ORE GRADE ESTIMATION OF A LIMESTONE DEPOSIT IN INDIA USING AN ARTIFICIAL NEURAL NETWORK

S. Chatterjee, Department of Mining Engineering, Indian Institute of Technology
A. Bhattacherjee, Department of Mining Engineering, Indian Institute of Technology
B. Samanta, Department of Mining Engineering, Indian Institute of Technology
S.K. Pal, Department of Mining Engineering, Indian Institute of Technology

This study describes a method used to improve ore grade estimation in a limestone deposit in India. Ore grade estimation for the limestone deposit was complicated by the complex lithological structure of the deposit. The erratic nature of the deposit and the unavailability of adequate samples for each of the lithogical units made standard geostatistical methods of capturing the spatial variation of the deposit inadequate. This paper describes an attempt to improve the ore grade estimation through the use of a feed forward neural network (NN) model. The NN model incorporated the spatial location as well as the lithological information for modeling of the ore body. The network was made up of three layers: an input, an output and a hidden layer. The input layer consisted of three spatial coordinates (x, y and z) and nine lithotypes. The output layer comprised all the grade attributes of limestone ore including silica (SiO_2), alumina (Al_2O_3), calcium oxide (CaO) and ferrous oxide (Fe_2O_3). To justify the use of the NN in the deposit, a comparative evaluation between the NN method and the ordinary kriging was performed. This evaluation demonstrated that the NN model decisively outperformed the kriging model. After the superiority of the NN model had been established, it was used to predict the grades at an unknown grid location. Prior to constructing the grade maps, lithological maps of the deposit at the unknown grid were prepared. These lithological maps were generated using indicator kriging. The authors conclude by suggesting that the method described in this paper could be used for grade-control planning in ore deposits.

1.0 INTRODUCTION

The quality of limestone ore plays an important role in the cost-effectiveness of an integrated cement plant. In the dry process of cement manufacturing, the consistent supply of quality raw material is necessary for the steady operation of the plant and to achieve optimum plant output. It is therefore essential that all the constituents of the raw material, such as calcium oxide (CaO), alumina (Al_2O_3), ferric oxide (Fe_2O_3), and silica (SiO_2), supplied to the plant be uniform or within permissible limits of variation. In particular, the kiln is one of the most decisive units for the performance of the cement plant. It is an established fact that SiO_2 has an adverse effect on kiln properties. It is monitored in the form of the silica ratio (SR) which is the ratio between silica and weighted sum of Al_2O_3 and Fe_2O_3. As the silica ratio increases, more heat is required to run the kiln. Nayak and Mallik (2002) have observed that the kiln is more effective where the silica ratio is less than 2.3. Another important parameter that makes a cement plant cost effective is the lime saturation factor (LSF): a ratio of CaO and weighted sum of alumina, silica and ferric oxide. The LSF plays a vital role for cement production because it contains CaO, the primary constituent of cement. It has been found (Ingram and Daugherty, 1991) that kiln operation and cement quality is improved where the CaO in limestone is more than 44%. Thus, both high LSF and a low silica ratio contribute significantly to the cost-effectiveness of a cement plant.

To maintain a consistent supply of suitable limestone grades to a cement plant, an ore grade model is generated, based on which a grade control plan can be produced to guide the excavation of the limestone. A good model aims to achieve reliable estimates of ore grades, representing the deposit as accurately as possible (Abdullah, 1990). The development of a good grade model is not an easy task in relation to Indian limestone deposits. Formed under complex geological structures, the limestones, typically, occur in multiple geological units. For example, in the present study there were as many as nine lithological units. The co-existence of various geological units side by side breaks up the spatial continuity of the deposits. In addition, the occasional presence of folds and faults further disrupts the deposit's spatial continuity. Limestone deposits therefore exhibit wide spatial variations (vertical and horizontal). Ideally wide patterns of variations in grade attributes would be captured by a suitable exploratory drilling program. However, in most cases, a drilling program adequate to gather enough borehole data to produce a reliable ore grade model would be prohibitively expensive. In most cases therefore, the data available for deposit modelling is limited. For example, in this particular study, the data indicated that borehole spacing varies from 50-100 meters: quite inadequate for modeling such complex patterns as under scrutiny here.

Mining engineers and geologists facing the challenge of delivering reliable ore grade models using limited resources apply a variety of grade estimation techniques. Geostatistics has established itself as the most versatile of these. However, of late, the neural network (NN) modeling technique has emerged as a popular alternative to geostatistics for spatial input-output mapping. For example, Samanta et al. (2004a & b) applied NN in ore grade modeling of a gold deposit. They found that the NN modeling approach outperformed geostatistics for gold grade estimation. Samanta et al. (2005) also investigated NN modeling in the context of a bauxite deposit. The outcome of this study indicated that both NN and ordinary kriging worked equally well for this deposit. Yama and Lineberry (1999) also demonstrated improved precision in ore grade estimation developed using NN in comparison with geostatistical techniques. Koike and Matsuda (2003) have used the NN technique for estimating content impurities in a limestone deposit.

In this paper, the NN modeling approach is investigated in the context of estimating the constituents of limestone ore. To justify the suitability of this technique for the deposit, a comparative evaluation of the effectiveness of NN versus ordinary kriging is also conducted. After the superiority of the NN modeling approach was established, grade models were prepared for each of the grade attributes CaO, Al_2O_3, Fe_2O_3 and SiO_2. Using this information, SR and LSF maps were prepared for the deposit. Because the NN model requires lithotype information at a particular location as an input variable, preparation of the grade models required spatial interpolation of lithological units using the lithological information provided by the exploratory borehole data. The lithological maps were prepared using an indicator kriging approach.

2.0 STUDY AREA AND DATA

Data for the present study refers to a limestone mine. The area of the mine is more than 6 square km with a maximum reduced level of 444 m. The terrain is very gently undulatating, the ground level rising about 10 m towards the northern and southern ends of the deposit as compared to the central part. Most of the area is covered with soil except where limestone outcrop is predominant. The area is drained by seasonal nalahs flowing across it. Geological mapping and a study

of structural details revealed that the limestone deposit was subjected to structural disturbances of moderate intensity as evidenced by the many minor and major folds present. Data from 212 exploratory boreholes were available for this study. The boreholes are located in a grid pattern; however, the spacing of the boreholes varies from 50-100 m. The average length of the boreholes is about 35 m. Samples from the boreholes were collected at intervals of less than 1 m. The ore is extracted from nine different lithologies: PPL (pink limestone), GGL (greenish gray limestone), DGL (dark gray limestone), LGL (light gray limestone), WTH (weathered limestone), UGL (upper gray limestone), shale (SHL), clay, and over burden soil (OBS). For the present study, raw borehole data were composited based on lithology. This approach avoided mixing up of grades from different lithologies. It produced 5149 composite values altogether. The mine has six working benches. Each has a height of 9 m. For the purposes of grade estimation, a separate estimation grid was created for each of the benches with a spacing of 50 m x 50 m, and estimation was done for each of the grid points of the entire grid.

3.0 STATISTICAL ANALYSIS AND DATA SUBDIVISION

For grade estimation purpose, all the major constituents of the limestone (CaO, Al_2O_3, Fe_2O_3 and SiO_2) were chosen. Prior to modeling, detailed statistical analyses of the composite data were carried out, because these data are the basis for the ore grade estimation. Table 1 presents the descriptive statistics of the data sets containing all the composite values. The table shows that grade attributes are highly erratic in nature, especially for the variables Al_2O_3 and Fe_2O_3, (co-efficient of variation: 86.55 and 138.7 respectively).

