Orebody modeling with Uncertainty: a Bayesian Neural Network Approach

Snehamoy Chatterjee
Asst. Professor
Department of Mining Engineering,
NIT Rourkela, Orissa, India
E-mail: snehamoy@gmail.com

Sukumar Bandopadhyay
Professor of Mining Engineering
Dept. of Mining and Geological Engineering
University of Alaska Fairbanks
Fairbanks, Alaska 99775-5880
E-mail: sbandopadhyay@alaska.edu

Abstract
A Bayesian Neural Network (BNN) based spatial modeling technique is proposed here for orebody modeling. The Bayesian method for posterior probability calculation of the output parameter (grades) helps to calculate the uncertainty associate with the estimate. The paramatrers of the BNN model is selected by grid search algorithm. The expected value and the variance of block support are calculated by Markov chain Monte Carlo (MCMC) sampling from the posterior distribution at discretize points within the block. The BNN model is validated by applying the method in Walker Lake data set and comparing with ordinary kriging results. The results revealed that the proposed BNN method perform marginally better than ordinary kriging results. The variance map is less smooth than ordinary kriging. The proportional effect is also less in BNN-based model than ordinary kriging model.
Keywords: Spatial modeling, posterior distribution, ordinary kriging, uncertainty.

1. Introduction

Precious materials like gold, platinum etc. frequently occur in narrow, discontinuous concentrations. The precious metal deposits are characterized by the presence of localized and erratic high grades. The high variability of the data due to erratic grade variation presents one of the most challenging tasks for a resource estimator.

In geostatistical ore reserve-estimation methods using ordinary kriging or sequential Gaussian simulation, it is assumed that the relationship of ore grade in two different locations is a function of distance, with both mean and covariance stationary. However, because of such factors as geological structure, deposition environment, type of deposit, type of ore, and degree of mineralization, stationarity assumptions may not be valid. Apart from that, due to sparseness of data, fitting a theoretical variogram is generally difficult. As an outcome,

geostatistical techniques may produce poor interpolation results when no variogram shape is observed from sample data. Also, the spatial continuity modeling needs an in depth knowledge; the practitioner who have very limited knowledge in spatial continuity modeling may come up with wrong variogram parameters and direction of anisotropy which may leads to wrong interpolated results.

To handle sparseness of data and variogram free spatial modeling, several researchers have proposed different non-linear estimation techniques using neural network (NN) for ore grade estimation (Samanta et al., 2004 Samanta et al., 2005; Chatterjee et al., 2006). Samanta et al. (2004) used MLP (multi-layered perceptron) for NN modeling in the Nome offshore deposit and proved the superiority of the NN technique over the ordinary kriging technique. The attractiveness of the NN technique is that it is flexible and can capture complex non-linear relationships between input and output patterns. In addition, other information (e.g., rock types, stratigraphy, time, type of formation, etc.) can be easily incorporated in a NN model.

There are limitations, nevertheless, to conventional NN techniques. Neural network training, which is data-driven, has a tendency to fit noisy data. This can result in over-fitting, which needs to be avoided for better generalization of a model. In addition, it is important to address the issue of choosing the most suitable values of hidden node size, learning parameter, etc., for a given problem. Finally, unlike geostatistical simulation, there is no direct and reliable method in the conventional NN model for calculating uncertainties in estimates.

The aim of this study was to utilize the advantage of NNs in non-linear data modeling, with special emphasis on uncertainty calculation and model generalization. The focus of this work is on the Bayesian approach to NN training (Denison et al., 2002; Bernardo and Smith, 1994), because it takes into account uncertainty of estimates through the use of probability distributions.

2. Bayesian Neural Network Model for Orebody modeling

2.1 Brief overview of Bayesian Neural Network

A generalized technique, NN is used for input-output mapping of many systems including ore deposits. In orebody modeling, it is assumed that grade values in an ore deposit vary from one location to another. There is a complex input-output relation between spatial coordinates (northing, easting) and grade values; hence, an output grade estimate is considered to be a function of this complex relationship between input grade values at sampled spatial coordinates.

