

Stochastic simulation and spatial estimation with multiple data types using artificial neural networks

Lance E. Besaw¹ and Donna M. Rizzo¹

Received 9 September 2006; revised 23 June 2007; accepted 16 July 2007; published 8 November 2007.

[1] A novel data-driven artificial neural network (ANN) that quantitatively combines large numbers of multiple types of soft data is presented for performing stochastic simulation and/or spatial estimation. A counterpropagation ANN is extended with a radial basis function to estimate parameter fields that reproduce the spatial structure exhibited in autocorrelated parameters. Applications involve using three geophysical properties measured on a slab of Berea sandstone and the delineation of landfill leachate at a site in the Netherlands using electrical formation conductivity as our primary variable and six types of secondary data (e.g., hydrochemistry, archaea, and bacteria). The ANN estimation fields are statistically similar to geostatistical methods (indicator simulation and cokriging) and reference fields (when available). The method is a nonparametric clustering/classification algorithm that can assimilate significant amounts of disparate data types with both continuous and categorical responses without the computational burden associated with the construction of positive definite covariance and cross-covariance matrices. The combination of simplicity and computational speed makes the method ideally suited for environmental subsurface characterization and other Earth science applications with spatially autocorrelated variables.

Citation: Besaw, L. E., and D. M. Rizzo (2007), Stochastic simulation and spatial estimation with multiple data types using artificial neural networks, *Water Resour. Res.*, 43, W11409, doi:10.1029/2006WR005509.

1. Introduction

[2] Imaging, joint interpolation and inversion play important roles in the discovery, characterization, simulation, and performance evaluation for subsurface mineral, energy, and environmental remediation projects. These methods are used to combine information, whether qualitative or quantitative, into a composite site description to interpret and/or predict responses of the geologic system. The oil and gas exploration business routinely uses minimally invasive or noninvasive data collection methods (e.g., three-dimensional seismic analysis) to conduct three-dimensional geophysical investigations producing enormous data sets that must be managed and processed. As a result, imaging and inversion of geophysical data have a long history (see *Menke* [1984] for a good review of common practices).

[3] The environmental restoration business is turning, with increasing frequency and commitment, to such methods of collecting field data that can be easily and inexpensively measured in situ [*U.S. Environmental Protection Agency*, 2000]. The cone penetrometer testing (CPT) methods, used for certain soils investigations since the mid-1960s, have been aggressively extended with new sensors, video cameras, and probes for rapid collection of in situ data [*Ballard and Cullinane*, 1998; *Purdy and Beam*, 1998; *Gelb and Wonder*, 1998; *Rossabi et al.*, 2000] (see also <http://www.cpeo.org/techtree/ttdescript/scaps.htm>). A variety of

direct push groundwater sampling tools now provide continuous and densely discrete (subcentimeter spacing) vertical data of solute concentrations in permeable, unconsolidated deposits [*Pitkin*, 1998]. As a result, the number of data types, the required spatial resolution, and the cost effectiveness of these shallow in situ probes and sensors make them an attractive alternative for site characterization, environmental restoration, and long-term monitoring.

[4] The term “data fusion” has gained portent in the Earth sciences community [*Oldenburg et al.*, 1998]. This process also goes by the names “joint data assimilation,” “objective analysis,” “coinversion,” and “joint inversion.” In it, multiple signal types are interpreted as an ensemble to draw inferences about parameters, experiments, and so on [*Hohmann and Raiche*, 1987]. Numerous inversion methods, many of which are described by *Yeh* [1986], have been developed to address these issues. Joint inversion methods allow for the incorporation of secondary information and have been employed extensively in the exploration and recovery of oil and natural gas to incorporate geophysical information [*Kis*, 2002; *Bosch*, 2004] and to estimate the extent of subsurface contamination [*de la Vega et al.*, 2003]. More recently, universal function approximators have been used to perform inversion of groundwater modeling parameters [*Morshed and Kaluarachchi*, 1998; *Garcia and Shigidi*, 2006].

[5] Unfortunately, in the environmental sector, the ability to collect and store large volumes of data has increased dramatically over the last decade, but computational methodologies for extracting useful information from these vast and growing stores of data has lagged behind. Although it is now widely recognized that effective assimilation of dispa-

¹School of Engineering, University of Vermont, Burlington, Vermont, USA.

rate types of field data can both improve parameter estimates [Vozoff and Jupp, 1975; Copty *et al.*, 1993; Copty and Rubin, 1995] and increase the efficiency of field sampling programs, the production of high data acquisition rates has led to data that are not fully being assimilated. One reason is that in the environmental community, as opposed to manufacturing control or the oil or gas industry, the cost of collecting data is considered a cost of regulation. As a result, collected environmental data have not been viewed, in general, with the same eye for adding value as data collected for manufacturing control. Another reason is that, although geophysical data have been used sporadically in the water resources community, interpretation of these data have been left for geophysicists and the groundwater community generally has not internalized these methods.

[6] The joint assimilation of tremendous amounts of data and multiple signal types creates significant data management and computational burdens. Solutions associated with joint inversion or cokriging methods often become infeasible when incorporating increasingly large numbers of secondary variables. Although geostatistical methods are excellent for estimating spatially distributed processes with low-dimensional data (see Istok and Rautman [1996], Goovaerts [1998, 2001], Goovaerts *et al.* [2005], Li and Yeh [1999], Gloaguen *et al.* [2001], Patriarche *et al.* [2005], and Mouser *et al.* [2005] for recent environmental applications involving a single type of secondary data), new methods (e.g., nonparametric statistical methods, categorical classification methods, data-driven artificial neural networks) are currently being developed to deal with the spatial analysis of large numbers of multiple data types with both continuous and categorical responses.

[7] In this paper (section 2), we introduce artificial neural networks (ANNs), provide an overview of the original counterpropagation algorithm as developed by Hecht-Nielsen [1987] and the modifications and extensions needed to use the algorithm for stochastic simulation of distributed parameters that exhibit spatial autocorrelation. In section 3, we present the simulation results and use two ANNs in series to demonstrate the computational ease with which this method can assimilate multiple data types.

[8] We compare and evaluate the method with traditional geostatistical (kriging) methods because of their well-developed theoretical foundation. However, our goal is to find new robust, nonparametric methods capable of dealing with large volumes and multiple types of environmental data. Such environmental applications often consist of data that are nonnormal and/or mixtures of categorical or continuous and discrete-valued variables that cannot be analyzed using the traditional geostatistical or categorical data analysis techniques.

2. Methodology

2.1. Artificial Neural Networks

[9] Artificial neural networks (ANNs) are dynamic, computational systems built from a large number of interconnected processing units that interact in some prescribed, parallel manner. Simply, they can be viewed as universal approximators that map one space (the input space) into some other space (the output space). The network architecture is presented with a set of examples, called

training patterns (input vectors and corresponding output vectors); while a training algorithm iteratively adjusts the internal parameters to better simulate (or “learn”) the desired mapping. The training process is complete when the error between the computed output and the desired output is minimal for all patterns in the training set. Once trained, the ANN may be used for classification or prediction. The ability to retrieve the functional relationship between the input and the output space through the presentation of data patterns suggests that these data-driven ANNs may be useful in cases where a clear understanding of the physics may be lacking. The architecture of the interconnected processing units, and how information is passed and updated between processing units, defines different ANN algorithms. The most popular algorithms include the feed forward back-propagation network [Rumelhart and McClelland, 1988], the Hopfield Network [Hopfield, 1982], the self organizing map (SOM) [Kohonen, 1989], and the radial basis function neural network [Aizerman *et al.*, 1964; Bashkirov *et al.*, 1964]. A comprehensive introduction to ANNs is provided by Wasserman [1989].

2.2. Counterpropagation

[10] Hecht-Nielsen [1987] first introduced the counterpropagation network, which sequentially combines the Kohonen self organizing map (SOM) and the Grossberg outstar structure. Although the networks listed in section 2.1 account for the majority (~95–99%) of all ANN applications, the counterpropagation ANN has found uses in a wide range of applications including the classification of soil samples [Fidencio *et al.*, 2001], assessment of ecological status and prediction of ecosystem water quality [Park *et al.*, 2003], modeling of nonlinear pH processes [Nie *et al.*, 1996] and prediction of spatial and temporal rainfall variation [Hsu *et al.*, 1995, 1999]. Rizzo and Dougherty [1994] and Rizzo [1994] introduce the concept of applying the counterpropagation algorithm to develop maps of discrete spatially distributed fields (e.g., log-hydraulic conductivity fields given estimates of hydraulic conductivity from pumping tests), and classifying soil lithology given soil sample descriptions from drillers’ well logs [Rizzo *et al.*, 1996]. Although details and pseudocode for this method have been presented by Rizzo and Dougherty [1994], an overview is provided in section 2.3.

2.3. Counterpropagation Algorithm

[11] The counterpropagation algorithm (Figure 1) is a supervised learning algorithm that self-adapts to cluster data and map a set of input vectors $\mathbf{x}^n = (x_1^n, x_2^n, \dots, x_1^n)$ into a prespecified number of classes represented by output vectors $\mathbf{y}^n = (y_1^n, y_2^n, \dots, y_k^n)$, where $n = 1, 2, \dots$ to the total number of training patterns N . The counterpropagation algorithm requires that the network output (parameter of interest) be categorized into some prespecified number of classifications (K), where the number of classes and corresponding accuracy of estimation is defined by the user for the given process of interest. Assume, for example, that $K = 6$ classes; then class $k = 1$ may be represented by the classified output vector $y = [0, 0, 0, 0, 0, 1]$. Likewise, $k = 2$ corresponds to $y = [0, 0, 0, 0, 1, 0]$ and so on. This subtle feature of the network is important in that this algorithm may be used to predict categorical variables (i.e., soil lithology), as will be demonstrated in the following example applications.

