Pure appl. geophys. 166 (2009) 2059–2090
0033–4553/09/122059–32
DOI 10.1007/s00024-009-0533-y

Pure and Applied Geophysics

A Hybrid Monte Carlo Method Based Artificial Neural Networks Approach for Rock Boundaries Identification: A Case Study from the KTB Bore Hole

Saumen Maiti[1] and R. K. Tiwari[2]

Abstract—Identification of rock boundaries and structural features from well log response is a fundamental problem in geological field studies. However, in a complex geologic situation, such as in the presence of crystalline rocks where metamorphisms lead to facies changes, it is not easy to discern accurate information from well log data using conventional artificial neural network (ANN) methods. Moreover inferences drawn by such methods are also found to be ambiguous because of the strong overlapping of well log signals, which are generally tainted with deceptive noise. Here, we have developed an alternative ANN approach based on Bayesian statistics using the concept of Hybrid Monte Carlo (HMC)/Markov Chain Monte Carlo (MCMC) inversion scheme for modeling the German Continental Deep Drilling Program (KTB) well log data. MCMC algorithm draws an independent and identically distributed (i.i.d) sample by Markov Chain simulation technique from posterior probability distribution using the principle of statistical mechanics in Hamiltonian dynamics. In this algorithm, each trajectory is updated by approximating the Hamiltonian differential equations through a leapfrog discrimination scheme. We examined the stability and efficiency of the HMC-based approach on "noisy" data assorted with different levels of colored noise. We also perform uncertainty analysis by estimating standard deviation (STD) error map of a *posteriori* covariance matrix at the network output of three types of lithofacies over the entire length of the litho section of KTB. Our analyses demonstrate that the HMC-based approach renders robust means for classification of complex lithofacies successions from the KTB borehole noisy signals, and hence may provide a useful guide for understanding the crustal inhomogeneity and structural discontinuity in many other tectonically critical and complex regions.

Key words: KTB boreholes, Hybrid Monte Carlo (HMC), lithofacies, well log, petrophysics, and uncertainty analysis.

1. Introduction

Information concerning the composition of rocks and structural features inside the earth is obtained by making continuous measurements of geophysical properties with depth using well logging methods. Such well log data have been extensively used for exploration of oil and gas, mineral mining, geothermal energy and, radio-active waste deposits (Pechnig *et al*., 1997). It is known that compositional and textural

[1] Indian Institute of Geomagnetism, Navi-Mumbai 410218, India. E-mail: saumen_maiti2002@yahoo.co.in
[2] National Geophysical Research Institute, Hyderabad 500007, India.

characteristics (e.g., grain shape, size, and organization or rock structure) are important factors for rock-facies classification. However, in complex geologic situations, metamorphism plays a significant role in producing varying composition and structural variations of crystalline rocks. Several factors, such as pore fluid, effective pressure, fluid saturation, pore shape, grain size, and shape, etc. affect geophysical well log signals (Pechnig *et al*., 1997; Maiti and Tiwari, 2008; Maiti *et al*., 2007; Maiti and Tiwari, 2005). Classification of lithology/lithofacies boundary from geophysical well log data is, therefore, a complex and nonlinear geophysical problem. To solve the problem, often a misfit function is defined to relate between model (here, lithology/lithofacies) and observed well log response (data). Usually, the misfit function is optimized through an iterative gradient-based process generally using some form of Newton's method. However, for a nonlinear problem the misfit function has many local minima. Optimizing/minimizing the misfit function by a gradient-based matrix inversion method results in some inescapable local minima. Sometimes a regularization term with an error function is used to prevent overtraining. In that case, physical interpretation may be ambiguous because of the underestimation. This problem could be avoided by using the Bayesian paradigm approach with a sampling-based global inversion method. It is, therefore, imperative to apply such a robust nonlinear global optimization scheme which could evade these difficulties.

The Bayesian inference (Tarantola, 1987; Mosegaard and Tarantola, 1995; Sambridge and Mosegaard, 2002) based on Bayes theory to approximate *posterior* probability distribution via data likelihood and prior information using the Monte Carlo algorithm has been proved to be very useful because of the non-unique solutions of a geophysical inverse problem. Monte Carlo inversion techniques were first used by earth scientists more than 30 years ago. Since then the method has been applied to a wide range of geophysical problems. The phrase "Monte Carlo Method" was first coined by Metropolis and Ulam (1949). Before also these concepts were used in numerical experiments to determine the value of π by an injured officer during the American Civil War (Sambridge and Mosegaard, 2002). The classic paper by Backus and Gilbert (1967) established the foundation of the geophysical inverse theory. Another remarkable work by Hastings (1970) drew wide attention to the Markov Chain simulation for its application for numerical integration by the Monte Carlo method. The practical use of the sampling-based inversion scheme (e.g., HMC/MCMC for neural network training was first described by MacKay, 1992 and Neal 1993). The HMC algorithm employed here is based on the work of MacKay (1992) and Neal (1993). Lampinen and Vehtari (2001) have presented a thorough review of the subject, which introduced the HMC algorithm for drawing samples from posterior probability distribution. Here, we apply the HMC based neural network optimization scheme to infer lithofacies boundaries from superposing/overlapping well log signals from KTB in anticipation that the results would provide a better understanding of the crustal inhomogeneity and structural discontinuity in the context of geodynamical collision in the KTB region.

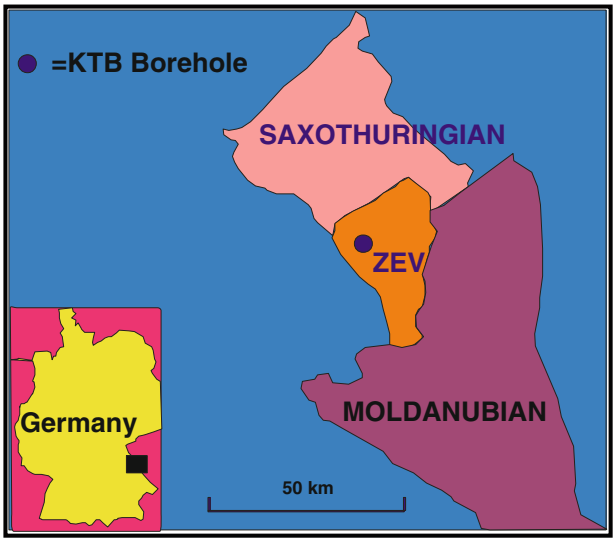


Figure 1
Location map of the German Continental Deep Drilling Program (KTB) boreholes, Saxothuringian, Moldanubian, and ZEV (Zone of Erbendorf-Vohenstrauss) represent the main geological units of the region (after LEONARDI and KÜMPEL, 1999).

2. *KTB Geology*

The drill site of the KTB lies within the Zone of Erbendorf-Vohenstrauss (ZEV). The rocks types mainly comprised of three main facies units: paragneisses, metabasites and alternations of gneiss-amphibolites, with a minor occurrence of marbles, calcsilicates, orthogeneisses, lamprophyres and diorites. Franconian Lineament, a thrust fault marking the boundary between the crystalline outcrop of the Bohemian Massif and the Permo-Mesozoic cover of southeast Germany, is situated to the west of the KTB drill hole (PECHNIG *et al.*, 1997). The lithologic and metamorphic characteristics, especially the abundant metabasites and the medium pressure metamorphism, suggest that the Zone of Erbendorf-Vohenstrauss (ZEV) is a part of the Bohemiam (FRANKE, 1989) (Figs. 1 and 2). The internal structure of the Zone of Erbendorf-Vohenstrauss (ZEV) is dominated by a steeply dipping foliation and a NW-SE striking stretching lineation which is parallel to the axes of at least two generations of folds generated during amphibolities facies metamorphism. The stretching lineation is well developed in equigranular, fine-grained, high temperature mylionite, amphibolites and paragneisses. Along the boundary of the Zone of Erbendorf-Vohenstrauss (ZEV) with the surrounding imbrication zone are amphibolites of oceanic affinity and isolated bodies of serpentinites (BERCKHEMER *et al.*, 1997; EMMERMANN and LAUTERJUNG, 1997; PECHNIG *et al.*, 1997; LEONARDI and KUMPEL, 1998, 1999). The detailed information about the

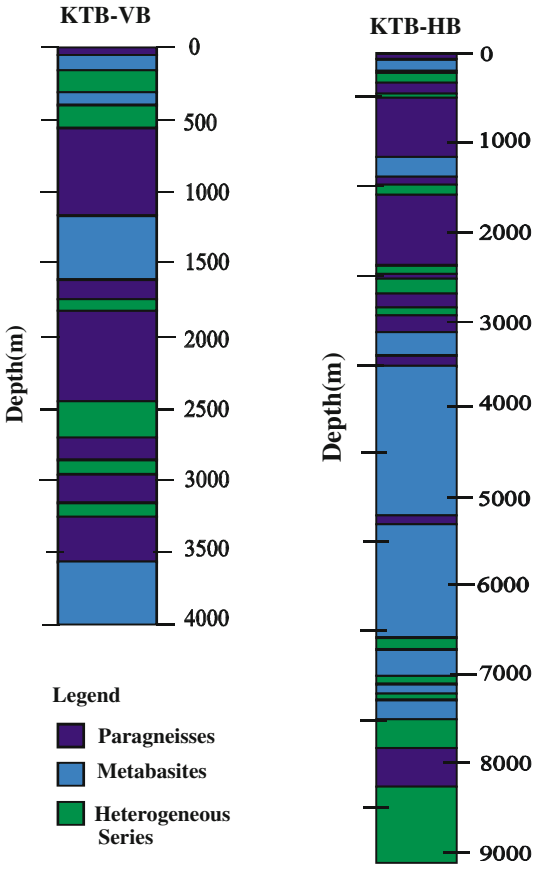


Figure 2
Reference diagram of litho facies section of the German Continental Deep Drilling Program (KTB). (left) pilot hole (KTB-VB), (right) main hole (KTB-HB) (after Emmermann and Lauterjung, 1997).

KTB data and its geophysical significance can be found in several earlier papers (Berckhemer *et al*., 1997; Pechnig *et al*., 1997; Emmermann and Lauterjung, 1997; Leonardi and Kumpel, 1998, 1999). We study here three types of well log data (viz. density, neutron porosity and gamma ray intensity) by using the HMC based algorithm for neural networks to constrain the lithofacies boundaries of the KTB (Maiti *et al*., 2007). The total depth of the main hole and pilot hole are 9101 m and 4000 m, respectively. The 3-D cross plot of density (g/cc), neutron porosity (%) and gamma ray (API) taken from selected portions of the KTB main hole and pilot hole are displayed in Figure 3, which apparently shows overlapping well log signal content indicating complex and nonlinear characteristics.

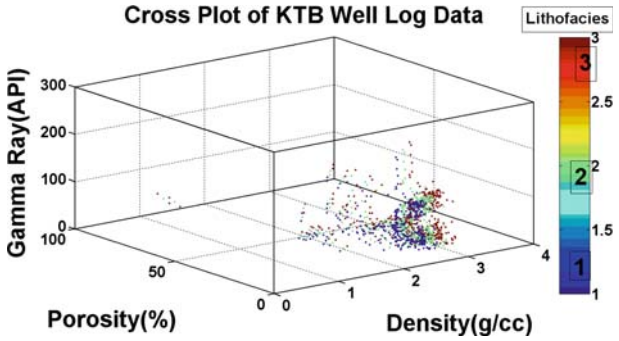


Figure 3
Cross plot of density (g/cc), porosity (%) and gamma ray intensity (API) of the German Continental Deep Drilling Program (KTB) well log data showing strong nonlinearity and very difficult to establish parametric boundary.