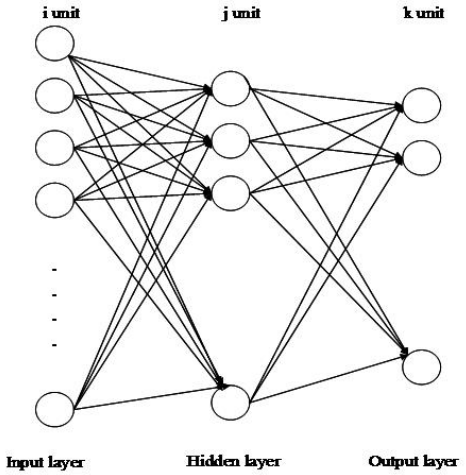


Fig 1 Architecture of a 3-layerd neural network model

A simple three-layer NN (Fig. 1) consists of an input layer, a hidden layer, and an output layer, interconnected by modifiable weights represented by links between layers. Each input vector is presented to the input layer, and the output of each input unit equals the corresponding component in the vector.

The basic procedure in NN learning is to start with an untrained network, present a training pattern to the input layer, pass the signals through the net, and determine the output, which is a function of weights. The outputs obtained from the model are compared with target output values of the same training pattern; any observed difference corresponds to an error. The error function is some scalar function of the weights and is minimized so that network outputs match the target output. Thus, the weights are adjusted to reduce this measure of error. The error (training error) on a pattern to be summed over the output units is the squared difference between the target output and network output obtained from the model:

$$E_D(W) \equiv \sum_{j=1}^{N} \frac{1}{2}\sum_{k=1}^{K}(t_k^j - z_k^j)^2 = \sum_{j=1}^{N} \frac{1}{2}\left|t^j - z^j\right|^2 \qquad (6)$$

where K represents the number of output units and, without loss of generality, we can assume only one output unit is present ($K=1$); t and z represent the target and the network output vector; $D=(t^1,t^2,...,t^N)$ represents the target output data from the training set, N represents number of training patterns; $W \equiv (w_1, w_2, ..., w_B)$ represents weight vector in the network; and B is the total number of weights and biases in the network.

In traditional NN models, networks are trained by the maximum-likelihood method (Bishop, 1996) for minimizing distance between the target output (t) and the NN model output (z) according to Eq. (6). Traditional NN models do not provide a

confidence interval of estimated values of output variables. Probabilistic interpretation in the maximum-likelihood method can be used for the NN learning process. Unlike the initialization of single weights during learning in the traditional NN model, the BNN model initializes the distribution of weights (Radford, 1996). Initialized weight distributions, known as prior distributions, are updated by Bayesian rule using training data. Suppose that patterns in a training set are independently drawn from a distribution $p(X,t)$, where $X \equiv (x_1, x_2, ..., x_d)$, and target output t is a deterministic non-linear function $z(X)$, plus zero-mean Gaussian noise. The probability of observing a single datum t for a given input vector X can then be defined as

$$p(t \mid X, W) \alpha \exp(-\frac{\beta}{2}(t - z(X|W))^2) \tag{7}$$

where $\beta = 1/\sigma_v^2$ is a hyper-parameter that controls the noise variance.

Before Bayesian learning of a MLP neural network, the prior probability distribution of the network weights W needs to be defined. A Gaussian prior probability distribution of networks can be defined as

$$p(W) = \frac{1}{V_B(\alpha)} \exp(-\alpha E_B) \tag{8}$$

where $V_B(\alpha)$ is a normalizing constant given by

$$V_B(\alpha) = \int \exp(-\alpha E_B) \tag{9}$$

Where, $E_B = \frac{1}{2}\|W\|^2 = \frac{1}{2}\sum_{i=1}^{B} w_i^2$

where B is the total number of weights and biases in the network.

After choosing the Gaussian model of prior probability distribution of network weights and an expression for the likelihood function, the Bayes' theorem can be used to find the posterior probability distribution of network weights. The posterior probability distribution of network weights can be obtained as:

$$p(W|D) = \frac{1}{V_S} \exp(-\beta E_D - \alpha E_B) = \frac{1}{V_S} \exp(-S(W)) \tag{10}$$

where $S(W) = \beta E_D + \alpha E_B$ and $V_S(\alpha, \beta) = \int \exp(-\beta E_D - \alpha E_W) dW$ and

$S(W) = \frac{\beta}{2}\sum_{c=1}^{N}(t^c - z(X^c|W))^2 + \frac{\alpha}{2}\sum_{i=1}^{B} w_i^2$