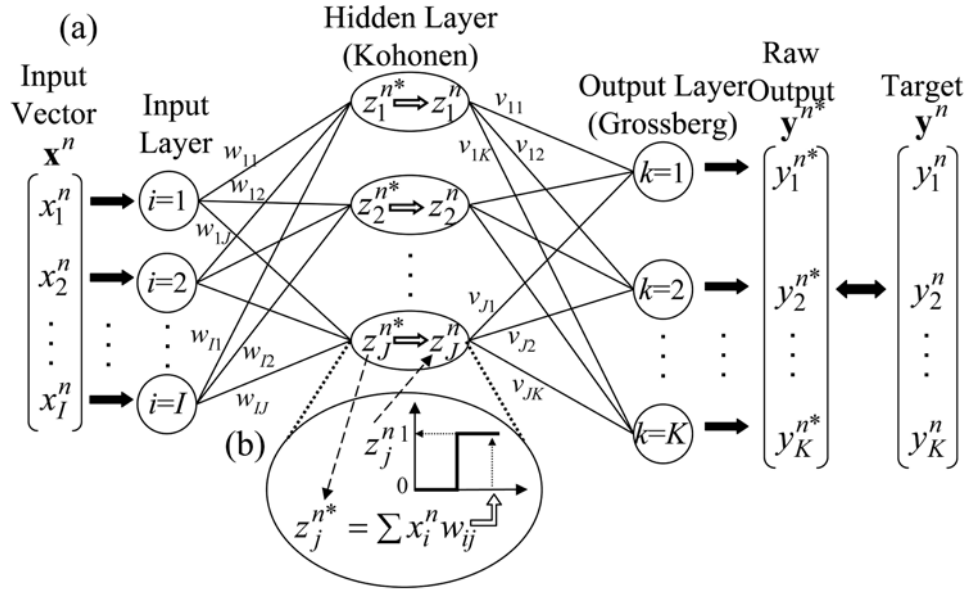


Figure 1. General schematic showing (a) architecture and notation and (b) activation function of the counterpropagation ANN.

[12] The counterpropagation architecture consists of an input layer, a “hidden” Kohonen layer, and an output Grossberg layer. Figure 1 shows a schematic of a counterpropagation ANN with I input nodes, J hidden nodes, and K output nodes. The arcs connecting the input and hidden layers have internal parameters known as Kohonen weights, \mathbf{w} . The arcs connecting the hidden and output layers have associated Grossberg weights, \mathbf{v} . Dual subscripts (e.g., w_{ij}) denote the weight connecting input node i and hidden node j . Therefore each of the J hidden nodes have an associated weight vector \mathbf{w}_j (single column of the Kohonen weight matrix, \mathbf{w}) comprising I components. Likewise, the Grossberg weight matrix \mathbf{v} comprising J vectors, each with K components.

[13] The counterpropagation algorithm is executed in two phases: a training phase and an operational phase (classification/prediction). Prior to training, the algorithm requires the input vectors $\mathbf{x}^{n=1,2,\dots,N}$ be normalized such that each vector will increase by one dimension, have an overall vector length equal to 1.0, and lie on a unit hypersphere; see *Rizzo and Dougherty* [1994] for pseudocode describing the normalization procedure. The Kohonen and Grossberg weights are initially set to random values between 0 and 1 and are renormalized for each iteration. This ensures that the input vectors and the Kohonen weight vectors lie in the same dimensional space.

2.3.1. Training Phase

[14] After normalization, the training patterns are sequentially presented to the network. The n^{th} input vector, comprising I predictor variables, $\mathbf{x}^n = (x_1^n, x_2^n, \dots, x_I^n)$, is presented to the input layer and the corresponding classified output (target) vector \mathbf{y}^n is simultaneously assigned to the output layer; see Figure 1. The input vector \mathbf{x}^n passes to the hidden layer where each hidden node computes a sum as follows:

$$z_j^{n*} = \sum_{i=1}^I (x_i^n \cdot w_{ij}) \text{ for } j = 1, 2, \dots, J, \quad (1)$$

where n would equal 1 for the first training pattern. This dot product, computed for each of the J hidden nodes, represents a measure of similarity between the input vector \mathbf{x}^n and the weight vector \mathbf{w}_j associated with the j^{th} hidden node. Since both the input vectors \mathbf{x}^n and the Kohonen weight vectors \mathbf{w}_j have been normalized to $\|\mathbf{x}\| = 1$ and $\|\mathbf{w}\| = 1$; the dot product of equation (1), by standard vector calculus, $z_j^{n*} = \|\mathbf{x}^n\| \cdot \|\mathbf{w}_j\| \cos\theta_j$, reduces to $z_j^{n*} = \cos\theta_j$, where θ_j is the angle between \mathbf{x}^n and \mathbf{w}_j .

[15] Next, the product associated with each of the J hidden nodes z_j^{n*} is passed through a “winner-take-all” activation function such that the maximum component of \mathbf{z}^{n*} (a.k.a. the winner) is assigned a value of 1 and all other components are assigned a value of 0; see Figure 1. The hidden node with the maximum value $z_{j=\text{winner}}^{n*} = \max(\cos\theta_j)$ is defined to be the winner, ensuring that the vector \mathbf{w}_j associated with this winning hidden node is the most similar, of all the hidden nodes, to the current input vector \mathbf{x}^n ($\mathbf{w}_{j=\text{winner}}$ has the smallest angle between \mathbf{w}_j and \mathbf{x}^n).

[16] This winning Kohonen weight vector is then adjusted (moved in the direction of the input vector \mathbf{x}^n) according to the Widrow-Hoff rule [*Widrow and Hoff*, 1960]:

$$\mathbf{w}_j^{\text{new}} = \begin{cases} \mathbf{w}_j^{\text{old}} + \alpha(\mathbf{x}^n - \mathbf{w}_j) & \text{if } j = \text{winning node} \\ \mathbf{w}_j^{\text{old}} & \text{otherwise } j \neq \text{winning node,} \end{cases} \quad (2)$$

where $\alpha \in [0, 1]$ is a learning coefficient (for this work $\alpha = 0.7$) that controls the magnitude of vector adjustment and thus rate of convergence [*Hecht-Nielsen*, 1988]. The Widrow-Hoff adjustment of equation (2) ensures that only one Kohonen weight vector becomes more similar to the input vector that caused it to become activated; all others remain the same.

[17] The activation vector \mathbf{z}^n (the absence of the * superscript indicates that the vector has been passed through the winner-take-all activation function) passes to the Grossberg

output layer where a second dot product is computed between \mathbf{z}^n and each of the Grossberg weight vectors \mathbf{v}_j . The raw ANN output is computed as

$$y_k^{n*} = \sum_{j=1}^J (z_j^n \cdot v_{jk}) \text{ for } k = 1, 2, \dots, K. \quad (3)$$

Therefore during training, the output \mathbf{y}^{n*} equals the Grossberg weight vector \mathbf{v}_j associated with the winning hidden node. The Grossberg weight vectors are adjusted according to:

$$\mathbf{v}_j^{\text{new}} = \begin{cases} \mathbf{v}_j^{\text{old}} + \beta(\mathbf{y}^n - \mathbf{y}^{n*}) & \text{if } j = \text{winning node} \\ \mathbf{v}_j^{\text{old}} & \text{otherwise } j \neq \text{winning node,} \end{cases} \quad (4)$$

such that updated Grossberg weight vectors equal the old Grossberg weight vectors plus some fraction, β , of the difference between the raw output vector \mathbf{y}^{n*} and the desired target vector \mathbf{y}^n (for this work the learning coefficient $\beta = 0.1$). A total of N training patterns (\mathbf{x}^n and the associated classified target vectors \mathbf{y}^n for $n = 1, 2, \dots, N$) are used to train the ANN. Equations (2) and (4) iteratively adjust the Kohonen and Grossberg weights so that subsequent network output \mathbf{y}^{n*} more closely approximates the desired target vectors \mathbf{y}^n . One passage of all N training patterns through the ANN completes one training iteration. The root-mean-square error (*RMSE*) is calculated for every training iteration as

$$RMSE = \sqrt{\frac{\sum_{n=1}^N \sum_{k=1}^K (y_k^{n*} - y_k^n)^2}{N * K}}$$

and provides a single quantitative description of how well all of the network outputs \mathbf{y}^{n*} match the known response vectors \mathbf{y}^n . When the network *RMSE* converges to the user-defined threshold (in this work 10^{-6}), training is complete and the ANN is ready for the classification/prediction phase.

2.3.2. Classification/Prediction Phase

[18] Once trained, the Kohonen and Grossberg weight matrices become fixed and the network may be used for normal operation (classification/prediction). New input vectors $\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^M$ are presented to the network to attain predictions/classifications $\hat{\mathbf{y}}^1, \hat{\mathbf{y}}^2, \dots, \hat{\mathbf{y}}^M$ where M is the total number of predictions. This prediction phase produces a raw output $\hat{\mathbf{y}}^{m*}$ for the m^{th} input vector that is subsequently passed through a winner-take-all activation function to create a classified prediction, $\hat{\mathbf{y}}^m$ (e.g., $\hat{\mathbf{y}}^{m*} = [0.08, 0.20, 0.10, 0, 0.70, 0.01]$ would yield $\hat{\mathbf{y}}^m = [0, 0, 0, 0, 1, 0]$).

[19] A number of modifications can be made to the original counterpropagation algorithm of *Hecht-Nielsen* [1987] that allow greater flexibility for predicting spatially distributed parameters. *Rizzo and Dougherty* [1994] showed that for the special case where the input vectors are normalized to values between 0 and 1, and the Kohonen weights are preprocessed to equal the normalized input training vectors, an L2 normalization procedure can be applied and the network acts as a nearest-neighbor classifier. In this special case, the Euclidean distance may be used to compute the similarity between the normalized input vector

and each of the normalized Kohonen weight vectors. The minimum value of $z_j^{n*} = \sqrt{\sum_{i=1}^I (x_i^n - w_{ij})^2}$ for $j = 1, 2, \dots, J$ is then used to select the winning hidden node as opposed to the dot product of equation (1). Preprocessing the Kohonen weights in this manner eliminates the need to generate initial random Kohonen weights by setting the total number of hidden nodes J equal to the number of input patterns N . Although, this alternative is convenient when the input predictor variables have the same dimensions (e.g., comprising only spatial locations), the strength of the counterpropagation algorithm is the sequential combination of the Kohonen (SOM) layer acting as a nearest k -means classifier and the Grossberg layer's ability to approximate a Bayes classifier that enables the counterpropagation to statistically approximate the nonlinear relationships inherent in the sample data [*Rizzo and Dougherty*, 1994]. The k -means classifier is a relatively flexible classifier (compared to discriminant analysis and logit models) and does not require an assumption of multivariate normality [*Cover and Hart*, 1967].

