GEOPHYSICS, VOL. 75, NO. 1 (JANUARY-FEBRUARY 2010); P. E67–E78, 12 FIGS., 5 TABLES.
10.1190/1.3298501

Automatic discriminations among geophysical signals via the Bayesian neural networks approach

Saumen Maiti[1] and Ram Krishna Tiwari[2]

ABSTRACT

The precise classification of changes in rock boundaries/facies from well-log records is a complex problem in geophysical data processing. Observed well-log data are a complex superposition of nonstationary/nonlinear signals of varying wavelengths and frequencies, shaped by the heterogeneous composition and structural variation of rock types in the earth. This impairs our ability to use traditional statistical techniques, which in most cases fail to discriminate and/or, at best, do not precisely extract facies changes from complex well-log signals. We propose a new method, set in a Bayesian neural network (BNN) framework and using a powerful hybrid Monte Carlo simulation scheme to identify facies changes from complex well-log data. We first construct a complex, composite, synthetic time series using the data from three simple models: first-order autoregressive, logistic, and random white noise. Then we attempt to identify individual signals in the pooled synthetic time series. We use the autocorrelation and the spectral characteristics of the individual signals as input vectors for training, validating, and testing the artificial neural network model. The results show that the Bayesian separation scheme provides consistently good results, with accuracy at more than 74%. When the method was tested using well-log data from the German Continental Deep Drilling Program (KTB), it was able to discriminate boundaries of lithofacies with an accuracy of approximately 92% in validation and 93% in test samples. The efficacy of the BNN in the presence of colored noise suggests that the designed network topology is robust for up to 30% correlated noise; however, adding more noise (say, 50% or more) obscures the desired signals. Our method provides a robust means for decoding finely detailed successions of lithofacies from complex well-log data, better describing the nature of the underlying inhomogeneous crust.

INTRODUCTION

Sharp changes in rock properties recorded in the form of well logs reflect physical boundaries and facies changes within various rock types. The well logs recorded in such rocks exhibit complex signal characteristics comprising nonlinear/nonstationary and random behavior. Discrimination of different rocks types/facies from such complex well-log signals is therefore an important challenge in geophysical signal analysis. Graphical crossplotting (Pickett, 1963; Gassaway et al., 1989) and other statistical techniques (multivariate statistical methods such as principal component and cluster analyses [Wolff and Pelissier-Combescure, 1982] and discriminant function analysis [Busch et al., 1987; Delfiner et al., 1987]) frequently have been used to study borehole data.

However, in complex geologic situations, such as in the presence of crystalline rocks where metamorphism leads to facies changes, it is not easy to extract accurate information from well-log data using these conventional methods. Moreover, inferences drawn by such methods are ambiguous because of strong overlapping of nonlinear/nonstationary well-log signals, which are also tainted with deceptive colored noise. Traditional techniques, which are semiautomated and require a large amount of data, are costly and not always easily available (Rogers et al., 1992). Further, these methods are very tedious and time consuming, particularly when dealing with noisy, complex borehole data.

Leonardi and Kumpel (1999) examine the fractal behavior of well-log signal variability, presuming that well-log signals in the superdeep German Continental Deep Drilling Program (KTB) borehole display nonlinear characteristics. Their study implies that nonlinear well-log data reveal signatures of crustal heterogeneities. However, it is difficult to differentiate individual nonlinear signals from composite and overlapping well-log signals arising from vari-

Manuscript received by the Editor 1 December 2008; revised manuscript received 9 July 2009; published online 12 February 2010.
[1]Indian Institute of Geomagnetism, Navi-Mumbai, India. E-mail: saumen_maiti2002@yahoo.co.in.
[2]National Geophysical Research Institute, Hyderabad, India. E-mail: rk_tiwari3@rediffmail.com.

ous physical sources. The problem becomes even more acute with individual signals characterized by varying degrees of complexity (e.g., nonlinearity, stochasticity, and randomness). Thus, it is imperative to search for a better, alternative perspective that takes care of the nonlinearity as well as the nonstationarity of well-log signals and can yield an appropriate mean for distinct classification of composite signals.

The artificial neural network (ANN) technique is used extensively to classify complex nonlinear signals because of its inherent ability to approximate the functional relationship between the input and the output space/domain by learning through examples, even if there is no deterministic relationship between the input and the output space/domain (Raiche, 1991; Bishop, 1995; Devilee et al., 1999; van der Bann and Jutten, 2000; Poulton, 2001). ANN has been applied to almost all branches of geophysics (van der Bann and Jutten, 2000; Poulton, 2001; Meier et al., 2007). Maiti and Tiwari have developed multiple linear and nonlinear algorithms to identify rock boundaries from the KTB borehole signal (Maiti and Tiwari, 2005; Maiti et al., 2007). However, one problem with the very popular ANN-based back-propagation algorithm (Rumelhart, 1986) is that it does not converge to a global minimum during optimization.

We therefore propose a more powerful approach known as Bayesian inference (Tarantola, 1987; Sambridge and Mosegaard, 2002) to approximate the a posteriori probability distribution from data likelihood and a priori information using a Monte Carlo algorithm in the case of the KTB well-log data. These methods have proven very useful in several other contexts because they yield nonunique solutions of complex geophysical inverse problems. The practical use of a sample-based inversion scheme, e.g., hybrid Monte Carlo (HMC)/Markov-chain Monte Carlo (MCMC), for neural network training can be found in several works (MacKay, 1992; Bishop, 1995; Neal, 1996; Lampinen and Vehtari, 2001; Maiti and Tiwari, 2009). However, the efficacy and applicability of these theoretical developments are not well explored for the case of complex and noisy signals.

Hence, we explored the stability of the method on noisy, synthetic, nonlinear models of varying complexity and then applied the method to the real KTB data. We tested the method on various synthetic data generated from well-known models: (1) first-order autoregressive [Ar (1)], (2) complex/logistic, and (3) white noise. Our experiment was intended to provide useful guidelines and confidence in classifying complex data sets and thereby to help make sound physical interpretations of actual well-log data. We applied the method to density (RHOB), neutron porosity (NPHI), gamma-ray intensity (SGR), seismic P-wave transit traveltime (DTCO), and electrical resistivity (LLD) KTB borehole data to discriminate among three lithofacies in a complex metamorphic region of central Europe.

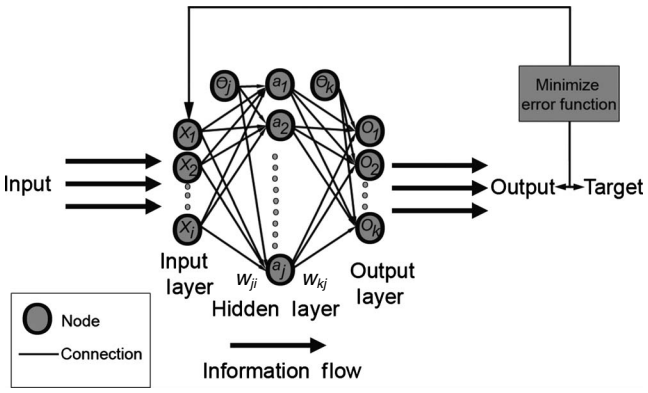

Figure 1. Layout of MLP with a three-layer neural network: x represents input, subscript i represents the number of nodes in the input layer, w_{ji} represents the connection weight between the ith node in the input layer and the jth node in the hidden layer, and w_{jk} represents the connection weight between the jth node in the hidden layer and the kth node in the output layer.

THEORY

Multilayer perceptrons

Multilayer perceptron (MLP) networks are parallel computational units composed of many simple processing elements that mimic biological neurons (Figure 1). Processing elements/nodes are interconnected layer by layer, and the functions of each node are determined by connections, by weights and biases, and by the topology of the network (Bishop, 1995; Poulton, 2001). In the popular back-propagation method, the error is usually minimized by adjusting the weights and biases using a gradient-based iterative chain rule from output to input layer (Rumelhart, 1986). The main drawback of the method is that it often becomes stuck in local minima on an error surface.

To avoid the latter problem, we use HMC simulations (also known as a leapfrog discretization scheme) in conjunction with Bayesian probability theory, which is naturally parsimonious, thus suiting our needs. Complete details of the neural network topology and learning rules can be found in Bishop (1995) and Poulton (2001). Here, we relate the geophysical observations to the model through the following forward equation:

$$\mathbf{x} = f(\mathbf{d}) + \boldsymbol{\varepsilon}, \tag{1}$$

where f is a nonlinear function relating the model space and data space, $\boldsymbol{\varepsilon}$ is an error vector, $\mathbf{x}$ is the data vector, and $\mathbf{d}$ is the model vector. A common way of inverting for $\mathbf{d}$ in equation 1 is via an iterative least-squares method. This, however, does not provide uncertainty measures, which are essential for sound physical interpretation of geophysical observations (Tarantola, 1987).