3. Bayesian Neural Network Approach

The forward relationship between model (here lithofacies) and the data (well log) can be given as

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

where x is data and d is model. The term f is a nonlinear function relating to the model space and data space and ε is error. It is noteworthy, to mention here that since f establishes a nonlinear functional relationship between the data space and model space, hence the direct estimation of the model is not possible. A common way of inverting the model d in equation (1) is via an iterative least-squares method. However, this approach does not provide uncertainty measures. To solve equation (1) in the Bayesian sense, we first considered here the total 702 representative realization samples (model/data pairs) from a finite data set $s = \{x_k, d_k\}_{k=1}^{N}$, conditioned upon the explicit limit of well log response (Table 1) (Maiti *et al.*, 2007). For this we have the following equation

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

where d is a desired model, f_{NN} is the output predicted by the network and w is the network weight parameter. In a conventional ANN approach, an error functions $E_S = 1/2 \sum_k^N (d_k - o_k(x_k; w_k)^2$ is measured in order to know how close the network output $o(x; w)$ is to the desired model d from the finite data set $\{x, d\}$. Often, regularization is included to modify the misfit function,

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

where, $E_R = 1/2 \sum_{i=1}^{R} w_i^2$ and R is the total number of weights and biases in the network. λ, μ are two controlling parameters known as hyper parameters. The forward functions

Table 1

A priori information on model parameter to generate forward model for neural network training indicating that gamma ray intensity value most crucial factor to categorize lithofacies unit in metamorphic area

Lithofacies unit	Density [g/cc]	Neutron porosity [%]	Gamma Ray Intensity [A.P.I]	Desired Output/binary code
Paragneisses	2.65–2.85	5–15	70–130	100
Metabasites	2.75–3.1	5–20	0–50	010
Heterogeneous Series	2.60–2.9	1–15	40–90 & 120–190	001

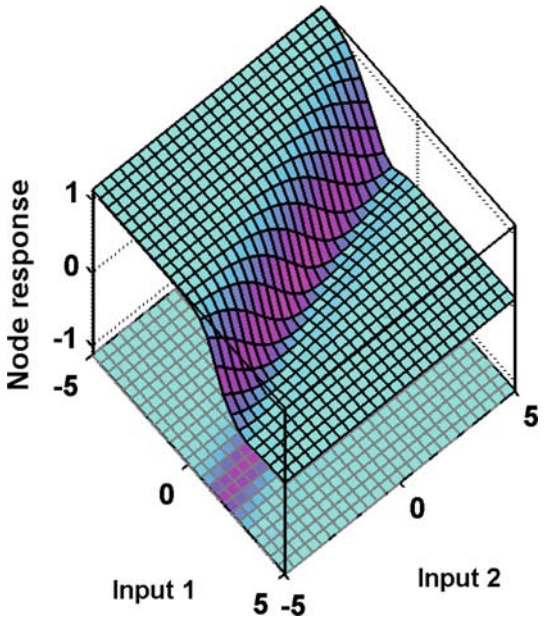


Figure 4
Output of a hidden node as a function of its input through sigmoid transfer functions.

used in the node are nonlinear tan sigmoids (Fig. 4) which eases to solve the nontrivial problems. In the traditional 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, given that the objective function is optimized.

In the Bayesian approach, a suitable prior probability distribution $P(w)$ of weights is considered before observing the data, instead of a single set of weights. Using the Bayes' rule, *a posteriori* probability distribution for the weights $P(w|s)$ can be given as (KHAN and COULIBALY, 2006),

$$P(w|s) = \frac{P(s|w)P(w)}{P(s)}, \tag{4}$$

where $P(s|w)$ and $P(s)$ are the data set likelihood function and normalization factor respectively. The denominator $P(s)$ is intractable, so direct estimation of posterior $P(w|s)$ is not possible. Using the rules of conditional probability, the probability distribution of outputs for a given input vector x can be written in the form (KHAN and COULIBALY, 2006),

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

The major problem in Bayesian computation is evaluating integrals for a *posteriori* probability density function of weights (equation (4)) and network output (equation (5)). Fortunately numerical methods based on MCMC sampling play an important role for evaluating posterior integrals. The equation (5) can be approximated as

$$P(d|x,s) = \frac{1}{N}\sum_{n=1}^{N} P(d|x,w_m), \tag{6}$$

where $\{w_m\}$ represents an MCMC sample of weight vectors obtained from the probability distribution $P(w|s)$, N is the number of points and w samples from $P(w|s)$ (Khan and Coulibaly 2006).

3.1. Markov Chain Monte Carlo (MCMC)/Hybrid Monte Carlo (HMC)

It is well known that a plain gradient descent algorithm can become trapped at shallow local minima. One way of overcoming this problem is to define the error surface in terms of a Hamiltonian statistical mechanics that accounts for the approximation errors and the momentum term of each trajectory. This is the basis of the HMC/MCMC algorithm (DUANE *et al.*, 1987). In this algorithm, each trajectory is updated by approximating the Hamiltonian differential equations by a leapfrog discrimination scheme. The MCMC algorithm draws an independent and identically distributed (i.i.d) sample $\{w^{(i)}; i = 1, 2, \ldots, N\}$ from the target probability distribution $P(w|s)$ (BISHOP, 1995). Markov process forms a sequence of "state" to draw samples from posterior probability distribution. It is noted that the chain converges to $P(w|s)$ given enough space (BISHOP, 1995). The states are represented by a particle in the high dimensional network parameter space whose positions are defined by $q \in R^w$. Thus the equation (4) can be written in the form of $\pi(q) \propto \exp\{-E(q)\}$, where π is a generic symbol and $E(q)$ is the potential energy functions/cost function for the optimization problem (DUANE *et al.*, 1987). By introduction of momentum variables p with corresponding kinetic energy functions $V(p) = \frac{1}{2}\sum_{i=1}^{N} p_i^2$ to efficiently explore a large region of phase space by simulating the Hamiltonian dynamics in fictitious time. The particle "mass" and "Boltzmann constant k_B" can be re-scaled to unity. Full Hamiltonian energy function on a fictitious phase space (DUANE *et al.*, 1987),

$$H(q,p) = E(q) + V(p). \tag{7}$$

The canonical distribution of Hamiltonian is

$$\pi(q,p) = \frac{1}{Q_H}\exp\{-H(q,p)\}. \tag{8}$$

It is clear that if we sample (q, p) from the distribution $\pi(q,p) = \frac{1}{Q_H}\exp\{-H(q,p)\}$, then the marginal distribution of q is exactly the target distribution $\pi(q)$ (DUANE *et al.*, 1987; NEAL, 1993). Here, π is the generic symbol, q is the position and p is the momentum discussed before.

In practice, to simulate the Hamiltonian dynamics we need to discretize the equations of motion of particle in general. Discrimination of equations introduced some errors which destroy time reversibility and volume preservations principle which is required to employ the Metropolis algorithm (METROPOLIS and ULAM, 1949; METROPOLIS *et al.*, 1953). Fortunately the leapfrog discrimination process has the desired properties to preserve both. The leapfrog process updates positions and momentum coordinates (q_i, p_i) of particle in three basic steps (DUANE *et al.*, 1987; MACKAY, 1992; NEAL, 1993; BISHOP, 1995).

First, it takes a half step for momentum,

$$p_i(\tau + \theta/2) = p_i(\tau) + \frac{\theta}{2}\frac{\partial \log E}{\partial q_i}(q(\tau)). \tag{9}$$

Then it takes a full step for the positions,

$$q_i(\tau + \theta) = q_i(\tau) + \theta.p_i(\tau + \theta/2). \tag{10}$$

Finally it takes the other half step for the momentum,

$$p_i(\tau + \theta) = p_i(\tau + \theta/2) + \frac{\theta}{2}\frac{\partial \log E}{\partial q_i}(q(\tau + \theta)). \tag{11}$$

In above equations (9–11), θ is time step size, τ is time, q is position, p is momentum and E is potential energy. Leap-leapity–leap, that is why it is called leapfrog. At the end of the three steps we obtain an approximation to the values of position and momentum at time $\tau + \theta$ from their corresponding values at time τ (DUANE *et al.*, 1987; BISHOP, 1995; NEAL, 1993). As we can readily check by simple inspection, the leapfrog discrimination has the following necessary property (DUANE et al., 1987; MACKAY, 1992; NEAL, 1993; BISHOP, 1995).

1. It mostly preserves H, in fact to order $O(\theta^2)$.
2. It preserves volumes since the above mentioned are just shear transformations/ Liouville's theorem, i.e., conservation of phase-space volumes.
3. It is time reversible.

The practically pure Metropolis-Hastings algorithm (Metropolis *et al.*, 1953; Hastings, 1970) is proven to be very slow because the method makes no use of gradient information. Our proposed algorithm which is based on the HMC algorithm for sampling from target distribution, makes use of the gradient information. Our HMC algorithm is a sampling-based algorithm that takes into consideration certain gradient information. The algorithm follows the following sequence of steps once a step size θ and the number of iterations L have been decided upon. First our algorithm randomly chooses a direction τ (Bishop, 1995). The τ can be either -1 or $+1$ with the probability 0.5, simulating the dynamics forward or backward in time. Following the theory of statistical mechanics of Hamiltonian dynamics, the transition probability matrix satisfies microscopic reversibility, which means that the probability of these 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. Then the algorithm carries out the iterations starting with the current state $[q, p] = [(q(0), p(0)]$ of energy H (Duane *et al.*, 1987; MacKay, 1992; Neal, 1993; Bishop, 1995). The momentum term p is randomly evaluated at each step. The algorithm performs L steps with a step size of θ resulting in the candidate state, $[w^*, p^*]$ with energy H^* (Duane *et al.*, 1987). The candidate state is accepted with usual metropolis probability of acceptance, $\min\{1, \exp[-(H^* - H)]\}$ where $H(.)$ is the Hamiltonian energy (Duane *et al.*, 1987; MacKay, 1992; Neal, 1993; Bishop, 1995). If the candidate state is rejected then the new state will be the old state. These three steps, in essence, describe how the sampling is done from posterior probability distribution of the network parameter so that the summation of equation (6) can be accomplished to obtain the posterior probability distribution and thus allow the optimization of the network (Neal, 1993). The momentum term p can be randomly generated or it can be changed dynamically at each step within which there are different ways of doing this (Duane *et al.*, 1987; MacKay, 1992; Neal, 1993; Bishop, 1995).

The sets of weights are thus selected or rejected according to the three steps mentioned above and the numbers of samples that are retained are the number of weights retained. For each set of weights there is a corresponding neural network output.

4. Model Setup

Normalization of the raw data (input/output), before presenting it to the network is crucial to avoid saturation of the network. Hence, we scale all the input/output pair values between 0 and 1 [-1 and $+1$] by using a simple linear transformation algorithm (Poulton, 2001). Normalized input = 2 $\times$ (input-minimum input)/(maximum input-minimum input) $-$ 1. The initialization of the model parameters is performed by a distribution of model parameters. The initial values of model parameters (synaptic weight and biases) of MLP (Multi-Layer Perceptron) are formed by Gaussian prior probability distribution of zero mean and inverse variance λ (also known as regularization coefficient or prior hyper-parameter). Gaussian prior probability distribution has been preferred here

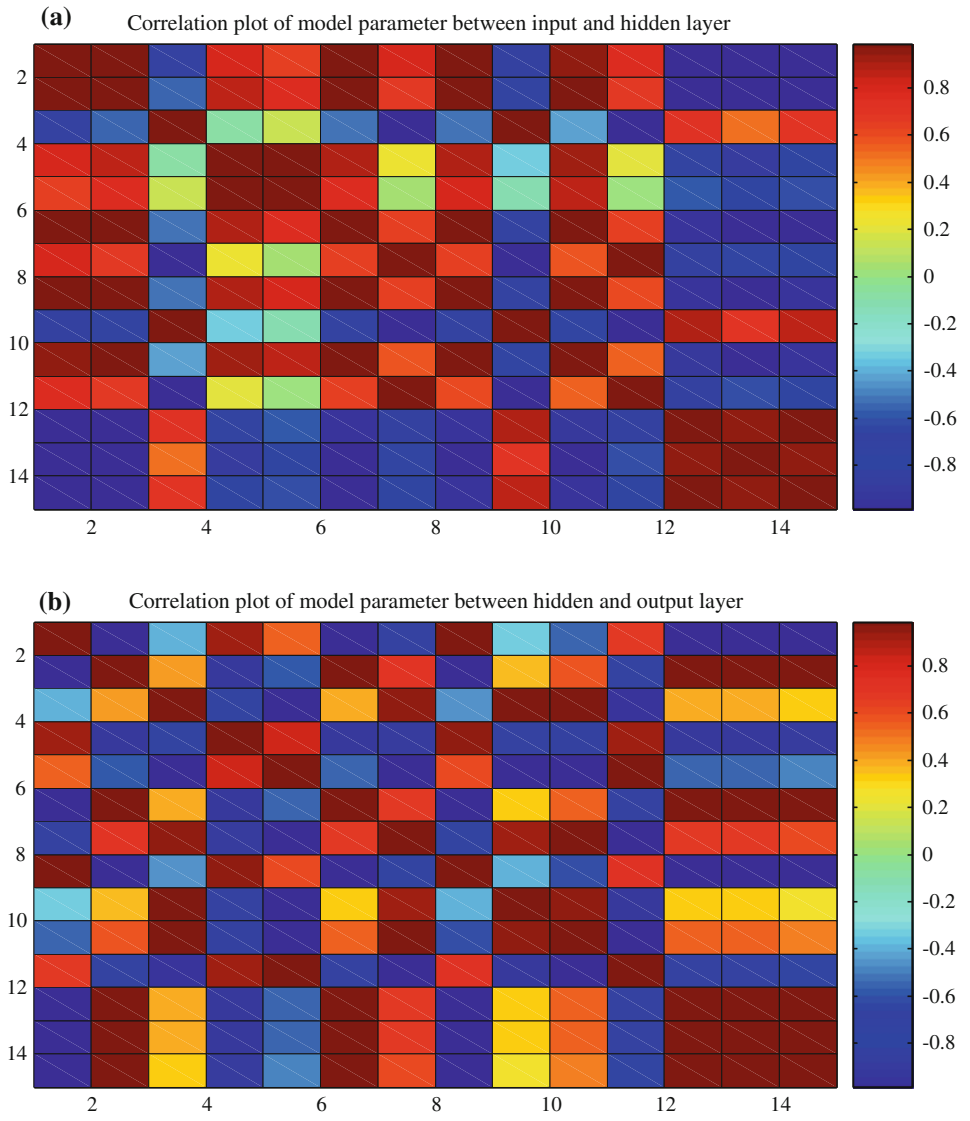


Figure 5
Correlation plot of each model parameter with all others between (a) input and hidden layer, (b) hidden and output layer.