[20] In addition, an alternative to the single winner-take-all activation function is to allow some select number of nodes greater than one (say k) to become activated during the prediction phase, and compute a network output based on a weighted combination of the k winners to produce a smoother estimation field (e.g., estimating values of concentration as opposed to classes of soil lithology).

2.4. Modifications/Extensions to the Counterpropagation Algorithm

[21] The original counterpropagation algorithm, as presented by *Hecht-Nielsen* [1987, 1988], is a stochastic estimation technique that is not well suited for applications containing variables that exhibit spatial autocorrelation. One of our goals was to modify the algorithm to generate equiprobable realizations to capture the spatial structure associated with the primary variable of interest. To accomplish this, we incorporate a radial basis function after the training phase and prior to the prediction phase. This modification involves an adjustment to the Grossberg weights prior to the prediction phase and requires that the spatial structure be analyzed a priori using traditional geostatistical methods (e.g., semivariogram analysis) to define the range of spatial correlation. If the semivariogram is well behaved and a pattern of anisotropy is identifiable from the available geologic information (e.g., secondary variables), one can incorporate the directions of maximum and minimum continuity using directional semivariograms.

[22] We incorporate these concepts into the ANN by centering a radial basis function over each of the winning (trained) Kohonen vectors. The radial basis function is defined as $\varphi(d) = \exp\left(-\frac{d^2}{2\sigma^2}\right)$, where d is a measure of distance from the center of the function and σ^2 is the variance and defines the spread of the function. In this work, we let $\varphi(d)$ equal the probability that "nearby" unused Kohonen weights (those not adjusted during the training procedure) be assigned the same value (output class) as with the winning Kohonen weight vector lying at the center of the radial basis function. This results in a probability of $\varphi(d) = 1$ when $d = 0$ and $\varphi(d) = 1/K$ when $d =$ the range of decorrelation (a), where K is the total number

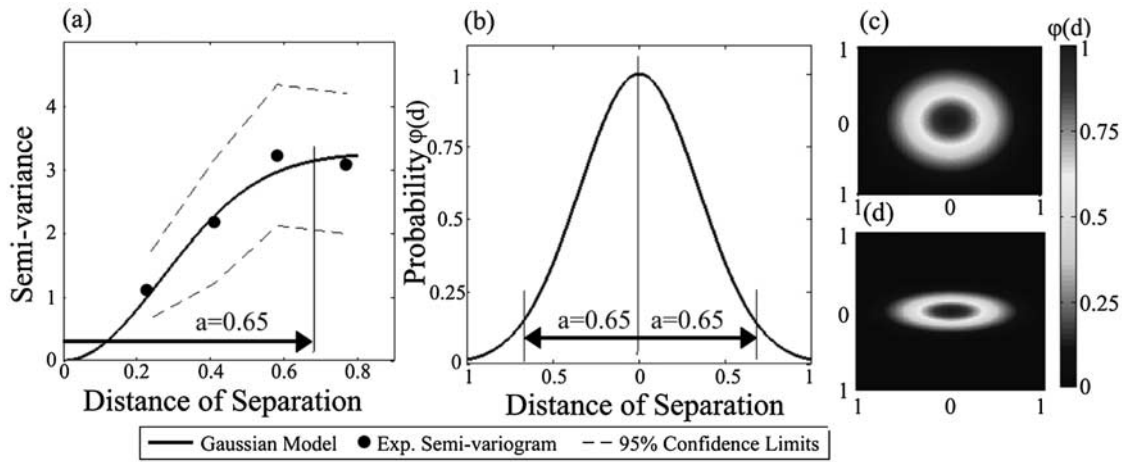


Figure 2. (a) Experimental semivariogram and best fit Gaussian model generated with 16 classified electrical resistivity observations of Figure 4a. Radial basis function with $a = 0.65$ shown in (b) cross-sectional and (c) plan view. (d) Example of radial basis function with axes of maximum and minimum continuity in the easting and northing directions, respectively ($a_{\text{easting}} = 0.65$ and $a_{\text{northing}} = 0.25$).

of classifications predefined by the user. The latter ensures that the classification, associated with untrained weights located beyond the range of decorrelation, will be a random outcome. Thus, by letting $\phi(d = a) = 1/K$, one may solve the radial basis function for the variance σ^2 . With σ^2 defined, the probability of classification $\phi(d)$ may be determined for any distance of separation d . Figure 2 presents a semivariogram and corresponding radial basis function that will be used in a later example.

[23] If evidence suggests the spatial continuity is not the same in all directions, the radial basis function may be rewritten to account for anisotropy along any perpendicular

axis as $\phi(d) = \exp\left(-\frac{1}{2}\left[\frac{a_{\text{major axis}}^2}{\sigma_{\text{major axis}}^2} + \frac{a_{\text{minor axis}}^2}{\sigma_{\text{minor axis}}^2}\right]\right)$. Figure 2d plots the radial basis function where the axes of maximum (major axis) and minimum (minor axis) continuity are defined as $a_{\text{major axis}} = 0.65$ and $a_{\text{minor axis}} = 0.25$. The directional variance terms are determined as presented previously with recognition that the two axes are independent (i.e., solve for $\sigma_{\text{major axis}}^2$ by setting $a_{\text{minor axis}} = 0$ and vice versa).

[24] Once defined, the radial basis function is used to assign values (output classifications) to the Grossberg weight vectors \mathbf{v}_j associated with Kohonen weight vectors that were not adjusted during the training phase. One can think of the radial basis function as a moving window centered on successive activated Kohonen weights whose size is defined by the spatial structure of the parameter of interest.

3. Results

[25] To investigate and test the ability of the modified counterpropagation ANN to generate equiprobable realizations and estimate highly resolved spatially distributed images using multiple data types, we use three types of geophysical data collected on a slab of Berea sandstone. Berea sandstone has been used by other researchers to investigate spatial prediction methods and properties of porous media. *Tidwell and Wilson* [1999, 1997] used it to investigate the upscaling of permeability measurements.

Journal and Alabert [1989] used Berea sandstone to compare multiple prediction methods including kriging, indicator simulation and Gaussian simulation. *Goovaerts* [1999] compared several conditional simulation techniques using a limited number of air permeability measurements from a slab of Berea sandstone.

[26] A $33 \times 35 \times 14.5$ cm slab of finely bedded Berea sandstone (Figure 3), from the Amherst quarry, Ohio, was collected and analyzed for several geophysical properties. Air permeability (milliDarcy, mD), compressional wave velocity (m/s) and electrical resistivity (ohm-m) were measured at very fine resolution (~ 3 mm grid spacing) and are displayed in Figures 3a, 3c and 3e. The corresponding cross scatterplots are shown in Figures 3b, 3d and 3f. This high-resolution data provides the unique opportunity to apply and compare stochastic simulation and spatial estimations to “reality” (the observed measurements). For details of the data collection methods, the reader is referred to G. N. Boinott et al. (Physical based upscaling of heterogeneous porous media: An illustrated example using Berea sandstone, submitted to *Petrophysics*, 2007) and New England Research, Inc., http://newenglandresearch.com/pdf/NER_AutoScanII.pdf.

3.1. Stochastic Simulation With Original Counterpropagation

[27] As mentioned in subsection 2.3.2, *Rizzo and Dougherty* [1994] showed that for the special case of $J = N$ (e.g., the number of hidden Kohonen weight vectors equals the number of input training vectors), the counterpropagation algorithm acts as a nearest-neighbor classifier. However, when $J > N$, the counterpropagation algorithm, approximates a nearest k -means classifier for which the generation of stochastic conditional simulations appears to be a natural by-product. This is demonstrated on a simple two-dimensional spatial approximation problem. Figure 4a depicts a portion of the original electrical resistivity (ohm-m) field as measured on the Berea sandstone (Figure 3e, region of interest (ROI)). The measured values have been preclassified into $K = 6$ categories. Sixteen point observations, open circles of Figure 4a, were selected randomly from the

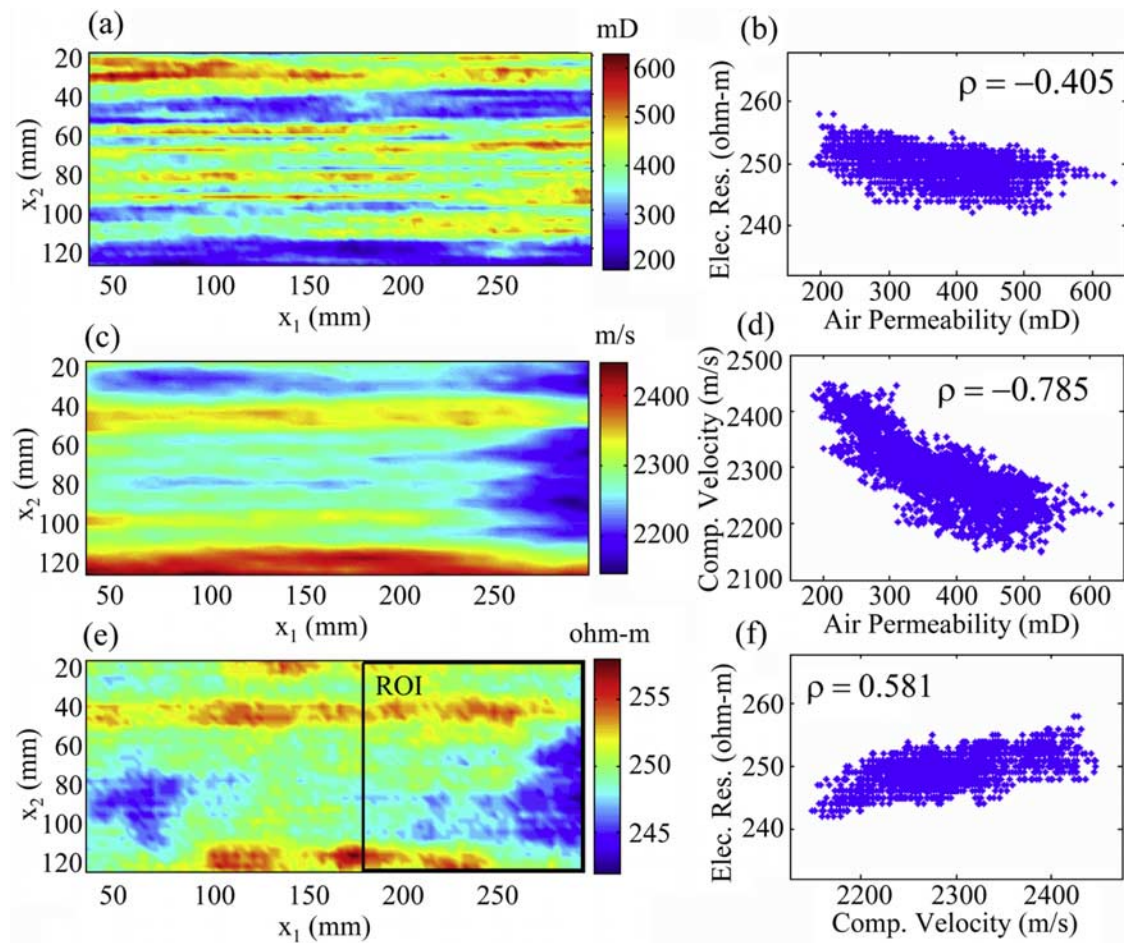


Figure 3. Geophysical properties of Berea sandstone: (a) air permeability (mD), (c) compressional wave velocity (m/s), and (e) electrical resistivity (ohm-m) along with the corresponding cross scatterplots for (b) electrical resistivity–air permeability, (d) compressional wave velocity–air permeability, and (f) electrical resistivity–compressional wave velocity.

classified electrical resistivity field to be used as training patterns.

