# Particle Size Distribution (PSD) from wireline logs and core photographs: Maximizing the value of legacy data and optimizing high-density coring operations

*John G Manchuk*
*Centre for Computational Geostatistics*

*Hong Long*
*Total E&P Canada*

*Clayton V Deutsch*
*Centre for Computational Geostatistics*

*Rebecca Haspel*
*Total E&P Canada*

*Xavier Mathieu*
*Total E&P Canada*

## Summary

Laboratory data from core analysis, particularly Dean Stark (DS) for bitumen grade and PSD (particle size distribution) carries high value in an oilsands geological model for its critical role in determining downstream hydrotransport, plant processing and tailings requirements. Unsampled intervals, lost core and missing laboratory core analysis are not uncommon in legacy oilsands mining corehole datasets. Historically, production drilling densities less than 200m and twinning of legacy wells adquately reduced data uncertainties and supplemented data gaps. Despite (a) appreciable drilling and lab testing costs, and (b) the time-delay assumed due to the full field-to-lab-to-model lifecycle, laboratory core analysis for PSD is still considered invaluable and irreplaceable.

In the context of maximizing the value of existing data sets, a new methodology for PSD prediction has been developed to predict PSD from wireline logs, core photos and bitumen grade or a combination thereof depending on legacy data availability. A training database for prediction was compiled from available core photos, geological information, laboratory results, and log data from a subset of data from the Fort Hills Oilsands Lease. Core photos were cropped to existing sample intervals and attributed with numerous features including image metadata. Features extracted from the core photos included the mean values of several color channels and textural information. Available variables were used to predict PSD curves that consisted of 20 particle sizes ranging from 2000 μm down to 1.3 μm. Prediction models, that were either linear regression or neural networks, were designed for the three units of the McMurray Formation and trained using a 25% hold-out. The coefficient of determination ($R^2$) for the 44 μm size, exceeded 0.9 for all units. Intermediate particle sizes ranging from 355 μm down to 1.3 μm also had high $R^2$ values exceeding 0.8. Only in the Lower McMurray were very fine sizes below 2.8 μm difficult to predict. Error distributions were approximatly Gaussian for intermediate sizes. Given the high $R^2$ values, PSD prediction could provide valuable and informative data for reducing the uncertainty of geological models of the McMurray Formation.

For example, the global standard deviation of prediction errors for the 44 μm bin was 5%, 7.4% and 9.3% passing for the Upper, Middle and Lower units respectively.

## Introduction

Resource prediction in an oil sands mining context depends highly on understanding the material properties of the ore that is sent to the processing plant and consequently, to tailings. Bitumen grade and fines are two of the most important variables for understanding process performance, both are sampled from core using Dean Stark analysis for grade and PSD analysis for particle size. PSD analysis provides much more information than fines, often consisting of 20 or more particle size bins. Clays (< 3 μm) and coarse sands are also important for hydrotransport, plant performance and tailings, as are additional variables including connate water chemistry (Masliyah *et al*, 2004; O'Carroll, 1999; Sanford, 1983). Building geological models of the pertinent variables at a given time can be a challenge due to the variation in sampling practices and degradation of databases that occurs over time. Data can be lost through mergers and acquisitions; interpretations can change with changes to personnel or corporate structure; and sampling practice can change depending on company philosophy, laboratory methods or changes to the industry norm. Ultimately, databases become more heterotopic with time.

PSD sampling is relatively expensive and is one of the reasons that that PSD may be sampled less frequently; hence, this work focused on predicting complete PSD curves from other available data including core photographs, bitumen grade, and wireline logs. Completing a database in this sense is a type of imputation problem since the objective is to obtain isotopic information. The ability to accomplish imputation with a low prediction error is invaluable for reducing the uncertainty in geologic models, especially those with a high degree of heterogeneity like the McMurray Formation. Without imputation, spatial estimates of PSD are based entirely on surrounding data and are often highly uncertain due to the spatial correlation of McMurray deposit. This is true even at locations that are collocated with wells that have missing PSD because integrating all other information like core photographs is challenging and software limitations prevent the use of heterotopic databases.

## Method

Predicting a PSD curve is a multivariate problem with numerous available sources of data and numerous particle sizes to predict. Available sources of data include bitumen grade and water content, core photographs, facies logging from core, and wireline logs. The availability of DS analysis is prolific since data aquisition for oil sands mines requires it, and the availability of core photos is similar due to the ubiquity of cored drillholes. Wireline logs are also quite common. Using core photographs was challenging because they are not available in convenient numeric databases and there are practically an unlimited number of features that could be extracted from an image for prediction purposes. Standard practice for core photograph preparation by labs is the placement of depth markers and DS sample interval markers on the core boxes, which were essential for the developments made in this work. Regions of the photos associated with each DS sample were cut and saved into a library of images and cross referenced with spreadsheets of lab results to obtain depth information, bitumen grade and water content. An example of an image from the image library is shown in Figure 1. A limitation with core photographs is they show a 2D slice of information, whereas DS and PSD data are obtained from a 3D sample volume notched into the core. Features within the notch such as a mud clast may not show on the image.

