Paper 26

Geophysical Data Assimilation for Grade Modelling at Different Scales

Gloaguen, E. [1], Tirdad, S. [1], Perozzi, L. [1, 3], Bouchedda, A. [1], Caté, A. [1, 3], Blouin, M. [1, 3], Giroux, B. [1], Dupuis, C. [2]

1. Institut National de la Recherche Scientifique, centre Eau, Terre et Environnement, Québec
2. Département de genie géologique, Université Laval, Québec
3. geoLEARN (www.geoLEARN.ca)

ABSTRACT

In the mining industry, the accurate assessment of the grades of ore is obviously a priority for resources and reserve estimation. For decades, geophysical data have been used as a qualitative tool for exploration. Indeed, deterministic least-squares approaches that have been used to analyze those data only permit their qualitative interpretation. However, these data have a much richer potential if one can use appropriate tools to analyze them. Hudbay Minerals and its partners have acquired multiple airborne, surface and in-hole geophysical data permitting the inference of quantitative estimates of mining parameters. Here, in the first example, we used machine learning algorithms to translate geophysical log data into gold grades, while in the second example we present a workflow allowing the inference of the 3D density model and its uncertainty from surface gravity data at the resolution of the log data. In the first case, we showed that machine learning allows one to predict the probability of gold to be above a given cut-off. In the second on, it is demonstrated that one can use raw gravity data to model the density in 3D at the resolution of a smallest mining unit.

INTRODUCTION

In the mining industry, building an interpretative model of a mineral deposit is the first step in moving from resource to reserve, as it forms the basis for the resource estimation process. Since the primary source of information comes from drill-hole cores, the interpretative model of the deposit is highly dependent on the spatial distribution of these cores. However, many new discoveries are located at depths between 500 m and 2000 m. At these depths, the costs of drilling increase dramatically and the ability of the drill-holes to accurately sample the mineralization decreases, such that increased depth of drilling limits the quality of the information. High resolution geophysical methods such as radio-frequency ($0.1-5$ MHz) electromagnetic (EM) methods, borehole radar tomography or seismic tomography provide the geologist with new information that can be incorporated into the process of orebody modelling. In addition, it has been shown that, using an appropriate stochastic workflow, logs can be used to constrain seismic tomography between drill-holes, resulting in a significant increase in the accuracy of the tomographic images.

Integrated or assimilation approaches are increasingly used for orebody modelling and mine planning in both open-pit and underground mining ventures (Bellefleur et al., 2014). Drill-hole data are often complemented with other secondary, or so-called "soft" data (e.g., geophysical, geotechnical, and geochemical data) to improve the understanding of the deposit model. In this framework, an optimal estimate of the mineralization grades, and thus the available resources, is likely achieved by integrating these different yet complementary data types (Schodde, 2011). The importance of integrating "hard" and "soft" spatial data has long been recognized in the petroleum industry, where reservoir properties such as permeability and porosity need to be inferred from a limited number of drill-holes (Grana et al., 2012). Integrated modelling has also been used in the mining industry in ore reserve estimation. Recently, integrated techniques have been used to merge core log data with crosshole tomographic data for orebody modelling (Dimitrakopoulos and Kaklis, 2001).

However, data assimilation techniques must be adapted to the scale of the task. Indeed, one would like to take advantage of the surface coverage of airborne or ground-based geophysical techniques to better infer the extent and spatial heterogeneity of the properties a deposit, whereas one would like to infer the grades along the entire length of a diamond drill-holes using log data. In this paper we present two assimilation techniques adapted to the scale of two different mining purposes.

GEOLOGICAL CONTEXT

The first set of experiments were carried out at the Zn-Cu-Au Lalor deposit, a volcanogenic massive sulphide deposit recently discovered by Hudbay Minerals (Hudbay) and located in central-north Manitoba in the Snow Lake mining camp. Read Caté et al. (2015) for a complete review of Lalor deposit history. Thanks to Hudbay Mineral and its different partners (DGI, GSC, Geotech, …), the area of the mine site was extensively investigated using multiple airborne, surface and in-hole geophysical data sets.