[28] For this example application, the network inputs consist of two-dimensional, spatial coordinates (x_1 = easting and x_2 = northing) normalized between 0 and 1 for each of the 16 point observations. The ANN architecture (Figure 4b), comprising two input nodes ($I=2$), 24 hidden nodes ($J=24$), and six output nodes ($K=6$), is used for stochastic simulation of the classified electrical resistivity. The two-dimensional Kohonen weight vectors, associated with each of the 24 hidden nodes have been initialized to random values between 0 and 1 and are plotted (as crosses in Figure 5a) along with the 16 point observations.

[29] During each training iteration, the set of 16 training patterns are presented to the network in series to map the relationship between spatial coordinates (x_1 , x_2) and the associated classified electrical resistivity. The initial and final configuration of the Kohonen weight vectors are displayed in Figures 5a and 5b, respectively. Note that only 14 of the 24 Kohonen weight vectors have been adjusted (moved closer) to the 16 point observations.

[30] The first simulation (Figure 5c) shows the location of the final 14 adjusted weight vectors and resulting classified electrical resistivity field that respects the point observations. The 10 unadjusted weight vectors, located further

from the point observations, result in random estimates when activated during the prediction phase. Different equiprobable realizations (Figure 5d) are generated by changing the initial random Kohonen and Grossberg weight matrices and training the network as previously described. The expected classified electrical resistivity field (Figure 5e) was generated by averaging the estimates for each spatial location over the (39×45) grid across 100 simulations. The probability associated with each point being classified as shown in Figure 5e, is provided in Figure 5f. The probability is high (shown in blue) near the 16 point measurements and decreases with distance from the point measurements (toward red).

[31] The selection of the number of hidden nodes J plays an important role in this stochastic estimation procedure. In the example of Figure 5 (where $J=24$), 14 of the 24 hidden nodes are adjusted (moved) closer to the 16 point measurements; and the remaining 10 unadjusted hidden nodes are available to produce random estimates of classified electrical resistivity. Increasing the number of hidden nodes (say from $J=24$ to $J=64$), results in realizations that are less correlated (more variable) in space because of the increased number of unadjusted hidden nodes (those that produce random estimates when activated). Although the resulting

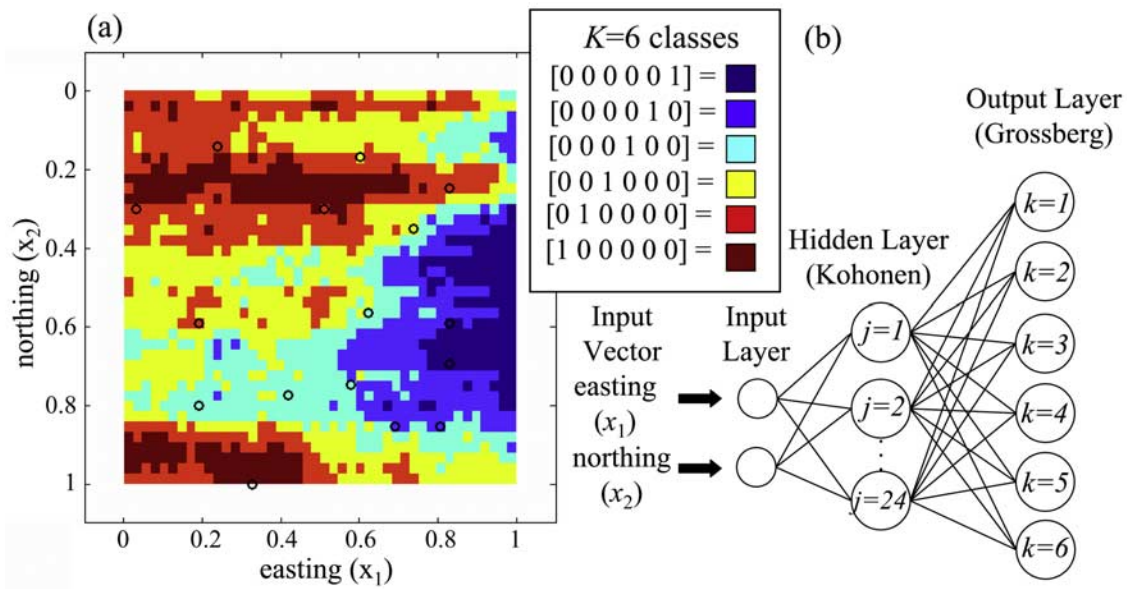


Figure 4. (a) Magnified representation of the electrical resistivity region of interest (Figure 3c) classified into six categories and (b) ANN architecture for performing stochastic conditional simulation. The network contains 2 input nodes, 24 hidden nodes, and 6 output nodes.

100 simulations have approximately the same frequency distribution as the observed data (results not shown) and honor the measured data values at known observation points, individual realizations (e.g., Figure 5c) do not exhibit the spatial structure associated with the observed

data (Figure 4a). Because of our focus on environmental subsurface characterization and Earth science applications that deal with spatially autocorrelated data, as well as, our desire to eliminate the need for an a priori user-defined number of hidden nodes, we have modified the original

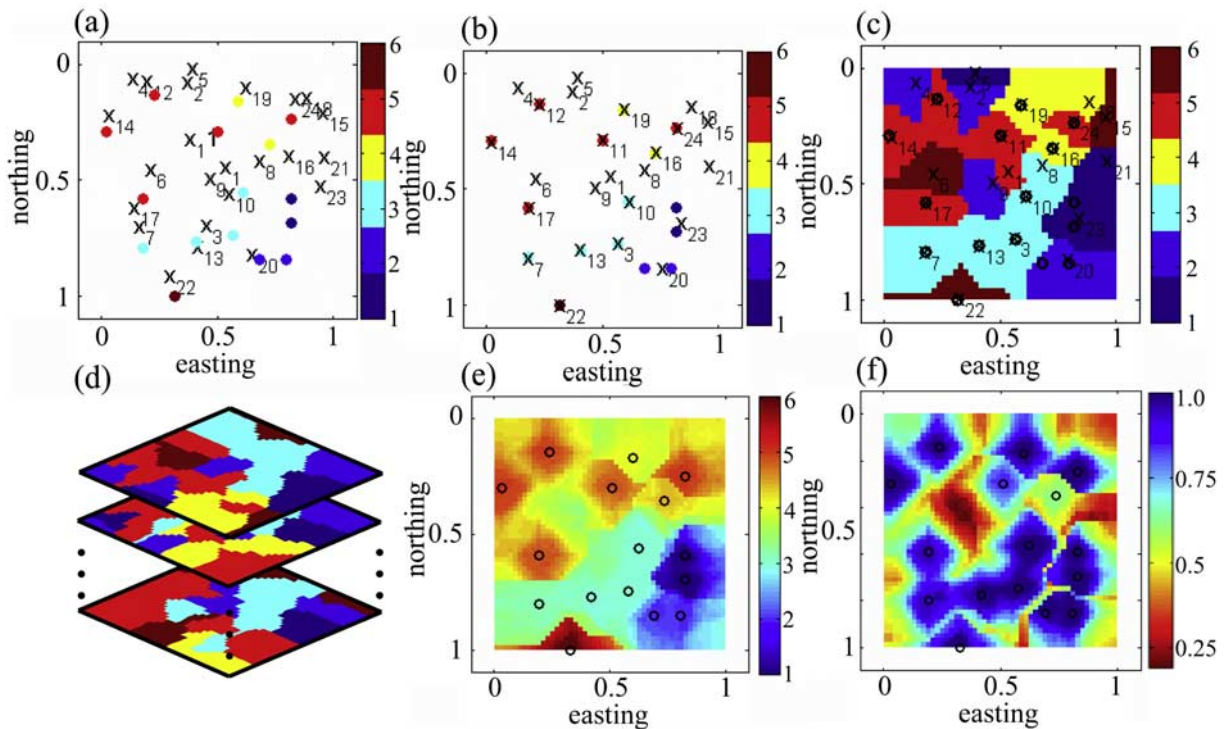


Figure 5. Spatial configuration of the 16 randomly selected classified electrical resistivity observations and the 24 Kohonen weight vectors (a) before and (b) after the training phase. (c) Single simulation associated with the Kohonen weight configuration of Figure 5b. (d) Diagram showing that 100 simulations were generated to produce (e) the expected field. (f) Probability associated with each point being classified as shown in Figure 5e.

