Predicting Copper Concentrations in ACID Mine Drainage: A Comparative Analysis of Five Machine Learning Techniques

Getnet D. Betrie[1], Solomon Tesfamariama[1], Kevin A. Morin[2], Rehan Sadiq[1]
[1] Okanagan School of Engineering, UBC, Kelowna, BC, Canada getnet.betrie@ubc.ca
[2] Minesite Drainage Assessment Group, Vancouver, BC, Canada contact@mdag.com

Abstract
This study presents machine learning techniques to develop models that predict acid mine drainage (AMD) quality using historical monitoring data of a mine site. The machine learning techniques explored in this study include artificial neural networks (ANN), support vector machine with polynomial (SVM-Poly) and radial base function (SVM-RBF) kernels, model tree (M5P), and K-nearest neighbors (K-NN). Input variables (physico-chemical parameters) that influence drainage dynamics are identified and used to develop models to predict copper concentrations. For these selected techniques, the predictive accuracy and uncertainty were evaluated based on different statistical measures. The results showed that SVM-Poly performed best, followed by the SVM-RBF, ANN, M5P and KNN techniques. Overall this study demonstrates that the machine learning techniques are promising tools for predicting AMD quality.

Key Words: acid mine drainage, machine learning, artificial neural network, support vector machine, model tree, K-nearest neighbors.

Introduction
Predicting the future drainage chemistry is important to assess potential environmental risks of AMD and implement appropriate mitigation measures. However, predicting the potential for AMD can be exceedingly challenging because its formation is highly variable from site-to-site depending upon mineralogy and other operational and environmental factors (USEPA 1994). For this reason, laboratory test, field test and a variety of modeling approaches have been used for predicting the potential of mined materials to generate acid and contaminant (USEPA 1994; Maest et al. 2005; Price 2009). Laboratory and field tests are often undertaken for short periods of time with respect to the potential persistence period of AMD; hence, they may inadequately mimic the evolutionary nature of the process of acid generation (USEPA 1994). Predictive modeling approaches have been used to overcome the uncertainties inherent in short-term testing and avoid the prohibitive costs of very long-term testing.

Predictive models for AMD can be classified as empirical and deterministic models (USEPA 1994; Perkins et al. 1995; Maest et al. 2005; Price 2009). Empirical models describe the time-dependent behavior of one or more variables of a mine waste geochemical system in terms of observed behavior trends. These empirical models are site specific and based on years of monitoring at a mine site. Thus the AMD prediction accuracy of the empirical models depends heavily on the quality of available data. On the other hand, deterministic models describe system in terms of chemical and/or physical processes that are believed to control AMD. Deterministic models often require intensive specific studies and data but collecting those data with sufficient accuracy is often difficult and expensive. In this study, the empirical modeling approach is investigated to make use of monitoring data collected at a mine site. An example of empirical model was provided by Morin and Hutt (1993, 2001). These researchers developed an empirical model, named empirical drainage-chemistry model (EDCM), and applied it for the prediction of drainage quality using historical data from mine sites. The EDCM approach involves defining correlation equations using least-linear fitting between concentrations and other geochemical parameters typically pH and sulphate.

In this paper, advanced types of empirical approach, named machine learning techniques have been explored to develop models that predict future drainage quality using existing data. These techniques are

useful to develop predictive models but their use requires insight into the learning problem formulation, selection of appropriate learning methods, and evaluation of modeling results to achieve the stated goal of the modeling activity (Rech and Barai 1997; Cherkassky et al. 2006). This paper compares the predictive accuracy and uncertainty of five selected machine learning techniques using rigorous statistical tests. The selected machine learning techniques are artificial neural networks (ANN), support vector machine with polynomial (SVM-Poly) and radial base function (SVM-RBF) kernels, model tree (M5P), and K-nearest neighbors (K-NN). The prediction accuracy refers to the difference between observed and predicted values. On the other hand, predictive uncertainty refers to the variability of the overall error around the mean error. The detailed description of machine learning methods and the approach are presented in the following sections.

Machine Learning Techniques

Machine learning is an algorithm that estimates an unknown dependency between mine waste geochemical system inputs and its outputs from the available data. The machine learning consists of input variables x, mine waste geochemical system that return output y, for each input variable, and a machine learning algorithm that selects mapping functions (i.e., $\bar{y} = f(x_i, y_i)$), which describes how the mine waste geochemical system behaves. The available mine waste geochemical data are usually represented as a pair (x_i, y_i), which is called an example or an instance. The goal of learning (training) is to select the best function that minimizes the error between the system output (y) and predicted output $(\bar{y})$ based on examples data. These examples data used for training purpose are called a training data set. The process of building a machine learning model follows general principles adopted in modeling: study the problem, collect data, select model structure, build the model, test the model and iterate (Solomatine and Ostfeld 2005). There are various types of machine learning techniques but artificial neural networks, support vector machine, model tree and K-neighbors are explored in this study.