The deposit is composed of at least twelve stacked ore lenses divided into base-metal (Zn-rich), gold and copper-gold ore lenses (Caté et al., 2015). Base metal and copper-gold lenses are mainly composed of massive to semi-massive sulphides and are easily distinguishable in drillcore. Gold lenses are composed of disseminated sulphides, which can be difficult to distinguish from the hydrothermally-altered sulphide-bearing wallrocks.

In "Proceedings of Exploration 17: Sixth Decennial International Conference on Mineral Exploration" edited by V. Tschirhart and M.D. Thomas, 2017, p. 379–384

The nugget effect and high variance in gold mineralization make the identification of mineralized bodies and mapping of their continuity in space challenging. Similarly to many other gold deposits, gold-bearing mineral assemblages at Lalor can be difficult to discriminate in drillcore, which can introduce errors in the process of selecting core intervals for assaying metal content. These errors can lead to an underestimation in the volume of ore and lead to the overlooking of economic zones during the exploitation of the deposit.

EXAMPLE 1 AT THE CORE SCALE

In this section, we will show how one can use the geophysical logs to infer gold grades through the use of machine learning algorithms.

Data

Combined drill-hole rock physical properties and metal assay data are available in a total of fourteen drill-holes intersecting lenses of the deposit. A typical dataset along a portion of a drill-hole is presented in Figure 1.

Assay data were collected by Hudbay and analyzed for metals by Hudbay and ACME Laboratories. During core-logging, assays were collected in 0.2 to 1 m-long intervals considered to be potentially metal-bearing. A total of eight elements were analyzed (Ag, As, Au, Cu, Fe, Ni, Pb and Zn), and a significant part of the measurements are below the detection limit. Only gold values are used in this study. Analysis and QA/QC methods are provided in Carter et al. (2012).

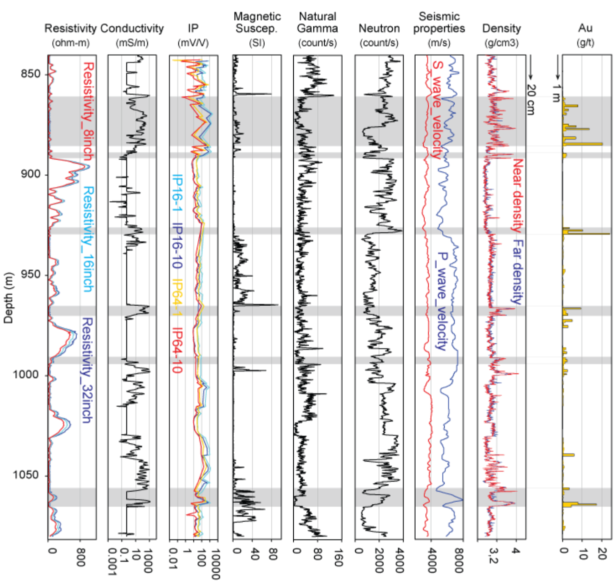


Figure 1: Typical geophysical logs at Lalor and gold grades.

Physical rock properties were logged by DGI Geoscience for Hudbay. A total of 15 rock properties were collected at a 10 to 20-cm spacing (Figure 1).

Classification Algorithms

Machine learning is an application of artificial intelligence, which identifies patterns in data and then makes predictions from those patterns (Pedregosa et al. 2011). If the data are in sufficient amount, the machine can learn complex relationships between the explicative data (here the geophysical logs) and the target data (here the probability of having gold grades above a given threshold). When the algorithm has sufficiently learned, it can predict the target variable using only the explicative ones. The machine learning algorithm also learns in time and increases its predictive power when new data are collected. Among the three types of machine learning methods (supervised, unsupervised, and reinforcement learning) supervised learning is the best suited for this work as its main goal is to learn a model from labelled training data that allows us to make prediction (Raschka, 2015). Here, the term supervised refers to a set of samples where both the desired output signals (label) and the predictive variables (logs and derived statistics) are already known. In this case, the label is a binary classification of samples having a gold value higher (positive class) or lower (negative class) than 1 g/t. A total of six machine learning algorithms are tested here:

