Subsurface Characterization With Support Vector Machines

Brendt Wohlberg, Daniel M. Tartakovsky, and Alberto Guadagnini

***Abstract*—A typical subsurface environment is heterogeneous, consists of multiple materials (geologic facies), and is often insufficiently characterized by data. The ability to delineate geologic facies and to estimate their properties from sparse data is essential for modeling physical and biochemical processes occurring in the subsurface. We demonstrate that the support vector machine is a viable and efficient tool for lithofacies delineation, and we compare it with a geostatistical approach. To illustrate our approach, and to demonstrate its advantages, we construct a synthetic porous medium consisting of two heterogeneous materials and then estimate boundaries between these materials from a few selected data points. Our analysis shows that the error in facies delineation by means of support vector machines decreases logarithmically with increasing sampling density. We also introduce and analyze the use of regression support vector machines to estimate the parameter values between points where the parameter is sampled.**

***Index Terms*—Data analysis, geologic facies, geostatistics, machine learning, support vector machine (SVM).**

I. Introduction

OUR knowledge of the spatial distribution of the physical properties of geologic formations is often uncertain because of ubiquitous heterogeneity and the sparsity of data. Geostatistics has become an invaluable tool for estimating such properties at points in a computational domain where data are not available, as well as for quantifying the corresponding uncertainty. Geostatistical frameworks treat a formation's properties, such as hydraulic conductivity $K(\mathbf{x})$, as random fields that are characterized by multivariate probability density functions or, equivalently, by their joint ensemble moments. Thus, $K(\mathbf{x})$ is assumed to vary not only across the physical space (coordinate $\mathbf{x}$), but also in probability space (this variation may be represented by another coordinate $\boldsymbol{\xi}$, which is usually suppressed to simplify notation). Whereas spatial moments of K are obtained by sampling $K(\mathbf{x})$ in physical space (across $\mathbf{x}$), its ensemble moments are defined in terms of samples collected in probability space (across $\boldsymbol{\xi}$). Since in reality only a single realization of a geologic site exists, it is necessary to invoke the ergodicity hypothesis in order to substitute the sample spatial statistics, which can be inferred from field and/or laboratory scale data, for the ensemble statistics, which are actually required. Ergodicity cannot be proved, and requires a number of modeling assumptions, e.g., [1, Sec. 2.7] and references therein. One of the most popular geostatistical approaches to lithofacies delineation employs discontinuous geostatistical models, such as indicator Kriging (IK) [2]–[4]. IK has also found its way into image processing [5].

Machine learning provides an alternative to geostatistics by allowing one to make predictions in the absence of sufficient data parameterization, without treating geologic parameters as random and, hence, without the need for the ergodicity assumptions. Closely related to the field of pattern recognition, machine learning refers to a family of computational algorithms for data analysis that are designed to automatically tune themselves in response to data. Neural networks are an example of such algorithms that have been used in hydrologic modeling. While versatile and efficient for many important applications, including the delineation of geologic facies [6], neural networks usually do not provide bounds on expected classification errors.

We recently introduced [7] another subset of machine learning techniques, the support vector machine (SVM), for application in facies delineation. While similar to neural networks in its goals, the SVM is firmly grounded in the rigorous mathematical analysis of Vapnik's statistical learning theory (SLT) [8], which allows one to assess its performance and bound the corresponding errors. Like other machine learning techniques, SVMs enable one to treat the subsurface environment and its parameters as deterministic. Uncertainty associated with insufficient data parameterization is then represented and quantified by treating sampling locations as a random subset of all possible measurement locations. Such a formulation is ideally suited for subsurface imaging.

In [7], we used linear SVMs to locate a boundary between two materials in a perfectly stratified geologic formation. Such a boundary is by definition either a straight line (in two dimensions) or a plane (in three dimensions), so that available data are always linearly separable. Here we consider the general (nonlinear) form, which is referred to simply as an SVM, to delineate highly irregular boundaries between two heterogeneous geologic facies, based on a sparsely sampled parameter.