Artificial neural network (ANN)

Artificial Neural Network (ANN) is one of machine learning techniques that consist of neurons with massively weighted interconnections (Bishop 1995). These neurons are arranged as input layer, hidden layer and output layer as displayed in Figure 1. The task of input layer is only to send the input signals to the hidden layer without performing any operations. The hidden and output layers multiply the input signals by set of weights and either linearly or nonlinearly transforms results into output values. These weights are optimized during ANN training (calibration) process to obtain reasonable predictions accuracy.

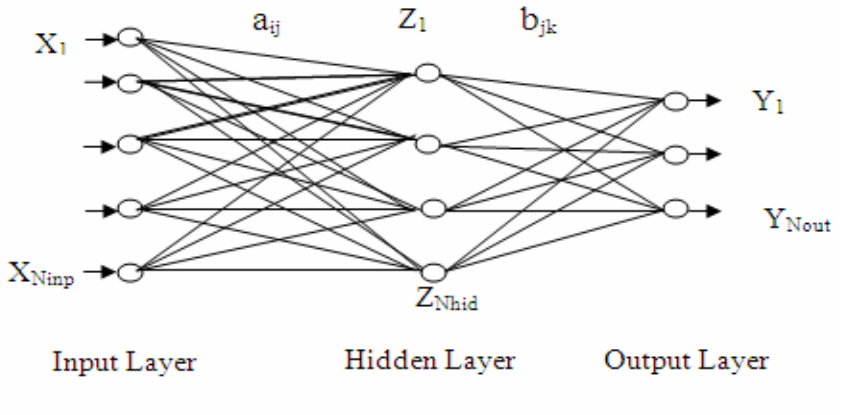

Figure 1. Neural networks

Support vector machine (SVM)

The Support Vector Machine was mainly developed by Vapnik and co-workers (Vapnik 1998; Cherkassky and Mulier 2007). Its principle is based on the Structural Risk Minimization that overcomes the limitation of the traditional Empirical Risk Minimization technique under limited training data. The Structural Risk Minimization aims at minimizing a bound on the generalization error of a model instead of minimizing the error on the training dataset. The SVM algorithm was first developed for classification problems and then adapted to address regression problems. In this study, the basic idea of SVM regression is illustrated since a regression problem is solved. Given a training dataset (x_i, y_i), where x_i is the $i-th$ input pattern and y_i is corresponding target value $y_i \in \Re$. The goal of SVM regression is to find a function $f(x)$ that has at most ε deviation from actually obtained targets y_i for all training data (Vapnik 1995). The variable ε is called loss function, and measures the cost of the errors on the training data. The function f is represented using a linear function in the feature space as shown in Equation 1.

$$f(x) = \langle w, x \rangle + b \text{ with } w \in x, b \in \Re \qquad \text{(Equation 1)}$$

where $\langle .,. \rangle$ denotes the dot product in x. Nonlinear regression problems are very common in most engineering applications. In such case, a nonlinear mapping kernel K is used to map the data into a higher-dimensional feature space or hyperplane by the function Φ. The kernel function, $K(x_i, x_i) = \langle \Phi(x_i, \Phi(x)) \rangle$ can assume any form. In this study, the polynomial (SVM-Poly) and radial basis function (SVM-RBF) kernels are used. These kernels are presented in Equation 2-3.

Polynomial Kernel: $$K(x_i, x) = (\gamma \langle x_i, x \rangle + \tau)^d, \gamma > 0 \qquad \text{(Equation 2)}$$

Radial Basis Function Kernel: $$K(x_i, x) = \exp(-\gamma \| x_i - x \|^2), \gamma > 0 \qquad \text{(Equation 3)}$$

where $\gamma, \tau, and\ d$ are kernel parameters.