1) The k-Nearest Neighbors method uses labelled neighbouring points in the Euclidian space formed by the input features to predict classes,

2) The Naïve Bayesian method uses Bayes theorem to evaluate the probability of an event (class) to occur given the value of the input data,

3) Support Vector Machine is a discriminative classifier formally defined by a separating hyperplane,

4) Classification trees are decision trees built by using thresholds on input features at each split, and

5) Ensemble algorithms. The latter is particularly well suited for this kind of problem as they combine the predictions of several base estimators built with a given learning algorithm in order to improve robustness over a single estimator. These algorithms are particularly resistant to noisy data and to outliers (Breiman, 2001). Here, the Random Forest and the Gradient Tree Boosting algorithms are tested. Both algorithms use decisions trees as base estimators.

All algorithms can be tuned with a series of various algorithm-specific parameters that significantly contribute to the robustness, variance and bias of the classification. The choice of the best parameters is done through the training/validating process. Here, the algorithms and the tuning tools proposed by ScikitLearn (Pedregosa et al., 2011) with Python have been used in the implementation.

Feature Engineering

One of the main issues in machine learning for geosciences is the so-called feature engineering, or, in other words, the transformation of data to more meaningful variables with regards to the task. In this case, the gold assay measurements have been composited into 1 m-long intervals using weighted averages in order to homogenize the dataset and avoid biased statistics during modelling. Physical properties were logged with 10 to 20 cm intervals, such that between 5 and 10 measurements were taken within each composited gold assay interval. Part of the physical properties were log-transformed to unskew their distribution. This compositing allowed for the computation, within each interval, of derived features from the summary statistics (minimum and maximum value, mean, median, standard deviation and variance) of each physical property in order to enhance the high resolution information contained in the geophysical logs.

Results of Training

In the case of Lalor, with the available data, the results on the training set showed that the ensemble methods, and particularly the gradient tree boosting presents the most accurate reproduction of the true data. It must be stated here that this is not a general rule. The user must be aware that the various algorithms must be retested repeatedly for applicability during the entire mine life. Indeed, as the number of data increase, one algorithm might appear to perform better, particularly for the semi-supervised algorithms like k-means.

Prediction Along Drill-holes

The prediction results obtained on one of the test drill-holes are presented in Figure 2. All zones with high gold values according to assays have been classified as gold-bearing by the prediction model in these drill-holes. The probability of an interval being gold-bearing is distributed as high probability value intervals centred on the actual high-grade gold zones, but with a more smoothed distribution than measured gold grades (presented on a log scale in Figure 2). Note that a few intervals detected by the model as potentially gold-bearing have not been assayed or have been assayed and include only gold values below 1 g/t.

Feature Importance

Ensemble algorithms allow the evaluation of the importance of the features used for the classification (Figure 3). The most informative features are derived from neutron, natural gamma and resistivity logs. Both neutron and natural gamma give insights on the elemental composition of the host rocks. Their classification powers are probably derived from the variations in rock composition, in part due to the presence of alteration associated with gold mineralization (Caté et al., 2015). Resistivity is a proxy for the presence of conductive minerals such as sulphides (e.g., pyrrhotite and chalcopyrite), which are associated with the presence of gold.