to favor small values for the network weights because a network with large weights will usually give rise to a mapping with large curvature (NABNEY, 2004). Moreover, Gaussian prior also provides computational simplicity. For prior hyper-parameter $\lambda = 0.02$, a single initial value has been considered for hidden layer and output layer weights. To define an objective function in a Bayesian framework, our program requires an error

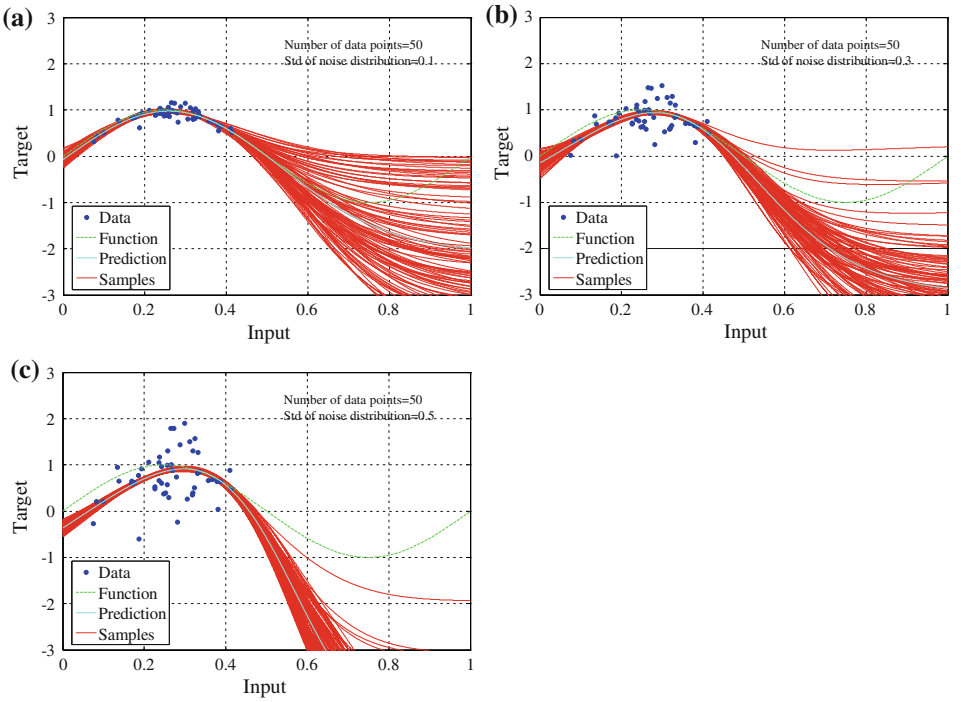


Figure 6

The graph shows the underlying function, 100 samples generated by Hybrid Monte Carlo simulation from the function given by the posterior probability distribution of the weights, and the average prediction weighted by the posterior probabilities using the 50 number of data points when standard deviations, (std) of noise distributions are (a) 0.1 (b) 0.3 (c) 0.5.

model for the data likelihood. It is assumed that target data are formed from a smooth function with additive zero mean Gaussian noise. Accordingly, hyper-parameter $\mu = 50$ is estimated for both hidden and output layer weights. After defining prior and likelihood functions, posterior has been estimated by using the Bayes' rule.

We have examined a different run with a different model initiation parameter and a different network parameter and presented the most consistent and optimized one. During the run, thousands of inverted network parameters are obtained. Providing enough space and time and very small step size (theoretically equal to 0) setting, our HMC-based algorithm draws samples from posteriors probability distributions but requires enough time. We believe that these values may not be the unique ones, rather problem dependent. However, the chosen parameters (e.g., weight decay, coefficient of a data error, etc.) enabled the fast and efficient training for the present case. The model parameters considered for the present work are: (i) number of input node = 3. (ii) Number of hidden node = 15, (iii) number of output node = 3, (iv) number of steps in a trajectory = 100, (v) number of Monte Carlo samples return = 100, (vi) steps size =

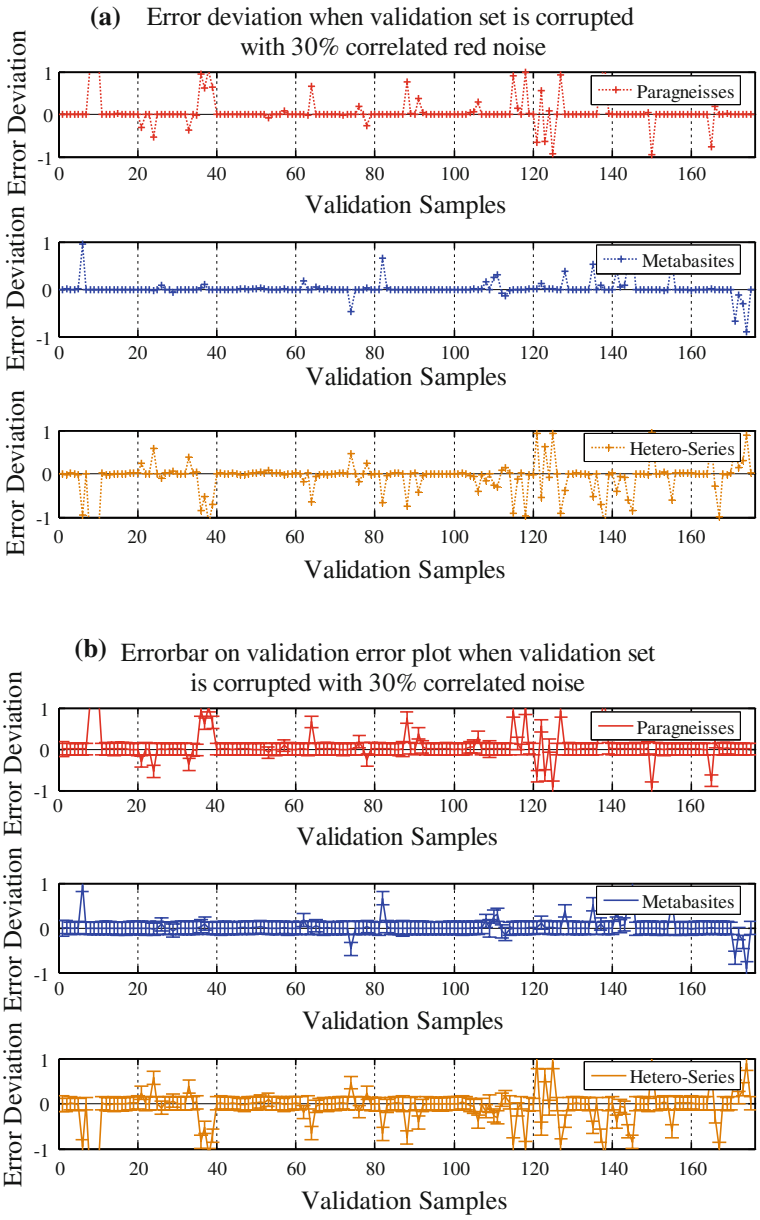


Figure 7

Error deviation and error bar map of validation and test data pertaining to paragneisses, metabasites and heterogeneous series 7(a)–(d), when the input generalization set is corrupted with 30% red noise 7(e)–(h). When the input generalization set is corrupted with 50% red noise. Error bar defines 90% confidence limit.

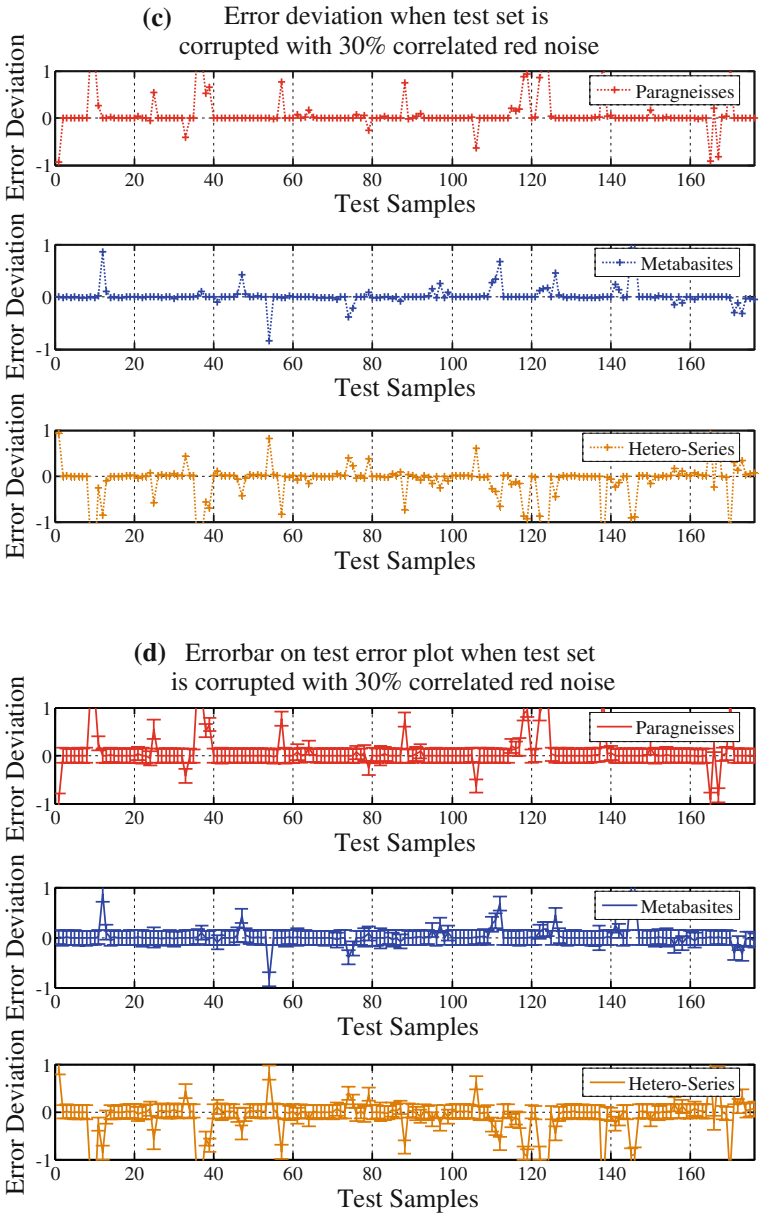
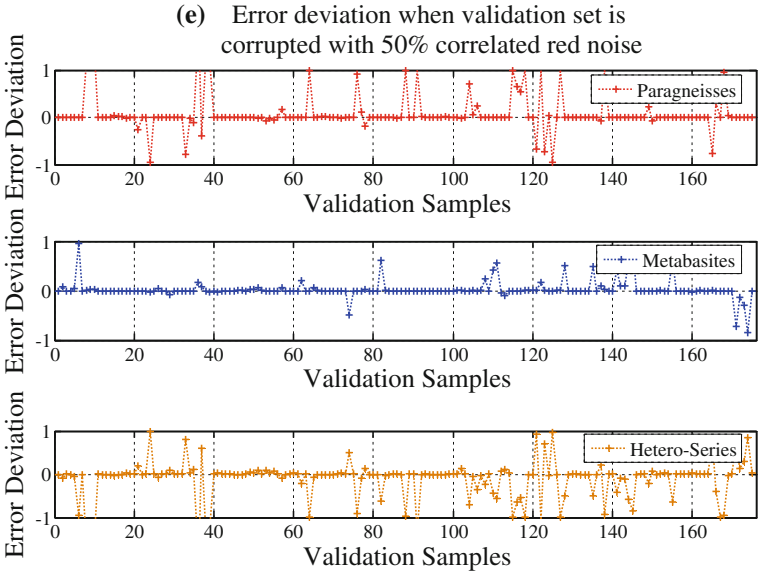


Figure 7
contd.

0.002, (vii) coefficient of weight decay prior = 0.02, (viii) coefficient of data error = 50. We determine correlation coefficients between each network weights and all other input to the hidden layer and hidden layer to output layer, which describes the quality of the training results (Figs. 5a and 5b).