Model trees (M5P)

Model trees are tree-based models for dealing with continuous-class learning problems with piecewise linear functions, originally developed by Quinlan (1992). The schematic representation of mode tree is depicted in Figure 2. Given a training set T, this set is either associated with a leaf or some test is chosen that splits T into subsets corresponding to the test outcomes and the same process is applied recursively to the subsets. For a new input vector, (i) it is classified to one of the subsets and (ii) the corresponding model is run to produce the prediction. The steps to build the M5P model trees are building the initial tree, pruning and smoothing. In the building tree procedure, the splitting criterion in each node is determined. The splitting criterion is based on treating the standard deviation of the class values that reach a node as a measure of the error at that node, and calculating the expected reduction in error as a result of testing each attribute at that node (Wang and Witten, 1997). The pruning procedure makes use of an estimate of the expected error that will be experienced at each node for the test data. First, the absolute difference between the predicted value and the actual class value is averaged for each of the training examples that reach that node. The smoothing process is used to compensate for the sharp discontinuities that will inevitably occur between adjacent linear models at the leaves of the pruned trees. This is a particular problem for models constructed from a small number of training instances. The smoothing procedure in M5P first uses the leaf model to compute the predicted value, and then filters that value

along the path back to the root, smoothing it at each node by combining it with the value predicted by linear model for that node.

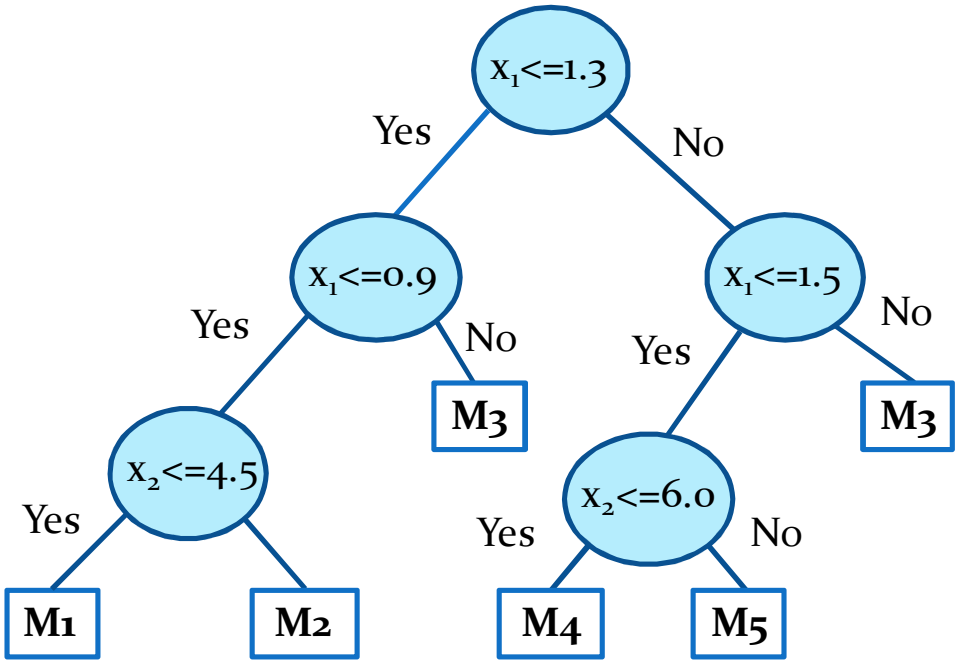

Figure 2. Schematic representation of model tree, where X_i and M_i represent the input parameters and local models

K-nearest neighbors (K-NN)

K-nearest neighbor (K-NN) technique is an instance-based learning, where training examples are stored and the generalization is postponed until a prediction made (Mitchell, 1997). The K-NN classifies an unknown input vector x_q by choosing the class of the nearest example x in the training set as measured by a Euclidean distance. For real valued target functions, the estimate is the mean value of the K-nearest neighboring examples. However, this method is slow for a bigger test set because it involves finding which member of the training set is closest to an unknown test instance (x_q) is calculated the distance from every member of the training set and select the smallest (Witten and Frank, 2005). One way of improving this limitation is to consider weighted distance. Thus, the distance weighted K-NN algorithm was used in this study. In this algorithm, each K-neighbor x_i is weighted according to their distance from the query point x_q using Equation 4.

$$f(x_q) = \frac{\sum_{i=1}^{k} w_i f(x_i)}{\sum_{i=1}^{k} w_i} \qquad \text{(Equation 4)}$$

where weight w_i is a function of distance $d(x_q, x_i)$ between x_q and x_i.