Model solution in the Bayesian framework

To solve equation 1 in the Bayesian framework, we recast it as

$$\mathbf{d} = f_{NN}(\mathbf{x};\mathbf{w}), \tag{2}$$

where f_{NN} is output predicted by the neural network, $\mathbf{x}$ is a given input vector, and $\mathbf{w}$ is the network weight parameters. In the conventional approach for solving equation 1, regularization is often included to minimize the misfit function:

$$E(w) = \mu E_S + \lambda E_R, \tag{3}$$

where $E_S = 1/2\Sigma_k^N\{d_k - o_k(x_k;w_k)\}^2$ and $E_R = 1/2\Sigma_{i=1}^R w_i^2$; R is the total number of weights and biases in the network, o_k is the neural network output at the output layer from a finite data sets, and λ and μ, which control other parameters (synaptic weight and biases), are known as hyperparameters.

In this approach, the training of a network starts with an initial set of weights and biases and ends up with the single best set of weights and biases to optimize the objective function. In the Bayesian approach, a suitable prior distribution, say, $P(\mathbf{w})$ of weights, is considered before observing the data instead of considering only a single

set of weights. Using Bayes' rule, an a posteriori probability distribution for the weights, say, $P(\mathbf{w}|s)$, can be defined as (Khan and Coulibaly, 2006)

$$P(\mathbf{w}|s) = \frac{P(s|\mathbf{w})P(\mathbf{w})}{P(s)}. \quad (4)$$

Here, $P(s|\mathbf{w})$ is a data-set likelihood function and the denominator $P(s)$ is a normalization factor. Because $P(s)$ is intractable, direct estimation of a posteriori $P(\mathbf{w}|s)$ is impossible. Using the rule of conditional probability, the distribution of the output for a given input vector $\mathbf{x}$ can be defined in the form (Khan and Coulibaly, 2006)

$$P(\mathbf{d}|\mathbf{x},s) = \int P(\mathbf{d}|\mathbf{x},\mathbf{w})P(\mathbf{w}|s)\mathbf{d}\mathbf{w}. \quad (5)$$

The major problem in Bayesian computation is evaluating the integrals for the a posteriori weights (equation 4) and for the network output (equation 5). In this regard, the MCMC sampling-based method plays an important role in evaluating a posteriori integrals. Equation 5 can be approximated as

$$P(\mathbf{d}|\mathbf{x},s) = \frac{1}{N}\sum_{n=1}^{N} P(\mathbf{d}|\mathbf{x},\mathbf{w}_m), \quad (6)$$

where $\{\mathbf{w}_m\}$ represents an MCMC sample of weight vectors obtained from the distributions $P(\mathbf{w}|s)$ and N is the number of points $\mathbf{w}$ sampled from $P(\mathbf{w}|s)$.

Hybrid Monte Carlo (HMC)

In the HMC algorithm, each trajectory is updated by approximating the Hamiltonian differential equations by a leapfrog discretization scheme (Duane et al., 1987). The MCMC algorithm draws an independent and identically distributed (IID) sample $\{\mathbf{w}^{(i)}; i = 1, 2,\ldots,N\}$ from the target distribution $P(\mathbf{w}|s)$. The Markov process forms a sequence of states to draw samples from the posterior probability. The chain converges to $P(\mathbf{w}|s)$ if given enough space to do so. The complete mathematical details can be found in Bishop (1995) and Nabney (2004). Unfortunately, the pure Metropolis-Hastings algorithm (Metropolis et al., 1953; Hastings, 1970) is very slow because it does not use gradient information. Contrary to this, the HMC-based algorithm for sampling from the target distribution uses gradient information.

The following steps are needed once a step size θ and the number of iterations L have been decided upon. First, we randomly choose a direction τ, which can be -1 or $+1$ with probability 0.5 to simulate a forward or backward step in time. Second, following the theory of Hamiltonian statistical mechanics, the transition probability matrix should satisfy microscopic reversibility, which means that the probability of the two transitions from q_j to q_i or from q_i to q_j be the same at all times and each pair of points maintains a mutual equilibrium (Sambridge and Mosegaard, 2002). Third, we iterate starting with the current state $[q,p] = [(q(0),p(0)\ ]$ of energy H, where p is a momentum term that is randomly evaluated at each step. Fourth, we let the algorithm be applied L times with a step size of θ, resulting in the candidate state $[\mathbf{w}^*,p^*]$ with energy H^*. The candidate state is accepted with the usual Metropolis probability of acceptance $\min\{1,\exp[-(H^* - H)]\}$ (Bishop, 1995). If the candidate state is rejected, then the new state will be the old state.

In essence, these steps describe how the sampling is done from a posterior distribution of network parameters so that the summation of equation 6 can be accomplished and the posterior distribution can be found, thus optimizing the network. The main idea of the algorithm is that the acceptance probability is evaluated at each step to produce the necessary number of realizations (Rabben et al., 2008). Thus, we estimate the posterior distributions that are not easily tractable analytically. The desired statistics can be estimated from the same realizations.

Neural network implementations

Following Cybenko (1989), the network topology for our purposes has one input layer and one output layer, each consisting of three nodes (output nodes are binary coded), and a single hidden layer consisting of 15 nodes. This number of hidden-layer nodes is suitable for our work (Table 1). The raw data (input/output) must be normalized before presenting it to the network to avoid saturation while mapping a nontrivial problem (Maiti et al., 2007). Hence, we scale all input/output pair values between zero and one /[−1 and +1] by using a simple linear transformation algorithm (Poulton, 2001): normalized input = 2 × (input − minimum input)/(maximum input − minimum input) − 1.

The initialization of the model parameters is performed using a distribution of model parameters. The initial values of model parameters of the MLP networks (synaptic weight and biases) are formed by Gaussian prior distributions of zero mean and inverse variance λ (also known as regularization coefficient or prior hyperparameter). We prefer a Gaussian prior distribution because it provides computational simplicity and favors small values for the network weights. Networks with large weights usually result in a map with large curvature (Nabney, 2004).

We consider a hyperparameter $\lambda = 0.02$, a single initial-value hidden layer, and output-layer weights as detailed in Table 1. To define the objective function in a Bayesian framework, an error model for the data likelihood is required. We assume that target data are formed by a smooth function with additive zero-mean Gaussian noise. Accordingly, hyperparameter $\mu = 50$ is estimated for the hidden-layer and output-layer weights (Table 1). After defining prior and likelihood functions, we estimate the a posteriori distribution using Bayes' rule. The parameter λ controls the tendency of the model to overfit the noise in the training data. The uncertainty of the network output is controlled by the value of μ (Tables 1 and 2).

We performed several experiments with different model setups (see Table 1). The number of hidden nodes ($N_h = 15$), hyperparameters ($\lambda = 0.02$; $\mu = 50$), step size ($\theta = 0.002$), and total iterations ($L = 100$) were found to be suitable, producing an average uncertainty value of approximately 0.14 at network output (see Table 1). Our experimental results also indicate that neither a very small step size nor a very large number of iterations improves accuracy of the network for our work (see Table 1).

It should be noted that this is a sampling-based algorithm. The leapfrog scheme updates the candidate state. The new state is accepted if the threshold value is greater than the Metropolis acceptance probability, which is a random number between zero and one decid-

ed at each step (Figure 2a). Control of the acceptance probability and, thus, the acceptance rate is very important. If it is equal to one, all states will be accepted and the result will be linear inversion (Bishop, 1995; Rabben et al., 2008). However, if it is close to zero, it will produce very few updates and will lead to poor convergence (Bishop, 1995; Rabben et al., 2008). In the present case, the acceptance rate is quite satisfactory (0.92) (Figure 2a). The algorithm retains a defined number of samples (here, 100) in the Markov chain.

Table 1. Noise-analysis test results.