Statistics	CaO	Al_2O_3	Fe_2O_3	SiO_2
Sample No	5149	5149	5149	5149
Mean	40.227	3.0582	1.8927	19.915
Variance	116.01	7.066	6.892	179.848
Coefficient of variation	26.77	86.55	138.7	67.33
Skewness	-1.987	2.600	7.512	2.094
Kurtosis	3.390	7.867	97.948	4.373

Table 1 Basic statistics of the composited data set

The frequency distributions of the data sets were determined using histogram plots. Figure 1 presents the histogram plots of CaO, Al_2O_3, Fe_2O_3 and SiO_2. From these plots it can be seen that Al_2O_3, Fe_2O_3 and SiO_2 are skewed towards the right side, whereas CaO is skewed towards the left side. Investigations were done to determine whether or not Al_2O_3, Fe_2O_3 and SiO_2 values are lognormally distributed, since they are skewed towards right and visually it appears that they might follow lognormal distribution. However, K-S tests conducted on the log-transformed data of these variables indicated that the data follow a log-normal distribution and it was statistically significant at 5% level of significance.
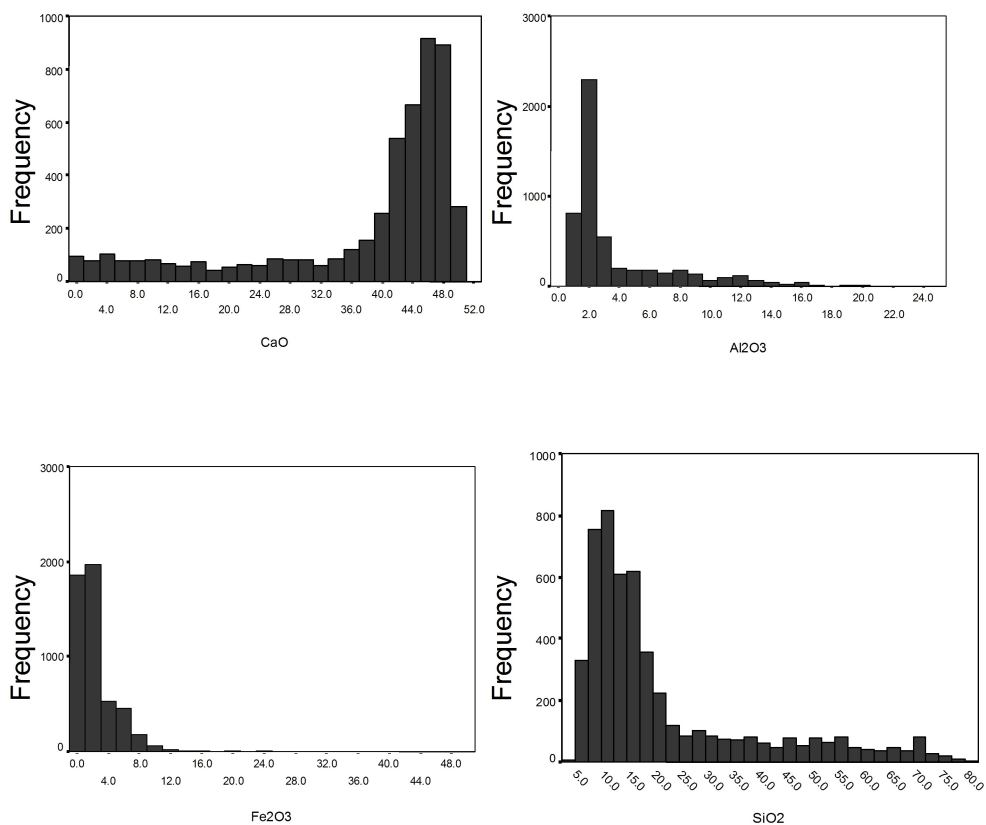
Figure 1 Histogram plots of CaO, Al$_2$O$_3$, Fe$_2$O$_3$ and SiO$_2$

The authors explored the possible reasons for the skewed patterns of the histogram plots. At first glance it was thought that these patterns might be due to the multiple lithologies and the distinct nature of ore grade variations for separate lithological units. To get further insight into this aspect, a separate box plot of the data (lithology-wise) was prepared for each of the grade attributes. Figure 2 presents the lithological box plots of the variables CaO, Al$_2$O$_3$, Fe$_2$O$_3$ and SiO$_2$. Box plots display the data variation in a concise form by indicating the minimum and maximum, 1st quartile (25% of the data less than or equal to this value), 2nd quartile (50% of the data less than or equal to this value) and third quartile (75% of the data less than or equal to this value) of a data set.
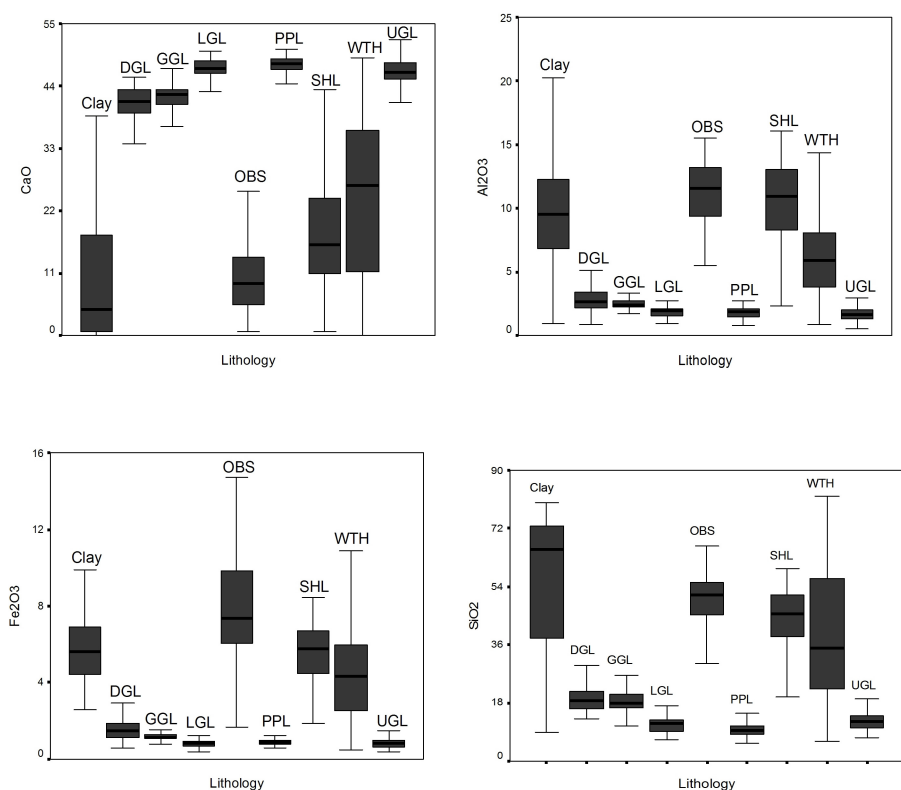
Figure 2 Box plots of CaO, Al$_2$O$_3$, Fe$_2$O$_3$ and SiO$_2$

The box plots reveal a distinct pattern of lithological grade variations. For example, for the variable CaO, grade variations for the lithological units, Clay, OBS, SHL and WTH, are relatively high compared with the other five lilthological units. In fact, these four lithological units show lower CaO values than the other five lithological units. The frequency distribution of the composite ore values from different lithologies thus resulted in skewed distributions of the grade attributes. This finding also holds for other grade attributes. It provides some initial clues about why it is important to incorporate detailed lithological information into grade modeling.

