Quantitative phase analysis of clay minerals by X-ray powder diffraction using artificial neural networks. I. Feasibility study with calculated powder patterns

D. T. GRIFFEN

Department of Geology, Brigham Young University, Provo, Utah 84602, USA

(Received 11 April 1997; revised 24 December 1997)

ABSTRACT: An artificial neural network with back-propagation architecture has been applied to the problem of the quantitative analysis of simulated clay mixtures from synthetic X-ray diffraction (XRD) data. A 'clay characterization function' (CCF) has been devised that combines information from two clay XRD peaks into a single peak that simplifies the problem which the neural network must solve. In addition, it eliminates peaks from non-layered minerals. A neural network with 17 neurons in the hidden layer and log-sigmoid transfer functions in each layer is sufficiently successful at predicting the compositions of binary and ternary mixtures of the three model clays to demonstrate the potential of the method. Moreover, training is accomplished in relatively short times using a Levenberg-Marquardt algorithm.

Although the problem ‘solved’ by this neural network is rather simple, the approach has the potential for the much more complex problem of the quantitative phase analysis of mixtures of real clay minerals. In this study, only the CCF peak heights have been used, but the positions and peak widths also contain important information. It may be practical to include as variables not only the concentrations of clay phases, but also compositional information to which XRD is sensitive.
FIG. 1. Generalized architecture of a neural network. In a back-propagation network, as used in this study, information flows only from left to right between neurons during training, until outputs are compared with target values, and corrections are ‘back-propagated’ through the network.

FIG. 2. The structure of a neuron, or processing element, of an artificial neural network. $x_1...x_n$ represent inputs, and $y_1...y_m$ represent outputs. The transfer function is described in the text.
function on the interval \( \{0,1\} \). This is the case here, of course, inasmuch as outputs are to be weight fractions in a mixture.

Because so-called back-propagation neural networks with a sigmoidal hidden layer are capable of closely approximating any non-linear function that has a finite number of discontinuities (Demuth & Beale, 1992), that type of architecture was chosen for this study. In a back-propagation network, each neuron in layer \( i \) (including the input layer) is connected to every neuron in layer \( i+1 \) (including the output layer), as indicated schematically in Fig. 1.

Artificial neural networks are particularly adept at solving problems involving pattern recognition or function approximation, and can be quite successful in spite of noisy data. For these reasons they have been extensively employed in monitoring industrial processes for years. The relatively few applications in geology have involved such diverse fields as remote sensing (e.g. Bischoff et al., 1992), well-log analysis (Huang et al., 1996), and the classification of rocks (Carr & Hibbard, 1991). Griffen et al. (1995) have given a progress report on the use of artificial neural networks for the modal analysis of igneous rocks using X-ray powder diffraction data.

The identification and quantitative analysis of mixtures and interstratifications of clay minerals using X-ray powder diffraction has recently involved the modelling of theoretical diffraction patterns for matching in some way against the observed diffraction patterns. For example, Bish (1993) has used the Rietveld method for well crystallized clay minerals; Pevear & Schuette (1993) have applied a genetic algorithm to the trial-and-error method of matching a mixture of four specific clay minerals to a diffraction pattern; Jones (1989) has developed a computerized curve-fitting/peak-decomposition method. Among the most widely used software packages for simulating XRD patterns of mixed-layer clays (as well as of pure clays, and for approximating the patterns of non-interstratified mixtures) are NEWMOD© (Reynolds, 1985) and NEWMOD2© (Reynolds & Reynolds, 1987). Walker (1993) has described the software in detail.

The purpose of this paper is to demonstrate the potential for using neural networks for the quantitative phase analysis of clay-mineral mixtures, given high quality X-ray powder diffraction data. Both preprocessing and neural network calculations were done with MATLAB and the MATLAB Neural Network Toolbox, although several commercial packages for neural computing are available, and no endorsement of any particular software is intended.

**METHODS**

With a back-propagation neural network (also known as a feed-forward neural network), learning is supervised, in that both the training input and the desired results (target data) are provided to the network through many cycles of training. Information flows in the direction from the input buffer to the output buffer; that is, there is no bidirectional flow or feedback, until the final output based on one cycle through the training input is compared with the desired output. Weights for each synapse are set randomly at first, so the initial outputs are far from the target data. The network is repeatedly presented with the sets of training input and target data, and in each training cycle (or epoch) the errors (differences between outputs and target data) are used by an algorithm external to the network to adjust the weights associated with each neuron, in order to produce output data more closely matching the target data. (The name ‘back-propagation’ comes from the fact that information used to adjust the weights is propagated back down the network after each complete training cycle.) When the sum of the squared errors reaches some predetermined value deemed to be acceptable, the network is said to have converged on the solution, and the weights are frozen. Thereafter the trained network should provide reasonable results when presented with similar data not used in the training set.