Parameters	Red-noise level	Accuracy in validation set (%)			Accuracy in test set (%)		
		PG	MB	HS	PG	MB	HS
$N_h = 15;\ \lambda = 0.02;$ $\mu = 10;\ \theta = 0.002;$ $L = 100;$ AU = 0.31	10%	81.25	83.81	84.94	72.73	75.00	75.85
	20%	75.00	75.00	76.14	70.74	69.89	69.60
	30%	66.19	63.90	70.74	62.22	61.93	61.65
	40%	60.23	61.08	63.92	55.68	56.53	57.10
	50%	55.40	54.83	54.26	51.42	52.27	51.99
$N_h = 15;\ \lambda = 0.02;$ $\mu = 50;\ \theta = 0.002;$ $L = 100;$ AU = 0.14	10%	87.22	87.78	86.93	75.57	78.13	78.41
	20%	83.24	82.95	83.52	75.00	75.57	75.28
	30%	81.53	78.69	79.55	71.59	72.44	71.88
	40%	76.42	78.41	77.84	67.61	68.18	69.32
	50%	72.73	73.58	73.58	65.91	66.19	66.19
$N_h = 15;\ \lambda = 0.02;$ $\mu = 100;\ \theta = 0.002;$ $L = 100;$ AU = 0.10	10%	87.78	87.50	87.40	75.57	76.42	76.14
	20%	82.67	81.82	81.53	72.44	74.43	73.86
	30%	77.56	78.13	76.70	67.90	70.45	68.75
	40%	73.30	73.01	72.44	65.63	67.90	64.77
	50%	70.17	69.60	69.89	64.49	65.63	63.07
$N_h = 5;\ \lambda = 0.02;$ $\mu = 50;\ \theta = 0.002;$ $L = 100;$ AU = 0.14	10%	78.98	79.55	79.55	71.31	71.31	71.02
	20%	76.70	77.56	78.13	69.60	70.45	69.89
	30%	75.28	76.14	77.27	69.32	70.17	68.75
	40%	73.58	73.86	74.72	67.61	68.18	67.71
	50%	72.44	72.73	73.30	65.91	67.33	67.33
$N_h = 20;\ \lambda = 0.02;$ $\mu = 50;\ \theta = 0.002;$ $L = 100;$ AU = 0.14	10%	82.67	86.36	86.65	75.57	76.99	75.85
	20%	77.27	78.98	79.26	68.75	72.44	71.02
	30%	69.32	71.31	73.01	62.50	64.77	63.64
	40%	63.92	65.63	64.49	57.39	57.67	56.53
	50%	57.95	60.80	57.10	53.13	55.40	52.56
$N_h = 15;\ \lambda = 0.1;$ $\mu = 50;\ \theta = 0.002;$ $L = 100;$ AU = 0.14	10%	86.66	87.78	84.38	75.28	77.56	74.43
	20%	82.39	82.10	80.97	74.15	75.85	74.43
	30%	78.69	78.69	78.41	70.45	72.44	70.74
	40%	73.30	75.28	73.86	65.63	67.90	67.33
	50%	70.45	71.31	69.32	63.07	65.06	63.92
$N_h = 15;\ \lambda = 0.001;$ $\mu = 50;\ \theta = 0.002;$ $L = 100;$ AU = 0.14	10%	85.80	87.50	87.50	77.27	79.26	76.70
	20%	82.39	82.67	82.95	75.57	76.42	74.15
	30%	79.83	79.83	80.11	70.74	72.44	71.31
	40%	75.57	75.00	74.43	67.05	69.89	66.76
	50%	71.59	72.73	69.89	62.78	66.76	63.07
$N_h = 15;\ \lambda = 0.001;$ $\mu = 50;\ \theta = 0.0002;$ $L = 200;$ AU = 0.14	10%	57.67	54.55	55.66	52.27	53.41	55.11
	20%	51.14	50.57	49.43	45.74	49.43	49.15
	30%	46.02	42.90	41.48	38.64	41.76	43.47
	40%	41.48	38.92	35.51	36.08	36.65	35.23
	50%	36.08	35.80	33.24	31.53	32.39	32.10

AU = average uncertainty

Thus, during the run, hundreds of inverted network parameters are obtained (Figure 2b). The output of the Bayesian approach can be interpreted as the a posteriori mean of hundreds of model realizations and, thus, as the solution to the inverse problem (Tarantola, 1987)

NETWORK MODEL FOR SYNTHETIC DATA

Before these methods can be used on actual well-log data, we test the efficiency of the proposed techniques on three theoretical models: autoregressive [Ar (1)], logistic, and a white-noise process.

Autoregressive model

An Ar (1) model (Fuller, 1976) takes the form $X_t = AX_{t-1} + \varepsilon_t$; where $t = 1,2,3,\ldots,N$ denotes the discrete spatial increment. Here, A is a maximum-likelihood estimator and ε_t denotes a purely random process (an uncorrelated normal distribution uniformly distributed in the interval between zero and one). An autocorrelation coefficient describes the degree of signal correlation in the noise and is calculated from the data. Its value ranges from zero to one. The term X_t depends partly on X_{t-1} and partly on the random distribution ε_t. The Ar (1) model exhibits a tendency to cluster toward low values (Figure 3a).

Logistic model

A complex system can be represented by the logistic model equation (May, 1976), which is of the form $X_{t+1} = BX_t(1 - X_t)$, where X_t and X_{t+1} are the present and future values of a generating process with relative values ranging from zero to one and where B is a coefficient (control parameter) between zero and four. Theoretically, complexity reaches the maximum for a B value near four. A 3D phase-space characteristic is displayed in Figure 3b, which plots present observations $X(t)$ on the x-axis, one step ahead $X(t + 1)$ on the y-axis, and $X(t + 2)$ on the z-axis. The data based on the logistic model evolve toward a well-behaved set (Figure 3b).

Random white-noise model

Random white noise is uncorrelated and has zero mean. Such a process is unpredictable because of its uncorrelated nature. A random 3D phase plot shows that the values scatter equally in all directions (Figure 3c).

Spectral and autocorrelation characteristics of synthetic data

Figure 4 shows the relationship among three data sets — logistic, stochastic, and random — in 3D phase space. The plot shows scattering that indicates a complex and nonlinear relation among the observations. It is not easy to draw any parametric boundary to classify the three sets of time-series data using linear methods (Figure 4). The spectral and statistical characteristics of the individual signals are used to distinguish between random, stochastic and deterministic components from the complex signals. We used these characteristics as input to an initial ANN model. This is because the spectral power spectra of chaotic and stochastic models exhibit broadband characteristics that distinctly classify random, red-noise, and deterministic components (Figure 5).

Figure 5 shows Fourier power spectra for all three models. The autocorrelation indicates the similarity between observations as a function of the time separation between them. More precisely, it is

Table 2. Uncertainty analysis for nonlinear series.

Network parameter	Control parameter	Average uncertainty
$N_h = 15;\ \lambda = 0.02;\ \mu = 100;\ \theta = 0.002;\ L = 100$	$A = 0.5;\ B = 3.8\ \sim N(0,1)$	0.10
$N_h = 15;\ \lambda = 0.02;\ \mu = 10;\ \theta = 0.002;\ L = 100$	$A = 0.5;\ B = 3.8\ \sim N(0,1)$	0.31
$N_h = 15;\ \lambda = 0.02;\ \mu = 50;\ \theta = 0.002;\ L = 100$	$A = 0.5;\ B = 3.8\ \sim (0,1)$	0.14
$N_h = 15;\ \lambda = 0.02;\ \mu = 50;\ \theta = 0.002;\ L = 100$	$A = 0.5;\ B = 4.0\ \sim N(0,1)$	0.14
$N_h = 15;\ \lambda = 0.02;\ \mu = 50;\ \theta = 0.002;\ L = 100$	$A = 0.75;\ B = 4.0\ \sim N(0,1)$	0.14

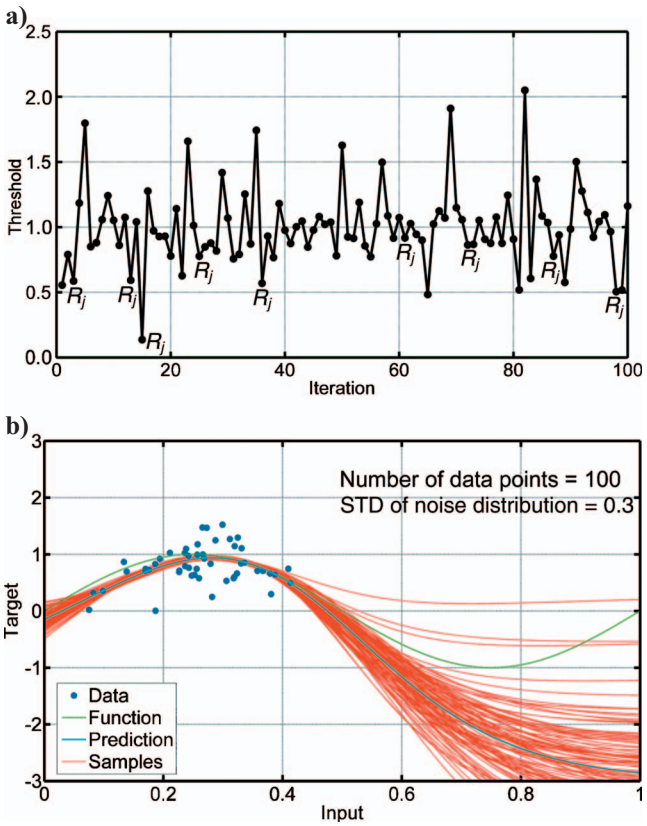

Figure 2. (a) The acceptance threshold history against iteration. R_j denotes the rejected threshold. Fraction of sample rejection is 8%. (b) Simulation of a synthetic underlying function using 100 data points in the presence of noise via HMC method. The standard deviation (STD) of the noise is 0.3.

the crosscorrelation of a signal with itself. Autocorrelation functions (ACFs) with lag computed for three models show distinct characteristics.