The spatial variability of the deposit was also explored using variogram modeling. Figure 3 shows the omni-directional variogram models for all the four attributes of the deposit. The variogram models indicate reasonably good spatial correlation with low nugget values. The anisotropy of spatial correlation was also examined through directional variogram modeling. However, no discernable anisotropy was found in the deposit.
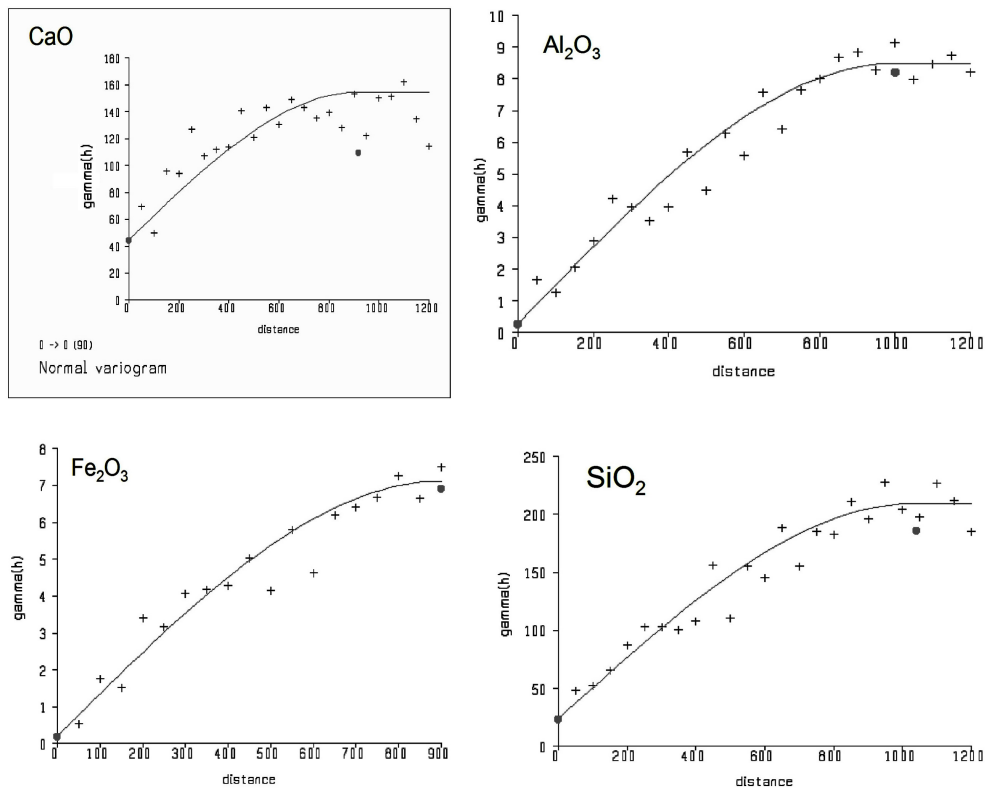
Figure 3 Omnidirectional variograms for CaO, Al$_2$O$_3$, Fe$_2$O$_3$ and SiO$_2$ respectively

The conventional way of assessing the suitability of a particular ore grade modeling technique is to collect ground-truth data from various locations in the area to be mined: the estimated grade of a model is then compared with the true grade. However, although this is accepted practice, it is rare to have anything other than exploratory borehole data at the initial stage of mine planning. This was the situation for the case study mine. Even though the mine is currently at the production stage, no routine grade-related information about mined material was made available to test the model's performance. In such cases, researchers usually opt for cross-validation to measure model performance. Although the cross-validation exercise is a good tool for measuring the performance of many kinds of models, it is a biased measure when used to test NN modeling. This is because the NN is a very flexible model and this flexibility makes the model liable to ac-commodate the 'noise' associated with real data. Therefore, although the NN model performs well with training data, when it is applied to real life problems its performance is often substand-ard. This situation is called "overfitting.

To derive an effective measure of model performance, the data available for this study were divided into two sets: training and testing. The model was trained using training data, and its performance was validated using testing data. It is worth pointing out that these two data sets should have similar statistical characteristics: there is no point training the model with data that has no relevance to the testing data. This would also lead to biased model performance.

This study followed the conventional practice of randomly dividing the data into training and testing sets. Instead of simple random data division, the data were randomly divided in terms of their lithology. For a lithological unit, 2/3 rd of the data were randomly put into the training set, with the remaining 1/3 of the data put into the testing set. This process was repeated for all the lithological units. Finally, all the data were mixed up, and a single training set and a single testing set were prepared. This resulted in 3455 units of data in the training set, and 1694 units of data in the testing set. This approach ensured that each data set contained a representation of all the lithological units: the training and testing data sets each therefore had similar statistical properties, as illustrated in Table 2. Both mean and standard deviation values are close for all the attributes. Further, the statistical similarity of the two data sets was checked using F test. The F test result confirmed that both the data sets are statistically similar at 0.05 level of significance.

Statistics	Data set	Attribute						
		Y	X	Z	CaO	Al_2O_3	Fe_2O_3	SiO_2
Mean	Training	2142.588	892.345	416.416	40.156	3.094	1.864	20.023
	Testing	2027.436	880.956	414.992	40.393	2.972	1.960	19.656
Standard Deviation	Training	1324.374	452.332	12.439	10.868	2.722	2.118	13.636
	Testing	1320.759	476.448	13.669	10.544	2.499	3.550	12.871

Table 2 Statistics of training and testing data

4.0 NEURAL NETWORK APPROACH FOR ESTIMATION

Recent advances in NNs, especially new insights into learning algorithms, have facilitated the development of a decidedly different approach to the estimation of ore grade and subsequently the estimation of ore reserve. The attractiveness of NNs lies in the fact that they are very flexible models and can capture any type of non-linear relationship between input and output through proper training. In spatial mapping, the network is presented with the field assay data at borehole locations for training, and given sufficient data and appropriate training, the network can be taught to recognize the relationship between input (coordinates and lithotypes) and output patterns (e.g. ore grade). In this process, the network acquires the generalised properties of a spatial relationship between the input and output, making it possible to interpolate ore grades for areas of unknown locations.

Among the various NNs, the feedforward NN is widely used for its learning ability. A feedforward NN is a biologically inspired regression/classification algorithm, which can be used for spatial input and output mapping. It consists of a large number of simple neuron-like processing units, organized in layers. Every unit in a layer is connected with all the units in the previous layer. Usually, a network consists of one input layer, one output layer, and one or two hidden layers. Each connection is associated with a connection weight. During the learning phase, the network is presented with a set of known input and output values called patterns. Using an optimal learning algorithm (gradient descent back-propagation algorithm for this study) the weights are modified iteratively, and after a number of iterations they get adjusted in such a way that when the input values are presented, the network produces outputs which are close to their actual

output values. Detailed descriptions of the NN models can be found in a number of text books such as Haykins (1999) and Hagan et al. (1995). A brief description of the NN working principle is presented in Appendix A.