One of the most common ways of adjusting the synaptic weights is the method of gradient descent, wherein the weights (and biases) are shifted in a direction opposite to the gradient in the error surface. The error is represented by

\[
E = 0.5 \sum (z_j - y_j)^2
\]  

(1)

where \( z_j \) and \( y_j \) are the \( j \)th target and output values in the target and output vectors \( z \) and \( y \) corresponding to an input vector \( x \). The error surface comprises the error values for all possible input vectors and all possible weights. The changes in the weights are calculated by the gradient descent algorithm, involving the input, output, and target values. This results in convergence, but is often very slow. A more sophisticated method of
error minimization is the Levenberg-Marquardt approximation:

$$\Delta W = (J^T J + \mu I)^{-1} J^T e$$  \hspace{1cm} (2)

where $\Delta W$ is the change in the weight vector, $J$ is the Jacobian matrix of derivatives of each error with respect to each weight, $\mu$ is a scaler, and $e$ is the error vector (Demuth & Beale, 1992). If $\mu$ is large, the method approaches gradient descent; if $\mu$ is small, it becomes the well known Gauss-Newton method of finding a minimum in the error-gradient surface (Prince, 1994). During training, $\mu$ is shifted so as to approach the Gauss-Newton method as quickly as possible, resulting in rapid convergence. Further discussion of network architecture, error analysis, and related topics, can be found in De Wilde (1997), Chester (1993), Carling (1992), and many other books on the subject.

The three model clays used in this study were kaolinite, illite with 0.6 K atoms per formula unit, and Fe-free smectite with one water layer, a cation exchange capacity of 0.36 Eq/100 g, and Mg as the exchange cation. Powder patterns for the three pure phases were calculated with NEWMOD2© (Reynolds & Reynolds, 1987), assuming the geometry of a Scintag XDS-2000 theta-theta X-ray diffractometer and ‘infinitely thick’ clay samples of 2.5 cm length. These choices were made because the next phase of this study will involve real XRD patterns, which will be collected with such an instrument. The diffraction patterns are shown in Fig. 3. The calculation of powder patterns of mixtures of the three phases was done by straightforward linear combination of the patterns of the pure phases. Inasmuch as the purpose of this paper is to investigate the potential of neural computing in the quantitative phase analysis of clay mineral mixtures, rather than to consider how accurately NEWMOD2© can be made to model real, non-interstratified, multi-clay mixtures, differential absorption (matrix effects) caused by different mass absorption coefficients was ignored — a simplification that will be re-evaluated in succeeding studies involving real clay data.

Fig. 3. Calculated X-ray powder diffraction patterns for the three model clay minerals used. Cu-K$\alpha$ radiation is assumed.
In general, the number of neurons needed in a hidden layer for a given multilayer network is not known, but is determined by trial and error. It is clear, however, that the larger the number of inputs and outputs, the greater the required number of neurons (De Wilde, 1997). In addition, networks with too many neurons have a high probability of being over-fit — that is, of predicting the values in the target data of the training set very well, but oscillating wildly between the values on which they were trained. In order to minimize the input data while maintaining a high information content, each clay in a mixture was represented by a single datum that combined information from the first two diffraction maxima; this is obtained from a function here referred to as the ‘clay characterization function’ (CCF). It is well known that successive peaks in the XRD pattern obtained from an oriented clay specimen are separated by approximately equal distances when intensity is plotted as a function of 2θ. Because 2θ is not linearly related to d-value, however, the peaks are not exactly equally spaced (see Fig. 4a). If the pattern is plotted as a function of sinθ instead of 2θ, the distances between successive peaks are precisely equal, and inversely proportional to d-values. The CCF is calculated as follows:

1. The raw X-ray data are corrected for background and the Kα₂ X-ray component is stripped from each peak. (This, of course, is irrelevant to the synthetic powder patterns used here, but very necessary for real samples, for which a background and Kα₂ contribution are present.) If the very-low-angle intensity from the incident X-ray beam is not adequately subtracted by the background correction, then this is done manually. Because diffraction patterns are not usually started at 2θ = 0°, but the data must begin there in order to take advantage of the equal spacing of basal reflections, the 2θ and intensity vectors are extended backwards from the beginning of the collected data to 0°2θ by prefixing the appropriate diffraction angles to the 2θ vector, and prefixing an intensity value of zero to the intensity vector for each 2θ value so added.