Network model for synthetic data

We train the MLP network coupled with HMC using the data of the three basic models as described above, their power spectra, and ACF coefficients. In all, 768 data sequences were used for network training. The results of a linear regression analysis of the training data are displayed in Figure 6 and Table 3. Parameters u' and v correspond to the slope and the y-intercept of the best linear regression fit relating to the network output A and target T, respectively. The network outputs are plotted against T, shown as open circles. A dashed line indicates the best linear fit (slope = 1; y-intercept = 0). The solid line in Figure 6 shows the perfect fit (output equal to target). The third variable is the correlation coefficient R between the network outputs and the targets, which is a measure of how well the trained network predicted the target. The number is equal to one if there is perfect correlation between targets and output.

A composite time series of three models (Figure 7) is used to test the trained network. Spectral discriminates are used to train the network, and the resulting network output for the presence or absence of a nonlinear sequence is indicated in the form of a probabilistic index (Figure 7). The trained network is able to discern individual signals from the test model sequence with 74% accuracy (Figures 6 and 7). Uncertainty analysis of the predicted output is performed by calculating standard deviation (STD) from the a posteriori covariance matrix of the network output. The discrimination results are presented in three gray bands, with black representing one and white representing zero (Figure 7).

We have experimented with different network parameters to estimate the uncertainty of the network output (see Table 2). The average uncertainty of the predictions is approximately 0.14 at network output, with a 90% confidence interval. The uncertainty of the network prediction depends on μ (see Table 2). It is interesting that clas-

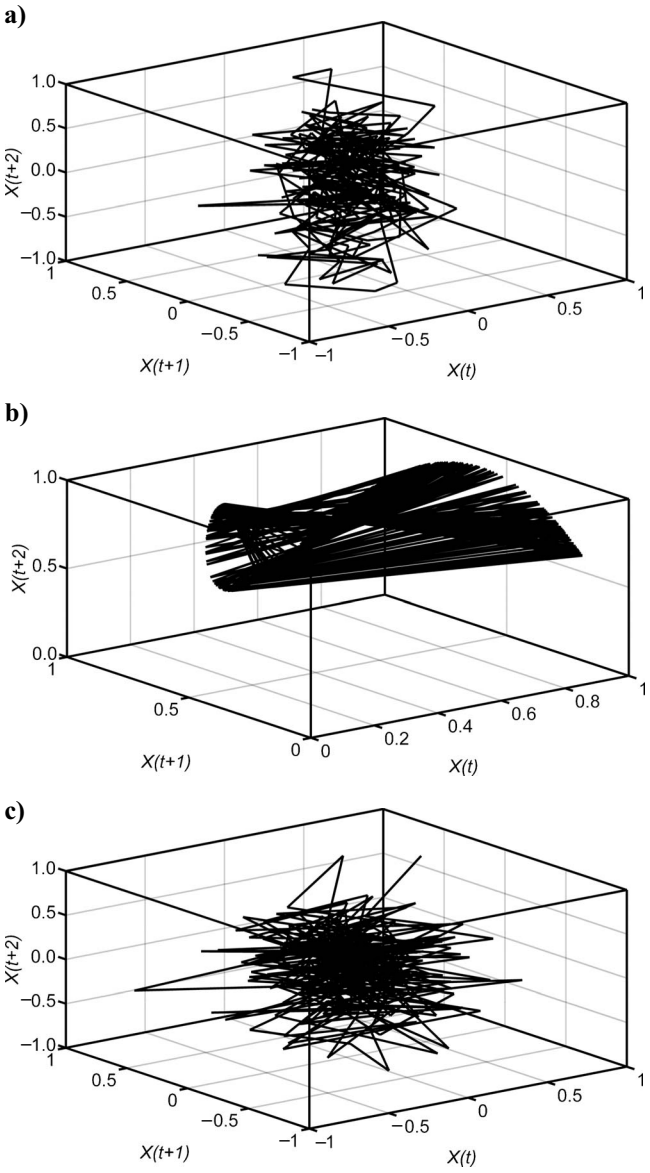

Figure 3. (a) Phase space plot of Ar (1)/stochastic series. The maximum-likelihood estimator A value used for the stochastic series is 0.5. (b) Phase-space plot of chaotic/logistic series. The control parameter B used for the chaotic/logistic series is 3.8. (c) Random white-noise processes.

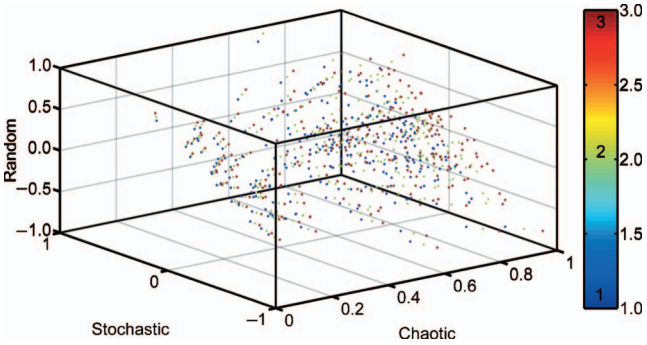

Figure 4. A 3D scatter plot of three types of nonlinear series, shows overlapping signal contents.

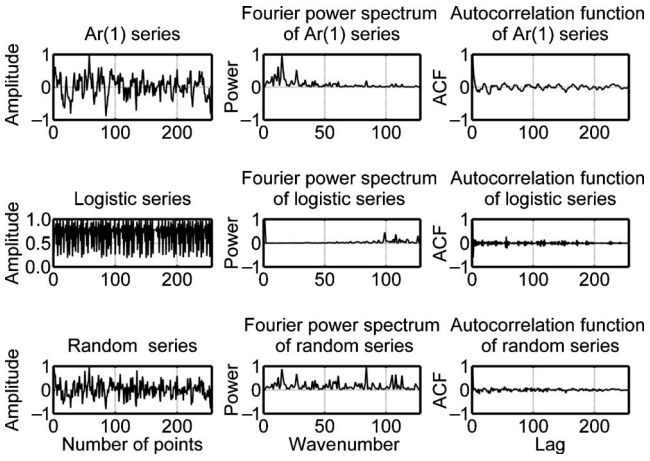

Figure 5. Plot of three types of nonlinear series; Fourier power spectrum and autocorrelation functions (ACFs) correspond to a nonlinear model.

sification results for the composite signals (comprising nonlinear, stochastic, and random sequences) by our method are very good. Thus, our method could be used as an alternative way to discriminate among complex signals.