Information that was extracted from each image included mean values of 7 color channels (blue, cyan, magenta, yellow, black, hue and saturation), the variance of the lightness channel, and five textures calculated from gray level co-occurrence matrices including entropy, contrast, correlation, energy and homogeneity. Pertinent log curves included the gamma ray (GR), bulk density (RHOB), neutron porosity (NPHI) and shallow resistivity (RS). Analysis of relationships and prediction performance was used for the selection of subsets of these 17 variables for each of the McMurray units. Facies information was used, but at a higher level than the refined picks provided by geologists. Based on the facies code, images were categorized as sand, muddy-sand, sandy-mud, or mud, where mud also contains coals. This approach increases the future utility of the prediction models since the advent of new facies codes, adjustment of existing facies, or the occurrence of unsampled facies in the area of interest would fall into one of these four categories.

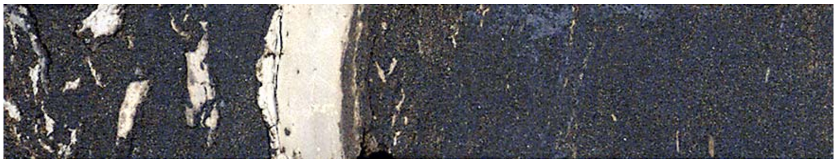

**Figure 1: 30 cm sample for Dean Stark (DS) analysis.**

PSD curves were characterized by 20 different particle sizes (bins) and accompanying percent passing values. Due to the high correlation between bins, especially the smaller particle sizes, curves were represented in a lower dimensional space using features extracted from the curves including the mean particle size (PSDM), standard deviation of particle sizes (PSDS), and sorting on the Krumbein-$\phi$ scale ($S\phi$) (Krumbeinn and Sloss, 1963). Complete curves are reconstructed with high accuracy using radial-basis function interpolation from these features, which are essentially lookup variables. Available information from core photos, lab results, and log data was used to predict these lookup variables, where PSDM and PSDS were used for the Upper and Middle McMurray units and PSDM and $S\phi$ for the Lower unit.

Prediction of the pairs of lookup variables was accomplished using neural networks (Bishop, 1995) with a single hidden layer with approximately twice as many nodes as the input layer that were chosen based on comparing error convergence of the training data versus the validation set, where validation was accomplished with a 25% hold-out. For the Upper McMurray unit, 15 inputs and 28 hidden nodes were used; for the Middle, 19 inputs and 30 hidden nodes; and for the Lower, 23 inputs and 40 hidden nodes. Gaussian activation functions were used for the Upper and Middle McMurray hidden and output layers; hyperbolic tangent and soft plus functions were used for the hidden and output layers respectively for the Lower. The complexity of networks necessary to minimize prediction error was related to the complexity of the geology of the three units. Networks were trained using back-propagation with the congugate-gradient algorithm with random restart. Multiple hold-out sets were considered and that with the minimum prediction error was retained. The early stopping approach was used to avoid overfitting, where stopping was based on observing the error convergence of the validation set during training.

The networks described involved all available data. In the absence of one data type such as wireline logs, different networks need to be designed and trained to handle image features and DS lab results. Prediction performance of the following combinations was assessed to determine the value of each data type: all sources; photos; photos+DS; photos+logs; DS; DS+logs.

## Examples

A small set of Fort Hills PSD curves was selected to cover a wide range of percent passing for the 44 $\mu$m size to compare true versus predicted curves, see Figure 2~~Figure 2~~. Predictions were made using a neural network with inputs from DS+logs+photos. Because predictions are constrainted to the [0,100]% interval, prediction errors are lowest in the upper/lower tails and highest through the central portion. Another set of curves selected to span a range of prediction errors associated with the 5%, 25%, 50%, 75% and 95% quantiles for the 44 $\mu$m size in the Middle McMurray is shown in Figure 3~~Figure 3~~. The interquartile range of prediction errors is within 5% passing in the Middle McMurray, where the bulk of mineable ore resides. A similar range is observed in the Upper McMurray and a slightly larger range for the Lower McMurray.