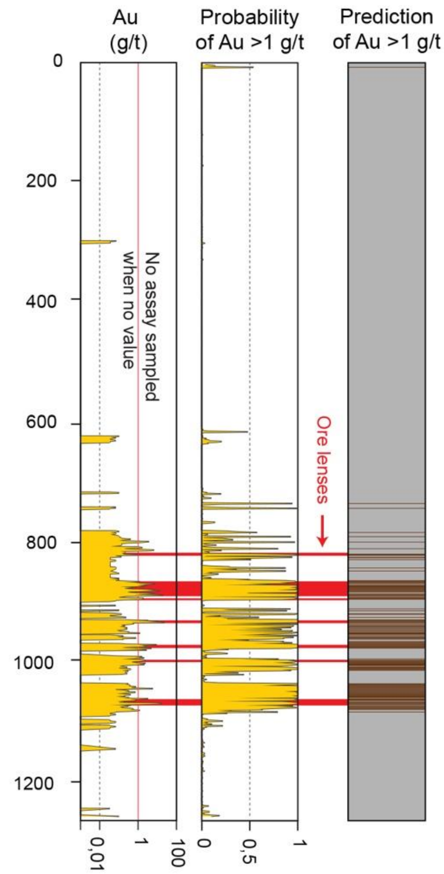
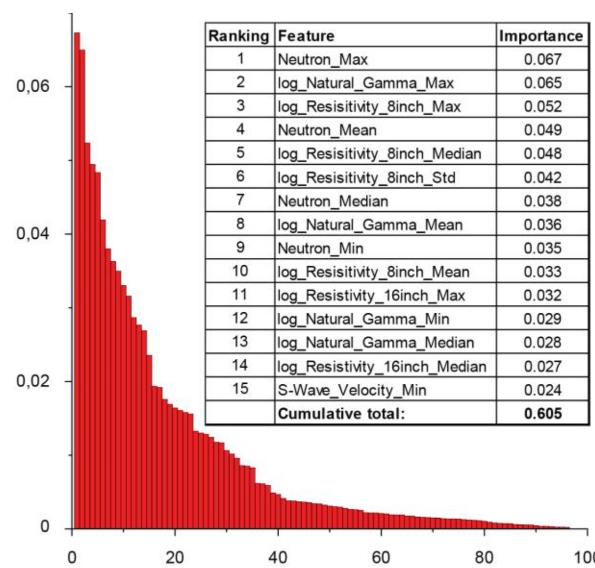


Figure 2: Left: gold grades from assays; Middle: probability of being above the cut-off; Right: pseudo-log.



Ranking	Feature	Importance
1	Neutron_Max	0.067
2	log_Natural_Gamma_Max	0.065
3	log_Resisitivity_8inch_Max	0.052
4	Neutron_Mean	0.049
5	log_Resisitivity_8inch_Median	0.048
6	log_Resisitivity_8inch_Std	0.042
7	Neutron_Median	0.038
8	log_Natural_Gamma_Mean	0.036
9	Neutron_Min	0.035
10	log_Resisitivity_8inch_Mean	0.033
11	log_Resistivity_16inch_Max	0.032
12	log_Natural_Gamma_Min	0.029
13	log_Natural_Gamma_Median	0.028
14	log_Resistivity_16inch_Median	0.027
15	S-Wave_Velocity_Min	0.024
	Cumulative total:	**0.605**

Figure 3: Feature importance from the training of the random forest algorithm.

EXAMPLE 2: AT THE MINING CAMP SCALE

Introduction

For many years, gravity has proven to be an efficient tool for mineral exploration for its capability to detect ore-related density contrasts in the ground. However, even though these data may have continuing value, they are almost always neglected in resources evaluation.

In order to infer density from gravity data, it is necessary to invert the gravity potential field, with its requisite ambiguities, in order to produce a meaningful density distribution of the subsurface (Sen and Stoffa, 1991). The conventional approach uses a deterministic least squares algorithm providing, by design, a single smooth and very low resolution 3D model of the spatial variability of the density in the ground. The downside of this method is that the least squares approach over samples the mean while under sampling the extreme low and high values that are often the most values important in resources evaluation, and over evaluates the spatial-extent of the anomalies. Consequently, this type of deterministic model can be only used for exploration and does not have the resolution required for further use in the mine design. To overcome this problem, the direct stochastic approach seeks the posterior probability density of the parameters, unlike the conventional inversion where a single model is sought. In real applications, this optimization step can be computationally expensive, which is the main impediment to using such stochastic inversion techniques. A secondary limitation of stochastic-inversion approaches is that the final model could overly depend on the initial model, as the solution of the inverse problem could have local minima. Differently optimized models with the same response could be achieved using different initial models (Grana et al., 2012). In many cases, stochastic inversion methods and their application on geophysical data (Chasseriau and Chouteau, 2003; Gloaguen et al., 2005; Shamsipour et al., 2010) have been useful to deal with the limitations of inversion problem (Shamsipour et al., 2013).