2. The 2θ values are replaced by corresponding sinθ values. Because 2θ and sinθ are not linearly related, the intervals of sinθ obtained at this stage are not equal, and this must be remedied. To do so, a convenient fixed interval of sinθ is chosen (for instance, the mean difference between successive sinθ values for the entire diffraction pattern), and the intensity values corresponding to these new sinθ values are determined by interpolation between the measured intensity values.

3. The CCF is computed by multiplying the intensity at sinθᵢ by the intensity at 2sinθᵢ for all i, and plotting the resulting values as a function of d-spacing. Only intensities at positions where sinθ<0.5(sinθmax) are used; for instance, if the powder pattern is run to 40°2θ, then the calculation for CCF is carried out only to sinθ values corresponding to 20°2θ. Because all clays have maximum basal spacings at relatively low Bragg angles, all will have at least two X-ray peaks at 2θ<40°.

While this calculation may seem involved, it can be carried out easily by computer as an essentially automatic process.

The CCFs for the three model clays used in this study are shown in Fig. 5. Note that the peaks corresponding to each clay are easily identified from their basal d-values. These powder patterns were calculated to 42°2θ, and both smectite and illite have four peaks within that range, resulting in two peaks on the CCF curves for each of them; the kaolinite CCF curve has only one peak in that range. To minimize the number of input data, only one peak was used for each mineral — the one plotted at the maximum d-value, because it yielded the highest CCF peak.

RESULTS

Initially, only the CCFs for mixtures lying on the edges and corners of the smectite-kaolinite-illite compositional triangle were used in the training set (Fig. 6). The CCFs were calculated for binary compositions at intervals of 5 wt%, yielding 60 values in the training set. The network was designed with one hidden layer, the Levenberg-Marquardt training algorithm was used, and several networks containing from 5–40 neurons in the hidden layer were tested. Networks with <10 neurons in the hidden layer did not converge. Those with between 11 and 15 neurons in the hidden layer were found to be underfitted; that is, they converged, but reproduced the compositions used in the training set poorly. Those with >20 neurons in the hidden layer did not converge. Those with >20 were overfitted; that is, they reproduced the values in the training set very accurately, but were poor at predicting compositions between those in the training set. A network with 17 neurons in the hidden layer and log-sigmoid transfer functions gave good convergence (i.e. a relatively rapid
FIG. 4. The calculated powder pattern for kaolinite, plotted against $2\theta$ (wherein peak separations are nearly equal) and against $\sin\theta$ (wherein peak separations are equal).
Phase analysis of clay minerals

FIG. 5. The clay characterization functions (CCFs) for the three model clays used.

decrease in the sum-squared errors between predicted and true compositions), and good predictions for binary compositions in the test set (which, by definition, do not include those used in the training set) except for mixtures containing <5 wt% of one phase.

With the edges of the compositional triangle modelled well, CCFs for 10 compositions within the triangle were added to the training set (Fig. 6). The same architecture was used, and the convergence history is shown in Fig. 7; convergence to a sum-squared error <0.001 was achieved in 648 training epochs (i.e. presentations of the training set to the network), which required ~12 min on a PowerMacintosh 8500/120. The fit to the training data, even though only a small percentage of the compositions were ternary, is excellent (Fig. 8). In order to test the network, it was used to predict the compositions of a test set of 27 mixtures (as required, none of them part of the training set); of these, 12 were binary mixtures and 15 were ternary. Figure 9 shows the agreement between predicted and true compositions. With the exception of one sample (symbols enclosed in dashed rectangles), the agreement is quite good; no explanation for the single exceptional sample is apparent. Because
Convergence of neural network using Levenberg-Marquardt method

Fig. 7. The convergence history of the final neural network. Convergence was achieved in 648 training epochs.

Fig. 8. The fit of the predicted wt fractions as a function of the true concentrations of the three model clay minerals for the training data.
using a previously trained neural network for prediction does not involve iteration, the computational time involved is trivial.

**DISCUSSION AND CONCLUSIONS**

This study has demonstrated the feasibility of using an artificial neural network for the quantitative phase analysis of mixtures of three clays. Extension to more than three clays is straightforward in principle, although the addition of many (or potentially many) clays to the problem may require a neural network with different architecture (say, two hidden layers, or something other than a back-propagation network). Work to investigate that is underway. It is clear that a denser distribution of ternary mixtures in the training set would have improved the agreement shown in Fig. 9, but this was unnecessary for the purpose of demonstrating feasibility. The important question now is not whether performance for calculated clay mixtures can be improved, but whether the method developed with data from calculated powder patterns can be applied with equal efficacy to mixtures of real clays. A second question of significance for the practical application of the method is whether the entire procedure can be automated to make it convenient for clay analysts who do not wish to become enmeshed in the details of CCF calculations, network architecture, etc.