2.2 Mean and Variance Calculation at Block Support

To calculate the volume estimation of a deposit, the grade value is required to be calculated at block support. To calculate expected block value of a two-dimensional block, the area is discretized into $M \times M$ points. For example, block K is discretized into 3×3, that is, 9 points (Fig. 2). The developed BNN model is then applied to all discretized points within the block. The Markov chain Monte Carlo (MCMC) sampling (MacKay, 1992) is performed from the posterior distribution of Eq. (26) at all discretized points.. To calculate the block mean and block variance of block K, the following equations are applied:

$$z_K = \frac{1}{M^2 * p} \sum_{i=1}^{M^2} \sum_{j=1}^{p} z_j^i \text{ and } \sigma_K = \frac{\sum_{i=1}^{M^2} \sum_{j=1}^{p} (z_j^i - z_K)^2}{(M^2 * p - 1)} \tag{11}$$

where z_K is the expected value of grade at block support, z_j^i is the j^{th} MCMC sample from the posterior distribution of network output the BNN model at discretized point i, M^2 is total discretized points within the block, p is the number of MCMC samples from posterior distribution and σ_K is the block variance.

3. Model validation

To validate the proposed Bayesian neural network-based model for spatial modeling with uncertainty, the model was applied on one exhaustive data set. The well known Walker Lake data set was selected for this purpose (Isaaks and Srivastava 1989). The sample data set consists of 470 data points (Isaaks and Srivastava 1989) and presented in Figure 2. Although, the regularized term in the BNN model helps to reduce model over-fitting, we have performed a cross validation study to assure that the developed method is not an over-fitted model. To perform the cross-validation study, the sample data set was divided into training, validation, and testing data sets. The BNN model was developed using the training data, the validation data was used to restrict model over-fitting, and the testing data, which was not used at any time during the model development, was used for testing the generalization ability of the model. The training data set, validation data set, and testing data set consists of 50% (235), 25% (118), and 25% (117) of the total sample data, respectively. The statistical similarity of these three data sets is verified by using the bivariate ANOVA F-test and by comparing cumulative distribution functions and basic statistics.

The BNN model was developed for spatial modeling the parameter *V* of Walker Lake data set. For every sample location, spatial positions (northing, easting) were used as input parameters, and *V* values were used as output for the network model. The input data were normalized in the range of zero to one before performing network training. Data normalization of input variables results in the individual components of the input vector being recognized as equally important by the network.

A three-layer NN model was used in this study. The tan sigmoid and linear activation functions (Bishop, 1996) were used at the hidden and output layers,

respectively. The weight vector was initialized with the Gaussian model of prior probability distribution, using hyper-parameters α and β. The initial values of α and β have a great impact on model performance, so it is important to select those values properly according to the algorithm discussed in Section 2.4. Grid pattern search method (Bazarra et al. 1993) was implemented to find out best initial values of α and β. But searching within such a vast search area is computationally expensive. So, coarse grid search was performed for finding the zone of minimum error. Then the fine grid search was performed within at minimum zone area. The initial values of α and β obtained using grid search algorithm are 0.2 and 22, respectively. Although the grid search algorithm is a computationally fast algorithm; however, it is not optimum one. The meta-heuristic algorithm can be applied for searching near optimum hyper-parameters in computationally efficient manner; however that is beyond the scope of the paper.

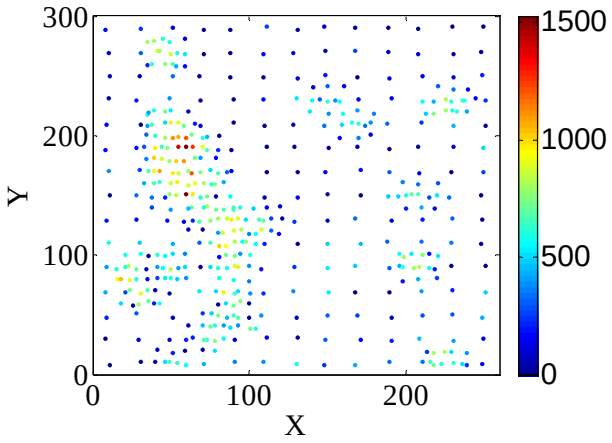


Figure 2 Sample data set of Walker Lake data

After initialization, the model was trained using the standard conjugate gradient descent algorithm to minimize total error function (Bishop, 1996). For posterior calculation, the NN was trained based on Bayesian theory. In each cycle, after a given number of epochs, hyper-parameters α and β were re-estimated. The number of epochs and number of cycles were kept constant at 100 and 3, respectively. In every cycle, the NN model was trained until it reached the convergence criteria. Convergence criteria in the BNN were determined by observing the mean squared error of the validation data set. The Bayesian regularized training algorithm stops the training of the neural network model immediately, when validation data error shows an increasing trend, which helps to restrict model over-fitting. The training algorithm is assigned with the condition that the algorithm will stop if the validation errors started showing an increasing trend or no further improvement of the validation error for 15 iterations. Therefore, the training stops after 27 epochs since validation error is constant after 12 epochs. In each cycle, α and β values were updated. Final values of α

and β were obtained after completion of all cycles. The posterior probability distribution of weights of the network along with α and β were then used to calculate the mean prediction and standard deviation at unknown points in Walker Lake area.