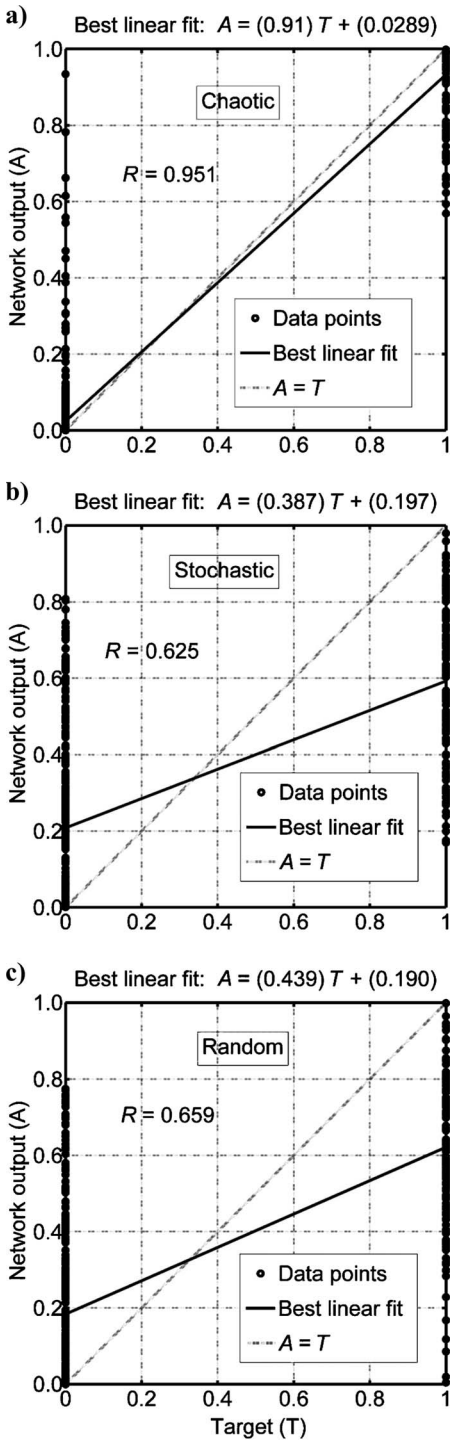

Figure 6. Linear regression analyses of total set of data corresponding to (a) chaotic/logistic model ($B = 3.8$), (b) Ar (1) series ($A = 0.5$), and (c) white-noise model.

NETWORK MODEL FOR FIELD DATA

After describing the successful test of the proposed algorithm on complex synthetic series, we now demonstrate the method on real KTB, main-borehole records — density (RHOB), neutron porosity (NPHI), gamma-ray intensity (SGR), seismic P-wave transit traveltime (DTCO), and electrical resistivity (LLD) — for discriminating among three lithofacies in a complex metamorphic region of central Europe.

About the KTB site and data

The KTB drill site is located near the western margin of the Bohemian massif, the largest surface exposure of crystalline rocks in central Europe. The drill site is located within the Zone of Erbendorf-Vohenstrauss, a small crustal unit consisting mainly of paragneisses (PG), metabasites (MB), and a heterogeneous series (HS). Detailed

Table 3. Linear regression analysis of nonlinear series.

Rock type	Correlation coefficient R between target T & network output A	Slope u	y-intercept of best linear regression relating target to network output v
Chaotic	0.95	0.91	0.02
Stochastic	0.62	0.38	0.19
Random	0.65	0.43	0.19

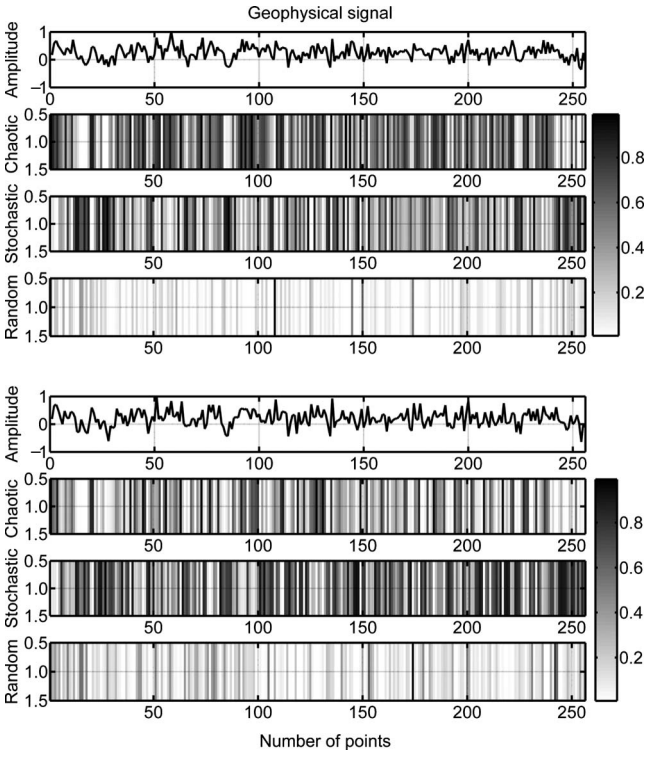

Figure 7. (a) HMC-based classification results of composite signals. In the chaotic model, the value for constant B is 3.8; for the stochastic model, the assigned value for A is 0.5. (b) Same as (a) with $B = 4.0$ for the chaotic model and $A = 0.25$ for the stochastic model. In both, the random white-noise signal is added point by point.

information about the KTB data is presented in several papers (Franke, 1989; Berckhemer et al., 1997; Emmermann and Lauterjung, 1997; O'Brien et al., 1997; Pechnig et al., 1997; Leonardi and Kumpel, 1998, 1999).

In a complex geologic setting such as that of the KTB, it is desirable to know the general well-log response of the various rocks to prepare an appropriate training set for ANN modeling. Spectral gamma ray (SGR) is the well-log parameter most effective in separating the different successions in the metamorphic rocks of the KTB. Because of their chemical compositions, SGR exhibits a general increase from the most mafic rocks (ultramafites) to the most acidic rocks (potassium-feldspar-gneiss) observed in the boreholes.

In general, amphibolites and metagabbros, the main rock types of the massive metabasite unit, are characterized by lower SGR and higher density (RHOB) than the rocks of paragneiss units. This is because the metabasites are composed of more mafic and dense minerals such as hornblende and garnet biotite than are the paragneisses, which are composed mainly of quartz, plagioclase, and micas. The highest gamma-ray activity is recorded in response to the potassium content. These paragneisses are characterized geochemically by very high aluminum values, bound to the feldspar minerals. The relation between aluminum content and amount of feldspar minerals also explains the low aluminum values recorded in the ultramafites, which are almost free of feldspar.

Enhanced NPHI generally reflects discrete zones of faulting and fracturing (Berckhemer et al., 1997). The rock types that are poor in phyllosilicates or amphibolites, such as quartz and feldspar-rich gneiss, exhibit very low NPHI; but rocks with high phyllosilicate and amphibole content produce striking increases in the NPHI log. In general, DTCO decreases with increasing density and consequently with the amount of mafic minerals. Significant differences, ranging between 165 and 200 μs/m, were established between the DTCO in metabasites with average values of 160 μs/m and the DTCO in paragneisses (Berckhemer et al., 1997; Pechnig, 1997). Unlike RHOB, NPHI, SGR, and DTCO, the LLD is affected little by rock composition (see Table 4). LLD values are generally very high in crystalline basement rocks. Variations between 10^3 and 10^6 ohm -m (LLD) are nearly independent of the main lithology penetrated by the KTB. One can see these distribution patterns in the well-log data of Figure 8.

Bayesian network model for the KTB borehole data

For the actual data modeling, we designed our network with a single input layer consisting of five nodes corresponding to RHOB, NPHI, SGR, DTCO, and LLD well-log records. We implemented binary discrimination codes in a single output layer that consisted of three nodes representing PG, MB, and HS (see Table 4). There was one intermediate hidden layer with 15 nodes. Before reaching any conclusion, we conducted several experiments by altering the number of nodes in the hidden layer, finding the present setup suitable for our problem (see Table 1). We parameterized five sets of log responses (considering the well-log response of RHOB, NPHI, SGR, DTCO, and LLD to the rock compositions, and the significant parameter limit as described above and in Table 4) and generated corresponding representative input/output pairs.

A total of 1408 input/target pairs were considered for the analysis. All available data sets were randomly partitioned for training (50%), validations (25%), and testing (25%). However, there is no rule of thumb for partitioning the data; the percentage depends on data and the problem. We rechecked different target units in each independent subset carefully and reshuffled all samples (1408 input and output pairs), partitioning them randomly to each independent subset. Early stopping is a common technique in conventional neural network learning. One avoids overfitting problems by monitoring the validation-set error during training (Poulton, 2001). We have adopted a powerful sampling-based Bayesian approach that is parsimonious enough to take care of the overfitting problem, even in a complex model (Bishop, 1995).

We considered even more realizations to obtain the generalized training set. The network parameters are well resolved through optimizing the HMC-based sampling (Figure 9). The robustness and stability of the method is examined by noise analysis (Maiti et al., 2007). To do that, maximum-likelihood-estimator (MLE) constants are estimated from the data series chosen for analysis. For the validation set, the MLE constants of RHOB, NPHI, SGR, DTCO, and LLD are 0.68, 0.64, 0.85, 0.91, and 0.89, respectively; for the test set, they are 0.67, 0.65, 0.84, 0.92, and 0.88, respectively. After successfully completing the network training, we input the test sets (validations and test) corrupted with different levels of correlated colored (red) noise. The noise-sensitivity analyses demonstrated that the network was stable for up to 30% correlated colored (red) noise mixed with the well-log data (Table 1; Figure 10). Further, we performed linear regressions analysis (Maiti et al., 2007) to examine overall performance of the trained networks for all of the data sets chosen for analysis. The linear regression analyses showed very good results (with ~99% accuracy) for resolving different rock types (Table 5; Figure 11).