Method

In the first step, a set of 99 stochastic realizations are computed from the conceptual geological model (also called the training image or TI) of the study area (Caté et al., 2015) and the facies identified on cores along the drill-holes (Figure 4). This TI contains nine lithological units showing different physical rock properties. The algorithm used to generate the realizations is the multiple point simulation technique (MPS). The objective is to simulate the facies on generated scenarios in such a way that the statistics and the patterns of the models are similar to those of the TI and conditional on the drill-hole facies data (Journel and Zhang, 2006). Figure 4 shows a selection of sections at A-A’ of these realizations within the 3D model shown in the right of the figure. We then compute the set of stochastic density models using conventional sequential Gaussian co-simulations. The density data were collected using density logs along 12 drill-holes and the vertical and horizontal variograms were computed on these data within each facies. During this step each geological unit is populated separately using its own histogram of density observed at drill-hole logs and its own variogram. At this stage, we have 99 density models, all equiprobable in terms of drill-hole data, conceptual training image and variogram. The scenarios are all mimicking the TI and are at very high resolution but none fit perfectly the raw gravity data.

The next step of the workflow consists of a global heuristic inversion combining all the density scenarios in a unique model that best fit the raw gravity data. The gravimetric response of the body, discretized into rectangular prisms, is calculated by summing the contribution of each of these prisms using the approach of Li and Chouteau (1998). In our case, the 3D volume of the ground is discretized into a grid of 50 x 50 x 50 cells along the x, y and z-axis, respectively. The dimensions of the study area are 974 x 1929 x 1663 m^3.

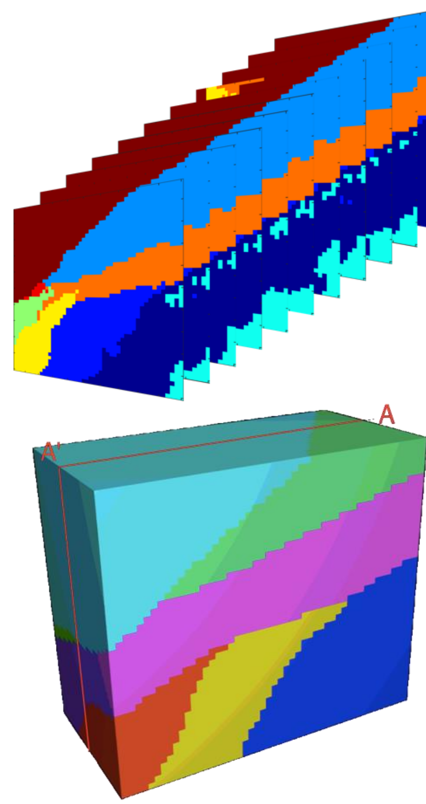


Figure 4: (top) Slices of the simulated realizations from the geological model using MPS; (bottom) 3D TI adapted from Caté et al. (2015)

In the next step, as shown in Figure 5, each realization within the set is iteratively combined with the others using a gradual deformation (GD) optimization technique (Hu, 2000) in order to minimize the difference between the measured and calculated gravity data. The technique is based on the linear combinations of two unconditionnal Gaussian random functions (Y_1 and Y_2) showed in Equation 1 (Hu, 2002). Such combination results in a Gaussian random model preserving the same mean and covariance as the two initial models (Feller, 2008).

$$Y(r) = Y_1 \cos(r) + Y_2 \sin(r) \qquad \text{(Equation 1)}$$

In Equation 1, r is the gradual deformation coefficient. In this work, we used a modified version of gradual deformation method that permits combining conditional realizations and it is called gradual conditioning (GC) (Hu, 2002). In this case, the realization built from combining three conditional realizations ($\alpha_1 Y_1 + \alpha_2 Y_2 + \alpha_3 Y_3$) is also conditional to the logged density data. Equation 2 shows how the weights are calculated for gradual conditioning:

$$\alpha_1 = \frac{1}{3} + \frac{2}{3}\cos r \qquad \text{(Equation 2)}$$
$$\alpha_2 = \frac{1}{3} + \frac{2}{3}\sin\left(-\frac{\pi}{6} + r\right)$$
$$\alpha_3 = \frac{1}{3} + \frac{2}{3}\sin\left(-\frac{\pi}{6} - r\right)$$

where r belongs to $[-\pi,\pi]$. The constraint on the weights forces the sum and the sum of the square of the weights to be equal to one. These constraints are called the conditioning and the covariance constraints, respectively (Hu, 2002).