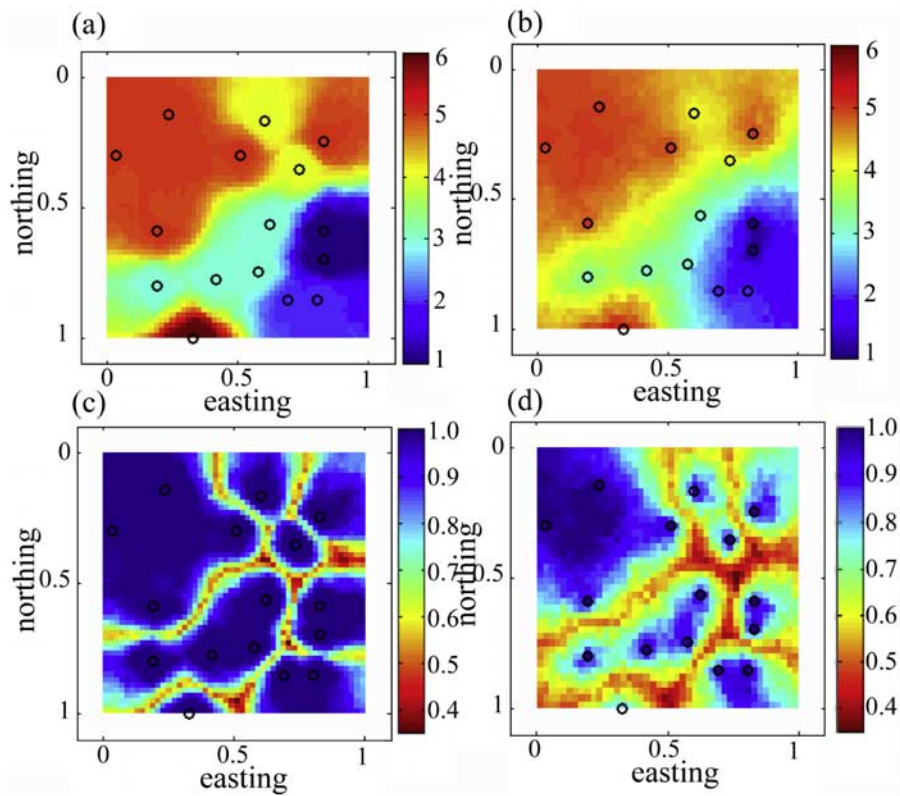


Figure 6. Expected classified electrical resistivity fields produced by (a) ANN and (b) sequential indicator simulation. Fields displaying the probability of classification as determined by (c) ANN and (d) sequential indicator simulation.

counterpropagation algorithm by incorporating the radial basis function prior to the prediction phase.

3.2. Conditional Simulation Using the Modified Counterpropagation

[32] We reuse the same 16 classified electrical resistivity observations (Figure 4a) to demonstrate the ability of the counterpropagation algorithm, modified with the radial basis function, to perform stochastic conditional simulation. A semivariogram analysis has been used to describe the spatial structure of the classified data. The omnidirectional, experimental semivariogram using the 16 classified point observations and the best fit ($R^2 = 0.959$) Gaussian model with range $a = 0.65$, nugget = 0 and sill = 3.25 are shown in Figure 2a. The 95% confidence limits are shown as dashed lines.

[33] Figures 2b and 2c display the corresponding radial basis function $\varphi(d)$ in cross section and plan view, respectively. The variance of the radial basis function is determined by substituting the range $a = 0.65$ for d and $\varphi = 1/K = 1/6$ into the radial basis function equation and solving for σ^2 . After training, and just prior to the execution of the prediction phase, this radial basis function is centered over each of the adjusted Kohonen weight vectors (crosses Figure 5b that have moved closer and become associated with a known observation point). The function is used to assign each of the unadjusted Kohonen weight vectors (crosses in Figure 5b lying within the radial basis function window) a probability value $\varphi(d)$. The trained (mapped) Grossberg weights v_j associated with the unadjusted Koho-

nen weights (those not associated with a known observation and shown as crosses) are assigned (with probability $\varphi(d)$) the same value (or class) as the observation at the center of the radial basis function. This probability is based on the distance d from the observation at the center of the function. Weights located beyond the range of decorrelation will produce random classifications.

[34] The modified counterpropagation algorithm was used to produce an additional 100 equiprobable realizations, where the total number of hidden nodes J for each of the 100 realizations was selected randomly from the range $2N$ to $10N$. Each equiprobable realization is the result of different initial random weight vectors and subsequent training and prediction. The resulting expected classified electrical resistivity field (Figure 6a) and the associated probability field (Figure 6c) are compared with those generated using sequential indicator simulation (SIS). Figures 6b and 6d display the associated expected SIS electrical resistivity and probability fields (across 100 realizations) implemented using the geostatistical software GSLIB (Geostatistical Software Library and User's Guide, Statios, WinGslib 1.4).

[35] Table 1 shows the global descriptive statistics (i.e., mean, median, standard deviation), residual correlation coefficients (0.67 and 0.73) and RMSE values (1.1 and 1.0) to be similar for the expected ANN and SIS estimation fields. In addition, we performed a comparison with a field (image not shown) estimated using ordinary kriging (OK). All three estimation fields honor the data at the observation locations. We also quantified the spatial structure (e.g.,

Table 1. Global and Spatial Performance Measures for the Classified Real Field of Figure 4a, 16 Point Observations, and Fields Estimated Using the Method of OK, the Conditional Simulation ANN, and the SIS

	Reality	16 Observations	OK	ANN	SIS
<i>Global Measures</i>					
Mean	3.8	3.6	3.9	3.8	3.8
Median	4	3.5	4.1	4.1	4.2
Mode	4	5	5	5	5
Standard deviation	1.4	1.6	1.1	1.3	1.1
Skewness	-0.3	-0.3	-0.8	-0.6	-0.8
RMSE	0	NA ^a	1	1.1	1
Residual ρ	NA	NA	0.73	0.67	0.73
<i>Spatial Measures</i>					
Range(a)	0.7	0.65	0.67	0.72 ± 0.06	0.83 ± 0.12
Sill	2.6	3.25	1.75	2.9 ± 0.2	3.0 ± 0.8
Nugget	0.09	0	0.02	0	0

decorrelation lengths, variances and nuggets) for each of the 100 realizations to ensure they exhibit spatial structure similar to that of the 16 observation data. The ANN and SIS methods produce fields with a mean range of decorrelation of 0.72 ± 0.06 and 0.83 ± 0.12 respectively; while the field estimated by OK has a range of decorrelation of 0.67, which is most similar to the model describing the 16 observation data ($a = 0.65$).

[36] The expected ANN and SIS estimates in Figure 6 are represented as real numbers. In practice, this may have little physical meaning (i.e., for a categorical parameter such as soil lithology); and one might round values to the nearest integer. However, if the parameters are better represented by real numbers (i.e., concentration), postprocessing is not performed to preserve information. To evaluate bias with the original quantized field, all fields were quantized (post-processed by rounding, to the nearest integer) to represent the same 6 classes of the original reality field. These fields could be postprocessed into any number of classes depending on the application and the data at hand.

[37] An evaluation of bias between reality, the sampled data, and the estimated OK, ANN, and SIS fields is reported in Table 2. The 16 point observations that make up the sampled field were chosen at random with no attempt to reproduce the bias in the original reality field. As a result, the sample field is biased toward electrical resistivity classes 1, 3, and 5 with less of classes 4 and 6 than are present in the reality field. The OK estimate is biased toward classes 4 and 5 with fewer of class 1, 2 and 6 than the sample or

reality field. The SIS field is biased toward classes 2, 4 and 5 with fewer classes 1 and 6 than either the sampled data or real field. The ANN shows less bias for 5 out of the 6 classes than either OK or SIS when compared to the sampled field, and less bias for 4 of the 6 classes when compared with the original reality field.

3.3. Spatial Estimation Using Multiple Data Types

[38] In this section, we show the ease with which the counterpropagation ANN is able to estimate parameter fields using multiple data types. In the first application, we use all three types of geophysical data collected on the slab of Berea sandstone and provide a comparison with the method of ordinary cokriging. In the second application (section 3.3.2), we showcase the ability of the ANN methodology to assimilate a much larger number of data types by estimating the extent of landfill leachate using a total of seven data types.

3.3.1. Berea Sandstone

[39] We estimate the air permeability field associated with the Berea sandstone (Figure 3a) using a limited subset of air permeability measurements as the primary data and correlated electrical resistivity and compressional wave velocity as secondary data. To provide a realistic scenario of the challenges associated with the unique spatial distribution of data collected during real subsurface investigations (i.e., approximately continuous data in the vertical direction and relatively sparse data in the horizontal direction), ten hypothetical well borings were placed randomly along the cross section of Berea sandstone, see Figure 7a. Measured air permeability was assumed to be known only in the vertical direction (at ~ 3 mm intervals) along the well screens (black bars of Figure 7a) for a total of 46 known observations. Electrical resistivity and compressional wave velocity were assumed to be known along the entire vertical length of the well borings for a total of 380 measurements of each data type. These secondary variables might be considered analogous to the more abundant (and perhaps less expensive) information (e.g., descriptions of grain size or classifications of soil lithology from drillers' well logs). Error free measurements of the primary and secondary data are assumed. The autocorrelated and cross-correlated spatial structure of the three data types were analyzed using semivariograms and cross semivariograms.

[40] Two counterpropagation ANNs executed in series were used to estimate a parameter field using these multiple data. The estimates from the first network were used as the input training data for the second network. The first ANN in series, Figure 8b, is trained using the 46 air permeability measurements collected along the well screens of Figure 8a.

Table 2. Evaluation of Bias for Quantized Estimated Electrical Resistivity Fields

Class	Reality		16 Observations		OK		ANN		SIS	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	147	8.4	2	12.5	59	3.4	131	7.5	27	1.5
2	210	12.0	2	12.5	192	10.9	219	12.5	280	16.0
3	320	18.2	4	25.0	325	18.5	327	18.6	264	15.0
4	493	28.1	2	12.5	568	32.4	341	19.4	595	33.9
5	401	22.8	5	31.3	603	34.4	706	40.2	586	33.4
6	184	10.5	1	6.3	8	0.5	31	1.8	3	0.2

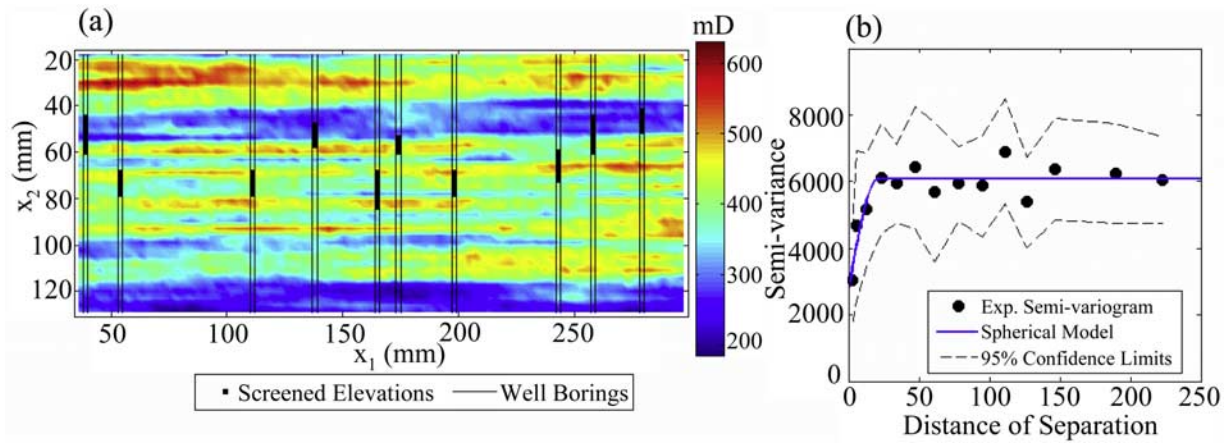


Figure 7. (a) Measurements of air permeability (mD) on Berea sandstone and (b) semivariogram describing spatial structure of the 46 air permeability observation at 3 mm intervals along the 10 vertical well screens (black bars of Figure 7a).