Parameter Selection for Drainage Quality

Modeling of AMD chemistry using machine learning methods requires defining control variables that dictate the process. According to Morin and Hutt (1994, 1997, 2010), the most important factors that control drainage chemistry are:

- geochemical production rates;
- infiltration of water;
- elapsed time between infiltration events;
- residence time of water within rocks;
- internal temperatures;
- pore-gas concentrations of oxygen and carbon dioxide;
- Solubility products; and
- Particle size.

Geochemical production rates refer to the production rates of elements, acidity and alkalinity under acid and pH-neutral conditions rock. Once produced, these reaction products are either flushed by flowing water or accumulate in the rocks. The drainage quality may be changed if the products are flushed out of waste rocks. The infiltration of water controls the amount of reaction products to be flushed. The importance of the elapsed time between infiltration events is that it provides opportunity for reaction products to accumulate in the flow channel. It is worth noting that both the volume of infiltrating water and the elapsed time between infiltration events affect the concentrations and loadings observed in the basal seepage. The residence time of water within waste rocks refers to time required for infiltrated water to pass through rocks. Thus residence time determines the time of reaction products to occur in the basal seepage. The internal temperatures and pore-gas concentrations of oxygen and carbon dioxide can affect the rates the rates of pyrite oxidation and acid generation, and the amount of reactions products.

In this study, the physico-chemical parameters monitored for over 25 years from waste rocks were obtained from Island Copper Mine, British Colombia, Canada (Morin et al. 1995). The chemical parameters include pH, conductivity, alkalinity, acidity, sulphate and metals. The physical parameters include flow rate, dissolved oxygen, temperature. The amount of precipitation infiltrated into waste rocks was estimated from climatic data. The climatic data such as precipitation, minimum and maximum temperature were obtained from Environment Canada (2011). The evapotranspiration of the site was calculated using Hargreaves method (Hargreaves and Riley 1985). The Hargreaves method uses minimum and maximum temperature, and solar radiation to estimate evapotranspiration. Then the site effective precipitation was estimated as difference between precipitation and evapotranspiration.

The input variables to machine learning techniques should consist of all relevant variables that influence the AMD's generation process. However, overlapping information of input variables should be avoided to simplify the task of the training algorithms. In order to make parsimonious selection of inputs, the linear correlations between input and output variables were examined. The correlation between the copper concentrations and the others variables with their time lags are shown in Table 1. It shows that the current copper concentration highly correlated to previous time copper concentrations (i.e., t-1 to t-5) and others variables except the effective precipitation. While the pH is negatively correlated to current time copper concentration, conductivity and acidity are positively correlated to it. Moreover, the table shows that the current time concentration has strong correlations with pH, conductivity and acidity at previous time state. Therefore, pH, conductivity, acidity and previous time copper concentrations were used as control variables and the current time copper concentrations were used as output.

Table 1. Correlation between copper concentration and other variables with time lags

Time(t)	pH	Conductivity	Acidity	Effective Precipitation	Cu
t	-0.74	0.52	0.81	-0.02	1.00
t-1	-0.69	0.51	0.78	-0.05	0.94
t-2	-0.68	0.50	0.76	0.01	0.90
t-3	-0.65	0.50	0.74	-0.01	0.89
t-4	-0.62	0.49	0.72	-0.01	0.86
t-5	-0.59	0.49	0.71	-0.01	0.84

The statistical summary of the input and output variables considered in this study are summarized in Table 2. These variables are pH, conductivity, acidity, and dissolved copper. The statistics of the data include minimum, maximum, mean, standard deviation and coefficient of variation. This table shows that pH dataset distribution has the lowest variability, followed by the conductivity, acidity and copper. In addition, the variability of the independent variables (i.e., pH, acidity, conductivity and effective precipitation) and the dependent variable (copper) are with reasonable range.

Table 2. Variables and summary of the data used in the study

Variables	Min	Max	Mean	SD	CoV[$]
pH*	3.56	6.46	4.53	0.45	9.93
Conductivity*	500	3140	1717.7	596.9	34.8
Acidity*	1.5	570	202.5	128.6	63.5
Cu**	0.01	2.50	0.93	0.48	51.74

[*] Used as inputs in model development
[**] Used as a model output
[$] Coefficient of variation

Model Development and Validation

The dataset was divided into training and testing sets following the k-fold cross-validation method (Mitchell 1997). In the k-fold cross-validation method, the dataset is subdivided into k subsets preferably of equal size. Next, the k-1 subsets are used to train the machine learning models and the remaining one subset are used for testing the models. In this study, each subset has the size of 128 values and ten-fold cross-validation with stratification was repeated 10 times. This exercise provided a total 100 independent models error for each machine learning techniques. This method is computationally very intensive; however, the authors strongly believe that it provided reliable results.