After successfully completing the training and the test for robustness of our method, we used the trained network to discriminate among lithofacies in the pilot borehole (KTB-VB) for a depth interval of 28–4000 m and in the main hole (KTB-HB) for a depth inter-

Table 4. Significant limits to generate forward model for neural-network learning.

Rock type	Density RHOB (g/cm^3)	Neutron porosity NPHI (%)	Gamma-ray intensity SGR (°API)	P-wave transit traveltime DTCO (μ/m)	Ln (resistivity) LLD (ohm-m)	Desired output(binary code)
PG	2.65–2.85	5–15	70–130	165–200	3–9	100
MB	2.75–3.10	5–20	0–50	143–196	4–10	010
HS	2.60–2.90	1–15	40–90	174–220	5–9	001
			120–190			

val of 3000–7000 m. The pilot borehole and main-hole data were sampled at a common interval of 15.24 cm (6 inches). The total depth of the main hole and the pilot hole were 9101 and 4000 m, respectively. It is important to mention that the well logs used here were continuous and uninterrupted throughout the depth intervals used.

The output of the networks is interpreted as a maximum a posteriori geologic section (MAPGS) derived from the BNN with HMC modeling for both KTB boreholes. Output is displayed as a gray-shaded matrix in Figure 12, along with real KTB well-log data, for critical and thorough examination. The STD of the error maps (average value ~0.14 within a 90% confidence interval) corresponding to the three lithofacies is estimated by Bayesian code to quantify the prediction uncertainties in the network output over the entire KTB lithosection. The result is presented in Figure 12 in the form of probability matrix with same gray shading. The maximum a posteriori

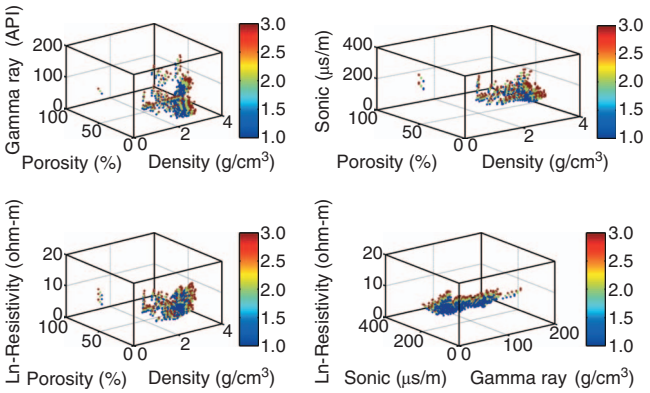

Figure 8. A 3D scatter plot showing the overlapping signal content of KTB data. It is very difficult to draw any parametric boundary.

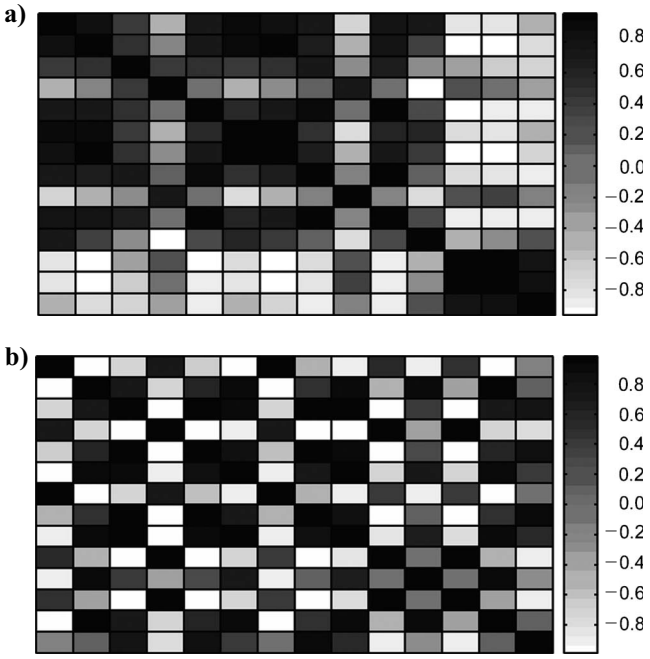

Figure 9. Correlation matrix plot (a) between input-layer and hidden-layer weights and (b) between hidden-layer and output-layer weights.

geologic section (MAPGS) corresponds to the class with maximum a posteriori probability. In the ideal case, if the lithofacies of a particular class exists, the output value of the node in the last layer is one or very close to one; if not, it is zero or very close to zero. The MAPGS is consistent with main-borehole lithosections (Maiti et al., 2007; their Figure 2). In addition, the proposed method detects some finer structures over a couple of depth intervals within the major successions of paragneiss, metabasites, and heterogeneous series. These finer structures appear to be geologically significant (Maiti et al., 2007).

Additional findings of heterogeneous series in between the paragneisses and metabasites were ambiguous in previous studies because of the strong superposition of well-log signals produced by the varying composition and structure of the crystalline rocks. For example, at depth intervals of 500–600 m, 1010–1040 m, 1100–1120 m, 1510–1550 m, 2500–2520 m, and 3400–3500 m for the pilot hole (KTB-VB) and 3210–3260 m, 4000–4010 m, 4100–4150 m, 4800–

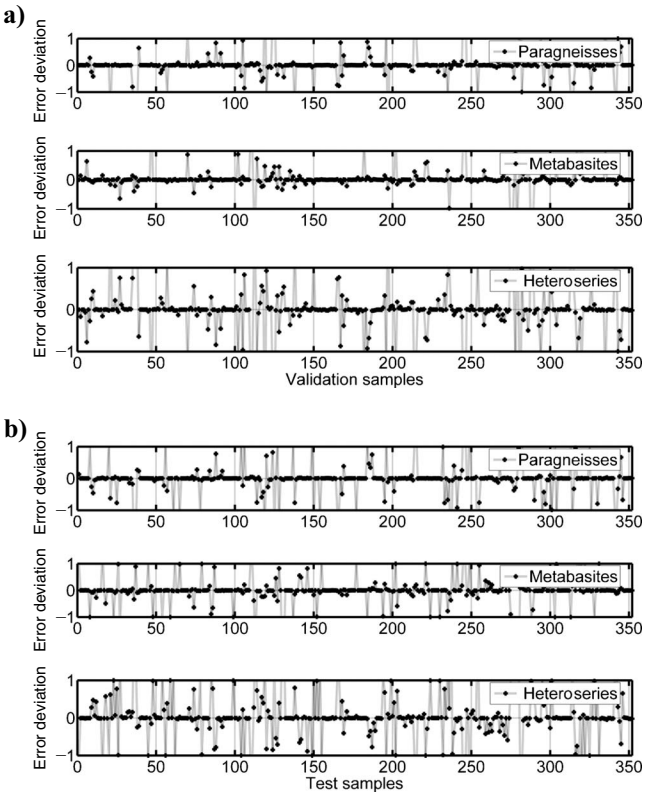

Figure 10. Network stability history for (a) validation samples and (b) test samples when both data sets are corrupted with 50% of correlated red noise.

Table 5. Linear regression analysis of major rock types.

Rock type	Correlation coefficient R between target T & network output A	Slope u	y-intercept of best linear regression relating target to network output v
PG	0.99	0.99	0.01
MB	0.99	0.99	0.00
HS	0.99	0.99	0.00

4810 m, 5300–5310 m, 5500–5520 m, 6000–6050 m, 6530–6550 m, 6750–6800 m, and 6900–6960 m for the main hole (KTB-HB), the HS lithotype is detected (Figure 12a and b). In the main hole, the change of MB to HS at a depth of 3413.15 m is also confirmed. It is geologically significant in view of the extension of the Franconian lineament, which cuts the KTB main hole (KTB-HB) at that particular depth (Emmermann and Lauterjung, 1997).