These training patterns consist of the spatial locations (x_1, x_2) and observed secondary variables (electrical resistivity and compressional wave velocity) and the corresponding output variable (air permeability). As a result, the architecture of the first ANN consists of 4 input nodes, 46 hidden nodes (Kohonen weights are preprocessed to equal the normalized input vectors) and 511 output nodes. The 511 output nodes correspond to the user-specified number of classifications selected to represent air permeability. We selected 511 classifications (1) because the largest air permeability value observed in the measured data set was 511 mD, and (2) to demonstrate that the algorithm is not constrained to a small number of output classes. In reality, we might expect the user to prespecify a number of classifications that seems reasonable for the problem and data type at hand (e.g., in the case of hydraulic conductivity, the user may wish to have significantly less output classifications as the distinction between 18 and 28 cm/hr may not be meaningful). The input data associated with the prediction phase of the first ANN contain $M = 380$ vectors corresponding to the spatial locations along the 10 well borings where x_1, x_2 , electrical resistivity and compressional wave velocity are known, and predictions of air permeability are desired (unscreened portion of the well boring). The resulting ANN estimates of air permeability along the well borings are shown in Figure 8c.

[41] The second ANN in the series is trained using the 380 estimates from the first ANN; and the corresponding architecture (Figure 8d) comprised 2 input nodes (x_1 and x_2), 380 hidden nodes (Kohonen weights are again preprocessed to equal the normalized input vectors) and 511 output nodes. Once trained to map the 380 spatial vectors to the corresponding estimates of air permeability, the ANN may be used to predict air permeability everywhere within the sandstone boundaries (Figure 8e).

[42] The estimated ANN field is compared to a field generated by the method of ordinary cokriging (Figure 8f) in GSLIB (Geostatistical Software Library and User's Guide, Statios, WinGslib 1.4). The spatial structure of the air permeability measurements was best fit ($R^2 = 0.83$) by an exponential model with range, $a = 18$ mm, nugget = 1,364 $\text{ohm}^2\text{-m}^2$ and sill = 6,075 $\text{ohm}^2\text{-m}^2$; see Figure 7b.

The 4:1 anisotropy ratio associated with the reality field could not be discerned from the 46 measurements; however, it was used with the method of cokriging to better estimate the layering of the Berea sandstone. The semivariogram and cross semivariogram analyses for the secondary data are not shown. The residuals of the two fields (the ANN and ordinary cokriging) with the reference field (measurements minus the estimates) were computed and statistically examined for normality, central tendency (similar means) and dispersion (similar variance). The residuals for each field were found to be normally distributed and showed no statistical difference with respect to measures of central tendency and dispersion (type I error rate α of 0.05).

3.3.2. Banisveld Landfill

[43] To showcase the ability of the counterpropagation ANN to assimilate large numbers of multiple data types, we estimate the extent of leaking landfill leachate using electrical formation conductivity as the primary data and six additional secondary types of data from the Banisveld landfill located outside of Boxtel, Netherlands. The landfill is undergoing long-term monitoring and is equipped with state of the art monitoring equipment. This site has well characterized leachate plumes and has been sampled for numerous microbiological and hydrochemical properties; see Rölting *et al.* [2000a, 2000b, 2001]; and Mouser *et al.* [2005] for details. A principle components analysis was used to reduce the dimensionality of the data from 24 hydrochemistry variables, 24 archaea variables and 29 bacterial variables to a total of six principle components (2 for each data type) that explain 70%, 86% and 68% of the total variance; see Mouser *et al.* [2005] for details. The extent of the contamination using the first hydrochemistry principle component and the observed 5:1 anisotropy ratio is estimated along a cross section using the method of ordinary kriging (Figure 9a). Two counterpropagation ANNs used in series combine this state of the art microbial information, hydrochemistry data and electrical formation conductivity measurements (mS/m) into an estimate defining the extent of subsurface contamination, Figure 9b. The training input vectors for the first ANN in series have dimensions $[8 \times 13]$, where 8 is the number of components in the input training vectors (comprising 2 spatial coordi-

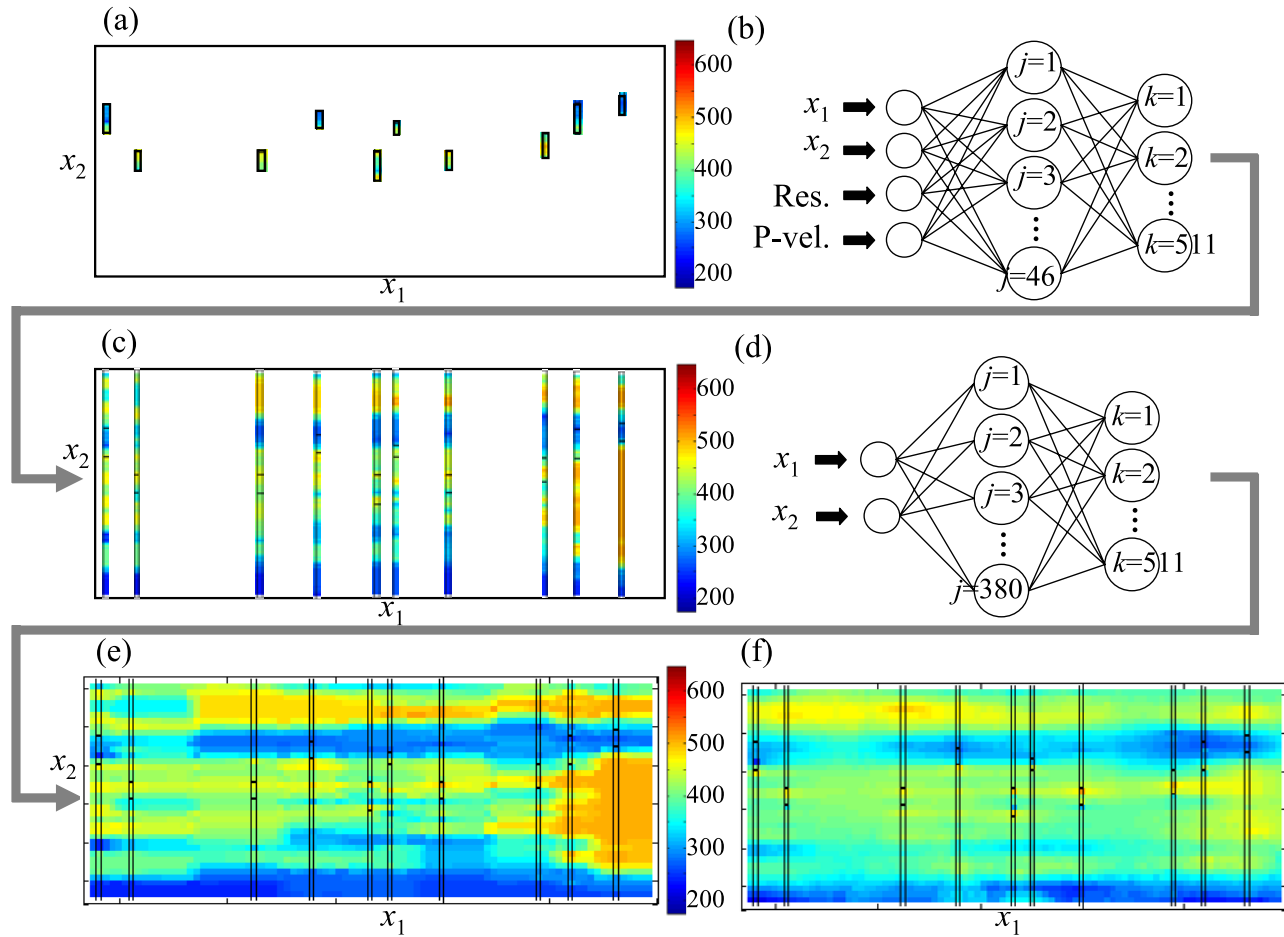


Figure 8. (a) Measured air permeability data from 10 randomly spaced well screens and (b) ANN architecture used to map the relationship between x_1 , x_2 , electrical resistivity, compressional wave velocity, and air permeability. (c) ANN predictions of air permeability at 380 spatial locations along the length of the well borings at which secondary data is considered known. (d) ANN architecture trained using predictions from first the ANN. Estimated air permeability fields (mD) using (e) sequential counterpropagation ANNs and (f) method of ordinary cokriging.

nates and the 6 above mentioned types of secondary data). Only 27 locations contain both primary and secondary data. Estimates are predicted for all 5,655 spatial locations at which only secondary information exists. Thus input training vectors for the second ANN in series have dimensions $[2 \times 5655]$, and estimates are predicted at 22,155 points in space to generate the cross section of Figure 9b. The total training time is ~ 6 min. (All timings were performed on an AMD Athlon 64 processor 3400+ (2.25 GHz) with 2.0 GB of RAM.)

4. Discussion

[44] In section 3.1, we use the original counterpropagation ANN as developed by *Hecht-Nielsen* [1987, 1988] to stochastically estimate maps of discrete spatially distributed fields (e.g., classified electrical resistivity). An analysis of the spatial structure revealed that the individual realizations do not possess the observed spatial autocorrelation and patterns of anisotropy associated with the primary variable of interest. To remedy this, and eliminate the need for an a priori user-defined number of hidden nodes, we modified

the original counterpropagation algorithm using a radial basis function.