Model Evaluation

The prediction accuracy helps to evaluate the overall match between observed and predicted values for each machine learning technique. The predictive accuracy of each machine learning technique was evaluated using the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Root Relative Squared Error (RRSE), Relative Absolute Error (RAE), where the smaller value indicates a better technique. Moreover, a paired t-test was used to determine whether the mean of error estimates of one machine learning technique is significantly different from another technique. The equations of the error estimates are given in Equation 5- 8.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_o - y_p)^2}{n}} \quad \text{(Equation 5)}$$

$$MAE = \frac{\sum_{i=1}^{n}\left|y_o - y_p\right|}{n} \quad \text{(Equation 6)}$$

$$RRSE = \sqrt{\frac{\sum_{i=1}^{n}(y_o - y_p)^2}{\sum_{i=1}^{n}(y_p - \overline{y_p})^2}} \quad \text{(Equation 7)}$$

$$RAE = \frac{\sum_{i=1}^{n}\left|y_o - y_p\right|}{\sum_{i=1}^{n}\left|y_o - \overline{y_p}\right|} \quad \text{(Equation 8)}$$

where y_o and y_p represent the observed and predicted outputs, $\overline{y_p}$ represents the mean the predicted output and n represents the number of examples presented to the learning algorithms.

Predictive uncertainty refers to the variability of the overall error around the mean error. The predictive uncertainty of each machine learning technique was evaluated using averaged error residuals of the models. Next the averaged residuals of the five techniques are assumed as random variable and 18 probability distributions were fitted using @Risk software (Palisade Corporation, 2005).

Results and Discussions

Performances of the five machine learning techniques in terms of predicting copper concentrations, four evaluation methods are presented in Table 3. This summary shows the Min, mean and the Max performance of selected five techniques. The best and worst values show the performance range of the techniques, whereas the mean value shows the average performance of the techniques over testing datasets. These indicators are important for making decision in environmental risk analysis. Comparison of the mean performances indicate that the SVM-Poly is the best technique, followed by SVM-RBF, ANN, M5P and K-NN techniques on all evaluation methods. The K-NN technique was found to be the least performed predictive method.

Table 3. Performance of models over testing sets

Models	MAE			RMSE			RAE (%)			RRSE (%)		
	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
ANN	0.07	0.17	0.56	0.09	0.22	0.62	11.67	41.05	269.37	13.73	43.19	232.81
SVM-Poly	0.06	0.14	0.42	0.08	0.18	0.46	9.06	32.47	152.07	10.83	36.11	135.91
SVM-RBF	0.06	0.15	0.38	0.09	0.20	0.41	11.32	35.59	218.01	15.16	38.86	188.86
K-NN	0.09	0.28	0.68	0.13	0.34	0.73	17.52	61.29	260.93	17.52	61.29	260.93

M5P	0.05	0.21	0.68	0.06	0.26	1.44	10.57	47.48	205.46	14.39	50.93	193.48

The observed and predicted copper concentrations using different models are presented in Figure 3. This figure shows that the overall predictions of the SVM-Poly fits best to the ideal line (i.e., the diagonal line), followed by SVM-RBF, ANN, M5P and K-NN. This confirms that the conclusion made above on the performance of the techniques. Most of the predicted values of the M5P technique are below the ideal prediction line, whereas the K-NN predictions are above the idea prediction line. This implies that the K-NN technique over estimates and the M5P underestimates the overall predictions. As can be seen in Figure 3, the SVM-Poly predicted the high values better, which is desirable as it is conservative for environmental risk analysis decision making. Whereas the higher values are under predicted by SVM-RBF, ANN and M5P techniques and the K-NN could not predict the higher values at all. This suggests K-NN should not be used for decision making in which the associated risk is high.