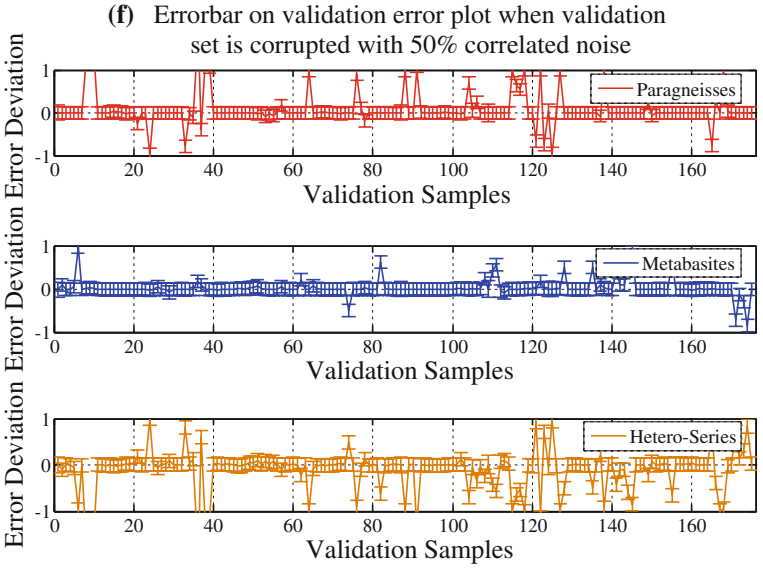


Figure 7
contd.

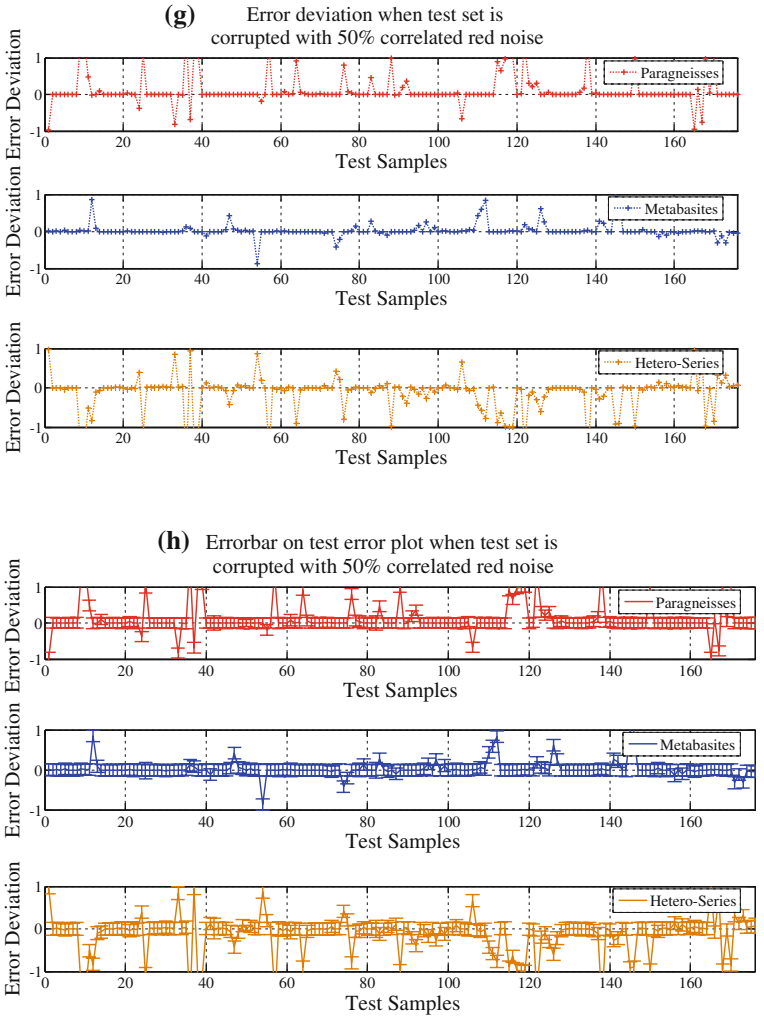


Figure 7
contd.

5. *Synthetic Example*

In the following example we performed a linear regression analysis to predict a sine waveform using an HMC simulation algorithm. Figures 6a-c show the underlying noise-free sine function with the 100 samples produced by the BNN. The experiment shows that BNN with HMC estimates true functions where the data samples are more condensed and less noisy. The uncertainty is more for predicting underlying sine functions where there are no data samples or data tainted with a high level of noise.

Table 2

Showing percentage of accuracy while validation and test data set are corrupted with different levels of red noise

Red Noise Level	Percentage of Accuracy in Generalization Data Set						
	Validation Data set			Test Data Set			Average Stability ± 5 % error limits
	Paragneisses	Metabasites	Heterogeneous Series	Paragneisses	Metabasites	Heterogeneous Series	
0%	86.29%	87.43%	77.14%	86.36%	84.09%	72.73%	82.34%
10%	86.29%	87.43%	73.71%	86.36%	84.09%	71.59%	81.57%
20%	83.43%	86.29%	70.29%	83.52%	83.52%	68.18%	79.20%
30%	81.71%	86.29%	68.00%	81.25%	83.52%	66.48%	77.87%
40%	80.00%	85.71%	65.71%	77.27%	82.39%	61.93%	75.5%
50%	74.29%	85.71%	62.29%	74.43%	81.25%	57.95%	72.65%

6. Sensitivity Analysis

It is well known that geophysical observations are corrupted with noise. To our knowledge, uncorrelated Gaussian white noise may be filtered using some modern filtering techniques. However, correlated noise/red noise is deceptive in nature often found to be mixed with observations. It is quite difficult to remove such noise from the data, even using powerful filtering techniques. The best way to deal with such problems is first to test the proposed algorithm on synthetic data in the presence of correlated red noise. Accordingly we have randomly divided the entire synthetic data into three subsets: training (50% of the total), validations set (25% of the total) and test set (25% of the total) and added different levels of correlated noise to the synthetic training sample following Maiti *et al*. (2007) and Fuller (1976). After network training, we have performed the sensitivity analysis of the network by HMC-based algorithm which shows very stable (up to 40% noise) and comparable results (Fig. 7).

Further we perform an uncertainty analysis by showing ± unit variance on the error deviation curve on validations and test set data which was corrupted with different levels of correlated red noise (Table 2). We note that most of the center of the vertical error bar passes through the zero line which shows a 90% confidence level (Fig. 7).

7. Regression Analysis

Linear regression analysis provides accuracy of the overall performance of the network. Our regression analysis code returns three parameters. The first two parameters *u* and *v,* respectively correspond to the slope and the intercept on the y axis. Figures 8a,b and c show the network outputs, which are plotted against the targets (T) and are shown as open circles. A dashed line indicates the best linear fit (slope 1 and y intercept 0). The solid line in the above figure shows the perfect fit (output equal to target). The third

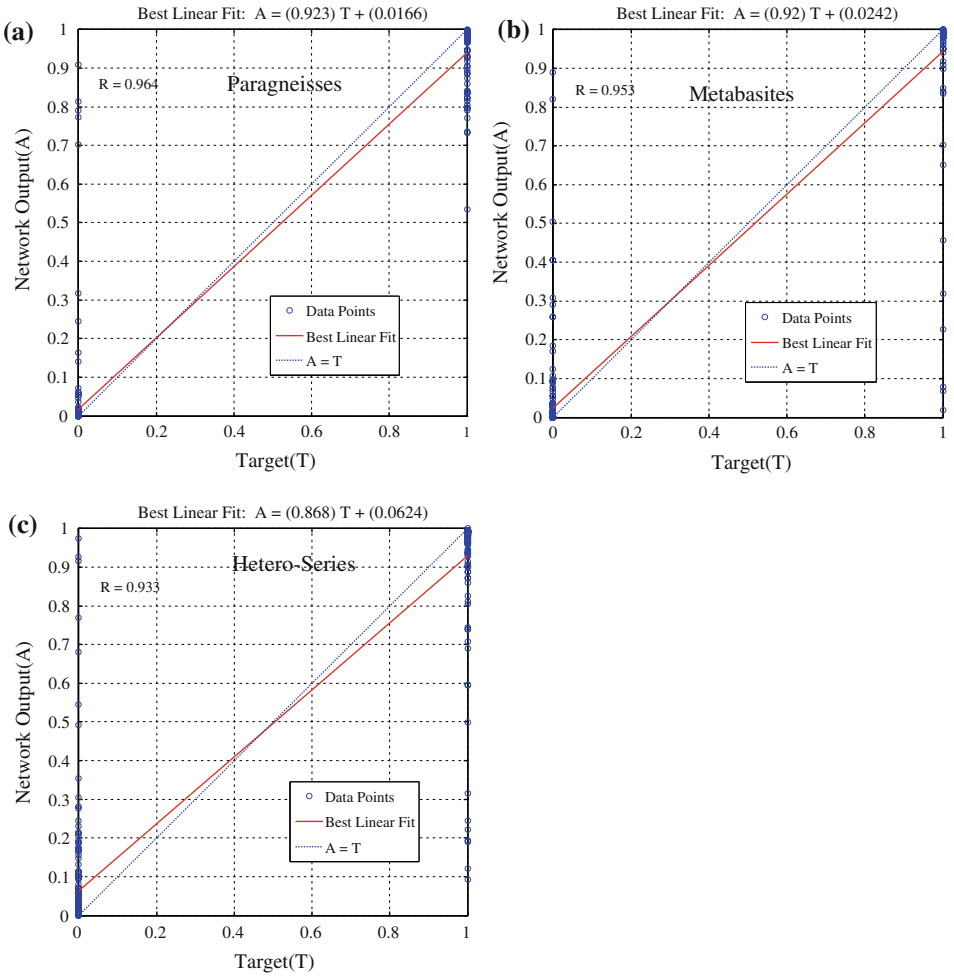


Figure 8

(a) Linear regression analysis of the total set of data corresponding to paragneisses. Correlation coefficient (R) between target (T) and network output (A) is 0.96, slope is 0.92, and y intercept of the best linear regression relating targets to network is 0.0166. The network outputs are plotted versus the targets as open circles. A dashed line indicates the best linear fit. The perfect fit (output equal to targets) is indicated by the solid line. (b) Same as (a) for metabasites yielding a correlation coefficient 0.95, slope 0.92, and y intercept 0.0242. (c) Same as (a) for heterogeneous series yielding a correlation coefficient 0.93, slope 0.868, and y intercept 0.0624.

parameter R is a correlation coefficient between the network outputs and the targets. The parameter "R" is a measure of how well trained HMC based network predicted values are related to the target values. The value of R closer to 1 implies that there is perfect correlation between targets and outputs. The results of the linear regression analysis for the total set of the data sets corresponding to paragneisses, metabasites and heterogeneous series are given in Table 3 and are displayed in Figures 8a-c. The results indicate that

Figure 9 ►

(a) Comparison of the maximum a *posteriori* geological section (MAPGS) obtained by the Bayesian Neural Network (BNN) with Hybrid Monte Carlo (HMC) approach with maximum likelihood geological section (MLGS) obtained by super self adaptive back propagation (SSABP) neural network, and published litho-facies subsection of German Continental Deep Drilling Program (KTB) pilot hole (KTB-VB) (right after EMMERMANN and LAUTERJUNG, 1997) and standard deviation (STD) error map estimated at network output by Bayesian Neural Network (BNN) approach at depth interval of 0–500 m. In this interval 0–28 m data is not available. (b) – (h) Same for depth range of 500–1000 m, …, 3500–4000 m in KTB pilot hole (KTB-VB).

correlation coefficient R and intercept y are respectively close to 1, 1 and 0. This finding thus justifies the application of the HMC-based algorithm for the analysis of a voluminous amount of noisy bore hole records.

8. *Uncertainty Analysis*

The data covariance matrix defines the uncertainties in the data. The error bars ± represent unit standard deviations derived from *a posteriori* covariance matrix. Figures 7, 9 and 10 show the result of the uncertainty analysis. The above figures show a plot of the square roots of the main diagonal elements of *a posteriori* covariance matrix. The Bayesian approach using the HMC-based algorithm provides the mean solutions of the posterior probability distribution. Along with the mean solutions, we also obtain the variability limit of the network predictions which show a 90% confidence interval (Figs. 7, 9 and 10). We estimated the standard deviation of output error maps for all three types of lithofacies over the entire KTB pilot hole (28–4000 m) and main hole (4000–7000 m) litho-section which show the confidence interval of prediction of the data of our HMC-based Bayesian neural network. The minimum, maximum and average values of STD of output error maps corresponding to the depths in the context of uncertainty analysis is discussed in the results and discussions sections.