To build up the NN model for this study, spatial positions (x, y, z coordinates) and various lithology types were used as inputs, and the grade attributes (CaO, Al_2O_3, SiO_2, Fe_2O_3) were used as outputs. Remember that the various lithological inputs were easily incorporated into the network model. Although lithogical units can be incorporated into geostatistical methods of orebody modeling as secondary information (Goovaerts, 1997), this makes the modeling exercise cumbersome. One of the biggest advantages of NN in orebody modeling lies in its capability to incorporate geology in a way that improves the precision of ore grade estimation. Figure 4 presents the network architecture used for this study:
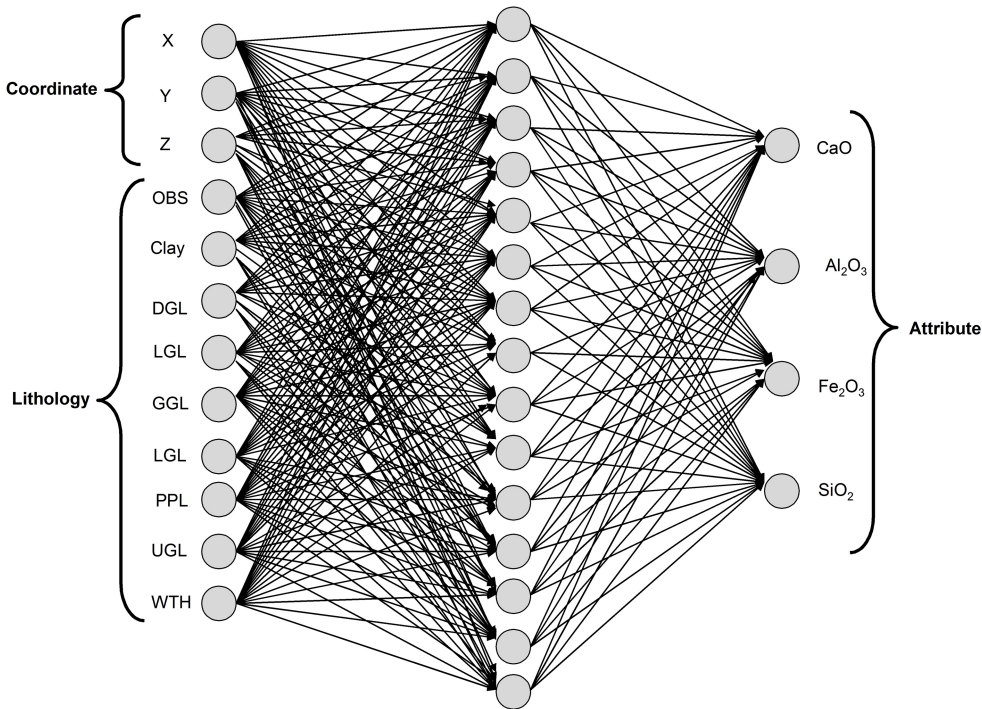

Figure 4 Architecture of multilayered artificial neural network for estimation of limestone

The network consisted of an input layer containing twelve input nodes (three for the spatial coordinates and nine for the lithology), an output layer consisting of four output nodes corresponding to four grade attributes, and a hidden layer composed of 15 nodes. Logistic activation was used both in hidden and output nodes. This particular network produced the best result for the data set. It can be noted that while the number of input and output nodes for a given problem are fixed, the user has flexibility to change the number of hidden nodes according to NN performance. After a trial and error method, 15 hidden nodes were chosen which resulted in the minimum error rates in the testing set.

The network was trained using the training data set consisting of 3455 sample values. Training was stopped when the network reached the mean squared error of 0.001, after 777375 iterations. It was noticed that when the network was trained beyond this point, the training errors decreased but the testing errors increased gradually. This is due to overfitting of the model in the presence of data 'noise'. When the training was complete, network performance was evaluated on the testing data set consisting of 1694 sample values. The Neural Network toolbox of the MATLAB software was used for all the NN studies.

5.0 COMPARISON OF NN AND KRIGING

The validity of the NN model was established through a comparative evaluation between NN and ordinary kriging. To derive a legitimate model comparison, the same data set which was used to train the NN model was used to developing the kriging model. Finally, the effectiveness of the two models was examined using the testing data set. For developing the kriging model, a 3-dimension kriging module of the SURPAC software was used. Table 3 shows the statistical summary of the comparative results of the NN and ordinary kriging techniques for all the grade attributes.

Statistics	CaO		Al_2O_3		Fe_2O_3		SiO_2	
	OK*	ANN**	OK	ANN	OK	ANN	OK	ANN
Mean error	-.3650	.2321	-0.038	-0.0829	-.1056	-0.0513	-.6045	-0.1359
Mean absolute error	3.3841	2.0192	.8051	.611	.7858	1.45	4.8824	2.2393
R^2	.874	.908	.750	.87	.740	.912	.851	.889
Mean Squared Error	37.8827	10.3722	2.1869	0.8536	7.0572	0.6714	76.5778	17.77535

* Ordinary kriging ** Artificial neural network

Table 3 Error statistics of Ordinary Kriging and neural network based estimation of testing data

The mean squared error as well as R^2 (co-efficient of determination) values indicate that the NN model provides superior performance in comparison with the ordinary kriging for all the attributes. Further, the R^2 values of the NN model for all the attributes are reasonably high, which indicates that a reliable grade model can be constructed for the deposit. Figures 5 through 8 present the scatter plot of the estimated versus observed values for the testing data set for the variables CaO, Fe_2O_3, Al_2O_3 and SiO_2 respectively.
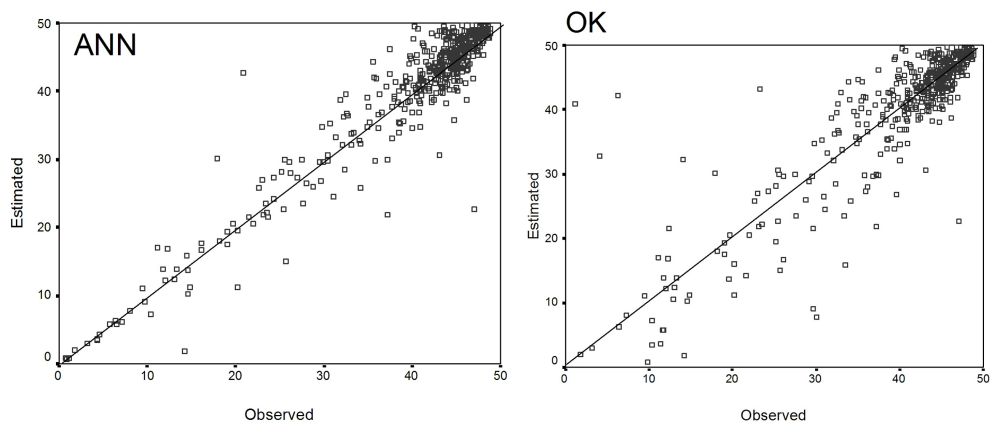
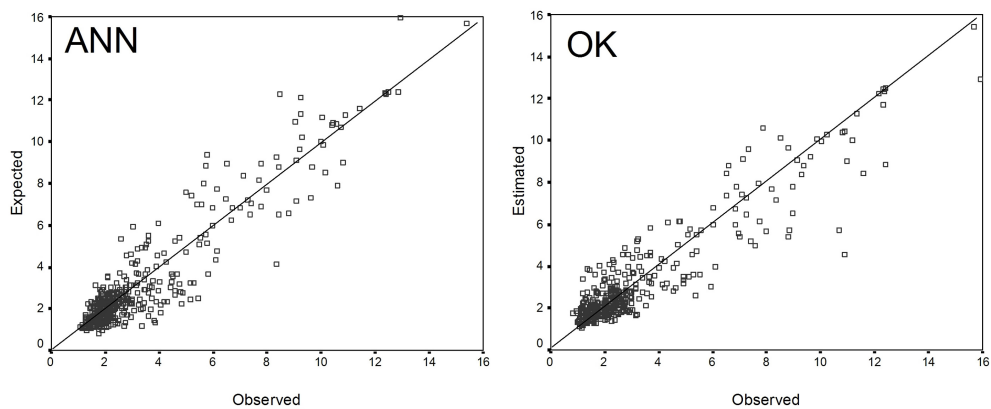**Figure 5** Scatter plot of testing data of true versus predicted values of CaO