The performance evaluation of BNN model was carried out by analyzing the error statistics of the testing data set. The error of estimates was calculated by comparing true test data value with estimated value at the test data location by developed BNN model. The mean error, the mean absolute error, the mean squared error (MSE), and the coefficient of determination (R^2) values were used as performance indicators for the model. The error histogram is nearly normally distributed with mean 0. Therefore, it can be conclude the proposed BNN method is globally unbiased. The error statistics of the testing data set is presented in Table 1. The mean error reveals that the proposed method slightly overestimated the true value. The coefficient of determination (0.63) and error variance (41102) reveals that the proposed BNN model can explain 63% of the test data variance (test data variance is 105051).

Table 1 Error statistics of testing data set using BNN model

	Bayesian neural network
Mean error	9.2
Means absolute error	148.2
Mean squared error	41278
Error variance	41102
R^2	0.63

The conditional un-biasedness of the BNN-based model was examined by studying the bivariate distribution of estimated and true values. Fig. 3 represents the scatter plot of the true and estimated values of the test data. It is observed from the figure the estimated values are reasonably matching the true values and they are scattered around the 45-line on the scatter plot. Therefore, it can be conclude that the BNN-based method for spatial modeling of the Walker Lake data is conditionally un-biased.

After developing the BNN model of Walker Lake data set, the model is applied for estimating the entire area. The area consists of 78,000 points on a 260 x 300 rectangular grid. At each point, we have calculated the BNN-based estimate, and ordinary kriging estimate. The variogram parameters and direction of anisotropy for ordinary kriging estimate are taken from Isaaks and Srivastava (1989). The summary statistics of true values and two estimated values are presented in Table 2. Both the methods produces estimates whose means are higher than the mean of the true value. The results revealed that both the estimates are globally biased for the Walker Lake date set. However, the globally biasdness is more in

case of BNN-based model than ordinary kriging. The estimated maps of BNN-based model and ordinary kriging are presented in Figure 4

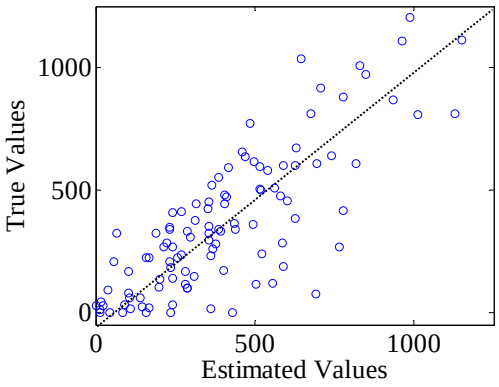
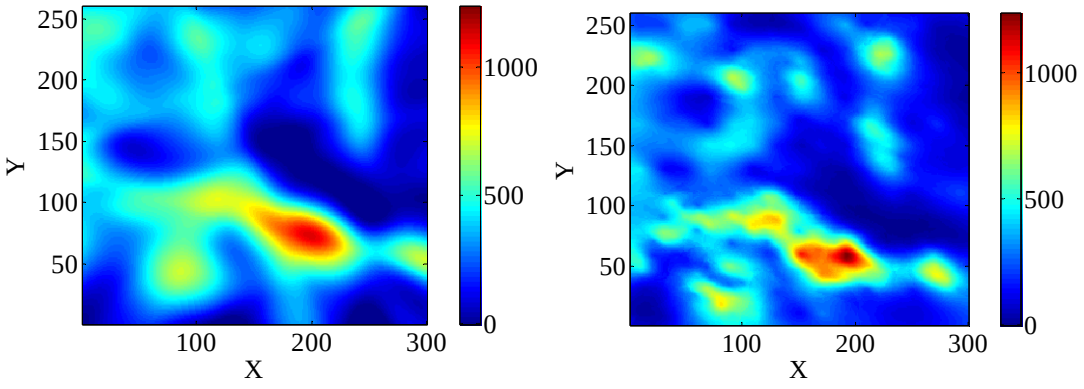


Figure 3 Scatter plot of true vs. estimated values of testing data set

The standard deviation of the estimated data reveals that the BNN-based model estimates variability close to the true values variability. The ordinary kriging map is smoother than BNN-based map. The reason of lees smoothing map in BNN modeling is that unlike ordinary kriging the BNN-based model doesn't combine several sample values for estimating at unknown points. It is known that the smoothing is a consequence of combining several samples to from an estimate.