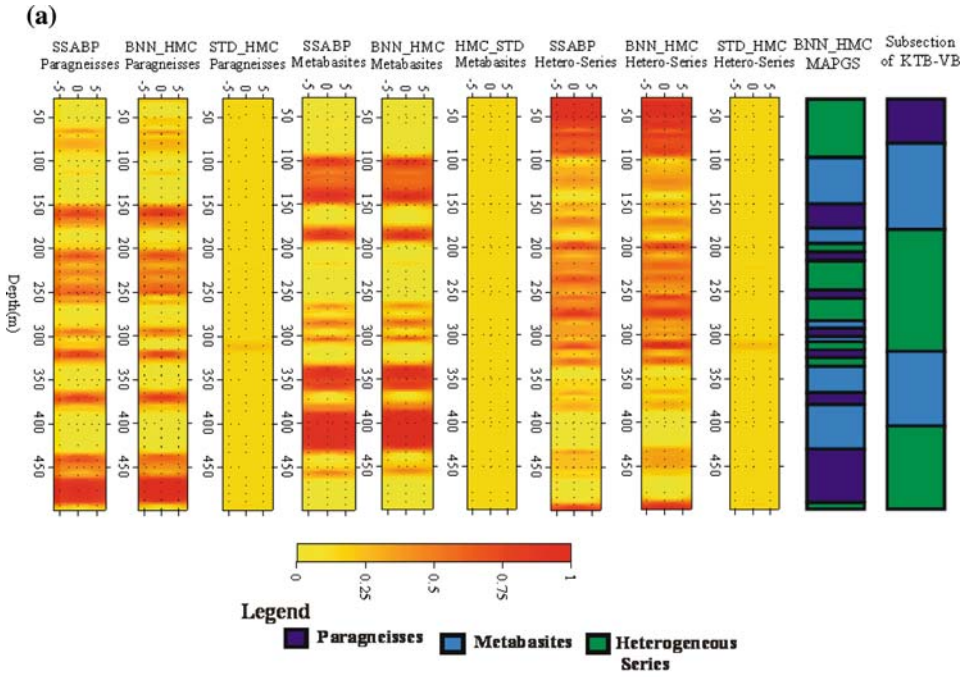
9. *Real Data Analysis*

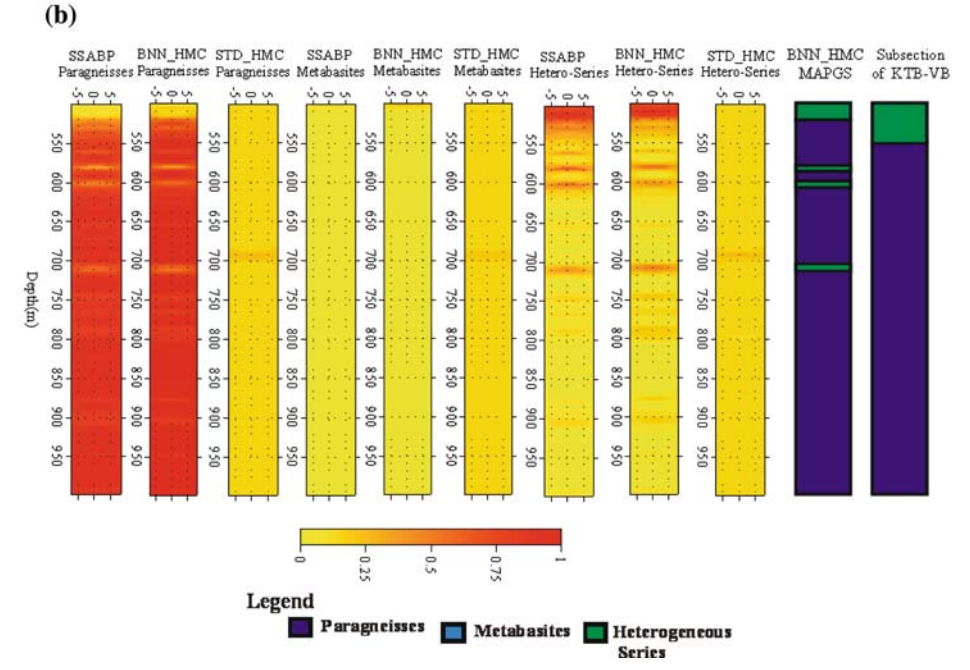
We performed the Principle Component Analysis (PCA) on the real KTB pilot bore hole (KTB-VB) from a depth of 28 m to 4000 m and the main bore hole (KTB-HB) from

Table 3

Illustrating the results of linear regression analysis of three major litho types

Litho types	Correlation coefficient (R) between target (T) and network output (A)	Slope (u)	Y-intercept of the best linear regression relating targets to network output (v)
Paragneisses Class	0.96	0.92	0.01
Metabasites Class	0.95	0.92	0.02
Heterogeneous Series Class	0.93	0.86	0.06


(a)
SSABP Paragneisses
BNN_HMC Paragneisses
STD_HMC Paragneisses
SSABP Metabasites
BNN_HMC Metabasites
HMC_STD Metabasites
SSABP Hetero-Series
BNN_HMC Hetero-Series
STD_HMC Hetero-Series
BNN_HMC MAPGS
Subsection of KTB-VB
Depth(m)
Legend
Paragneisses
Metabasites
Heterogeneous Series


(b)
SSABP Paragneisses
BNN_HMC Paragneisses
STD_HMC Paragneisses
SSABP Metabasites
BNN_HMC Metabasites
STD_HMC Metabasites
SSABP Hetero-Series
BNN_HMC Hetero-Series
STD_HMC Hetero-Series
BNN_HMC MAPGS
Subsection of KTB-VB
Depth(m)
Legend
Paragneisses
Metabasites
Heterogeneous Series

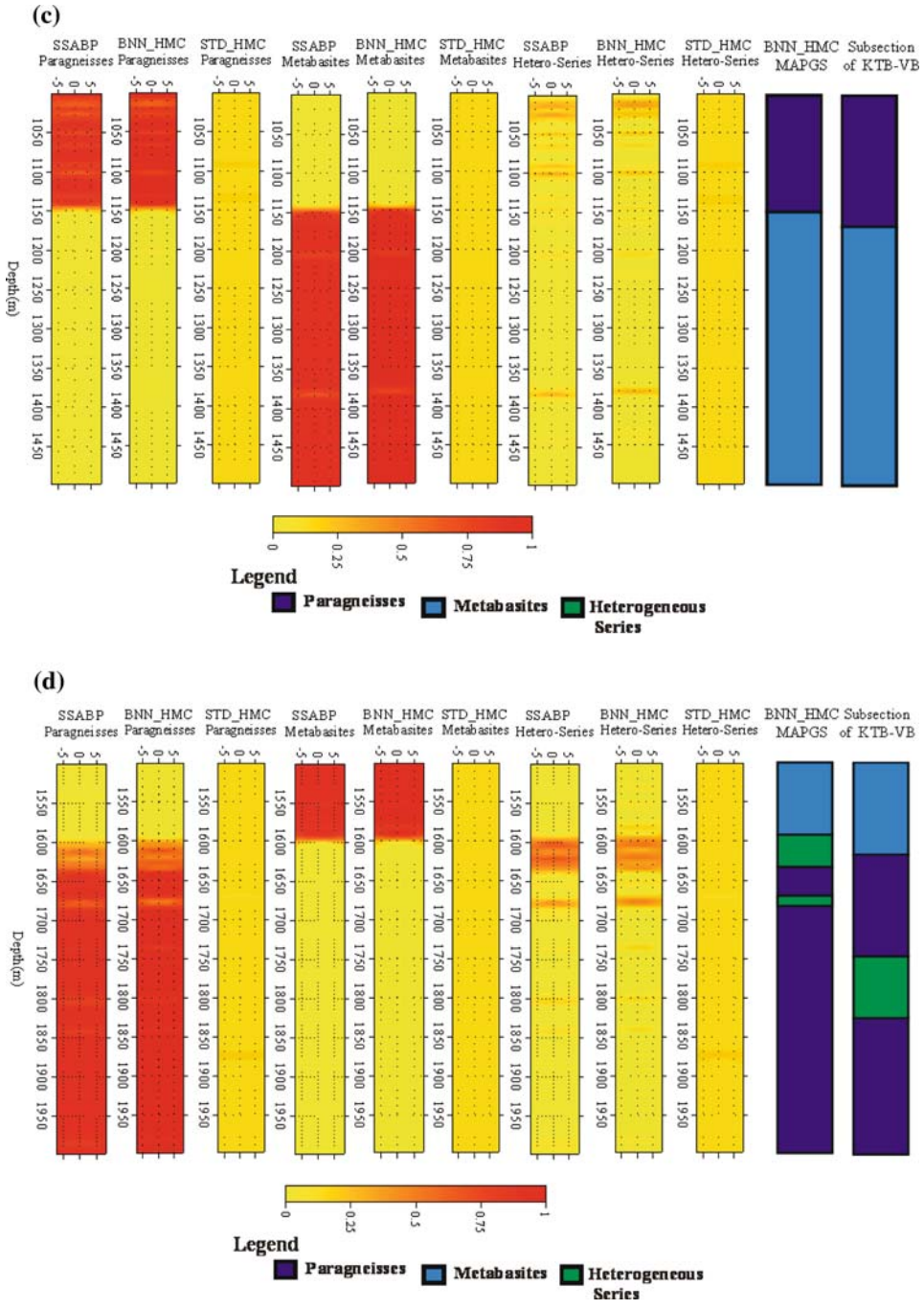


Figure 9
contd.

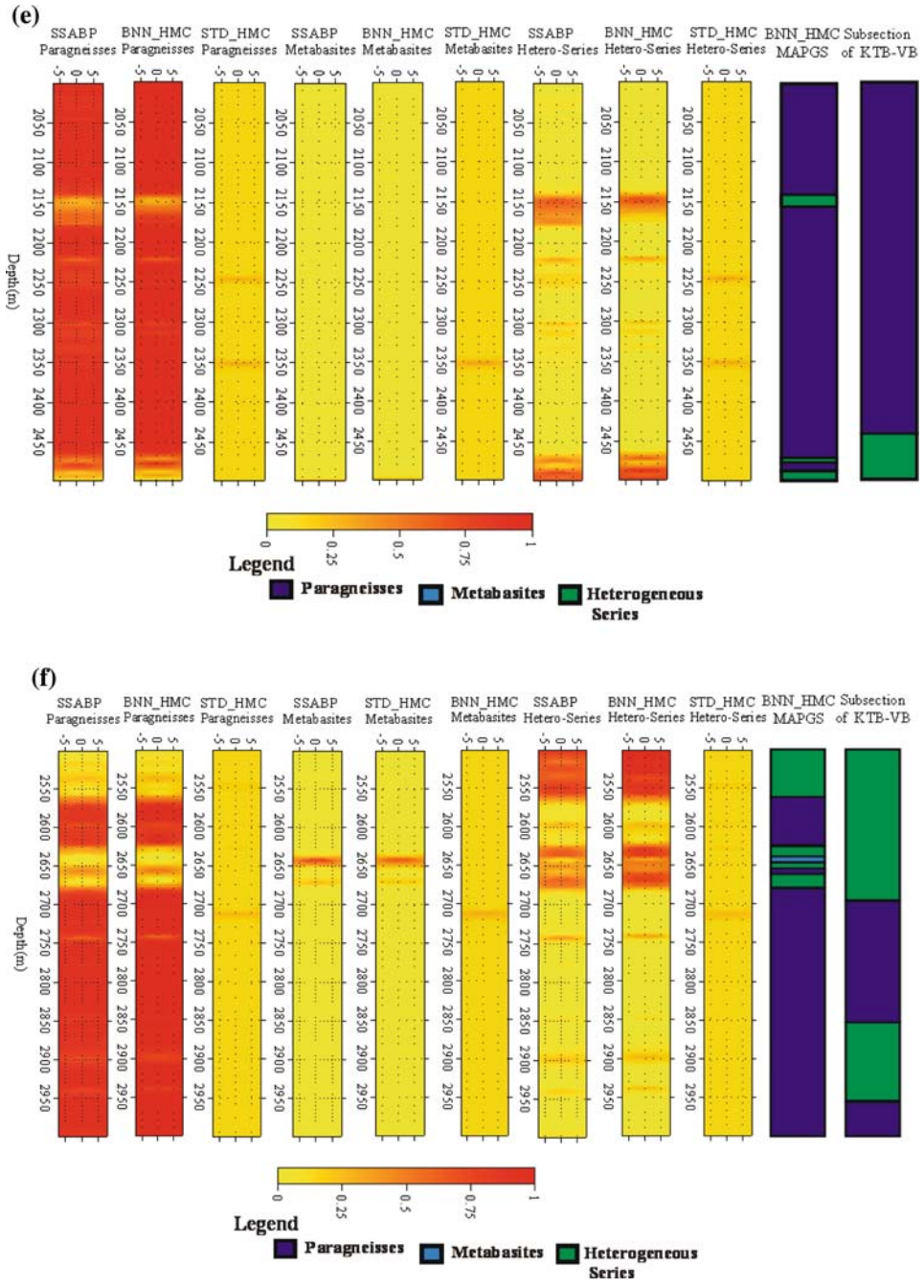


Figure 9
contd.