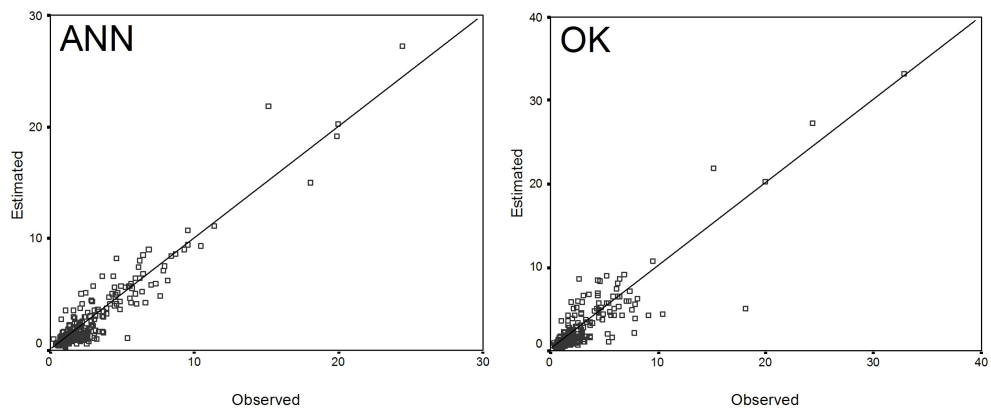**Figure 6** Scatter plot of testing data of true versus predicted values of Al_2O_3

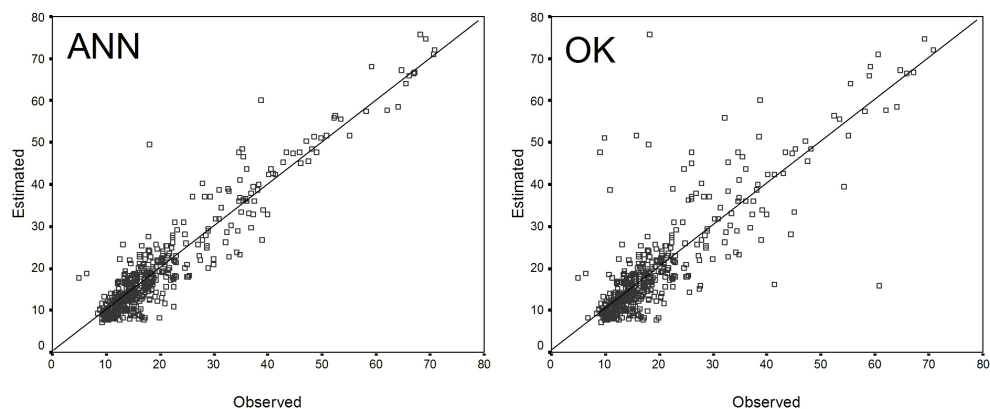**Figure 7** Scatter plot of testing data of true versus predicted values of Fe_2O_3

Figure 8 Scatter plot of testing data of true versus predicted values of SiO$_2$

6.0 PREPARATION OF ORE GRADE MAPS

The main objective of applying the neural network technique in this study was to generate grade maps based on which, the grade control plan can be constructed for subsequent mining and grade control operations. To construct a grade map, an estimation grid was defined within the leasehold area of the deposit. The grid spacing was chosen as 50m x 50m along northing (Y) and easting (X) directions. There were six benches in the deposit: six such grids were defined and grade estimation was exercised for each of the benches separately. Since the comparative evaluation of NN and kriging techniques indicated the superior performance of the neural network model, at this stage the decision was made to use the NN model for the grade estimation purpose. It should be noted that the NN model takes lithologies as input variables. Therefore, it is necessary to provide litho type information at each of the grid points. However, the lithological information was only available at the exploratory borehole locations. Hence, to use the NN model, a lithological map should be prepared from the exploratory borehole data at the estimation grid using a spatial interpolation technique. For this purpose indicator kriging was used.

6.1 CONSTRUCTION OF LITHOLOGICAL MAPS

Indicator kriging (see, for instance, Goovaerts, 1997; Deutsch and Journel, 1998) was used to construct the lithological map. For convenience, a brief description of the indicator kriging approach for categorical variables is presented here. Iindicator kriging, a non-linear geostatistical technique, is essentially a non-parametric counterpart of ordinary kriging. Indicator kriging works in the same fashion as the ordinary kriging; however, instead of using continuous data, it works on indicator binary data. When using an indicator kriging approach for categorical data such as lithological data, a series of indicator values corresponding to each lithological type is chosen. These indicator values are used to numerically build up the probability of occurrence for different lithologies at each of the estimation points. For each lithology, data in each of the observed locations are transformed into 0s and 1s: 1s if the data belong to the particular lithological type, and 0s if the data do not belong to that lithological type. Such indicator transformations are carried out for each of the lithological types. The estimates used in probability maps for the

lithological types are arrived at using the neighbourhood of transformed indicator data and a variogram model which was prepared from the indicator values.

The first step was to develop a categorical indicator transformation for each of the nine lithological units. An indicator transformation for a particular lithology at a sample location x in a mineral deposit is designated by following equation:

$$i_k = 1 \quad \text{if location x belongs to k}^{th} \text{ lithotype}$$
$$ 0 \quad \text{otherwise}$$

Where k varies from 1 to 9.

For each lithology, a different set of indicator transformations was defined. To quantify the spatial variability of the lithology, separate variogram models should be developed for each of the individual lithological units. However, this takes up excessive computational time. Instead, in this study, a single average variogram model was constructed for all the lithological units, assuming the same spatial continuity for the latter. Although this assumption is not entirely valid, it saved much computational time and was adequate for this study. It should be noted that Soares (1992) used this approach in determining lithofacies in a petroleum reservoir.

The typical variogram was calculated using the following formulae.

$$\gamma(h) = \frac{1}{2MN} \sum_{n=1}^{N} \sum_{m=1}^{M} [i_n(x_m) - i_n(x_m + h)]^2$$

Where, M is the number of pairs for lag h.

N is the number of lithological units.

The omni-directional variograms for the horizontal and vertical directions are shown in Figures 9 and 10. Although they show more or less the same sills, their range values are different. The range value in the horizontal direction is 1000 m; in the vertical direction it is 50 m. Indicator maps for each of the lithological units are produced in order to determine the probability of occurrence of a particular lithology at specified points in the estimation grid. Nine different indicator maps indicate the probability of occurrence for the respective lithotype. Using nine different indicator maps, a single lithotype map was produced, by assigning a lithotype at a particular point which had maximum probability of occurrence among the nine lithologies at that point. For example, browsing through all the indicator maps, it was found that at an unknown point x, the probabilities of occurrence of the lithological units PPL, GGL, DGL, LGL, WTH, UGL, SHL, clay, and OBS were 0.05, 0.02, 0.1, 0.2, 0.25, 0.15, 0.03, 0.5, .27 respectively. Clay was the litholigical unit assigned to that point.