Real clays present some problems not encountered in the present work. As in any method involving XRD of clay minerals, sample preparation is of utmost importance, and one of the initial challenges will be to ensure that NEWMOD2 properly models peak intensities and peak shapes for a variety of real clays prepared using standardized methods. In addition, there may be other variables available from the CCF that would be useful, besides peak height. For calculated powder patterns, using the CCF provides no advantage over using the heights of XRD maxima, because the calculations take no account of the sample preparation problems or random errors which plague clay analysts. For real clays, however, the use of the CCF should tend to ‘smooth out’ random errors in X-ray peak intensities and provide information from two
peaks rather than just one. In addition, because non-
layer minerals do not present peaks at uniform
spacings, peaks due to non-clay minerals in the clay
fraction will be eliminated by use of the CCF.

In addition, the use of other information available
from the CCF (e.g. using two CCF peaks per clay
mineral and peak widths and positions, as well as
heights) might make it possible to add additional
functionality to the neural network. As an example
of this type of enhancement, it might prove possible
to add the Fe content of smectites or the K content
of illites as a variable to be estimated by the neural
network. The quantitative characterization of
mixed-layer clays may also be amenable to analysis
by neural computing, although the CCF will
probably not be useful in that case, and some
architecture other than back-propagation may yield
superior performance for that very complex
problem.

ACKNOWLEDGMENTS

Very helpful reviews of the manuscript were provided
by Jeff Walker and Steve Hillier. In addition, the
author benefited from interaction with clay scientists
convened at the Macaulay Land Use Research Institute,
Aberdeen, Scotland, for the Golden Jubilee Meeting of
the Clay Minerals Group of the Mineralogical Society
of Great Britain and Ireland in April 1997, where an
early version of this work was reported.

REFERENCES

Multispectral Classification of Landsat Images
Sens. 30, 482–490.

Bish D.L. (1993) Studies of clays and clay minerals
using X-ray powder diffraction and the Rietveld
method. Pp. 79–121 in: Computer Applications to
X-ray Powder Diffraction Analysis of Clay Minerals,
(J.R. Walker & R.C. Reynolds, Jr., editors) CMS
Workshop Lectures, 5, The Clay Minerals Society,
Boulder, CO.

Carling A. (1992) Introducing Neural Networks. Sigma

Carr J.R. & Hibbard M.J. (1991) Open-ended mineral-
17, 1409–1463.


for use with MATLAB. The MathWorks, Natick, MA.


Toward an artificial neural network for the modal
analysis of rocks from X-ray diffraction data. Geol.

Improved Fuzzy Neural Network for Permeability
Estimation from Wireline Logs in a Petroleum
(TENCON) on Digital Signal Proc. Appl., 2,
912–917.

52–101 in: Quantitative Mineral Analysis of Clays,
(P.R. Pevear & F.A. Mumpton, editors), CMS
Workshop Lectures, 1, The Clay Minerals Society,
Boulder, CO.

Pevear D.R. & Schuette J.F. (1993) Inverting the
NEWMOD© X-ray diffraction forward model for
clay minerals using genetic algorithms. Pp. 19–41
in: Computer Applications to X-ray Powder
Diffraction Analysis of Clay Minerals, (J.R. Walker
& R.C. Reynolds, Jr., editors) CMS Workshop
Lectures, 5, The Clay Minerals Society, Boulder,
CO.

Prince E. (1994) Mathematical Techniques in
Crystallography and Materials Science. Springer-
Verlag, Berlin.

Reynolds R.C. (1985) NEWMOD©, a Computer
Program for the Calculation of Basal X-ray
Diffraction Intensities of Mixed-Layered Clays. 8
Brook Road, Hanover, NH 03755, USA.

Description of program NEWMOD2 for the calcula-
tion of the one-dimensional X-ray diffraction pat-
terns of mixed-layered clays. 8 Brook Road,
Hanover, NH 03755, USA.

modeling of X-ray diffraction patterns of clay
in: Computer Applications to X-ray Powder
Diffraction Analysis of Clay Minerals, (J.R. Walker
& R.C. Reynolds, Jr., editors) CMS Workshop
Lectures, 5, The Clay Minerals Society, Boulder,
CO.