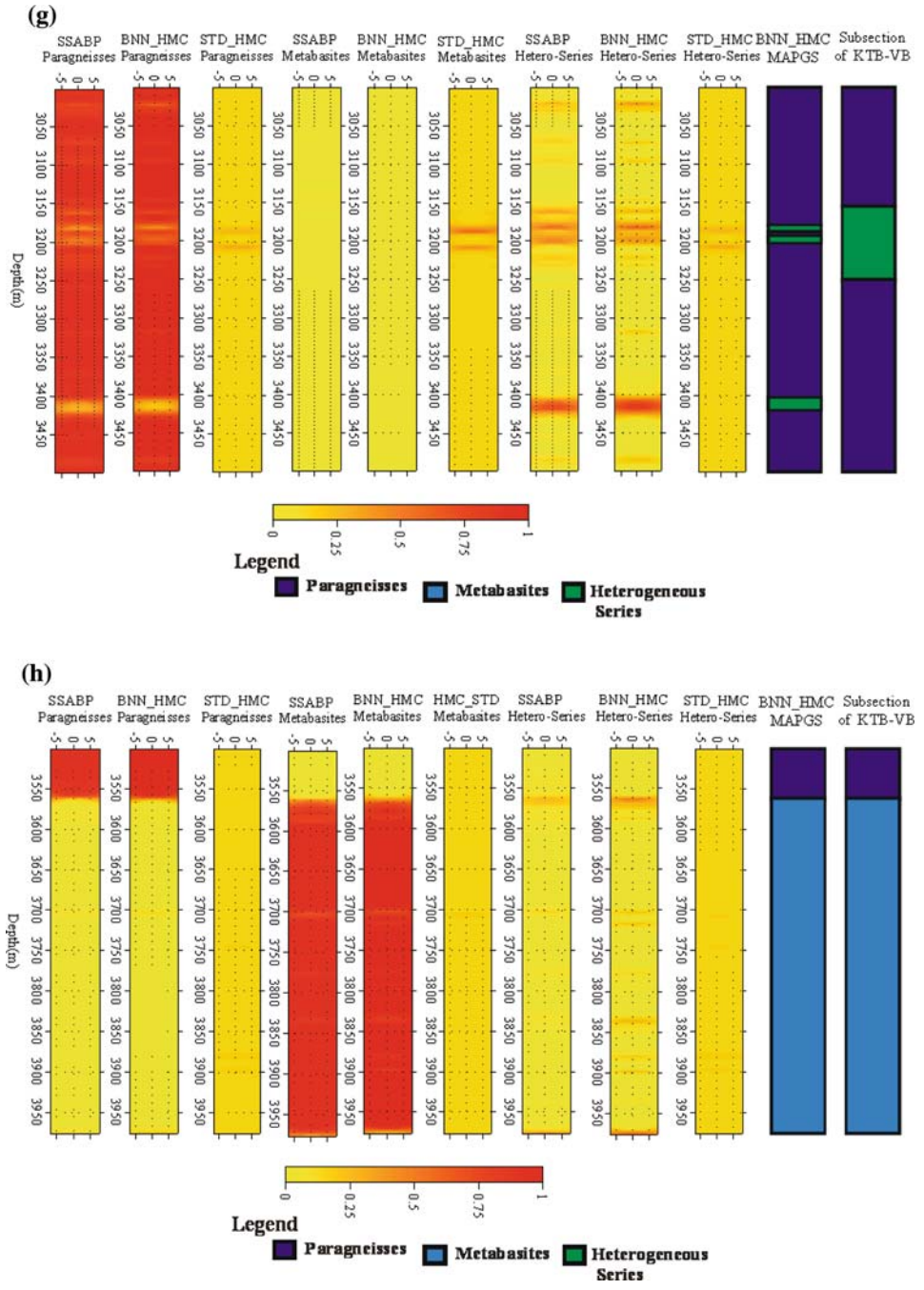


Figure 9
contd.

a depth of 4000 m to 7000 m to reduce the dimension. Preprocessed data are then used to train network and the output of the HMC based network is shown in a 3-colums color matrix with red representing 1 and yellow representing 0. STD of error maps corresponds to three lithofacies estimated by Bayesian code to quantify the prediction uncertainties at the network output. The result is compared with the published results of MAITI *et al.* (2007) in Figures 9 and 10. We note that the maxima of *a posteriori* geological section (MAPGS) correspond to the class of maxima of *a posteriori* probability. In the ideal case, if the lithofacies of a particular class exist, the output value of the node in the last layer is 1 or very close to 1 and if not, it is 0 or very close to 0.

9.1. Comparisons of the Hybrid Monte Carlo based Maximum a posteriori Model (MAP) Result with the Published Results

We have redrawn the published results of lithofacies successions (EMMERMANN and LAUTERJUNG 1997) (Fig. 2) and compared them with the maximum *a posteriori* geologic section (MAPGS) derived from the HMC modeling results and the SSABP results (MAITI *et al.*, 2007) for both KTB boreholes (Figs. 9 and 10). Evidently the HMC-based algorithm provides better insight into the analysis of borehole data. It is also clear that in most cases it matched well with the published geology. The results also indicate some additional finer beds that were not detected in earlier studies (Figs. 9 and 10).

10. Discussions

We have used ANN applying the concept of Bayesian statistics for classifying lithofacies boundaries from the KTB data. In order to do this, we have employed the MCMC chain to sample the network parameter/model parameter from target distribution (output). The advantage is that HMC is a sampling-based inversion technique unlike a gradient-based matrix inversion in which an algorithm results with inescapable local minima. We have used nonlinear hyperbolic tangent forward functions so that the underlying geophysical inverse problem can be treated in a fully nonlinear manner. Posterior probability distribution of the network parameter is obtained by using prior and likelihood functions. The output of the network gives mean value (averages over all possible network parameters), therefore, providing variability in the possible solutions rather than a single best solution using the conventional neural network. The Bayesian approach incorporates posterior data uncertainty caused by inherent noise (deceptive correlated noise) and inexact theory (modeling uncertainty). Comparison of MAPGS by the HMC-based algorithm with the published litho-species section of EMMERMANN and LAUTERJUNG (1997) (Figs. 9 and 10) exhibits more or less matching patterns that are well correlated too. In addition to this, the HMC-based Bayesian model also reveals some finer structural details, which seems to be geologically significant. We manually checked a few samples produced by the HMC-based trained network (see Table 4). Surprisingly, the

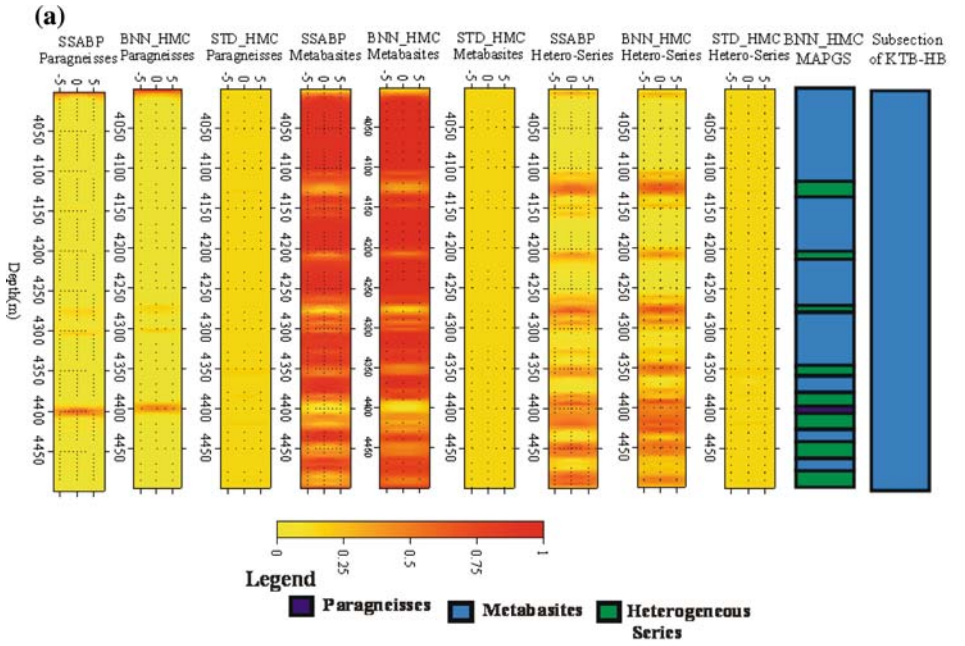
Figure 10 ►

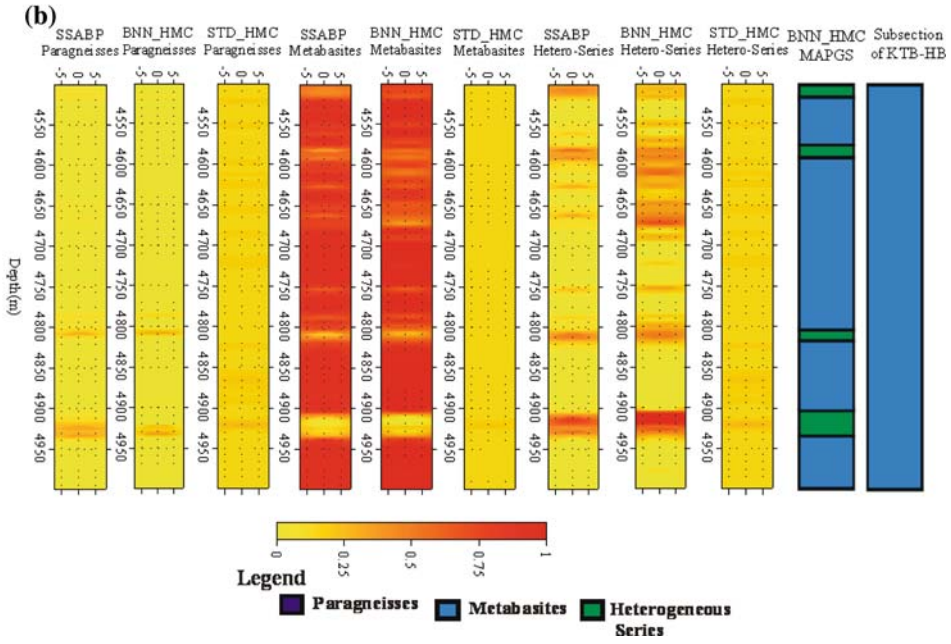
(a) Comparison of maximum a *posteriori* geological section (MAPGS) obtained by the Bayesian Neural Network (BNN) with Hybrid Monte Carlo (HMC) approach with maximum likelihood geological section (MLGS) obtained by super self adaptive back propagation (SSABP) neural network approach (left) with published litho-facies subsection of the German Continental Deep Drilling Program (KTB) main hole (KTB-HB) (right after EMMERMANN and LAUTERJUNG, 1997) and standard deviation (STD) error map estimated at network output by the Bayesian Neural Network (BNN) approach at depth interval of 4000–4500 m. (b) – (f) Same for depth range of 4500–5000 m, …, 6500–7000 m in German Continental Deep Drilling Program (KTB) main hole (KTB-HB).

Bayesian network approach produced more or less identical results that are consistent with the prior information (geological and geophysical). However, a mismatch with the published results could be explained by the inexact theory (modeling uncertainty). Observed data corrupted with deceptive 'red' noise and with non-zero mean might contain additional error. This might arise either due to limitations in the KTB data resolution or due to the poor sampling in the model space where more uncertainty can be expected (Figs. 9 and 10). Some deviations between the two results are, therefore, expected. We note, however, that while interpreting prediction of the network's output node in a probability sense with the maximum *a posteriori* value, the actual patterns show considerably high correlation coefficients ~ 0.97 (Table 3). Further regression analysis of both bore hole results between the two algorithms also reveals very consistent and sound agreements ($R \sim 0.94$) (Figs. 11a-f). Thus, it may be emphasized that the HMC-based Bayesian algorithm, developed here, combined with sensitivity analysis, uncertainty and regression analyses do provide credence to the authenticity of these results. Hence, maximum *a posteriori* values are not due to any other reason, rather they are, in fact, inter-bedded geological structures which remained unrecognized in previous visual interpretation.

11. Conclusions

We have developed an ANN technique utilizing the concept of an HMC algorithm for the analysis of the well log data. The method is applied to KTB well log data for classifying lithofacies. These results are also compared with published histogram models. Our analysis shows that the HMC based ANN is robust for detecting lithospecies boundaries. Modeling of the KTB data reveals some finer bed boundaries in addition to main lithofacies boundaries. In addition, uncertainty and non-uniqueness of the geophysical problem are discussed in conjunction with the Markov Chain Process where hundreds of acceptable solutions are generated to choose the mean model from the one hundred acceptable solutions. This solution approach is more precise conceptually and more defensive in a quantitative manner.


