geoconvention
Calgary • Canada • May 15-17 2023

# Predicting Facies Using Logged Core, a Sequence of Random Forests and a 3D Seismic Dataset

*Author information – Sam M. Muller*
*Affiliation – Parex Resources Ltd.*

## Summary

There are measurements and observations of the subsurface that exist only along the wellbore due to either the need of a sensor, or the need of human observation. Being able to extrapolate these observations through the subsurface would bring great value to the exploration and development of resources. This work presents a model that can predict both geological labels (ie. facies) and sensed subsurface quantities (ie. porosity) between well locations using a 3D seismic dataset and well data for training. The model is composed of two separate Random Forest models in sequence. First, a set of petrophysical log values are predicted for each seismic trace in the survey. Next, these petrophysical logs are used as input into a second Random Forest to predict the desired quantitative or qualitative value. Initial testing in the Llanos Basin of Colombia indicates that this method was effective at distinguishing between coarse and fined grained facies and at extrapolating porosity values throughout the subsurface.

## Workflow

This workflow was conducted in the Llanos Basin of Colombia, in a fluvial setting where the prediction of channel facies and a non-channel facies provides great value. The data used to create this workflow consisted of an inverted 3D seismic dataset and wells with petrophysical logs that lie within the seismic area. The depth increments along each well received facies picks from a geologist based on core logging and qualitative use of the petrophysical logs.

During a preliminary data exploration phase, it was found that the facies used throughout the area of interest have relatively unique combinations of petrophysical values (see figure 1 below). That is to say, if each depth increment has 4 petrophysical values (gamma, density, neutron porosity and resistivity for example), in a 4-dimensional space with axes defined by these petrophysical values, each facies would occupy a unique region in that 4D space. Figure 1 demonstrates a 2D view of these distinct regions.