We formulate the problem of facies delineation in Section II. Section III provides a brief description of the general theory of nonlinear SVMs, with an emphasis on their application in subsurface imaging. SVMs are then used in Section IV to reconstruct a boundary between two heterogeneous geologic facies

Manuscript received January 27, 2005; revised September 6, 2005. This work was performed under the auspices of the U.S. Department of Energy, under Contract W-7405-ENG-36 and was supported in part by the U.S. Department of Energy under the DOE/BES Program in the Applied Mathematical Sciences, Contract KC-07-01-01, and in part by the LDRD Program at Los Alamos National Laboratory. This work made use of shared facilities supported by Sustainability of semi-Arid Hydrology and Riparian Areas (SAHARA) under the STC Program of the National Science Foundation under Agreement EAR-9876800.

B. Wohlberg and D. M. Tartakovsky are with the Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545 USA (e-mail: dmt@lanl.gov; brendt@t7.lanl.gov).

A. Guadagnini is with the Department of Hydraulic, Environmental, Infrastructure, and Surveying Engineering, Politecnico di Milano, 20133 Milan, Italy (e-mail: alberto.guadagnini@polimi.it).

Digital Object Identifier 10.1109/TGRS.2005.859953

from a few data points extracted from a randomly generated porous medium. In Section IV we also introduce a regression SVM to estimate parameter values at points where parameter data are not available. Finally, we contrast the performance of the SVM with that of a geostatistical approach.

II. Problem of Facies Delineation

Consider the problem of reconstructing a boundary between two heterogeneous materials (geologic facies) from parameter data $K_i = K(\mathbf{x}_i)$ collected at N locations $\mathbf{x}_i = (x_i, y_i)^T$, where $i \in \{1, \ldots, N\}$. Such problems are ubiquitous in subsurface hydrology since the geologic structure of the subsurface plays a crucial role in fluid flow and contaminant transport. A typical example is the problem of locating permeable zones in the aquiclude that separates two aquifers, the upper aquifer contaminated with industrial pollutants, and the lower aquifer used for municipal water supplies [9]. Parameter data can include measurements of hydraulic conductivity, electric resistivity, cumulative thickness of relevant geologic facies, and grain sizes.

The first step to facies delineation consists of analyzing a data histogram to assign to each data point a value of the indicator function

$$I(\mathbf{x}_i) = \begin{cases} 1, & \mathbf{x}_i \in M_1 \\ 0, & \mathbf{x}_i \in M_2 \end{cases} \tag{1}$$

where M_1 and M_2 are the two facies. This step is often nontrivial, since a typical geologic facies is heterogeneous, so that parameter measurements vary from point to point. Here we assume that the available parameter data $\{K(\mathbf{x}_i)\}_{i=1}^N$ are well differentiated, so that the process of assigning the values of the indicator functions to points $\{\mathbf{x}_i\}_{i=1}^N$ does not introduce interpretive errors.

Let $\mathcal{I}(\mathbf{x}, \boldsymbol{\alpha})$ be an estimate of the "true" indicator field $I(\mathbf{x})$, whose adjustable parameters $\boldsymbol{\alpha}$ are consistent with, and determined from, the available data $\{\mathbf{x}_i, I(\mathbf{x}_i)\}_{i=1}^N$. One would like to construct an estimate that is as close to the true field as possible, i.e., to minimize the difference, $\|I - \mathcal{I}\|$, between the two. Since parameters are typically sparsely sampled, this problem is ill-posed and requires regularization. The most common regularization procedure treats the underlying deterministic but sparsely sampled spatial functions $K(\mathbf{x})$ and $I(\mathbf{x})$ as random fields. An added benefit of such approaches is that they allow one to estimate the uncertainty associated with insufficient data parameterizations.

Within a general probabilistic framework, both the indicator field $I(\mathbf{x})$ and the choice of sampling locations $\{\mathbf{x}_i\}_{i=1}^N$ can be viewed as random, and can be described by a joint probability distribution $P(I, \mathbf{x})$. Then the problem of obtaining the best estimate of the indicator field is equivalent to minimizing the functional

$$R = \int \|I - \mathcal{I}\| dP(I, \mathbf{x}). \tag{2}$$

Unfortunately, since in reality only a single realization of a specific geologic formation exists, there is no direct way to evaluate $P(I, \mathbf{x})$. Geostatistical and statistical learning techniques provide two alternatives for evaluating (2).