The validity of an indicator-based approach for lithological mapping was also tested on the testing data set. Table 4 presents the classification errors for each of the lithological units in terms of percentage of misclassification. On an average, the misclassification error is close to 5%. This result shows that WTH has the highest (6.704%) misclassification errors and UGL has the lowest (3.547%). The lithological maps for all the six benches are shown in Figure 11.
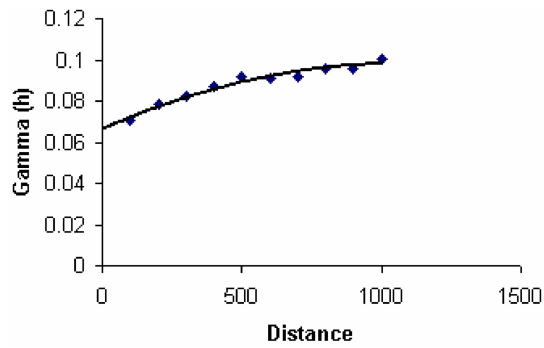
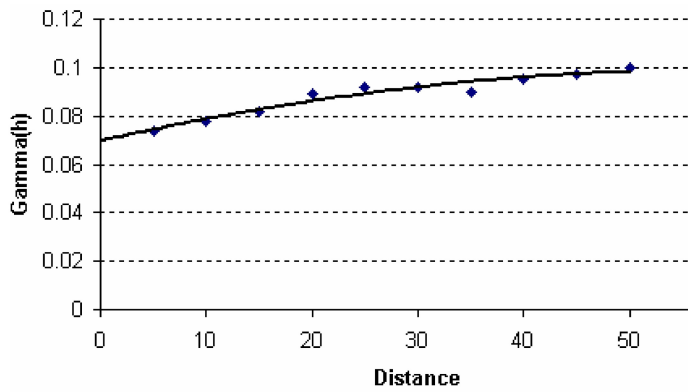**Figure 9** Omni-directional Variogram of lithotypes in the horizontal direction

Figure 10 Omni-directional Variogram of lithotypes in the vertical direction

Lithology	Percentage of misclassification
Clay	6.1224
DGL	5.785
GGL	4.22
LGL	6.493
OBS	6.382
PPL	5.479
SHL	5.263
WTH	6.704
UGL	3.547
Total	5.076

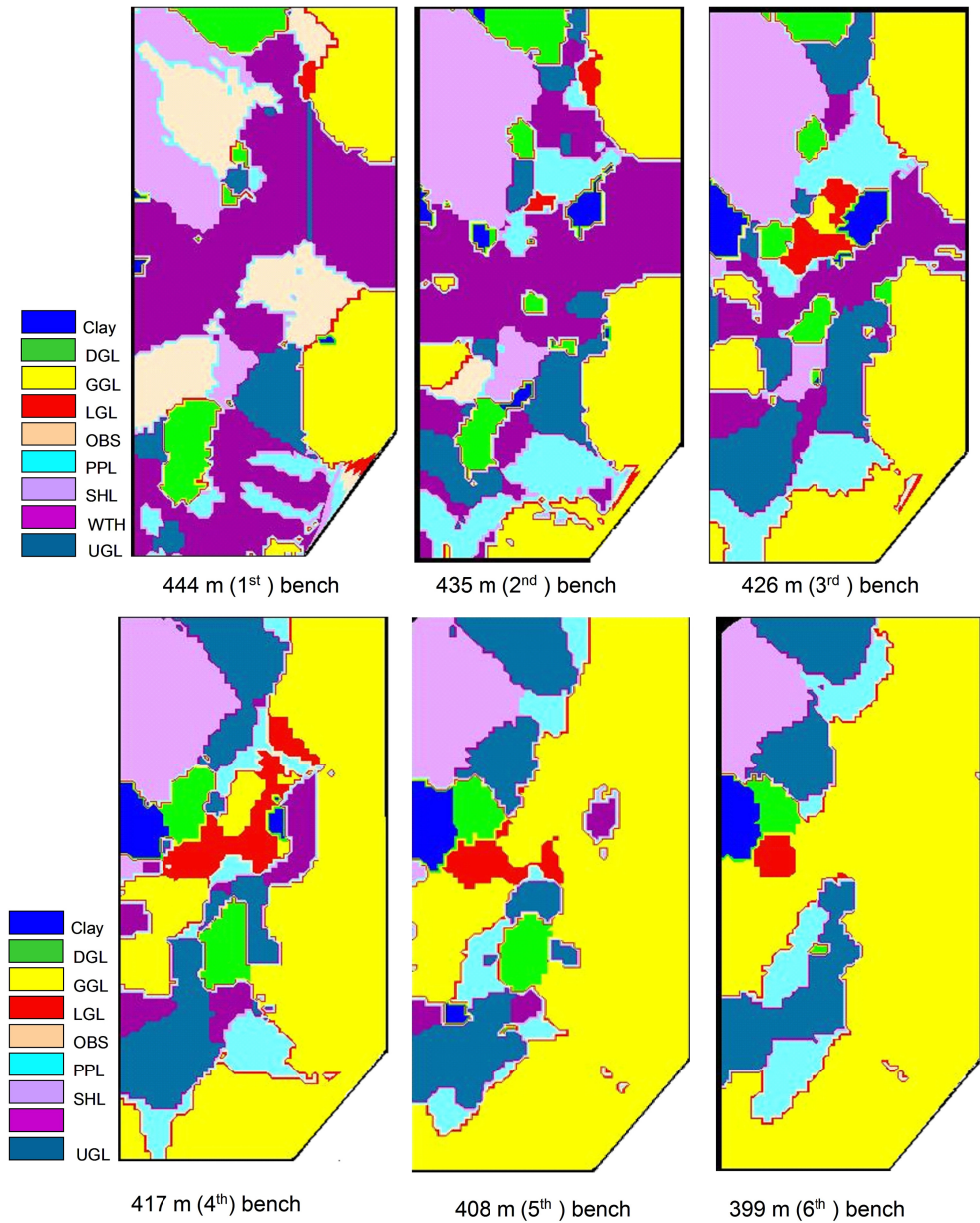**Table 4** Lithological classification results of testing data set

Figure 11 Lithological maps of all 6 benches (444 m, 435 m, 426 m, 417m, 408 m and 399 m)

6.2 ORE GRADE MAPS

Information from the lithological mapping was fed to the NN model in order to generate grade maps of the deposit. For example, Figure 12 shows the grade maps of all the grade attributes CaO, Fe_2O_3, Al_2O_3 and SiO_2 along with LSF and SR for the 417m bench. Note that the grade