(a)
SSABP Paragneisses
BNN_HMC Paragneisses
STD_HMC Paragneisses
SSABP Metabasites
BNN_HMC Metabasites
STD_HMC Metabasites
SSABP Hetero-Series
BNN_HMC Hetero-Series
STD_HMC Hetero-Series
BNN_HMC MAPGS
Subsection of KTB-HB
Depth(m)
Legend
Paragneisses
Metabasites
Heterogeneous Series


(b)
SSABP Paragneisses
BNN_HMC Paragneisses
STD_HMC Paragneisses
SSABP Metabasites
BNN_HMC Metabasites
STD_HMC Metabasites
SSABP Hetero-Series
BNN_HMC Hetero-Series
STD_HMC Hetero-Series
BNN_HMC MAPGS
Subsection of KTB-HB
Depth(m)
Legend
Paragneisses
Metabasites
Heterogeneous Series

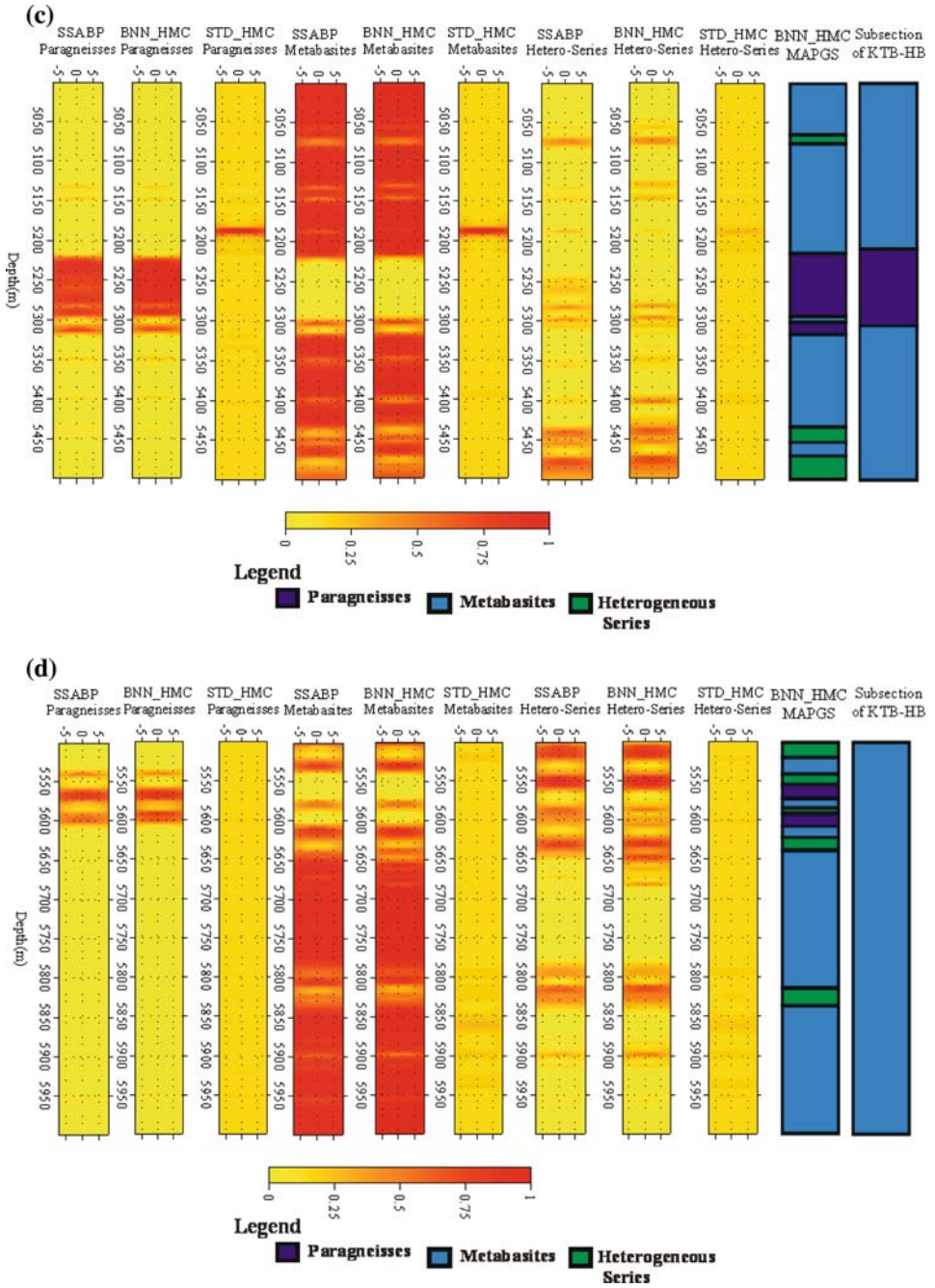


Figure 10
contd.

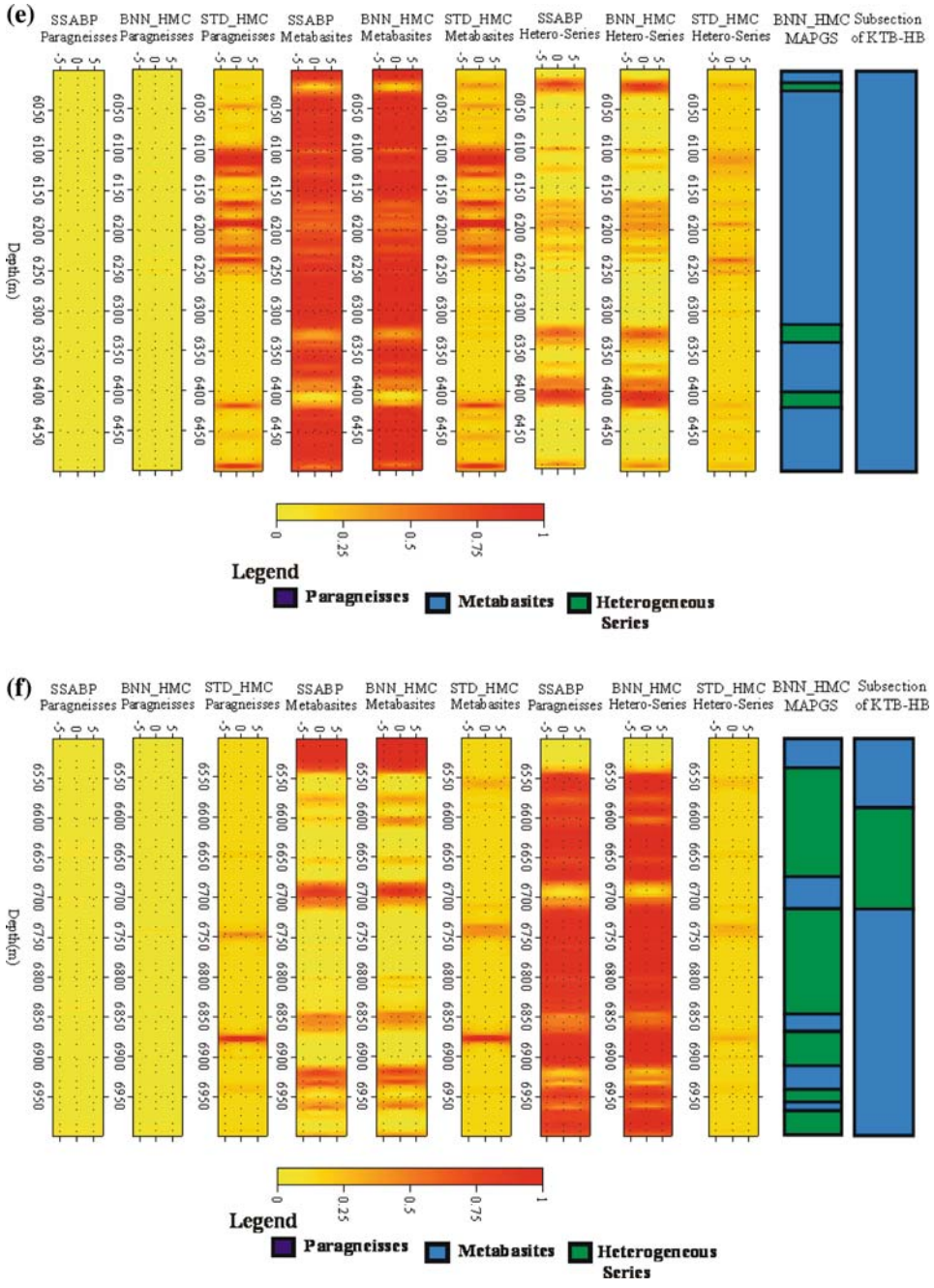


Figure 10
contd.

Table 4

Analysis of real data taken from both KTB pilot hole (KTB-VB) and KTB main hole (KTB-HB) from different depth with ± prediction error

Bore Hole (Samples of Data Taken)	Depth (m)	Density (g/cc)	Neutron porosity (%)	Gamma Ray Intensity (A.P.I)	Desired Output/ Binary Code			Neural Networks Actual Output ± STD		
KTB-VB	3119.17	2.71	10.80	107.1	1	0	0	1.02 ± 0.14	0.01 ± 0.14	0.03 ±0.14
KTB-VB	1574.29	2.94	12.34	45.00	0	1	0	0.00 ± 0.14	0.98 ± 0.14	0.01 ± 0.14
KTB-VB	89.00	2.73	10.46	82.05	0	0	1	0.05 ±.0.14	0.02 ± 0.14	0.96 ± 0.14
KTB-VB	305.86	2.94	14.06	32.61	0	1	0	0.00 ± 0.14	0.81 ± 0.14	0.18 ± 0.14
KTB-VB	893.82	2.82	12.94	128.41	0	0	1	0.67 ± 0.14	0.00± 0.14	0.31 ± 0.14
KTB-VB	1393.54	3.01	11.24	23.26	0	1	0	0.00 ± 0.14	0.99 ± 0.14	0.01 ± 0.14
KTB-VB	2252.47	2.80	14.94	119.17	1	0	0	0.75 ± 0.14	0.00 ± 0.14	0.22 ± 0.14
KTB-VB	3864.25	2.94	13.59	19.10	0	1	0	0.00 ± 0.14	0.98 ± 0.14	0.02 ± 0.14
KTB-VB	3559.45	2.74	12.11	104.35	1	0	0	1.00 ± 0.14	0.01 ± 0.14	0.03 ± 0.14
KTB-VB	771.60	2.68	5.33	112.73	1	0	0	0.96 ± 0.14	0.00 ± 0.14	0.00 ± 0.14
KTB-VB	1072.43	2.776	9.72	115.58	1	0	0	1.00 ± 0.14	0.0 ± 0.14	0.02 ± 0.14
KTB-VB	1145.74	2.745	12.03	106.69	1	0	0	0.99 ± 0.14	0.01 ± 0.14	0.00 ± 0.14
KTB-VB	1374.03	2.97	8.73	15.83	0	1	0	0.00 ± 0.14	1.00 ± 0.14	0.01 ± 0.14
KTB-VB	1715.26	2.70	10.08	105.55	1	0	0	1.00 ± 0.14	0.01 ± 0.14	0.00± 0.14
KTB-VB	1878.48	2.72	12.03	111.65	1	0	0	1.00 ± 0.14	0.01 ± 0.14	0.00 ± 0.14
KTB-VB	2084.83	2.75	16.51	120.17	1	0	0	1.00 ± 0.14	0.00 ± 0.14	0.01 ± 0.14
KTB-VB	2290.42	2.75	14.70	119.58	1	0	0	1.00 ± 0.14	0.01 ± 0.14	0.00 ± 0.14
KTB-VB	2697.17	2.65	10.00	103.20	1	0	0	1.02 ± 0.14	0.00 ± 0.14	0.01 ± 0.14
KTB-VB	2891.33	2.71	8.76	103.30	1	0	0	1.00 ± 0.14	0.01 ± 0.14	0.02 ± 0.14
KTB-VB	3177.84	2.77	9.69	96.31	1	0	0	0.77 ± 0.14	0.02 ± 0.14	0.20 ± 0.14
KTB-VB	3801.16	3.04	14.10	30.46	0	1	0	0.00 ± 0.14	0.99 ± 0.14	0.01 ± 0.14
KTB-VB	3889.55	3.00	8.677	29.21	1	0	0	1.00 ± 0.14	0.00 ± 0.14	0.14 ± 0.14
KTB-HB	6515.86	3.00	14.82	20.32	0	1	0	0.00 ± 0.14	0.98 ± 0.14	0.00 ± 0.14
KTB-HB	6807.09	2.74	11.17	55.44	0	0	1	0.00 ± 0.14	0.08 ± 0.14	0.91 ± 0.14
KTB-HB	6470.14	2.72	25.11	16.72	0	1	0	0.00 ± 0.14	0.99 ± 0.15	0.01 ± 0.14
KTB-HB	6999.12	2.85	12.47	49.45	0	0	1	0.00 ± 0.14	0.38 ± 0.14	0.61 ± 0.14
KTB-HB	5677.35	2.95	1.66	26.75	0	1	0	0.00 ± 0.14	1.00 ± 0.14	0.03 ± 0.14
KTB-HB	5372.25	2.97	15.23	16.83	0	1	0	0.00 ± 0.14	1.00 ± 0.14	0.02 ± 0.14
KTB-HB	5217.87	2.93	4.08	29.53	0	1	0	0.01 ± 0.18	1.00 ± 0.14	0.12 ± 0.18

Table 4
contd.