Geostatistical approaches use the L^2 norm in (2), and treat

1) the indicator function $I(\mathbf{x})$ as a random field, and
2) the choice of sampling locations $\{\mathbf{x}_i\}_{i=1}^N$ as deterministic.

Then the problem of minimizing (2) reduces to the minimization of the indicator variance

$$\sigma_I^2 = \int (I - \mathcal{I})^2 dP(I). \tag{3}$$

To approximate $P(I)$, geostatistical approaches assume ergodicity, i.e., that the sample statistics of I, including mean μ_I, variance σ_I^2, and correlation function ρ_I computed from spatially distributed data $\{I(\mathbf{x}_i)\}_{i=1}^N$ can be substituted for the ensemble statistics. Furthermore, it is necessary to assume that these sampling statistics are representative of the whole field.

SLT [8] often uses the L^1 norm in (2) and treats

1) the indicator function $I(\mathbf{x})$ as deterministic;
2) the choice of sampling locations $\{\mathbf{x}_i\}_{i=1}^N$ as random.

Then the problem of minimizing (2) reduces to the minimization of the *expected risk*

$$R_{\text{exp}} = \frac{1}{2} \int |I - \mathcal{I}| dP(\mathbf{x}). \tag{4}$$

Rather than attempting to estimate probability distribution $P(\mathbf{x})$ from spatially distributed data, statistical learning replaces the expected risk R_{exp} with the *empirical risk*

$$R_{\text{emp}} = \frac{1}{2N} \sum_{i=1}^{N} |I(\mathbf{x}_i) - \mathcal{I}(\mathbf{x}_i)|. \tag{5}$$

These two quantities are related by a probabilistic bound, $R_{\text{exp}} \leq R_{\text{emp}} + \phi$, where the known function ϕ depends on the *Vapnik—Chervonenkis (VC) dimension* (representing a measure of the complexity of the family of functions $\mathcal{I}$) and the number of data points N (see [10], [11, ch. 4], and [12, Sec. 1.3 and ch. 5]). Analysis of the tightness of the bound $R_{\text{exp}} \leq R_{\text{emp}} + \phi$ provides a useful theoretical motivation for the SVM described below. However, it should be noted that this bound is often too loose to be of much practical significance.

III. Support Vector Machines

The SVM is a relatively recent technique that has attracted a great deal of interest due to its excellent performance on a wide range of classification problems, e.g., [10], [11], and [13]. The theoretical foundation of this technique is provided by SLT, which provides a bound on the expected risk R_{exp} [11, ch. 6, remark 6.7], [12, ch. 7].

The simplest SVM deals with linearly separable data collected from perfectly stratified geologic media, where different geologic facies are separated by planes (in three dimensions) or straight lines (in two dimensions). It is generalized to accommodate arbitrary datasets by means of the kernel technique introduced in the following section.

From the outset, we wish to emphasize the novelty of the proposed use of SVMs. In the usual pattern classification applications of SVMs, each data point is of high (often very high) dimensionality. For example, in the classical example of handwritten digit classification, each image of a digit is considered

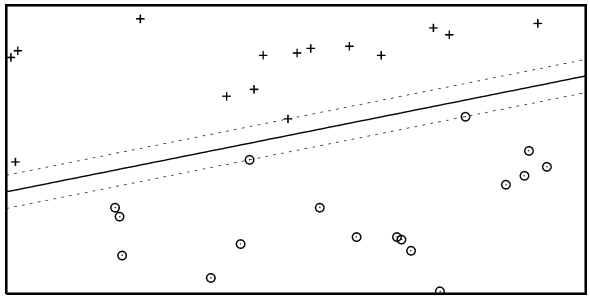

Fig. 1. Schematic representation of the boundary between two heterogeneous geologic facies M_1 and M_2 (located above and below the boundary, respectively) in a perfectly stratified geologic formation. The $+$ and $\odot$ signs indicate the locations where a parameter K is sampled and $I = 1$ or $= 0$, respectively. A maximum margin linear classifier for the displayed samples consists of the boundary estimate (solid line) and the margin (dotted lines).

a single high-dimensional data point, and the sampling density of the entire space is, by practical necessity, exceedingly small. In subsurface characterization, each sample point represents a location within two-dimensional (2-D) or three-dimensional (3-D) space, and the sampling density, while usually small, is much larger than in most pattern recognition applications.