The minimization resulted from GD depends only on the r parameter. For the optimal value of r the algorithm returns the minimum value of our objective function, which is the misfit between our measured and observed gravity data.

The objective function here is defined as the root mean square error between the observed gravity data and the measured data from the algorithm. The minimum value of data misfit obtained after the gradual deformation of all the simulated density models equals 0.5517 mGal. The corresponding density model to this minimum misfit is illustrated in Figure 3. As can be seen in this optimized density contrast model (Figure 6) there is a clear correspondence to the geological model and the lithological units.

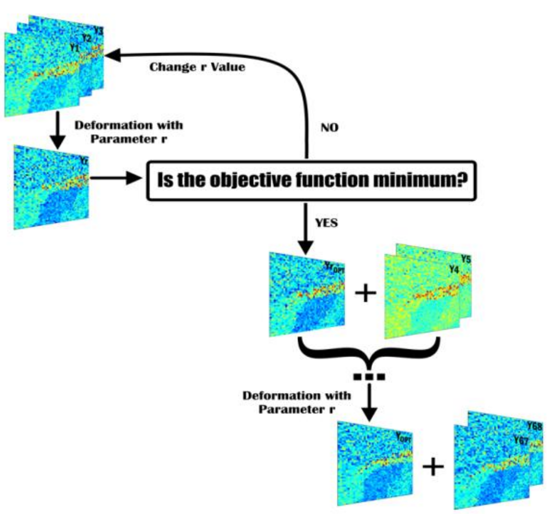


Figure 5: Workflow of the stochastic inversion methodology.

CONCLUSIONS

At the core scale:

Contrary to base metals (e.g., Zn and Cu), the presence of non-visible gold in a rock is difficult to assess, even for the trained eye of an experienced geologist. The combination of a set of rock physical properties measured at closely-spaced intervals along drillcore with machine learning allows the detection of gold-bearing intervals with an adequate success rate close to or better than that of an experienced geologist. The predicted recall, precision and accuracy show room for further improvement, which could be obtained by the collection of more training data or different data types, and/or improved feature engineering.

Because the resulting prediction is continuous along drillcore, this kind of tool may help the geologist when making decisions on which intervals to select for assay sampling. Missed gold-bearing intervals will then be significantly reduced, which in turns will potentially increase the reserve.

Along with predicting the presence of metals in rocks, physical properties combined with machine learning have the potential to classify lithologies, characterize hydrothermal alteration, and estimate exploration vectors and geotechnical information in drillcore (e.g., Ross et al., 2013). The combination of these predictions can significantly improve the work of logging geologists and the quality of geological interpretations as well as decisions taken during a drilling campaign and during the exploitation of a mine.

This method has been developed on a restricted set of data (14 drill-holes) and the success rate of predictions will increase with increasing data collected. This method should be applied from the very beginning of the exploration stage (i.e. starting from the discovery hole) so that the initial model could be trained and continuously updated with new drill-holes.

At the site scale:

The results show that the optimized model of density represents an improved structural similarity with the reference model compared to conventional simulations. Also, the algorithm allows accounting for the uncertainty related to each geological and geophysical measurement.

The proposed approach does not require use of a conventional least squares inversion method and each petrophysical model (geology, geophysics) retains its very high resolution, opening the door for a quantitative use of the density model in the resources evaluation.

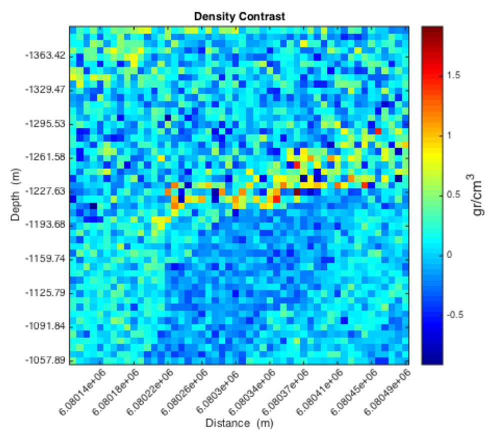


Figure 6: Optimized density contrast model.