maps for LSF and SR were prepared using the estimated values of the CaO, Fe_2O_3, Al_2O_3 and SiO_2, since they are the two critical parameters for optimising the cement manufacturing process. These grade maps can be used at the planning stage for grade control and other ancillary mining operations.
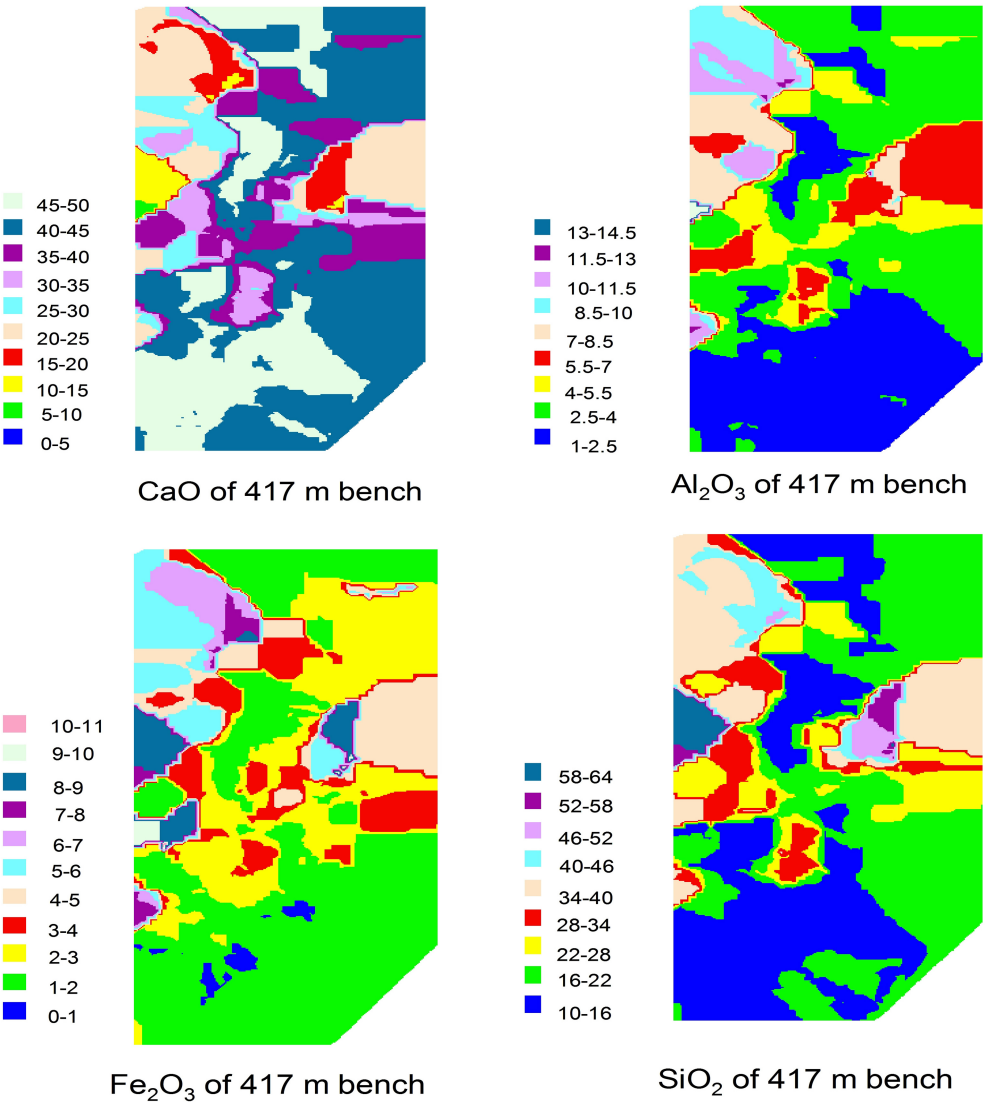

CaO of 417 m bench

Al_2O_3 of 417 m bench

Fe_2O_3 of 417 m bench

SiO_2 of 417 m bench

Figure 12 Ore grade maps, LSF map and SR map of 417 m bench using neural network

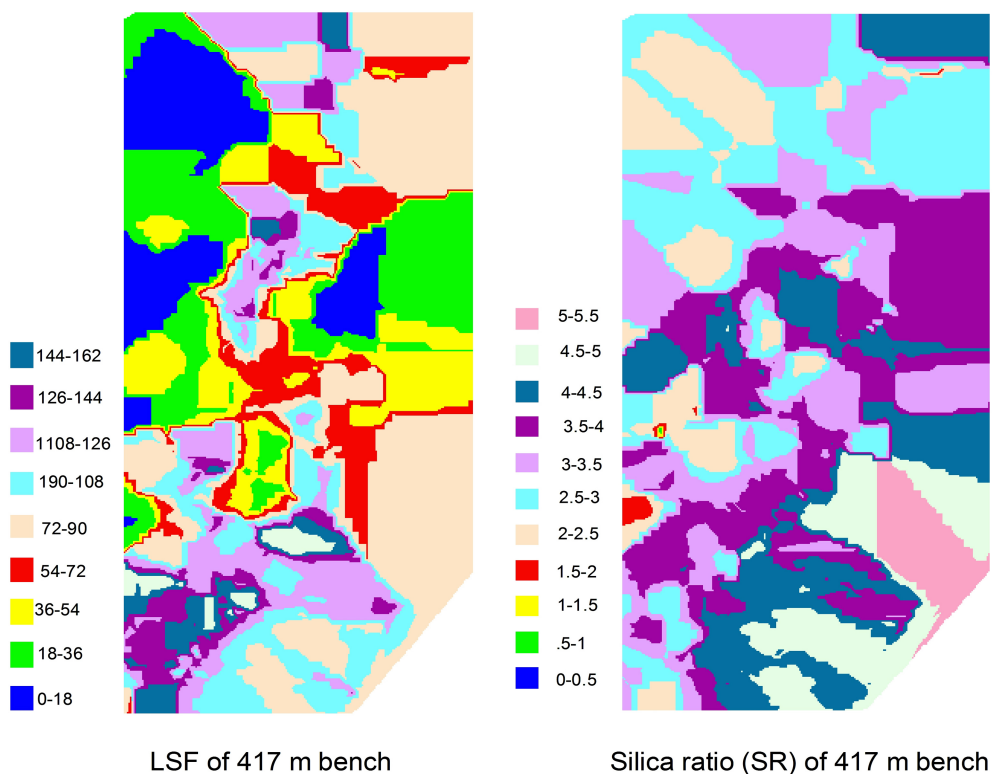LSF of 417 m bench   Silica ratio (SR) of 417 m bench

Figure 12 (cont'd) Ore grade maps, LSF map and SR map of 417 m bench using neural network

7.0 SUMMARY

The aim of this study is to develop a reliable ore grade model using a NN approach. For this purpose, a NN model was built which used spatial locations and lithotype information to estimate the grades of all the major ingredients of limestone such as CaO, Fe_2O_3, Al_2O_3 and SiO_2. The validity of the NN model for this deposit was tested through a comparative evaluation of the NN and kriging models. The mean error, mean absolute error, mean squared error (MSE) and R^2 were used as performance indicators of the two models for the testing data set. After the superiority of the NN model was established, it was further used for producing ore grade maps of the deposit. To prepare the ore grade maps, the NN model required the litho-type information at every point of the unknown estimation grid. For this purpose, indicator kriging was used to prepare the lithological maps of the deposit from the exploratory borehole data.

APPENDIX – A

A.1. WORKING PRINCIPLE OF THE NN MODEL

NN, is a generalised technique used for input-output mapping of many systems including ore deposit modeling. It has an artificial intelligence-like capability to capture the inter-relationship between input and output of a system. In orebody modeling, it is assumed that the grade attribute value in an ore deposit varies from location to location, and that this will be reflected in a complex input and output spatial relationship between spatial co-ordinates and grade values. Hence, output grade is considered to be a function of spatial co-ordinates such as X and Y co-ordinates. A NN is a nonlinear dynamic system consisting of a large number of highly interconnected processing units, which can be used to exploit this relationship. In a NN, there are basically two major physical components: processing units or neurons and connections between these processing units. Each processing unit, acting as an idealised neuron, receives input, computes activation, and transmits that activation to other processing units. A weight value, defined to represent connection strength, is associated with each connection between these processing units. The connection weight of each processing unit is optimally determined through the presentation of known examples, and application of a learning rule. Once the connection weight is determined through NN learning, the inter-connection between input and output embedded in the data is captured. Thus the NN acquires a generalised capability to predict outputs from a given set of inputs.
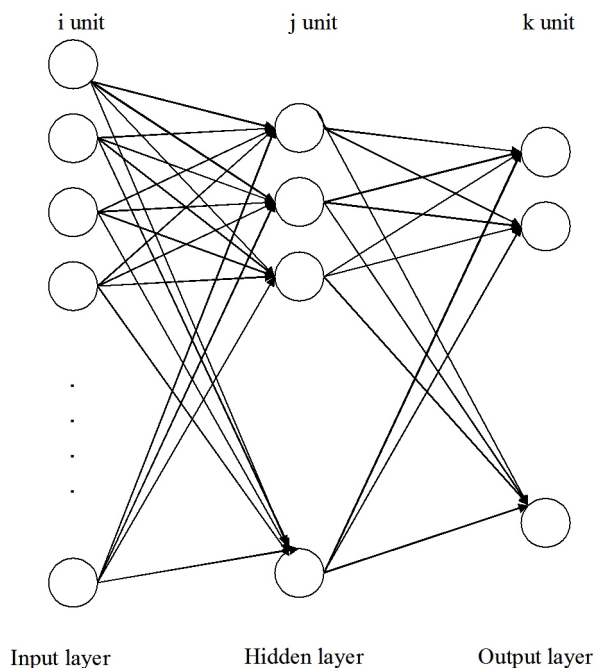