A. Linear SVMs

Consider a boundary given by the straight line

$$\mathbf{a} \cdot \mathbf{x} + b = 0. \tag{6}$$

We wish to determine the unknown coefficients $\mathbf{a} = (a_1, a_2)^T$ and b from the dataset $\{\mathbf{x}_i, I(\mathbf{x}_i)\}_{i=1}^N$. In machine learning, an algorithm for constructing such a boundary is known as a linear classifier.

A maximum margin linear classifier is illustrated in Fig. 1—the boundary estimate is indicated by the solid line, and the dotted lines indicate the extent of the *margin*, i.e., the region within which the boundary could be shifted orthogonally without misclassifying any of the data. If d_1 and d_2 designate the perpendicular distances from the estimated boundary (solid line) to the nearest data point(s) in materials M_1 and M_2, respectively, then the size of the margin (dotted lines) is $d = d_1 + d_2$, and the sample points determining the position of the margin are called the *support vectors*. Since the lines bounding the margin are parallel to the boundary (6), their normal is also $\mathbf{a}$. The SVM determines the coefficients $\mathbf{a}$ and b in (6) by maximizing the size of this margin. While any choice of straight line that lies within the margin provides the same empirical risk R_{emp}, the straight line at the center of the maximum margin is a principled choice for minimizing the expected risk R_{exp} [12, Sec. 7.2].

The SVM is constructed as follows. Let $\mathbf{a} \cdot \mathbf{x} + b = \pm 1$ be two equations for the dashed lines bounding the margin in Fig. 1. (Note that while constants $\mathbf{a}$ and b in (6) are defined up to a multiplicative constant, these equations ensure their uniqueness.) Since the margin separates the two materials, all data points satisfy either

$$\mathbf{a} \cdot \mathbf{x}_i + b \geq +1 \tag{7a}$$

or

$$\mathbf{a} \cdot \mathbf{x}_i + b \leq -1. \tag{7b}$$

These inequalities imply that the estimated boundary lies at the center of the margin, i.e., that $d_1 = d_2 = d/2$. Defining the indicator function $J(\mathbf{x}) = 2I(\mathbf{x}) - 1$, so that $J(\mathbf{x}) = -1$ whenever $I(\mathbf{x}) = 0$ and $J(\mathbf{x}) = 1$ whenever $I(\mathbf{x}) = 1$, and denoting $J_i = J(\mathbf{x}_i)$ allows one to combine the two inequalities (7) into one

$$(\mathbf{a} \cdot \mathbf{x}_i + b) J_i \geq 1, \qquad \text{for } i \in \{1, \ldots, N\}. \tag{8}$$

The inequalities (8) become equalities for the $\mathbf{x}_i$ that are support vectors. Let $\|\mathbf{a}\| \equiv \sqrt{a_1^2 + a_2^2}$ denote the Euclidean length of $\mathbf{a}$. Since the distances ρ_1 and ρ_2 from the coordinate origin to the lines $\mathbf{a} \cdot \mathbf{x}_i + b = 1$ and $\mathbf{a} \cdot \mathbf{x}_i + b = -1$ are respectively

$$\rho_1 = -\frac{b+1}{\|\mathbf{a}\|} \quad \rho_2 = -\frac{b-1}{\|\mathbf{a}\|} \tag{9}$$

and the distance between these two lines $\rho_2 - \rho_1$, i.e., the margin d, is given by

$$d = \frac{2}{\|\mathbf{a}\|}. \tag{10}$$

Thus, the SVM can be formulated as a problem of maximizing d (or, equivalently, minimizing $\|\mathbf{a}\|$) subject to the linear constraints (8). Introducing Lagrange multipliers $\gamma_i \geq 0$ for $i \in \{1, \ldots, N\}$ leads to the objective function

$$L(\mathbf{a}, b, \boldsymbol{\gamma}) = \frac{1}{2}\|\mathbf{a}\|^2 - \sum_{i=1}^{N} \gamma_i \big[(\mathbf{a} \cdot \mathbf{x}_i + b) J_i - 1\big]. \tag{11}$$

A solution of this optimization problem defines $\mathbf{a}$ and b and thus, in accordance with (6), the boundary between the two layers, located at the center of the margin.