[45] The 100 realizations of electrical resistivity were analyzed (section 3.2) to ensure that they approximate the known spatial correlation as well as the frequency distribution associated with the measured data. The resulting expected electrical resistivity field produced by averaging the 100 realizations (Figure 6) was compared with the methods of sequential indicator simulation and ordinary kriging. Each averaged field and each of the 100 individual realizations honors the measurement data at the known observation locations. An evaluation (Table 1) of the global statistics, spatial structure (section 3.2.) and bias (Table 2) show favorable comparisons between the three methods. We automated the process of best fitting the semivariograms for each of the realizations using the nonlinear regression platform in JMP 5.0.1.2. This method uses a least squares regression method that weights all semivariogram points equally instead of giving greater weight to points within the range of decorrelation as might be done if best fit by an expert. Despite this suboptimal method of fitting, the difference between the semivariogram parameters is not

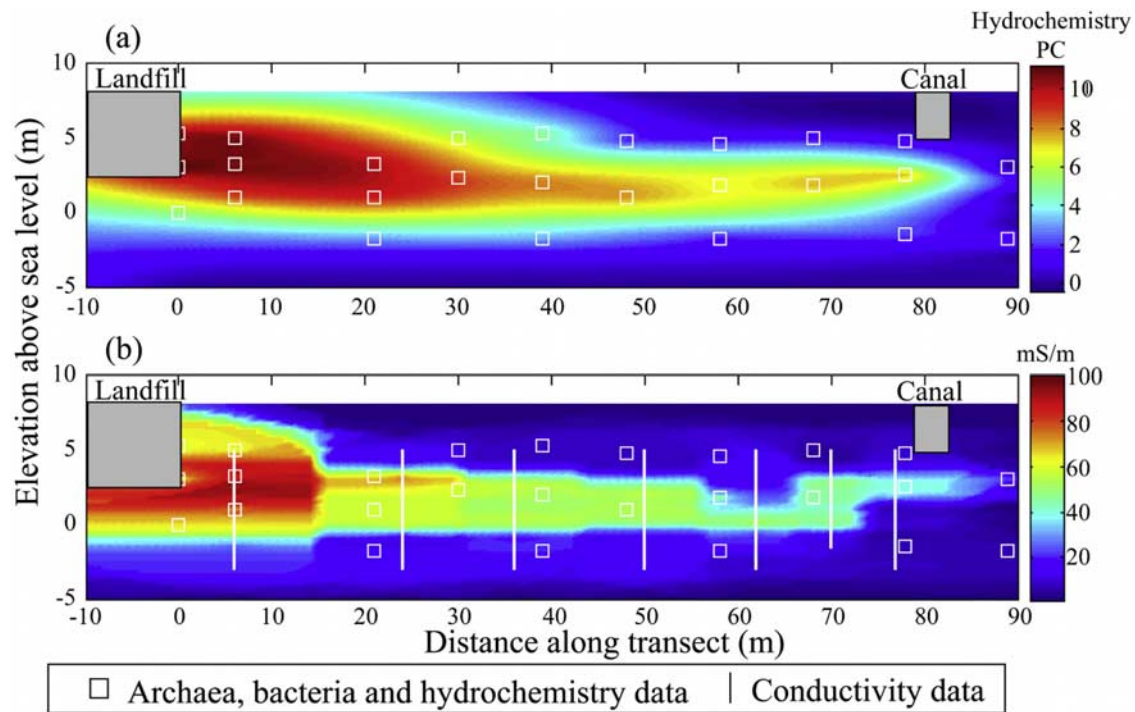


Figure 9. Cross-sectional estimate of landfill leachate contamination at the Banisveld landfill produced using (a) first-principle component of hydrochemistry data and the method of ordinary kriging and (b) first- and second-principle components of microbiological and hydrochemistry data and electrical formation conductivity (mS/m) with the proposed sequential ANNs.

statically significant. The ANN field is biased toward class 5 (the most abundant class sampled) with fewer of class 6 than the real or sampled field; however, it better captures the distribution of classes 1, 2, 3 and 6 than either OK or SIS.

[46] To further compare the three estimated fields, residuals (measurements minus the estimates) were computed and examined for normality, central tendency (similar means) and dispersion (similar variance). Normal probability plots reveal that the residuals associated with the OK and SIS fields were normally distributed, while the ANN field was not. As a result, nonparametric statistical tests were used to show that the residuals of the three estimation fields are not statistically different with respect to measures of central tendency (Wilcoxon Rank Sum test) and dispersion (Levene's and Brown-Forsythe tests) using a type I error rate α of 0.05.

[47] In addition to performing conditional simulation, we used two modified counterpropagation ANNs in tandem to estimate spatially distributed fields using multiple data types. The estimated air permeability field (Figure 8e), generated with 46 measured air permeability data and 380 observations of compressional wave velocity and electrical resistivity, respects the known data. In the vicinity of an observation, the estimates closely resemble a nearest-neighbor classifier, and the probability that a point will be classified similarly to its nearest-mean Kohonen weight vector decreases as the distance from the data approaches the maximum range of decorrelation found in the primary data variable. This estimated field compares well with the method of ordinary cokriging (Figure 8f). Although the limited number of observation data did not suggest the 4:1

anisotropy ratio or spatial structure observed in the reality field, we used the known structure in the method of cokriging to produce the best possible estimate for comparison. Which method provides a "better" estimate may depend on the type of data the user wishes to estimate. Despite our inability to find any statistical difference in the estimates provided by the two methods, the kriging estimates appear superior when smoothing of the data is preferred (e.g., concentration data). On the other hand, if there is a preference for the primary variable to be layered or blocky, the ANN estimates appear superior to the kriging methods.

[48] In section 3.3.2, we estimate leaking landfill leachate at the Banisveld landfill in Boxtel, Netherlands to demonstrate the ease with which the ANNs in tandem can assimilate large numbers of multiple data types. This example involved the assimilation of multiple types of hydrochemistry, microbiological and subsurface electrical conductivity information (total of 7 data types) to delineate the landfill leachate plume along a well characterized two-dimensional cross section. Despite the recognition that the assimilation of disparate field data can both improve parameter estimates, we cannot say with any degree of certainty whether the estimate of Figure 9b, which uses matrix multiplication to incorporate 7 data types, is better or worse than the ordinary kriging estimate using only the hydrochemistry data. We were unable to assimilate 7 data types with traditional geostatistical methods and present this application to showcase the ease (e.g., the user simply adds additional columns of data to the ANN input vector) with which multiple data may be incorporated into the parameter

estimate. Training times for these two ANNs in series were on the order of ~ 6 min.

5. Conclusions

[49] We present a novel method for performing stochastic simulation using a modified counterpropagation algorithm. The method is capable of estimating maps that (1) respect the measured data at the known observations locations, (2) have minimal conditional and/or global bias, and (3) approximate the known spatial structure, as well as the univariate statistical distribution (or frequency distribution in the case of categorical variables) associated with the primary data variable.

[50] In addition, a data-driven spatial approximation method using two modified counterpropagation ANNs in tandem has been introduced, applied and tested to show the ease with which a method (using matrix multiplication) can estimate spatially distributed fields directly from multiple data types without the computational burden associated with the construction and inversion of positive definite covariance and cross-covariance matrices. The strength of the proposed methodology lies in its ability to deal with large numbers of multiple data types (amounts that prohibit the use of traditional methods). The method does require the spatial structure of the primary data be analyzed a priori (e.g., semivariograms analysis); however, this structure is usually well defined (and easy to compute for only one variable type) when large amounts of data are available. Comparison of the global statistics, spatial structure, and bias with traditional kriging methods show no statistical difference between the methods. The training times required for the two-stage Berea sandstone estimation took on the order of seconds (and less than a second for each of the individual realizations in the conditional simulation section), and our coding of the algorithm in MatLab V. 7.3.0.267 (R2006b) does not take full advantage of the parallel nature of the ANN architecture. The method can be readily applied in one-, two-, or three-dimensional mappings; and the combination of simplicity and computational speed make the method ideally suited for performance assessment, environmental site characterization, and other Earth science applications that deal with spatially autocorrelated data.

Notation

\mathbf{x}^n	n^{th} input training vector of I components.
x_i^n	i^{th} component of n^{th} training vector.
\mathbf{y}^n	target vector associated with the n^{th} input training vector.
N	total number of training patterns.
I	total number of nodes in the input layer.
J	total number of nodes in the hidden layer.
K	total number of nodes in the output layer.
\mathbf{w}	Kohonen weight matrix.
\mathbf{v}	Grossberg weight matrix.
w_{ij}	Kohonen weight connecting i^{th} input node and j^{th} hidden node.
\mathbf{z}^{n*}	similarity vector comparing J hidden nodes and n^{th} input vector.
\mathbf{w}_j	Kohonen weight vector associated with j^{th} hidden node.

α	Kohonen weight learning coefficient.
\mathbf{z}^n	similarity vector after activation function.
\mathbf{v}_j	Grossberg weight vector associated with j^{th} output node.
β	Grossberg weight learning coefficient.
\mathbf{x}^M	M^{th} input predictor vector of I components.
$\hat{\mathbf{y}}^M$	M^{th} prediction/classification vector.
M	total number of vectors for which an approximation is desired.
$\varphi(d)$	probability of hidden nodes having similar output classifications.
d	distance between two Kohonen weights vectors.
σ^2	variance of radial basis function.
a	semivariogram range of decorrelation.

[51] **Acknowledgments.** This work was funded through the NSF Vermont EPSCoR (Experimental Program to Support Collaborative Research) grants for graduate research assistantship. We thank David E. Dougherty, P. Goff, P. Mouser, and L. Zhiqiang for comments on this manuscript. Special thanks go to New England Research, Inc., for supplying the geophysical data and their time in support of this study. We also thank Laurence Bentley of the University of Calgary for suggestions that strengthened the comparisons with the kriging methods.