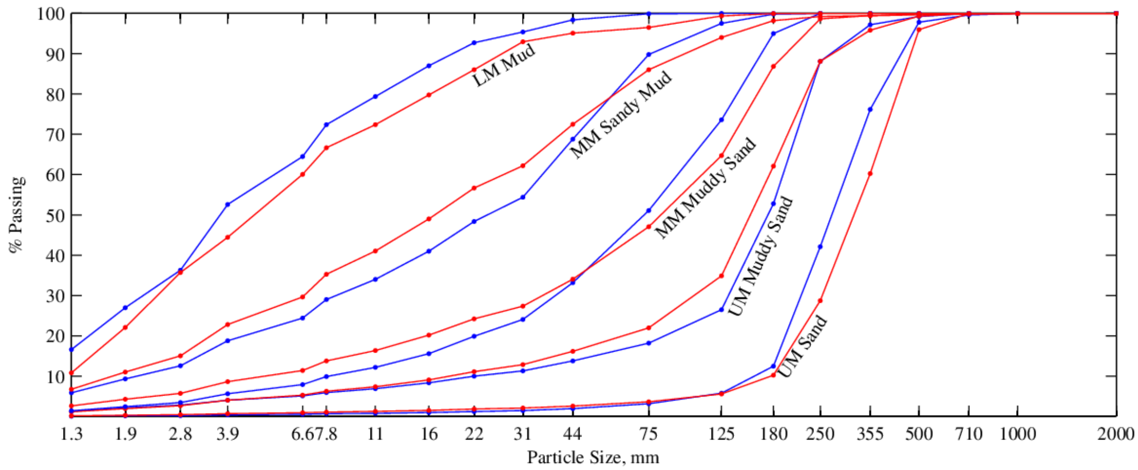


**Figure 2: Example predicted (red) versus true (blue) PSD curves for a wide range of 44 µm. UM – Upper McMurray; MM – Middle McMurray; LM – Lower McMurray.**

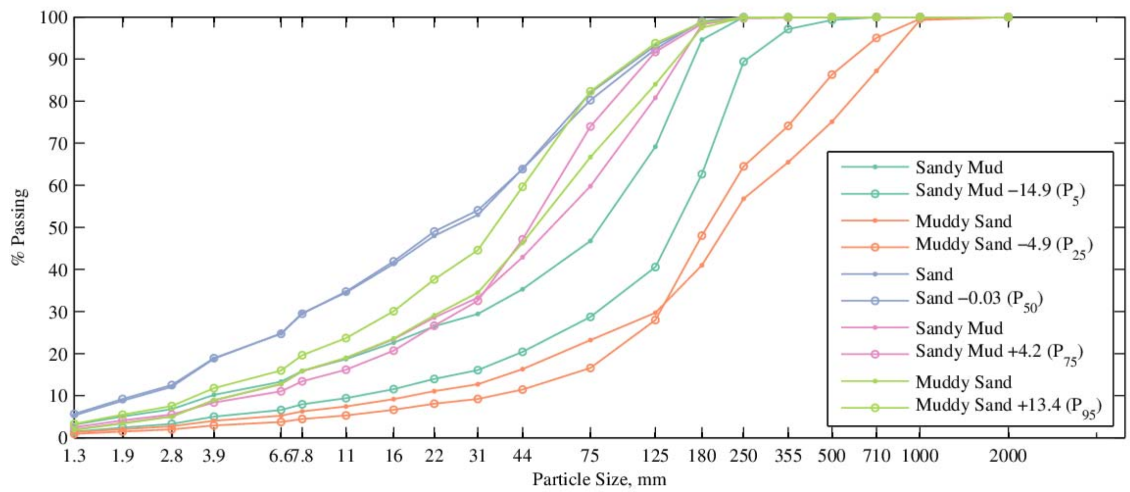


**Figure 3: Example predicted (circles) versus true (dots) curves for 44 µm errors that range from the P5 to P95 levels for the Middle McMurray.**

## Conclusions

Prediction models are a useful tool for assigning values to missing information from within available databases. In this study, a subset of core photographs, DS analysis, and wireline logs from the Fort Hills Lease were used to predict PSD, which is valuable information for oil sands mining. Training libraries of sample data and photographs were compiled from 113 Fort Hills wells covering an area of roughly 5 km by 9.5 km. The training librairies were then used to provide PSD predictions for 65 wells with missing PSD and to infill data gaps from 17 wells with predicted values. Although core photograph processing was time consuming and some effort was necessary in designing a successful prediction model, the potential time and cost savings would be insignificant compared to multiple years worth of redrilling legacy core holes where PSD was not obtained. Furthermore, in areas with little lab sampling, PSD prediction using existing datasets could help decrease uncertainty in long-range geological models or to give a preliminary indication of grain size distributions if drilling in a given area is delayed. To date, the predictions were made on a well-by-well basis only. Further work would be to upscale the interpreted data inside a geological model and to validate the results both in the model and in operations.

## Acknowledgements

The authors wish to acknowledge the continued support of Total E&P Canada in bringing this work to fruition. The authors thank Fort Hills Energy Corporation (FHEC) and Fort Hills Energy L.P. (FHELP) for their permission to use a portion of the Fort Hills dataset and to publish the results.

### References

Bishop, C.M., 1995. Neural networks for pattern recognition. Oxford University Press, 504 pages

Krumbein, W.C., Sloss, L.L., 1963. Stratigraphy and sedimentation. W. H. Freeman, San Francisco, 660 pages

Masliyah, J., Zhou, Z., Xu Z., Czarnecki, J., Hamza, H., 2004, Understanding water-based bitumen extraction from Athabasca oil sands. The Canadian Journal of Chemical Engineering, 82, 628-654

O'Carroll J.B., 1999. Factor affecting bitumen recovery from oil sands. Department of Chemical Engineering, University of Ottawa, 136 pages

Sanford, E.C., 1983. Processibility of Athabasca oil sand: interrelationship between oil sand fine solids, process aids, mechanical energy and oil sand age after mining. The Canadian Journal of Chemical Engineering, 61, 554-567