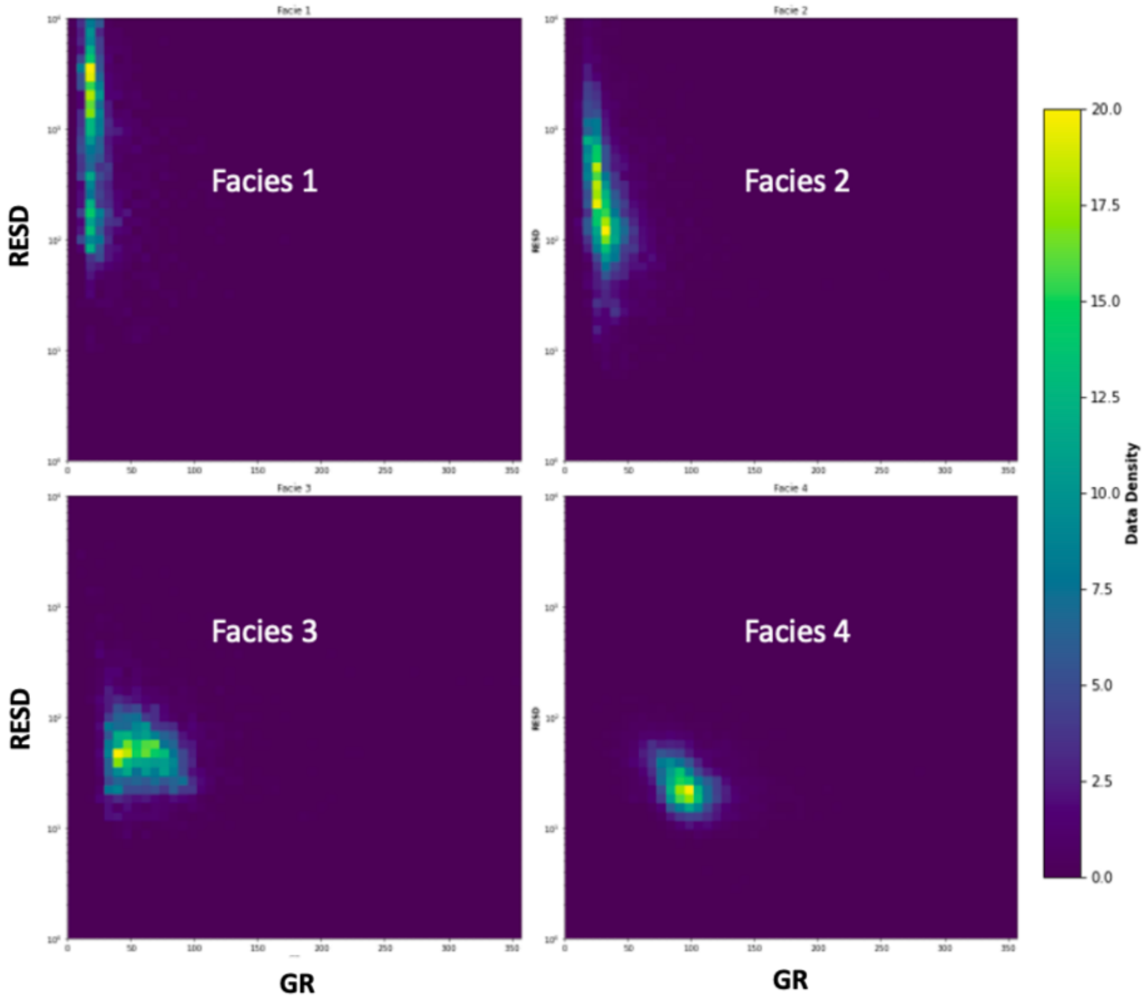


Figure 1. 2-Dimensional Histogram plots showing the localized regions that each Facies occupies when the Deep Resistivity (Y-Axis) is plotted again Gamma Ray (X-Axis) for each datapoint.

After recognizing that petrophysical values could provide insight on the facies at a given depth, a Random Forest Regressor model (referred to as Model 1) was created to predict petrophysical logs based on the data from a seismic trace. To increase the resolution and information content of each seismic datapoint (initially each datapoint was just an amplitude) several additional attributes were calculated for each trace. Each trace was convolved with the Ormsby wavelet to extract information about the response of the trace at low, medium and high frequencies. Additionally, the complex part of each amplitude value was calculated and added as an attribute at each depth step.

The input data for Model 1 was therefore composed of the raw seismic amplitude signal and the additional seismic attributes. The output data to train Model 1 was the set of petrophysical logs. In this workflow, a unique model was created for each petrophysical log (NPSS, RHOB, RESD, GR), resulting in 4 separately trained versions of Model 1. The set of wells used to train Model 1 was 70% of the total wells, leaving 30% to test Model 1. Figure 2 below shows an example set of petrophysical logs output from the 4 instances of Model 1 for a single well in the testing dataset.

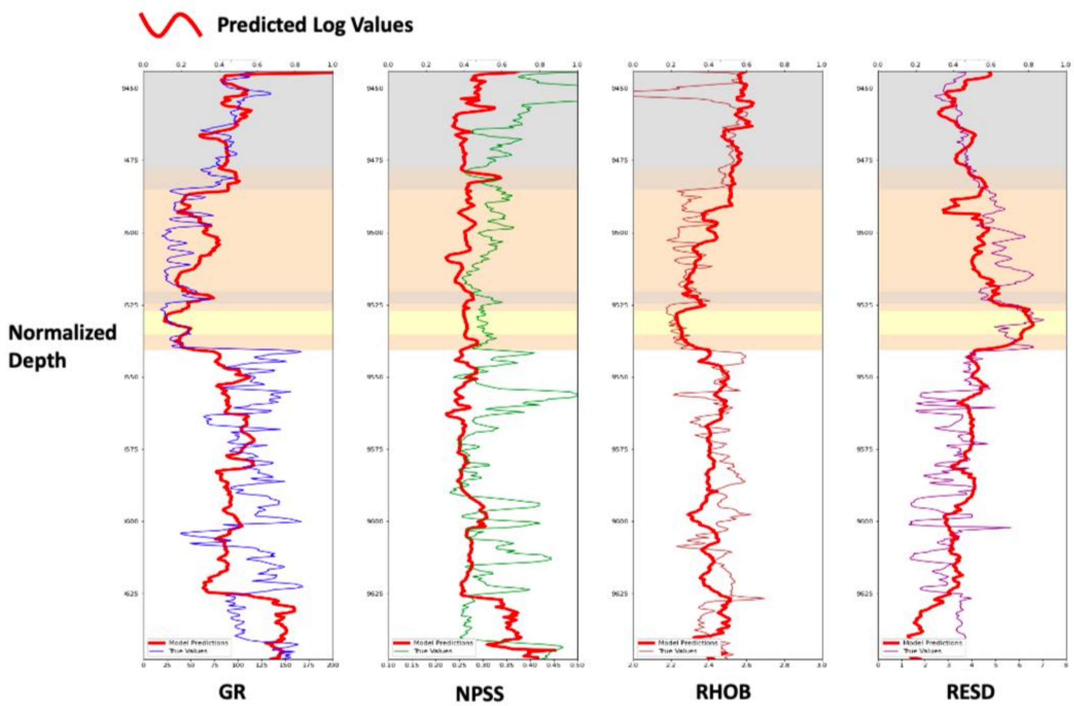


Figure 2: Predicted petrophysical logs (in red) for GR, NPSS, RHOB and RESD logs (from left to right respectively)

Model 2 was designed to take the 4 petrophysical outputs of Model 1, in addition to the seismic derived input values of Model 1, to output a facies prediction. The prediction from Model 2 is a vector of probabilities summarizing the likelihood that the given location is each of the possible facies. The remaining 30% fraction of the wells were split again into a training and testing dataset for Model 2. A Random Forest Classifier was trained to predict a facies label based on the seismic derived attributes and the outputs from Model 1.