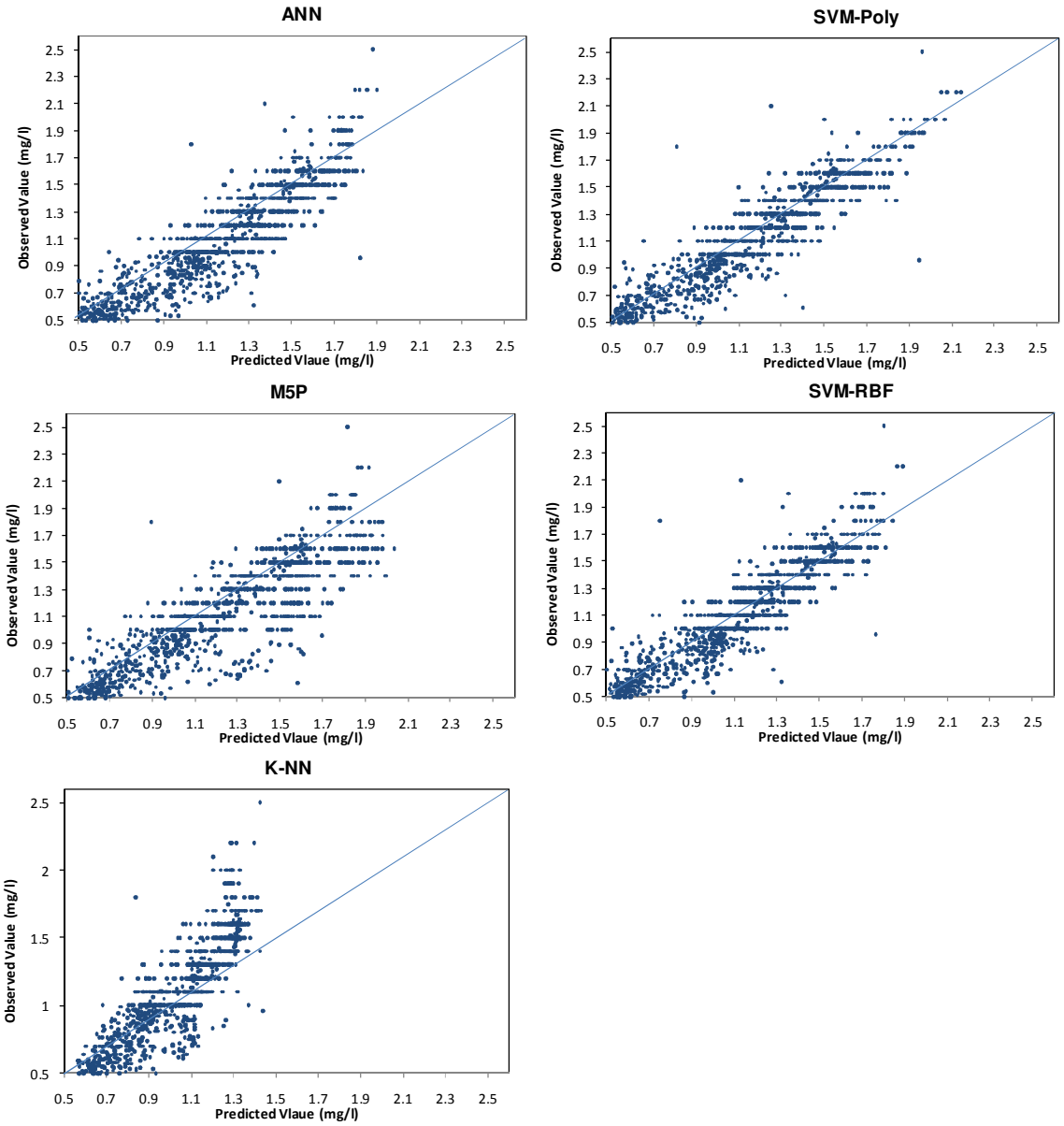

Figure 3. Scatter plots of observed and predicted copper concentrations

A paired t-test was used to determine whether the mean of error estimates of one machine learning technique is significantly different from another technique. This t-test is important to ensure that the obtained results are not because of a particular dataset used. The p-values, at a significance level of p = 0.05, of the paired t-test on prediction error residuals of the five techniques are shown in Table . The test results show that the obtained results are statistically significant except the SVM-Poly and SVM-RBF techniques predictions. Although the SVM-Poly performed better than SVM-RBF on all model evaluation methods, this test indicates that difference is not statistically significant.

Table 4. The p-values of the paired t-test on error residuals

	ANN	K-NN	M5P	SVM-Poly	SVM-RBF
ANN	1	0.000	0.009	0.000	0.005
K-NN		1	0.001	0.000	0.000
M5P			1	0.000	0.001
SVM-Poly				1	0.034
SVM-RBF					1

The predictive uncertainty of each machine learning technique was evaluated using error residuals. These error residuals were computed as a difference between measured and predicted copper concentrations. For each machine learning technique, the residuals of 100 independent models were calculated and averaged. Next the averaged residuals of the five techniques are assumed as random variable and 18 probability distributions were fitted. The lognormal probability distribution was the best fit to the residuals of the five techniques as shown in Figure 4. The best technique is the one that has residuals represented by narrowest, symmetrical and highest probability distribution. The SVM-Poly is the best in terms of the predictive uncertainty, followed by SVM-RBF and ANN techniques. The M5P shows a wider range of predictive uncertainty, whereas the K-NN showed the worst predictive uncertainty.