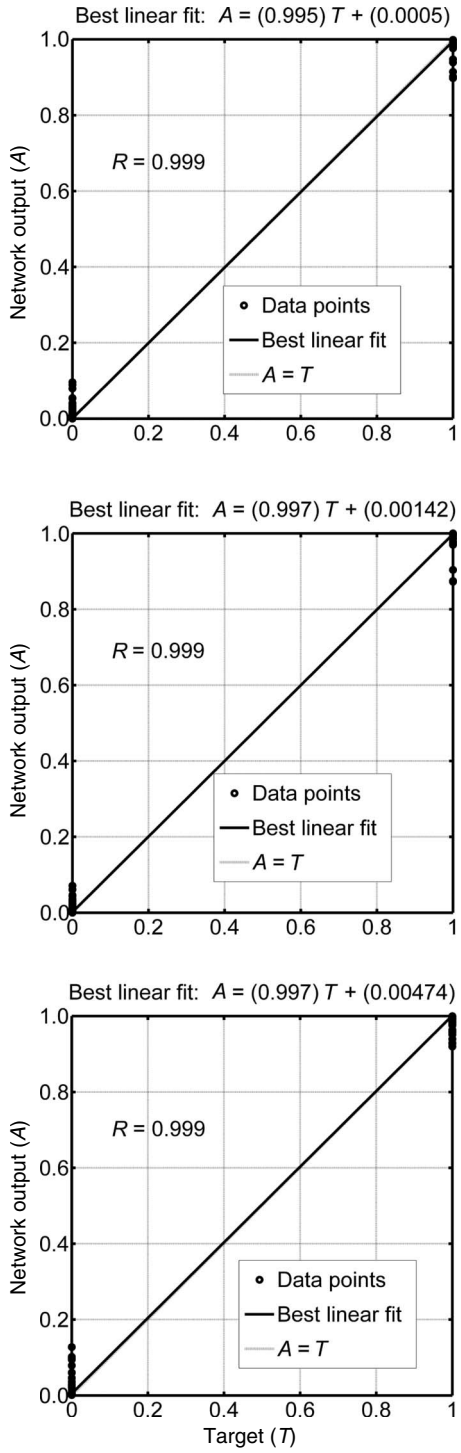

Figure 11. Linear regression analysis of the total set of data corresponding to (a) paragneisses, (b) metabasites, and (c) heterogeneous series.

Our analysis shows clear demarcation of the three types of lithofacies. HS contains cataclastic rocks, hornblende, and lamprophyre, which usually occur at the faulted and altered zones in the KTB deep hole. So the identification of the HS series with our enhanced interpretation technique certainly helps us develop a clearer picture of KTB heterogeneity as well as understand the degree of alteration of different stages of metamorphism over the entire KTB lithosection. In general, these findings must have great implications for understanding the complex tectonic setting during different phases of metamorphism. Thus, the HMC-based Bayesian network approach increases insight into the evolution and present heterogeneity of the European crust in a more quantitative and qualitative manner.

DISCUSSION

In demonstrating our HMC-based BNN algorithm for classifying lithofacies boundaries, we first used complex synthetic signals generated from the well-known simple models because these data represent, directly or indirectly, the inherent complexity of earth signals. We tested the potential of the discrimination scheme on synthetic data, resulting in an approximately 93% success rate. We also used several statistical parameters shown to have discrimination power for classifying complex signals. The impacts of the various control parameters were studied extensively (Tables 1 and 2). It is clear that uncertainty estimation is mostly controlled by μ (Tables 1 and 2). The hyperparameter λ controls the regularization of noise present in data (Table 1). Actual KTB log analyses provide more detailed information of geologic relevance, and our results correlate nicely with the existing geologic and geophysical findings available. Hence, our method could be applied for any type of complex data using various combinations of the statistical parameters; nothing in the method restricts its application to specific lithologies.

No geophysical field records are completely free from noise, but we do not know the exact level of noise present in the real well-log data. Keeping this in mind, we tested the stability of our scheme in the presence of different levels of correlated red signals. Often, this red noise is misread as data because of its correlation structure. Our test results on such noisy records show that the HMC-based scheme is stable even up to mixing 30% of red noise with the actual data. Hence, we believe our HMC-based scheme can reliably tackle the problem of correlated noise present in the data. In the same way, it provides more confidence in interpreting noisy geophysical records.

We performed uncertainty analysis and found an average uncertainty value of approximately 0.14 throughout the KTB lithosection. However, at some depths intervals, it varied significantly from that value. For instance, a somewhat anomalous uncertainty value of 0.30 was observed in the depth interval of 300–350 m in the pilot hole (KTB-VB), which may result from the poor sampling rate, as can be seen in the other three STD error maps. Again, looking at the depth interval of 6855–6865 m in the main hole (KTB-HB), we find anomalous uncertainty values that might arise from the presence of increased metabasites in the heterogeneous series.

The Bayesian approach incorporates a posteriori data uncertainty caused by inherent correlated noise and by inexact theory (modeling uncertainty) in a very natural way. The MAPGS by an HMC-based algorithm (Figure 12) demonstrates patterns that more or less match published lithology (Maiti et al., 2007). In addition, the HMC-based Bayesian model also reveals finer structural details that seem to be geologically significant. The HMC-based Bayesian network approach provides results that are consistent with prior geologic and

geophysical information. Some mismatch with published results could be explained by inexact theory (modeling uncertainty) or poor sampling. Our HMC-based Bayesian approach, combined with correlated noise sensitivity, uncertainty, and regression analyses, lends credence to these results.

The MAP probability value identifies a robust solution for the presence of lithofacies within the error limit of approximately 5%. Our result is further constrained by error analyses in estimating boundaries of lithofacies (Figure 12). Hence, we conclude that the maximum a posteriori (MAP) value identifies real, interbedded geologic structures that went unrecognized in previous visual interpretations.

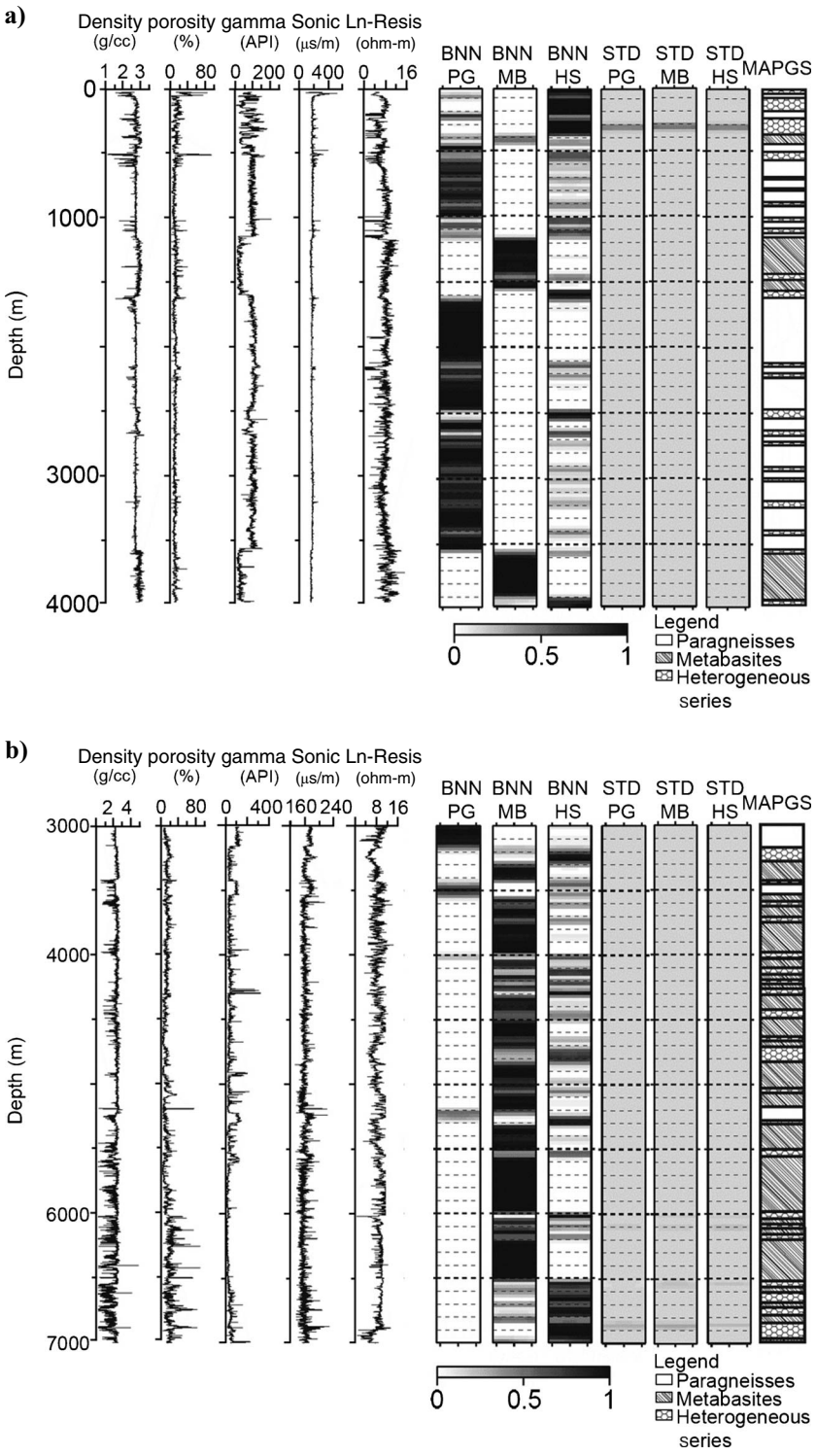

Figure 12. (a) MAPGS based on Bayesian neural networks with HMC simulations of pilot borehole KTB-VB from a depth of 28–4000 m. (b) Same for main borehole KTB-HB from a depth of 3000–7000 m.