Bore Hole (Samples of Data Taken)	Depth (m)	Density (g/cc)	Neutron porosity (%)	Gamma Ray Intensity (A.P.I)	Desired Output/ Binary Code			Neural Networks Actual Output ± STD		
KTB-HB	4547.00	2.82	8.67	35.92	0	1	0	0.00 ± 0.14	0.79 ± 0.13	0.22 ± 0.13
KTB-HB	4427.37	2.81	4.52	38.63	0	1	0	0.00 ± 0.14	0.88 ± 0.14	0.17 ± 0.14
KTB-HB	4433.16	2.75	4.89	42.33	0	1	0	0.00 ± 0.14	1.18 ± 0.14	0.01 ± 0.14
KTB-HB	4442.00	2.92	10.21	37.06	0	1	0	0.00 ± 0.14	0.99 ± 0.14	0.01 ± 0.14
KTB-HB	4950.56	2.90	12.09	33.75	0	1	0	0.01 ± 0.31	0.99 ± 0.31	0.03 ± 0.31
KTB-HB	6325.36	2.75	15.72	13.35	0	1	0	0.01 ± 0.14	0.87 ± 0.14	0.01 ± 0.13
KTB-HB	4006.59	2.95	15.73	36.147	0	1	0	0.00 ± 0.13	0.84 ± 0.14	0.15 ± 0.14
KTB-HB	4002.93	2.84	16.83	109.94	1	0	0	1.00 ± 0.14	0.00 ± 0.13	0.01 ± 0.14
KTB-HB	4206.24	2.90	14.51	44.58	0	1	0	0.00 ± 0.14	1.00 ± 0.13	0.05 ± 0.14
KTB-HB	4311.39	2.92	17.16	22.42	0	1	0	0.00 ± 0.13	0.98 ± 0.14	0.02 ± 0.14
KTB-HB	4548.22	2.87	7.96	34.12	0	1	0	0.00 ± 0.14	0.64 ± 0.13	0.41 ± 0.13
KTB-HB	4556.76	2.97	5.31	25.51	0	1	0	0.00 ± 0.14	1.01 ± 0.14	0.03 ± 0.14
KTB-HB	4630.67	2.89	2.73	24.59	0	1	0	0.01 ± 0.14	0.85 ± 0.14	0.14 ± 0.14
KTB-HB	4873.75	3.08	3.00	28.48	0	1	0	0.00 ± 0.14	1.02 ± 0.14	0.01 ± 0.13
KTB-HB	5054.34	2.87	10.66	32.95	0	1	0	0.00 ± 0.14	0.99 ± 0.14	0.00 ± 0.14
KTB-HB	5058.30	2.69	14.47	135.41	0	0	1	0.16 ± 0.14	0.22 ± 0.14	0.62 ± 0.14
KTB-HB	5149.90	2.97	9.91	19.61	0	1	0	0.00 ± 0.14	1.01 ± 0.14	0.03 ± 0.14
KTB-HB	5321.19	2.86	6.56	40.54	0	1	0	0.03 ± 0.14	0.64 ± 0.14	0.380 ± 0.14
KTB-HB	5743.95	3.06	6.10	18.39	0	1	0	0.00 ± 0.14	1.01 ± 0.13	0.00 ± 0.14
KTB-HB	5840.27	2.65	14.53	16.09	0	1	0	0.00 ± 0.14	0.98 ± 0.13	0.00 ± 0.14
KTB-HB	6271.56	2.96	18.45	13.75	0	1	0	0.00 ± 0.14	1.00 ± 0.13	0.00 ± 0.14
KTB-HB	6486.90	2.78	15.27	42.39	0	1	0	0.00 ± 0.14	0.77 ± 0.14	0.22 ± 0.14

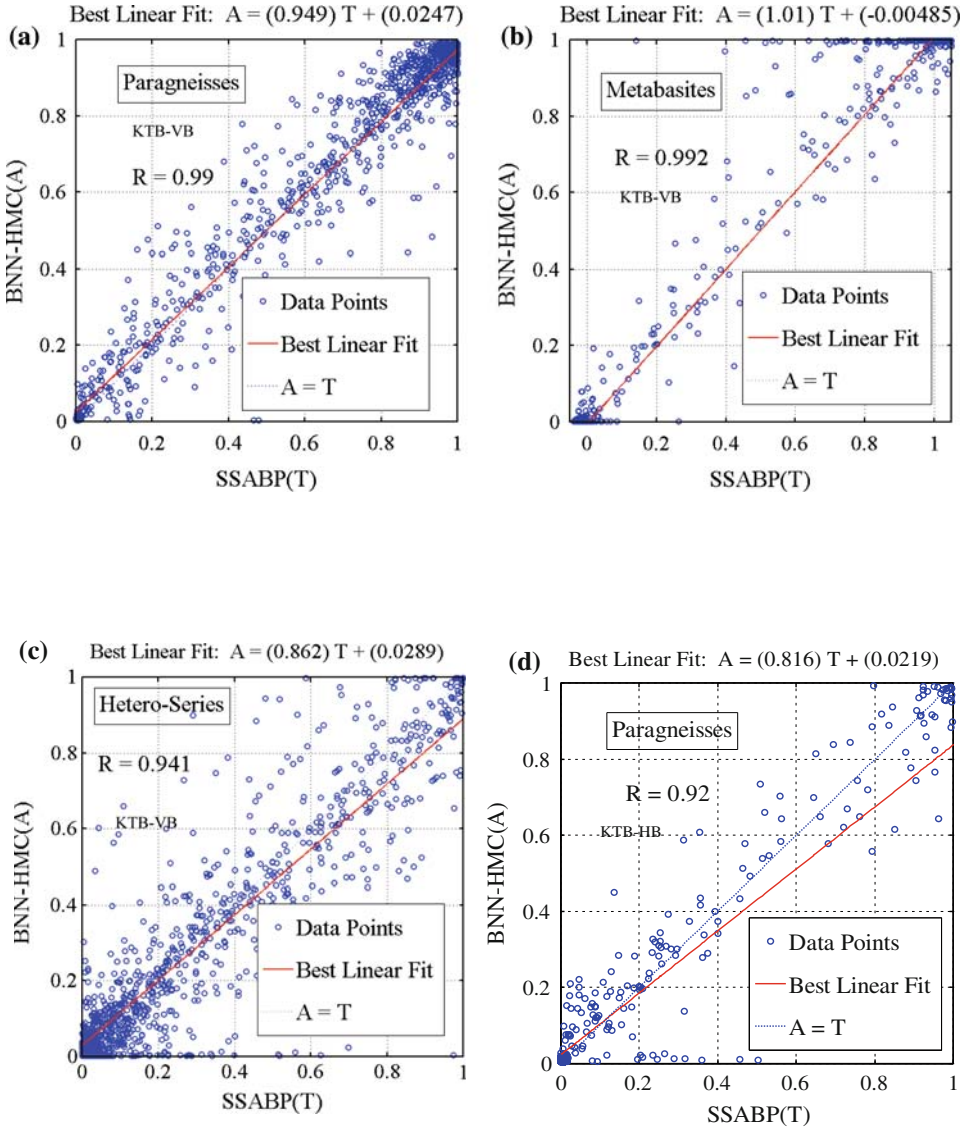


Figure 11

(a-f) Regression analysis of German Continental Deep Drilling Program (KTB) pilot hole (KTB-VB) and German Continental Deep Drilling Program (KTB) main hole (KTBHB) corresponds to paragneisses, metabasites and hetero-series showing very good agreement between two algorithm overall

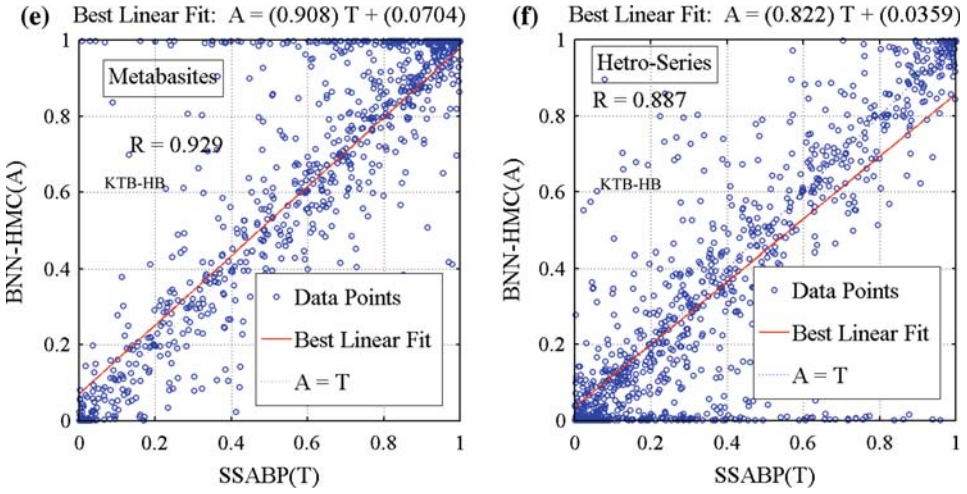


Figure 11
contd.

Acknowledgements

Saumen Maiti (SM) expresses sincere thanks to the Department of Science and Technology (DST), Govt. of India to carry out the research work**.** SM is also thankful to Prof. Archana Bhattacharyya, Director, Indian Institute of Geomagnetism (IIG), Navi-Mumbai for her kind permission to publish this work. SM expresses deep gratitude to Prof. S. G. Gokaran, and Dr. G. Gupta, IIG, Navi-Mumbai for their motivation. We are also thankful to Prof. Hans-Joachim Kumpel for providing the KTB data.

References

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

Bishop, C.M., *Neural Networks for Pattern Recognition*. (Oxford University Press, 1995).

Backus, G.E. and Gilbert, J.F. (1967), *Numerical applications of formalism for geophysical inverse problems*, Geophys. J. R. Astron. Soc., *13*, 247–276.

Duane, S., Kennedy, A.D., Pendleton, B., and Roweth, D. (1987), *Hybrid Monte Carlo*, Phys. Lett. B *195*, 216–222.

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

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M.N., Teller, A.H., and Teller, E. (1953), *Equations of state calculations by fast computing machines*. J. Chem. Phys. *21*, 1087–1091.

Metropolis, N. and Ulam, S.M., (1949), *The Monte Carlo method*, J. Am. Stat. Assoc. *44*,335–341.

Emmermann, R. and Lauterjung, J. (1997), *The German Continental Deep Drilling Program KTB: Overview and major results*, J. Geophys. Res. *102*, .18179–18201.

FRANKE, W. (1989), *The geological framework of the KTB drill site*. In Emmermann, R., Wohlenberg, J., (Eds.), The German Continental Deep Drilling Program (KTB). (Springer, Berlin, 1989) pp. 38–54.

FULLER, W.A., *Introduction to statistical time series* (John Wiley and Sons, New York, 1976).

KHAN, M.S. and COULIBALY, P. (2006), *Bayesian neural network for rainfall-runoff modeling*, Water Resour. Res. *42*, W07409, doi: 10.1029/2005WR003971.

LAMPINEN, J. and VEHTARI, A. (2001), *Bayesian Approach for neural networks-review and case studies*, Neural Networks *14*(3), 7–24.

LEONARDI, S. and KUMPEL, H. (1998), *Variability of geophysical log data and signature of crustal heterogeneities at the KTB*, Geophys. J. Int. *135*, 964–974.

LEONARDI, S. and KUMPEL, H. (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 Comput. *4*(3), 448–472.

MAITI, S. and TIWARI, R.K. (2005), *Identifying lithofacies boundaries using super self adaptive back propagation neural network (SSAB): A Case Study from the KTB Borehole*, Petrotech, Delhi, Extended Abstract.

MAITI, S., TIWARI, R.K., and KUMPEL, H.J., (2007), *Neural network modeling and classification of lithofacies using well log data: a case study from KTB borehole site*, Geophys. J. Int. *169*, 733–746.

MAITI, S. and TIWARI, R.K. (2008), *Classifications of lithofacies boundaries using the KTB borehole data: A Bayesian neural network modeling*, 7th Biennial International Conference and Exposition on Petroleum Geophysics, SPG Hyderabad, Extended Abstract.

MOSEGAARD, K. and TARANTOLA, A., (1995), *Monte Carlo sampling of solutions to inverse problem,* Jr. Geophys. Res. *100*, B7, 12,431–12,447.

NABNEY, I.T., *Netlab Algorithms for Pattern Recognition* (Springer, New York 2004).

NEAL, R.M, *Bayesian learning via stochastic dynamics*. In *Advances in Neural Information Processing Systems 5*, (eds) C.L.Giles *et al*.), pp 475–482, (Elsevier, New York 1993).

PECHNIG, P., HAVERKAMP, S., WOHLENBERG, J., ZIMMERMANN, G., and BURKHARDT, H., (1997), *Integrated interpretation in the German Continental Deep Drilling Program: Lithology, porosity, and fracture zones*. J. Geophys. Res., *102*, 18363–18390.

POULTON, M., Ed., *Computational Neural Networks for Geophysical Data Processing*, (Pergamon 2001).

SAMBRIDGE, M. and MOSEGAARD, K. (2002), *Monte Carlo methods in geophysical inverse problems*, Rev. Geophys., *40*, 3, 1–29.

TARANTOLA, A., *Inverse Problem Theory* (Elsevier, New York, 1987).

(Received August 6, 2008, revised February 10, 2009, accepted February 10, 2009)
Published Online First: August 20, 2009