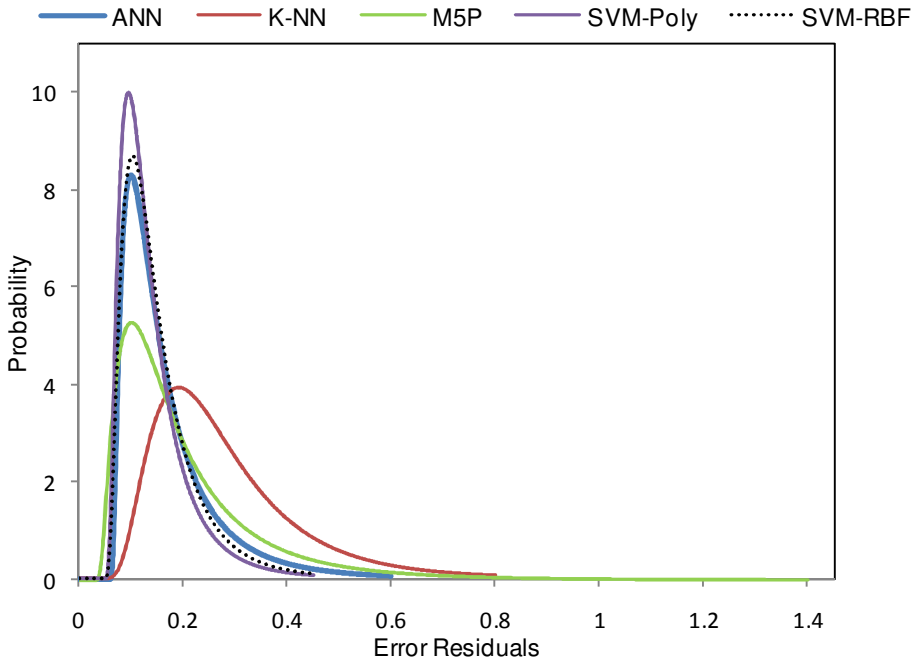

Figure 4. The probability distribution of the error residuals of five techniques

Summary and Conclusions

This study evaluated the accuracy and uncertainty of five machine learning techniques used for predicting acid mine drainage (AMD) chemistry from mine sites. The machine learning techniques investigated include artificial neural networks (ANN) with multilayer perceptrons, support vector machine with polynomial and radial base function kernels, model tree with M5P algorithm, and K-nearest neighbors. Physico-chemical parameters and time lag that influence the drainage chemistry were identified as important parameters and were used as inputs in the five techniques. Although the precipitation has been mentioned in the literature as the most controlling parameter of AMD generation, the correlation analysis revealed that it has not affected the AMD process for this case study area. The prediction results show that the identified inputs parameters represented the system dynamic. However, the predicted results likely improve if more parameters (e.g., flow rate, internal gases concentrations and temperature) are considered as more data become available in the future.

The experimental results showed that the support vector machine with polynomial kernel performed best both in terms of predictive accuracy and uncertainty evaluation methods. The support vector machine with radial base function kernel and artificial neural network provided better prediction results, followed by model tree and the K-nearest neighbors showed the worst performance in terms both evaluation measures. These results indicate that the process of AMD generation is highly nonlinear and could not be captured with techniques that build local linear models such as model tree and K-nearest neighbor techniques.

This study shows that machine learning techniques are promising tools for predicting AMD chemistry. Their prediction results could be used to evaluate and identify cost-effective AMD management alternatives for a given mine site.

Acknowledgements

This research has been carried out as a part of NSERC-DG (Discovery Grant) funded by Natural Sciences and Engineering Research Council of Canada (NSERC). We also like to thank Minesite Drainage Assessment Group (MDAG) for providing valuable data to test various models.

References

Bishop, C. M. (1995) *Neural networks for pattern recognition*, Clarendon, Oxford.

Cherkassky, V., Krasnopolsky, V., Solomatine, D. and Valdes, J. (2006) Computational intelligence in earth sciences and environmental applications: Issues and challenges. Neural Networks, Vol. 19, pp. 113–121.