CONCLUSIONS

We have developed an HMC-based Bayesian neural network method to discriminate among geophysical signals emanating from complex geologic sources. We examined the method's robustness and stability using complex synthetic data in the presence of different kinds of correlated noises. The important advantage of this HMC-based BNN approach is that it can discriminate accurately complex signals even in the presence of different kinds of noises that are encountered in many geologic situations. Our KTB data analysis using the method suggests that the method can accurately classify lithofacies boundaries. Comparative results confirm that the HMC-based model results corroborate favorably with existing results inferred from earlier geologic and geophysical studies. In addition to agreeing well with earlier findings, our model result suggests the presence of finer bed boundaries that were missed in previous studies. The presence of finer structures seems to have geologic significance for understanding the crustal inhomogeneity and structural discontinuities within the central European crust. Thus, our main contribution has been to demonstrate how our method could be used to generate boundaries of lithofacies from well-log signals decoded from the borehole. The method could be further exploited for modeling different kinds of geologic and geophysical well-log signals.

ACKNOWLEDGMENTS

We are deeply indebted to the editors and reviewers for their critical review of our manuscript and painstaking editorial corrections. We are also grateful to V. P. Dimri, Director, NGRI; and Archana Bhattacharyya, Director, Indian Institute of Geomagnetism (IIG), Navi-Mumbai, for their support and kind permission to publish this work. S. Maiti expresses sincere thanks to the Department of Science and Technology (DST), Government of India, for support of the research work and senior colleagues S. G. Gokaran and G. Gupta, IIG, Navi-Mumbai, for their support. We are also thankful to Hans-Joachim Kumpel for providing the KTB borehole data.

REFERENCES

Berckhemer, H., A. Rauen, H. Winter, H. Kern, A. Kontny, M. Lienert, G. Nover, J. Pohl, T. Popp, A. Schult, J. Zinke, and H. C. Soffel, 1997, Petrophysical properties of the 9-km deep crustal selection at KTB: Journal of Geophysical Research, **102**, B8, 18337–18361.

Bishop, C. M., 1995, Neural networks for pattern recognition: Oxford University Press.

Busch, J. M., W. G. Fortney, and L. N. Berry, 1987, Determination of lithology from well logs by statistical analysis: Society of Petroleum Engineers Formation Evaluation, **2**, 412–418.

Cybenko, G., 1989, Approximation by superpositions of sigmoidal function: Mathematics Control and Signals Systems, **2**, 303–314.

Delfiner, P., O. Peyret, and O. Serra, 1987, Automatic determination of lithology from well logs: Society of Petroleum Engineers Formation Evaluation, **2**, 303–310.

Devilee, R. J. R., A. Curtis, and K. Roy-Chowdhury, 1999, An efficient probabilistic neural network approach to solving inverse problems: Inverting surface wave velocities for Eurasian crustal thickness: Journal of Geophysical Research, **104**, 28841–28856.

Duane, S., A. D. Kennedy, B. Pendleton, and D. Roweth, 1987, Hybrid Monte Carlo: Physics Letters B, **195**, 216–222.

Emmermann, R., and J. Lauterjung, 1997, The German Continental Deep Drilling Program KTB: Overview and major results: Journal of Geophysical Research, **102**, 18179–18201.

Franke, W., 1989, The geological framework of the KTB drill site, *in* R. Emmermann and J. Wohlenberg, eds., The German Continental Deep Drilling Program (KTB): Springer-Verlag Berlin, 38–54.

Fuller, W. A., 1976, Introduction to the statistical time series: John Wiley & Sons, Inc.

Gassaway, G. R., D. R. Miller, L. E. Bennett, R. A. Brown, M. Rapp, and V. Nelson, 1989, Amplitude variations with offset: Fundamentals and case histories: SEG.

Hastings, W. K., 1970, Monte Carlo sampling methods using Markov chain and their applications: Biometrika, **57**, 97–109.

Khan, M. S., and P. Coulibaly, 2006, Bayesian neural network for rainfall-runoff modeling: Water Resources Research, **42**, W07409.

Lampinen, J., and A. Vehtari, 2001, Bayesian approach for neural networks-review and case studies: Neural Networks, **14**, 7–24.

Leonardi, S., and H. Kumpel, 1998, Variability of geophysical log data and signature of crustal heterogeneities at the KTB: Geophysical Journal International, **135**, 964–974.

———, 1999, Fractal variability in super deep borehole-implications for the signature of crustal heterogeneities: Tectonophysics, **301**, 173–181.

MacKay, D. J. C., 1992, A practical Bayesian framework for back-propagation networks: Neural Computation, **4**, 448–472.

Maiti, S., and R. K. Tiwari, 2005, An automatic method for detecting litholog boundary using Walsh transform, A case study from KTB borehole: Computers and Geosciences, **31**, 949–955.

———, 2009, A hybrid Monte Carlo method based artificial neural networks approach for rock boundaries identification: A case study from KTB borehole: Pure and Applied Geophysics, **166**, 2059–2090, doi: 10.1007/s00024-009-0533-y.

Maiti, S., R. K. Tiwari, and H. J. Kumpel, 2007, Neural network modeling and classification of lithofacies using well log data: A case study from KTB borehole site: Geophysical Journal International, **169**, 733–746.

May, R., 1976, Simple mathematical models with very complicated dynamics: Nature, **261**, 459–467.

Meier, U., A. Curtis, and J. Trampert, 2007, Global crustal thickness from neural network inversion of surface wave data: Geophysical Journal International, **169**, 706–722.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, 1953, Equations of state calculations by fast computing machines: Journal of Chemical Physics, **21**, 1087–1091.

Nabney, I. T., 2004, Netlab algorithms for pattern recognition: Springer-Verlag New York.

Neal, R. M., 1996, Bayesian learning for neural networks: Springer-Verlag New York.

O'Brien, P., J. Duyster, G. Grauert, W. Schreyer, B. Stockhert, and K. Weber, 1997, Crustal evolution of the KTB drill site: From the oldest relics to the late Hercynian granites: Journal of Geophysical Research, **102**, B8, 18203–18220.

Pechnig, P., S. Haverkamp, J. Wohlenberg, G. Zimmermann, and H. Burkhardt, 1997, Integrated interpretation in the German Continental Deep Drilling Program: Lithology, porosity, and fracture zones: Journal of Geophysical Research, **102**, 18363–18390.

Pickett, G. R., 1963, Acoustic character logs and their application in formation evaluation: Journal of Petroleum Technology, **15**, 659–667.

Poulton, M., ed., 2001, Computational neural networks for geophysical data processing: Pergamon Press.

Rabben, T. E., H. Tjelmeland, and B. Ursin, 2008, Non-linear Bayesian joint inversion of seismic reflection coefficients: Geophysical Journal International, **173**, 265–280.

Raiche, A., 1991, A pattern recognition approach to geophysical inversion using neural nets: Geophysical Journal International, **105**, 629–648.

Rogers, S. J., J. H. Fang, C. L. Karr, and D. A. Stanley, 1992, Determination of lithology from well logs using a neural network: AAPG Bulletin, **76**, 731–739.

Rumelhart, D. E., G. E. Hinton, and R. J. Williams, 1986, Learning representations by back-propagating errors: Nature, **323**, 533–536.

Sambridge, M., and K. Mosegaard, 2002, Monte Carlo methods in geophysical inverse problems: Reviews of Geophysics, **40**, no. 3, 1–29.

Tarantola, A., 1987, Inverse problem theory: Elsevier.

van der Bann, M., and C. Jutten, 2000, Neural networks in geophysical applications: Geophysics, **65**, 1032–1047.

Wolff, M., and J. Pelissier-Combescure, 1982, FACIOLOG: Automatic electrofacies determination: 23rd Annual Symposium, Society of Petrophysicists and Well Log Analysts, Paper FF.