green, blue and brightness

- XGBoost ML Models use mean/variance values averaged over ~0.5 cm (32 pixels)

- Wavenet (CNN) uses each pixel row individually, but bins the label to 32 pixel high sections
• Standard wireline data from UK OGA website

• Not depth shifted

• Standardizing on the entire dataset per curve
  • GR is normalized with GR
Labeling Data

- Used LabelImg, a graphical interface to label the core
- Allows for fine scale label divisions (<1cm)

Example of LabelImg
High quality representative labeled data is needed for supervised learning.

- 11 Wells, 500 meters of core material labeled
- 5 Lithologies, 4 of them discussed in Haughton et al. 2009
  - Sandstone
  - Clay-prone sandstone
  - Sandy Mudstone
  - Mudstone
  - Oil Stained
  - No core
- Labeling on the sub centimeter level
Machine Learning Models

• “A field of study that gives computers the ability to learn without being explicitly programmed” – Arthur Samuel

• XGBoost (Chen and Guestrin 2016)
  • Boosted Tree Algorithm
  • Flexible data input
  • Fast
  • Can take into account context, did not improve scores

• Bi-Directional WaveNet (Oord et al. 2016)
  • Specific type of Convolutional neural network
  • Developed for text-to-speech
  • Context is important

Upper: https://towardsdatascience.com/
Lower: https://deepmind.com/
• All models are run on a single NVIDIA 1080 GPU in a standard Linux desktop workstation

• Possible to run on a higher end laptop

• Each epoch (iteration) runs from 5s to 60s

• From data load to prediction for most wells is under 5 minutes

• Limited by memory for larger image datasets
• Well dependent, more laterally homogenous the better
  • Training data needs to be representative of testing data!

• Wireline ~ 20% Accuracy
  • As good as guessing!

• RGB-G Pseudo Gamma 60-75% Accuracy
  • Sand category is 5-10% more accurate than overall score
  • N:G overall is within ~5%

• Image 60-75% Accuracy
  • Similar results, but much more computationally expensive
  • Some wells image is better than RGB-G Pseudo Gamma
Current Work

• 70%+ Accuracy is a great start!

• Explore combined datasets more

• Explore different labeling schemes (facies, flow unit, spatial patterns, etc.)

• Natural extension to other data types like hyperspectral, CT, UV, image logs

Implications

- Reservoir property statistics
- Coming up to speed on data trades
- Re-examining legacy datasets
- Augmented interpretation
- Workflow used for other deposit types (carbonates, tidal, etc.)

Thomas Martin
thomasmartin@mines.edu
@ThomasM_geo

Ross Meyer
@rgmyr_

Zane Jobe
@ZaneJobe

DOI: 10.6084/m9.figshare.8023835