Cherkassky, V. S. and Mulier, F., 2007. *Learning from data: concepts, theory, and methods*, John Wiley & Sons, New Jersey.

Environment Canada (2011) Climate data online. Viewed 5 August 2011, http://climate.weatheroffice.gc.ca/climateData/canada_e.html

Hargreaves, G., and Riley, J. (1985) Agricultural benefits for Senegal River basin. Journal of Irrigation Drainage, E-ASCE, Vol. 111, pp. 113–124.

Maest, A. S., Kuipers, J. R., Travers, C. L., and Atkins, D. A. (2005) Predicting water quality at hardrock mines: methods and models, uncertainty and state-of-the-Art. Kuipers & Associate and Buka Environmental.

Mitchell, T. (1997) *Machine Learning*, McGraw Hill.

Morin, K.A., and Hutt, N. M. (1993) The use of routine monitoring data for assessment and prediction of water chemistry. In, Proceedings of the 17th Annual Mine Reclamation Symposium of Mining Association of British Columbia, May 4-7, 1993, Port Hardy, Canada.

Morin, K.A., and Hutt, N. M. (1994). An empirical technique for predicting the chemistry of water seeping from mine-rock piles. In, Proceedings of the international conference on the Abatement of Acidic Drainage, April 1994, Pittsburgh, USA.

Morin, K. A., Hutt, N. M. and Horne, I. A. (1995) Prediction of future water chemistry from Island Copper Mine's On-Land Dumps. In, 19th Annual British Columbia Mine Reclamation Symposium, June 19-23, 1995, Dawson Creek, Canada.

Morin, K. A. and Hutt, N. M. (2001) Prediction of minesite-drainage chemistry through closure using operational monitoring data. Journal of Geochemical Exploration, Vol. 73, pp. 123-130.

Morin, K. A. and Hutt. N. M. (1997) *Environmental Geochemistry of Minesite Drainage: Practical Theory and Case Studies*. MDAG Publishing, Vancouver, British Columbia.

Morin, K.A., Hutt, N.M. and Aziz M.L. (2010) Twenty-three years of monitoring minesite-drainage chemistry, during operation and after closure: The Equity Silver Minesite, British Columbia, Canada. MDAG.com Internet Case study # 35. Viewed 5 August 2011, http://www.mdag.com/case_studies/cs35.html

Palisade Corporation Inc. (2005) Guide to using @RISK. Advanced risk analysis for spreadsheets, Palisade Corporation, New York, USA.

Perkins, E. H., Nesbitt, H. W., Gunter, W. D., St-Arnaud, L. C. and Mycroft, J. R. (1995) Critical Review of Geochemical Processes and Geochemical Models Adaptable for Prediction of Acidic Drainage from Waste Rock, MEND Report 1.42.1

Price, W. A. (2009) Prediction Manual of Drainage Chemistry from Sulphidic Geologic Materials, MEND Report 1.20.1

Quinlan, J. R. 1992 Learning with continuous classes. In, Proceedings of 5th Australian Joint Conference on Artificial Intelligence, A. Adams & L. Sterling, World Scientific, Singapore.

Reich, Y. and Barai, S. V. (1999) Evaluating machine learning models for engineering problems. Artificial Intelligence Engineering, Vol. 13, pp. 257-272.

Solomatine, D. P. and Ostfeld, A. (2008) Data-driven modelling: some past experiences and new approaches. Journal of Hydroinformatics, Vol. 10, pp. 3-22.

USEPA (1994) *Technical document of acid mine drainage prediction*, Report EPA530-R-94-036, Office of Solid Waste, 44 p.

Vapnik, V. (1995). *The Nature of statistical learning theory*, Springer, New York.

Vapnik, V. (1998) *Statistical learning theory*, Willey, New York.

Wang, Y. and Witten, I. (1997) Inducing model trees for continuous classes. In, Proceedings of the 9th European Conference on Machine Learning, 128-137

Witten, I. H. and Frank, E. (2005) *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, San Francisco.

Acronyms

ANN	Artificial Neural Networks
EDCM	Empirical Drainage-Chemistry Model
K-NN	K-Nearest Neighbors
MAE	Mean Absolute Error
M5P	Model tree
RAE	Relative Absolute Error
RMSE	Root Mean Squared Error
RRSE	Root Relative Squared Error
SVM	Support Vector Machine
SVM-Poly	Support Vector Machine with Polynomial
SVM-RBF	Support Vector Machine with Radial Base Function