Table 2 Summary statistics of BNN and ordinary kriging models estimates

	True	Bayesian neural network	Ordinary Kriging
Estimated points	78000	78000	78000
Mean	278	281	286
Standard Deviation	249.8	207.6	193.2



(a) BNN (b) ordinary kriging

Figure 4 Estimated maps of Walker Lake date using (a) BNN; and (b) ordinary krging

The error statistics of both the estimates are presented in Table 3. The differences observed in the statistics of the estimates are also supported by the error statistics. The higher R^2 value supports that BNN based model performed slightly better than the kriging estimates. The error distributions (Fig 8) of both the methods are nearly following the normal distribution with close to zero mean.

Table 3 Error statistics of true values, estimated values of BNN and ordinary kriging models

	Bayesian neural network	Ordinary kriging
Mean error	0.22	8.05
Means absolute error	107	113
Mean squared error	21040	22380
Error variance	21041	22315
R^2	0.67	0.64

4. Summary

A Bayesian neural network-based spatial model is proposed in this paper. The spatial coordinates and any other secondary information can be directly used as inputs for the BNN model. The hyper-parameters α and β are selected using the grid search algorithm. The structure of the BNN is an important factor in developing a good model. The main network structure parameter—the number of hidden nodes—was determined by a rigorous exercise carried out by changing the initial values of the hyper-parameters. The values of selected hyper-parameters and the number of hidden nodes provide the near optimum network model for spatial estimation.

The model is validated by applying the algorithm in Walker Lake exhaustive data set and comparing with ordinary kriging results. The BNN model of Walker Lake data reveals that the model is globally unbiased and the model can explain the 64% of the total data variance. The results also show that the BNN model is conditionally unbiased. However, this is a case specific observation. The comparative result with ordinary kriging reveals that the BNN model performs marginally better than the ordinary kriging. The main advantage of BNN model is that it is easy to develop. Unlike kriging, no spatial continuity modeling or spatial anisotropy modeling need to be developed. Therefore, any mining engineer or geologist with limited exposure to spatial modeling can use the BNN model for their deposit evaluation. The results also revealed that the BNN model generates less smooth variance map as compared to ordinary kriging map.

References

Bazarra, M. S., Sherali, H. D. and Shetty, C. M. 1993, Nonlinear Programming: Thoery and Algorithms. 2nd edition. John Wiley & Sons, New York, NY.

Bernardo, J.M., Smith, A.F.M., 1994. Bayesian Theory, 1st edn., Wiley Publication, New York, NY, 608pp.

Bishop, C.M., 1996. Neural Networks for Pattern Recognition, 1st edn., Oxford University Press, New York, NY, 504pp.

Chatterjee, S., Bhattacherjee, A., Samanta, B., Pal, S.K., 2006. Ore grade estimation of a limestone deposit in India using an Artificial Neural Network. Applied GIS 2(1), 2.1–2.20.

Denison, D., Holmes, C., Mallick, B., Smith, A.F.M., 2002. Bayesian Methods for Nonlinear Classification and Regression, 1st edn., Wiley Publication, New York, NY, 294pp.

Koike, K., Matsuda, S., 2006. New indices for characterizing spatial models of ore deposits by the use of a sensitivity vector and an influence factor. Mathematical Geology 38(5), 541–564.

MacKay, D.J.C., 1992. A practical Bayesian framework for backpropagation networks. Neural Computation 4(3), 448–472.

Radford, M.N., 1996. Bayesian Learning for Neural Networks (Lecture Notes in Statistics), Springer, New York, 207pp.

Samanta, B., Bandopadhyay, S., Ganguli, R., 2004. Data segmentation and genetic algorithms for sparse data division in Nome placer grade estimation using neural network and geostatistics. Exploration and Mining Geology 11, 69–76.

Samanta, B., Ganguli, R., Bandopadhyay, S. (2005). Comparing the predictive performance of neural network technique with ordinary kriging technique in a bauxite deposit. Transactions of the Institute of Mining and Metallurgy 114(3), 129–139.