Often, the data are not perfectly linearly separable. A more general SVM formulation introduces slack variables $\xi_i \geq 0$ into the optimization, allowing for misclassification but penalizing the sum of the classification errors, so that the problem becomes the minimization of

$$\frac{1}{2}\|\mathbf{a}\|^2 + C \sum_{i=1}^{N} \xi_i \tag{12}$$

subject to the constraints [13, ch. 2], [14, Sec. 12.2.1]

$$(\mathbf{a} \cdot \mathbf{x}_i + b) J_i \geq 1 - \xi_i, \qquad \text{for } i \in \{1, \ldots, N\}. \tag{13}$$

As before, introducing Lagrange multipliers $\gamma_i, \delta_i \geq 0$ for $i \in \{1, \ldots, N\}$ gives the objective function

$$L(\mathbf{a}, b, \boldsymbol{\xi}, \boldsymbol{\delta}) = \frac{1}{2}\|\mathbf{a}\|^2 - \sum_{i=1}^{N} \gamma_i \big[(\mathbf{a} \cdot \mathbf{x}_i + b) J_i - 1 + \xi_i\big] + C \sum_{i=1}^{N} \xi_i - \sum_{i=1}^{N} \delta_i \xi_i. \tag{14}$$

Denoting the optimal values of $\mathbf{a}$ and b by $\mathbf{a}^\star$ and $b^\star$ respectively, the indicator function $J(\mathbf{x})$ is given by

$$J(\mathbf{x}) = \operatorname{sgn}(\mathbf{a}^\star \cdot \mathbf{x} + b^\star) \tag{15}$$

and is usually referred to as a *decision function* in the SVM literature.

To obtain $\mathbf{a}^\star$ and $b^\star$, it is often convenient to use the dual optimization problem [14, Sec. 12.2.1]

$$\max_{\boldsymbol{\gamma}} \left\{ \sum_{i=1}^{N} \gamma_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \gamma_i \gamma_j J_i J_j \mathbf{x}_i \cdot \mathbf{x}_j \right\} \tag{16}$$

subject to the constraints

$$0 \le \gamma_i \le C \quad \sum_{i=1}^{N} \gamma_i J_i = 0. \tag{17}$$

Let $\gamma_i^\star$ $(i \in \{1, \ldots, N\})$ be solutions of (16) and (17). Then the solutions of $\partial L / \partial a_k = 0$ for (14) give

$$\mathbf{a}^\star = \sum_{i=1}^{N} \gamma_i^\star J_i \mathbf{x}_i. \tag{18}$$

Let $\mathbf{x}_+$ and $\mathbf{x}_-$ denote arbitrary support vectors for which $J = 1$ and $J = -1$, respectively. Then the constraints (13), which at these points become equalities, give

$$b^\star = -\frac{1}{2} \mathbf{a}^\star \cdot (\mathbf{x}_+ + \mathbf{x}_-). \tag{19}$$

Thus, a solution for the indicator function (15) is

$$J(\mathbf{x}) = \operatorname{sgn}\left(\sum_{i=1}^{N} \gamma_i^\star J_i \mathbf{x}_i \cdot \mathbf{x} + b^\star \right). \tag{20}$$

B. Nonlinear SVMs

In most practical problems, boundaries between geologic facies are significantly more complex than a straight line or a plane. To account for this geometric complexity, one can generalize the linear SVM by noting that data which cannot be separated by a straight line or plane in the 2- or 3-D space of observation often become linearly separable (i.e., separable by a hyperplane) when projected onto another, usually higher dimensional space.