Figure A.1 An example of neural network model

A simple three-layer neural network is presented in Figure A.1. It consists of an input layer, a hidden layer and one output layer, interconnected by modifiable weights and represented by links between layers. Each input vector is presented with the input layer, and the output of each input unit equals the corresponding component in the vector. Each hidden unit's unit computes the weighted sum of its input to form its net activation.

$$net_j = \sum_{i=1}^{d} x_i w_{ij} + w_{j0} = \sum_{i=0}^{d} x_i w_{ij} \tag{A.1}$$

Where the subscript i indexes units in the input layer, j in the hidden layer, w_{ij} denotes the input to hidden layer weights at the hidden unit j, *net j* is the activation for hidden j.

Each hidden unit emits an output that is a non linear function of its activation, *f(net)*.

$$y_j = f(net_j) \tag{A.2}$$

Each output unit similarly computes its net activation based on the hidden unit signals as

$$net_k = \sum_{j=i}^{n_H} y_j w_{kj} + w_{k0} = \sum_{j=0}^{n_H} y_j w_{kj} \tag{A.3}$$

Where the subscript k indexes units in the output layer and n_H denotes the number of hidden units. It has mathematically treated the bias unit as equivalent to one of the hidden units whose output is always $y_0 = 1$.

An output unit computes the nonlinear function of its *net*, as

$$z_k = f(net_k) \tag{A.4}$$

Where,
z_k is the k output units.
So, the total network output for a three layers model can presented in this fashion

$$z_k = f(\sum_{j=1}^{n_H} w_{kj} f(\sum_{i=1}^{d} w_{ji} x_i + w_{j0}) + w_{k0}) \tag{A.5}$$

A.1.1 NETWORK LEARNING

The basic approach in 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 at the output layer. These outputs are compared to the target values; any observed difference corresponds to an error. The error function is some scalar function of the weights and is minimised when the network outputs match the desired output. Thus the weights are adjusted to reduce this measure of error. The error (training error) on a pattern is the sum over output units of the squared difference between the desired and actual output and is given by

$$E(w) \equiv \frac{1}{2} \sum_{k=1}^{c} (t_k - z_k)^2 = \frac{1}{2} |t - z|^2 \tag{A. 6}$$

where **t** and **z** are the target and the network output vectors of length c and **w** represent all the weights in the network. A standard gradient descent back-propagation learning algorithm is

generally used to select the final weights of the network. The weights are initialised with random values, and are then changed in a direction that will reduce the error:

$$\Delta w = -\eta \frac{\partial E}{\partial w} \qquad (A.7)$$

where η is the learning rate.

The iterative algorithm requires taking a weight vector at iteration m and updating it as

$$w(m+1) = w(m) + \Delta w(m) \qquad (A.8)$$

From the chain rule, it shows that the weight change is:

$$\frac{\partial E}{\partial w_{kj}} = \frac{\partial E}{\partial net_k} \frac{\partial net_k}{\partial w_{kj}} = -\delta_k \frac{\partial net_k}{\partial w_{kj}} \qquad (A.9)$$

Where the sensitivity of unit k is defined to be

$$\delta_k = -\frac{\partial E}{\partial net_k} \qquad (A.10)$$

it describes how the overall error changes with the unit's net activation. Assuming that the activation function f(.) is differentiable, the differentiate of equation A.6 for such an output unit is, δ_k is

$$\delta_k = -\frac{\partial E}{\partial net_k} = -\frac{\partial E}{\partial z_k} \frac{\partial z_k}{\partial net_k} = (t_k - z_k) f'(net_k) \qquad (A.11)$$

The last derivatives of equation A. 9 is found using equation A. 3

$$\frac{\partial net_k}{\partial w_{kj}} = y_j \qquad (A.12)$$

Taken together, these results give the weight update or learning rule of the hidden layer neurons to output weights

$$\Delta w_{kj} = \eta \delta_k y_j = \eta (t_k - z_k) f'(net_k) y_j \qquad (A.13)$$

The learning rule for the input to hidden units, using the chain rule is represented by

$$\frac{\partial E}{\partial w_{ji}} = \frac{\partial E}{\partial y_j} \frac{\partial y_j}{\partial net_j} \frac{\partial net_j}{\partial w_{ji}} \qquad (A.14)$$

All the three terms in equation A.14 can be solved in the same way as in the case of the hidde-to-output layer, and results are represented as:

$$\Delta w_{ij} = \eta x_i \delta_j = \eta \left[\sum_{k=1}^{c} w_{kj} \delta_k \right] f'(net_j) x_i \qquad (A.15)$$

Depending upon the activation function used in the input-to-hidden layer and hidden-to-output layer, the function $f(net_j)$ and $f(net_k)$ will be changed. In this case study, the sigmoidal

function has been used for both the layers, although other activation functions like tan-hyperbolic, Gaussian can also be used.

Equations A.13 and A.15 help to adjust the weights for both the layers. The adjusted weights are then used to calculate the error term for the next iteration. The process is repeated in the same fashion as outlined above and stops when the error reaches a threshold value defined by the user. When the training is complete, the model can be used for the purposes of prediction.

ACKNOWLEDGEMENTS

REFERENCES

Abdullah, Arik. 1990. 'Effects of search parameters on kriged reserve estimation'. *International journal of mining and geological engineering* 8: 319–331.

Deutsch, C. V.; Journel, A. G. 1998. *GSLIB: Geostatistical software library and user's guide*. 2nd edn: New York: Oxford University Press, 369p.

Goovaerts, P. 1997. *Geostatistics for natural resources evaluation*. New York: Oxford University Press, 483 pp.

Hagan, M.T.; Demuth, H.B.; Beale, M. 1995. *Neural network design*. Boston: PWS Publishing Company.

Haykins, S. 1999. *Neural networks: A comprehensive foundation*. 2nd edn. New Jersey: Prentice Hall, 824 pp.

Ingram, Kevin; Daugherty, Kenneth 1991. 'A review of limestone additions to Portland cement and concrete'. *Cement and concrete composites* 13 (3): 165–170.

Koike, K; Matsuda, S. 2003. 'Characterizing content distributions of impurities in a limestone mine using a feedforward neural network'. *Natural resources research* 12 (3): 209–222.

Nayak, B; Mallik, P. K. 2002. 'Characterisation of Portland cement clinker manufactured by down draft sintering and vertical shaft kiln processes'. *Advances in cement research* 14 (1): 1–7.

Samanta, B; Bandopadhyay, S; Ganguli, R. 2004 (a). '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; Bandopadhyay, S; Ganguli, R; Dutta, S. 2004 (b). 'Sparse Data division using data segmentation and Kohonen Network for neural network and geostatistical ore grade modeling in nome offshore placer deposit'. *Natural resources research* 13 (3): 189–200.

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 (accepted).

Soares, A. 1992. 'Geostatistical estimation of multi-phase structures'. *Mathematical geology* 24 (2): 149–160.

Yama, B; Lineberr, G. T. 1999. 'Artificial neural network application for a predictive task in mining'. *Mining engineering* 51 (2): 59–64.