References

- Azerman, M. A., E. M. Braverman, and L. I. Rozoner (1964), Theoretical foundations of the potential function method in pattern recognition learning, *Autom. Remote Control*, 25, 821–837.
- Ballard, J. H., and M. J. Cullinane Jr. (1998), Innovative Site Characterization and Analysis Penetrometer System (SCAPS): In-situ sensor and sampling technologies, paper presented at SAGEEP 1998: Symposium and the Application of Geophysics to Environmental and Engineering Problems, Soc. of Eng. and Miner. Explor. Geophys., Denver, Colo.
- Bashkurov, O. K., E. M. Braverman, and I. B. Muchnik (1964), Potential function algorithms for pattern recognition learning machines, *Autom. Remote Control*, 25, 629–631.
- Bosch, M. (2004), The optimization approach to lithology tomography: Combining seismic data and petrophysics for porosity prediction, *Geophysics*, 69(5), 1272–1282.
- Copt, N., and Y. Rubin (1995), A stochastic approach to the characterization of lithofacies from surface seismic and well data, *Water Resour. Res.*, 31(7), 1673–1686.
- Copt, N., Y. Rubin, and G. Mavko (1993), Geophysical-hydrological identification of field permeabilities through Bayesian updating, *Water Resour. Res.*, 29(8), 2813–2825.
- Cover, T. M., and P. E. Hart (1967), Nearest neighbor pattern classification, *IEEE Trans. Inf. Theory*, 13(1), 21–27.
- de la Vega, M., A. Osella, and E. Lascano (2003), Joint inversion of Wenner and dipole-dipole data to study a gasoline-contaminated soil, *J. Appl. Geophys.*, 54(1–2), 97–109.
- Fidencio, P. H., I. Ruisanchez, and R. J. Poppi (2001), Application of artificial neural networks to the classification of soils from Sao Paulo state using near-infrared spectroscopy, *Analyst*, 126(12), 2194–2200.
- Garcia, L. A., and A. Shigidi (2006), Using neural networks for parameter estimation in ground water, *J. Hydrogeol.*, 318, 215–231.
- Gelb, S., and J. Wonder (1998), ESC demonstration: D-area oil seepage basin—Savannah River Site: A case study, paper presented at SAGEEP 1998: Symposium and the Application of Geophysics to Environmental and Engineering Problems, Soc. of Eng. and Miner. Explor. Geophys., Denver, Colo.
- Gloaguen, E., M. Choteau, D. Marcotte, and R. Chapuis (2001), Estimation of hydraulic conductivity of an unconfined aquifer using cokriging of GPR and hydrostratigraphical data, *J. Appl. Geophys.*, 47(2), 135–152.
- Goovaerts, P. (1998), Geostatistical tools for characterizing the spatial variability of microbiological and physico-chemical soil properties, *Biol. Fertility Soils*, 27(4), 315–334.
- Goovaerts, P. (1999), Impact of the simulation algorithm magnitude of ergodic fluctuations and number of realizations on the spaces of uncertainty of flow properties, *Stochastic Environ. Res. Risk Assess.*, 13(3), 182–191.
- Goovaerts, P. (2001), Geostatistical modeling of uncertainty in soil science, *Geoderma*, 103(1–2), 3–26.

- Goovaerts, P., G. AvRuskin, J. Meliker, M. Slotnick, G. Jacquez, and J. Nriagu (2005), Geostatistical modeling of the spatial variability of arsenic in groundwater of southeast Michigan, *Water Resour. Res.*, *41*, W07013, doi:10.1029/2004WR003705.
- Hecht-Nielsen, R. (1987), Counterpropagation networks, *Appl. Opt.*, *26*(23), 4979–4984.
- Hecht-Nielsen, R. (1988), Applications of counterpropagation networks, *Neural Networks*, *1*, 131–139.
- Hohmann, G. W., and A. P. Raiche (1987), Inversion of controlled-source electromagnetic data, *Electromagn. Methods Appl. Geophys.*, *1*, 469–503.
- Hopfield, J. J. (1982), Neural networks and physical systems with emergent collective computational abilities, *Proc. Natl. Acad. Sci. U. S. A.*, *79*, 2554–2558.
- Hsu, K., H. V. Gupta, and S. Sorooshian (1995), Artificial neural network modeling of the rainfall-runoff processes, *Water Resour. Res.*, *31*(10), 2517–2530.
- Hsu, K., H. V. Gupta, X. Gao, and S. Sorooshian (1999), Estimation of physical variables from multichannel remotely sensed imagery using a neural network: Application to rainfall estimation, *Water Resour. Res.*, *35*(5), 1605–1618.
- Istok, J. D., and C. A. Rautman (1996), Probabilistic assessment of groundwater contamination 2: Results of case study, *Ground Water*, *34*(6), 1050–1054.
- Journel, A. G., and F. Alabert (1989), Non-Gaussian data expansion in Earth sciences, *Terra Nova*, *1*(2), 123–134.
- Kis, M. (2002), Generalised series expansion (GSE) used in DC geoelectric-seismic joint inversion, *J. Appl. Geophys.*, *50*(4), 401–416.
- Kohonen, T. (1989), *Self Organization and Associated Memory*, Springer, New York.
- Li, B. L., and T. C. J. Yeh (1999), Cokriging estimation of the conductivity field used variably saturated flow conditions, *Water Resour. Res.*, *35*(12), 3663–3674.
- Menke, W. (1984), The resolving power of cross-borehole tomography, *Geophys. Res. Lett.*, *11*, 105–108.
- Morshed, J., and J. J. Kaluarachchi (1998), Parameter estimation using artificial neural network and genetic algorithm for free-product migration and recovery, *Water Resour. Res.*, *34*(5), 1101–1113.
- Mouser, P. J., D. M. Rizzo, W. F. M. Röling, and B. M. Van Breukelen (2005), A multivariate statistical approach to spatial representation of groundwater contamination using hydrochemistry and microbial community profiles, *Environ. Sci. Technol.*, *39*(19), 7551–7559.
- Nie, J. H., A. P. Loh, and C. C. Hang (1996), Modeling pH neutralization processes using fuzzy-neural approaches, *Fuzzy Sets Syst.*, *78*(91), 5–22.
- Oldenburg, D. W., Y. Li, C. G. Farquharson, P. Kowalczyk, T. Aravanis, A. King, P. Zhang, and A. Watts (1998), Applications of geophysical inversions in mineral exploration, *Leading Edge*, *17*, 461–465.
- Park, Y. S., P. F. M. Verdonshot, T. S. Chon, and S. Lek (2003), Patterning and predicting aquatic macro-invertebrate diversities using artificial neural networks, *Water Res.*, *37*(8), 1749–1758.
- Patriarche, D., M. C. Castro, and P. Goovaerts (2005), Estimating regional hydraulic conductivity fields—A comparative study of geostatistical methods, *Math. Geol.*, *37*(6), 587–613.
- Pitkin, S. E. (1998), Detailed subsurface characterization using the Waterloo profiler, paper presented at Symposium on the Application of Geophysics to Environmental and Engineering Problems, Soc. of Eng. and Miner. Explor. Geophys., Denver, Colo.
- Purdy, C., and P. Beam (1998), Acceptance of the expedited site characterization (ESC) approach in environmental cleanups, paper presented at Symposium on the Application of Geophysics to Environmental and Engineering Problems, Soc. of Eng. and Miner. Explor. Geophys., Denver, Colo.
- Rizzo, D. M. (1994), Optimal groundwater remediation designs and characterization of aquifer properties using artificial neural networks, Ph.D. dissertation, Univ. of Vermont, Burlington.
- Rizzo, D. M., and D. E. Dougherty (1994), Characterization of aquifer properties using artificial neural networks: Neural kriging, *Water Resour. Res.*, *30*(2), 483–497.
- Rizzo, D. M., T. P. Lillys, and D. E. Dougherty (1996), Comparisons of site characterization methods using mixed data, in *Uncertainty in Geologic Environment*, edited by C. D. Shackelford, P. P. Nelson, and M. J. S. Roth, pp. 167–179, Am. Soc. of Civ. Eng., New York.
- Röling, W. F., B. M. Van Breukelen, and H. W. Verseveld (2000a), Linking microbial community structure to pollution: Biolog-substrate utilization in and near a landfill leachate plume, *Water Sci. Technol.*, *41*(12), 47–53.
- Röling, W. F., B. M. Van Breukelen, M. Braster, M. T. Groelton, J. Groen, and H. W. Verseveld (2000b), Analysis of microbial communities in a landfill leachate polluted aquifer using a new method for anaerobic physiological profiling and 16S rDNA based fingerprinting, *Microbiol. Ecol.*, *40*, 177–188.
- Röling, W. F., B. M. Van Breukelen, M. Braster, B. Lin, and H. W. Verseveld (2001), Relationships between microbial community structure and hydrochemistry in a landfill leachate-polluted aquifer, *Appl. Environ. Microbiol.*, *67*(10), 4619–4629.
- Rossabi, J., B. D. Riha, J. W. Haas, C. A. Eddy-Dilek, A. G. L. Kreeger, M. Carrabba, W. K. Hyde, and J. Bello (2000), Field tests of a DNAPL characterization system using cone penetrometer-based Raman spectroscopy, *Ground Water Monit. Rem.*, *20*(4), 72–81.
- Rumelhart, D. E., and J. L. McClelland (1988), *Parallel Distributed Processing*, Mass. Inst. of Technol. Press, Cambridge, Mass.
- Tidwell, V. C., and J. L. Wilson (1997), Laboratory method for investigating permeability upscaling, *Water Resour. Res.*, *33*(7), 1607–1616.
- Tidwell, V. C., and J. L. Wilson (1999), Permeability upscaling measured on a block of Berea sandstone: Results and interpretation, *Math. Geol.*, *31*(7), 749–769.
- U.S. Environmental Protection Agency (2000), Innovations in site characterization: Geophysical investigation at hazardous waste sites, Off. of Solid Waste and Emergency Response, Washington, D. C.
- Vozoff, K., and D. L. B. Jupp (1975), Joint inversion of geophysical data, *J. R. Astron. Soc.*, *42*, 977–991.
- Wasserman, P. D. (1989), *Neural Computing: Theory and Practice*, Van Nostrand Reinhold, New York.
- Widrow, G., and M. E. Hoff (1960), Adaptive switching circuits, Western Electronic Show and Convention, Inst. of Radio Eng., New York.
- Yeh, W. W.-G. (1986), of parameter identification procedures in groundwater hydrology: The inversion problem, *Water Resour. Res.*, *22*(2), 95–108.

L. E. Besaw and D. M. Rizzo, School of Engineering, University of Vermont, Votey Hall, 33 Colchester Avenue, Burlington, VT 05405-0156, USA. (lbesaw@cems.uvm.edu; drizzo@cems.uvm.edu)