Let $\mathcal{F} : \mathbb{R}^n \to \mathbb{R}^m$ be a mapping of the n-dimensional physical space onto an m-dimensional space (known as a feature space) in which the linear SVM can be applied. The equation for a hyperplane separating the two materials in the m-dimensional space is

$$\mathbf{a} \cdot \mathcal{F}(\mathbf{x}) + b = 0 \tag{21}$$

where the dimension m and parameters $\mathbf{a} \in \mathbb{R}^m$ and b are determined from the transformed dataset $\{\mathcal{F}(\mathbf{x}_i), J_i\}_{i=1}^N$ by solving the quadratic optimization of the linear SVM (14)

$$L(\mathbf{a}, b, \boldsymbol{\xi}, \boldsymbol{\gamma}, \boldsymbol{\delta}) = \frac{1}{2} \|\mathbf{a}\|^2 - \sum_{i=1}^{N} \gamma_i \left[(\mathbf{a} \cdot \mathcal{F}(\mathbf{x}_i) + b) J_i - 1 + \xi_i \right] + C \sum_{i=1}^{N} \xi_i - \sum_{i=1}^{N} \delta_i \xi_i. \tag{22}$$

In analogy to (15), the indicator function is given by

$$J(\mathbf{x}) = \operatorname{sgn}(\mathbf{a}^\star \cdot \mathcal{F}(\mathbf{x}) + b^\star). \tag{23}$$

While this indicator function is linear in the feature space, it corresponds to a nonlinear function in the physical (or input) space, the specific form being determined by the mapping $\mathcal{F}$.

The dual optimization problem (16) is now recast as

$$\max_{\boldsymbol{\gamma}} \left\{ \sum_{i=1}^{N} \gamma_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \gamma_i \gamma_j J_i J_j \mathcal{F}(\mathbf{x}_i) \cdot \mathcal{F}(\mathbf{x}_j) \right\} \tag{24}$$

subject to the constraints

$$0 \le \gamma_i \le C \quad \sum_{i=1}^{N} \gamma_i J_i = 0. \tag{25}$$

The key observation here is that the feature space vectors enter into the optimization only within an inner product. If a Mercer kernel

$$\mathcal{K}(\mathbf{x}, \mathbf{x}') = \mathcal{F}(\mathbf{x}) \cdot \mathcal{F}(\mathbf{x}') \tag{26}$$

is available for a specific mapping $\mathcal{F}$, the required inner products may be computed directly in the physical space, without explicitly performing the potentially computationally expensive mapping into the feature space. Hence, the dual optimization (24) may be expressed as

$$\max_{\boldsymbol{\gamma}} \left\{ \sum_{i=1}^{N} \gamma_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \gamma_i \gamma_j J_i J_j \mathcal{K}(\mathbf{x}_i, \mathbf{x}_j) \right\} \tag{27}$$

avoiding explicit computation of the mapping $\mathcal{F}$.

In analogy to (18)–(20), the indicator function is now given by

$$J(\mathbf{x}) = \operatorname{sgn}\left(\sum_{i=1}^{N} \gamma_i^\star J_i \mathcal{K}(\mathbf{x}, \mathbf{x}_i) + b^\star \right). \tag{28}$$

Here $\gamma_i^\star$ $(i \in \{1, \ldots, N\})$ is defined as a solution of the dual optimization problem (12)

$$\mathbf{a}^\star = \sum_{i=1}^{N} \gamma_i^\star J_i \mathcal{F}(\mathbf{x}_i) \tag{29}$$

and $b^\star$ is given by (19). Note that the decision function (28) is expressed in terms of the kernel $\mathcal{K}$, without the need for explicit mapping onto the feature space.

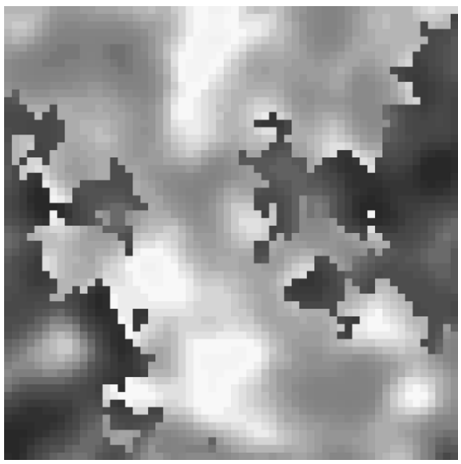

Fig. 2. Synthetic data on a 60×60 grid. Values range between -2.04 and 9.89.