ACKNOWLEDGEMENTS

We would like to express our gratitude to Hudbay Minerals Company and Mira Geoscience for providing us with the data of the Lalor deposit.

REFERENCES

Bellefleur, G., E. Schetselaar, and J.A. Craven, 2014, Targeted Geoscience Initiative 4: a synthesis of the methodology project at the Lalor VMS Deposit, Manitoba, Geological Survey of Canada, Open File 7863, 50 p.

Breiman, L., 2001, Random forests: Machine Learning, 45 (1), , 5–32.

Carter, R. P., T. Schwartz, S. West, and K. Hoover, 2012, Pre-feasibility study technical report, on the Lalor deposit, Snow Lake, Manitoba, Canada: HudBay Minerals, http://s1.q4cdn.com/305438552/les/doc_downloads/our_business/Lalor%20PreFeasandTechnicalReport_Mar30_v001_o9by3l.pdf, accessed 30 January 2017.

Caté, A., P. Mercier-Langevin, P.S. Ross, S. Du , M. Hannington, B. Dubé, and S. Gagné, 2015, Geology and Au enrichment processes at the Paleoproterozoic Lalor auriferous volcanogenic massive sulphide deposit, Snow Lake, Manitoba, in J. Peter and P. Mercier-Langevin, eds., Targeted Geoscience Initiative 4: Contributions to the understanding of volcanogenic massive sulphide deposit genesis and exploration methods development: Geological Survey of Canada, Open File 7853, 131–146.

Chasseriau, P., and M. Chouteau, 2003, 3D gravity inversion using a model of parameter covariance: Journal of Applied Geophysics, 52, 59–74.

Feller, W., 2008, An introduction to probability theory and its applications: John Wiley & Sons.

Gloaguen, E., D. Marcotte, M. Chouteau, and H. Perroud, 2005, Borehole radar velocity inversion using cokriging and cosimulation: Journal of Applied Geophysics, 57, 242–259.

Grana, D., T. Mukerji, J. Dvorkin, and G. Mavko, 2012, Stochastic inversion of facies from seismic data based on sequential simulations and probability perturbation method: Geophysics, 77, M53–M72.

Hu, L.Y., 2000, Gradual Deformation and Iterative Calibration of Gaussian-Related Stochastic Models: Mathematical Geology, 32, 87–108.

Hu, L.Y., 2002, Combination of dependent realizations within the gradual deformation method: Mathematical Geology, 34, 953–963.

Journel, A., and T. Zhang, 2006, The necessity of a multiple-point prior model: Mathematical Geology, 38, 591–610.

Li, X., and M. Chouteau, 1998, Three-dimensional gravity modeling in all space: Surv. Geophys., 19, 339–368.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. irion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, 2011, Scikit-learn: Machine learning in python: JMLR, 12, 2825–2830, http://www.jmlr.org/papers/v12/pedregosa11a.html.

Raschka, S., 2015, Python machine learning: Packt Publishing Ltd.

Ross, P.-S., A. Bourke, and B. Fresia, 2013, A multi-sensor logger for rock cores: Methodology and preliminary results from the Matagami mining camp, Canada: Ore Geology Reviews, 53, 93–111.

Schodde, R., 2011, Recent trends in gold discovery: Presented at NewGenGold Conference.

Sen, M.K., and P.L. Stoffa, 1991, Nonlinear one-dimensional seismic waveform inversion using simulated annealing: Geophysics, 56, 1624–1638.

Shamsipour, P., D. Marcotte, M. Chouteau, and P. Keating, 2010, 3D stochastic inversion of gravity data using cokriging and cosimulation: Geophysics, 75, I1–I10.

Shamsipour, P., D. Marcotte, M. Chouteau, M. Rivest, and A. Bouchedda, 2013, 3D stochastic gravity inversion using nonstationary covariances: Geophysics, 78, G15–G24.