To predict a facies label for each depth step throughout a given 3D seismic survey, the following steps were iterated for each individual trace. First, the necessary seismic derived trace attributes were calculated. Next, the data corresponding to each depth step was iteratively fed into each of the instances of Model 1 (remember that a unique Model 1 exists for each petrophysical log), resulting in predicted petrophysical values for each depth. These Model 1 predictions, in addition to the initial seismic derived attributes, are then combined to create the set of attributes fed into Model 2. Finally, this set of attributes is input into Model 2 to receive a facies prediction. After completing this process, with special care to organize these predictions, a volume of data will exist with dimensions equal to the seismic volume but with a vertical resolution equal to that of the raw petrophysical logs. In this output data volume each (X,Y,depth) location has a facies prediction.

To mitigate the effects of statistical bias in the sampling of training wells, a Monte Carlo simulation was run. In the Monte Carlo simulation, the workflow described above was iterated *n* times, with a new random sample of training wells both for Model 1 and Model 2 on each iteration. This process results in *n* distinct prediction volumes. As such, each (x, y, Depth) point in the defined data volume has a set of *n* predictions, and can therefore be represented by statistical measures such as the mean and standard deviation.

## Results

By calculating mappable metrics from the prediction volume (where each (x,y,Depth) location has a set of facies probabilities), maps of facies thickness can be created. Figure 3 below shows a map representing the total thickness of facies considered to be channel-like in nature (Facies 1 and 2). Figure 3, shows a map of the standard deviation of channel thickness predictions at each (X,Y) location in the study area. Understandably, the standard deviation generally increases away from the well locations, highlighting the importance of having an unbiased training dataset.

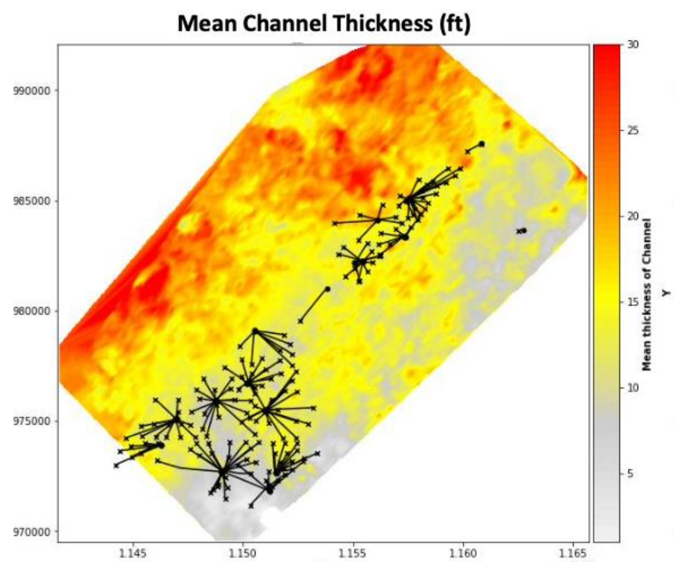


Figure 3. Map of Mean Channel Thickness (ft) predicted throughout the land base used to test this workflow. The wells used in the workflow are shown in black. Where Channel Facies are defined as either Facies 1 or 2.

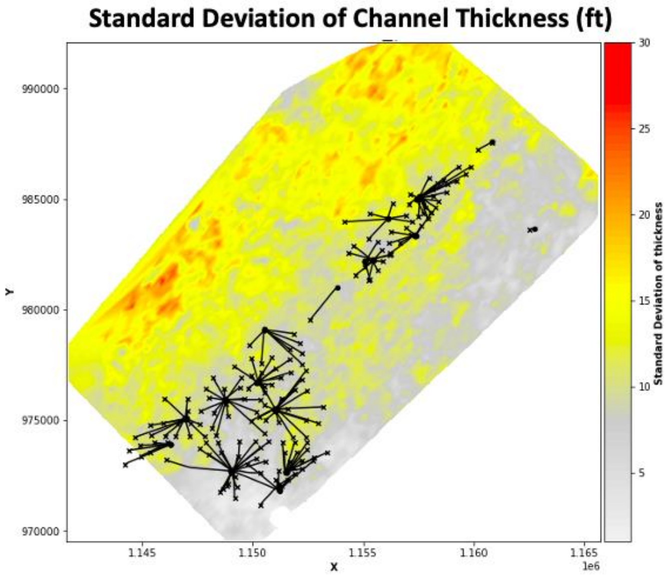


Figure 4. Standard Deviation of channel thickness prediction at each (X,Y) point in the study area.

## Additive Information

The workflow described can easily be manipulated to output a regressed prediction value instead of a probability of classifications. To achieve this change, one must swap the Random Forest Classifier used to create Model 2 out for a Random Forest Regressor. Additionally, a continuous variable (such as porosity or water saturation) must be used as the training variable. With these changes, a new workflow is achieved that predicts a continuous variable, that was measured or calculated along a well bore, throughout the seismic data volume.

## Acknowledgements

Marco Perez
Parex Resources Ltd.