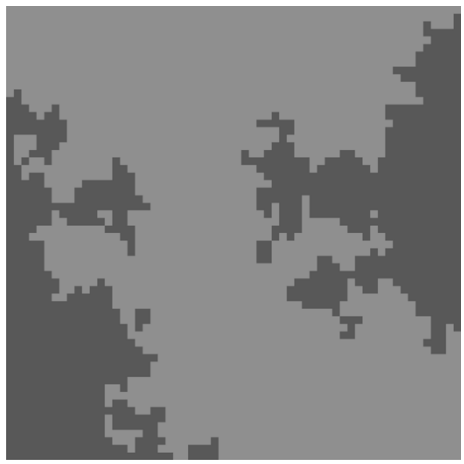

Fig. 3. Classification of data in Fig. 2, obtained by setting a threshold value of 4.0.

Among a wide variety of Mercer kernels, we will consider the polynomial kernel of order p

$$\mathcal{K}_{\mathrm{PLM}}(\mathbf{x}, \mathbf{x}') = (\mathbf{x} \cdot \mathbf{x}' + 1)^p \tag{30a}$$

the sigmoid kernel

$$\mathcal{K}_{\mathrm{SIG}}(\mathbf{x}, \mathbf{x}') = \tanh(\rho \mathbf{x} \cdot \mathbf{x}' + \varrho) \tag{30b}$$

the exponential radial basis function kernel

$$\mathcal{K}_{\mathrm{ERB}}(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|}{2\sigma^2}\right) \tag{30c}$$

and the Gaussian radial basis function kernel

$$\mathcal{K}_{\mathrm{GRB}}(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\sigma^2}\right). \tag{30d}$$

IV. Synthetic Example

To demonstrate the applicability of SVMs to subsurface imaging, and to elucidate their relative advantages with respect to a geostatistical approach, we reconstruct, from a few data points randomly selected according to a uniform distribution, the boundaries between two heterogeneous geologic facies in a synthetic porous medium shown in Fig. 2. This synthetic example is generated as follows.

We start by generating two autocorrelated, weakly stationary, normally distributed processes, representing two distinct spatial distributions of log hydraulic conductivity $Y = \ln K$ with ensemble means of -0.1 and 7.0. When hydraulic conductivities are expressed in centimeters per day, this corresponds to clayey and sandy materials, respectively. Both log-conductivity distributions have unit variance and Gaussian autocorrelation with unit correlation scale. We take these two fields to be mutually uncorrelated. The fields are generated by the SGSIM code [15] on a 60×60 grid, using a grid spacing of 1/5 of the log-conductivity correlation length.

Next, the composite porous medium in Fig. 2 is constructed by randomly choosing the shape of the internal boundary between the two materials and by assigning values of log-conductivity to cells in the domain. Using a threshold value of 4.0 for the indicator function (1) results in Fig. 3.

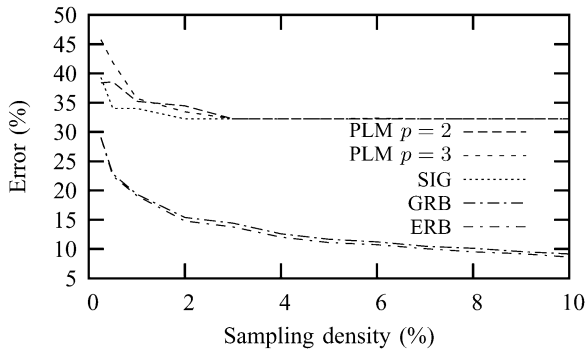

Fig. 4. Error rates corresponding to the SVMs with $C = 1.0$ and the following kernels: polynomial (PLM) with $p = 2$ and $p = 3$, sigmoid (SIG) with $\rho = 1.0$ and $\varrho = 0.0$, Gaussian radial basis (GRB) with $\sigma = 1.0$, and exponential radial basis (ERB) with $\sigma = 1.0$.

We used an SVM [13], [16] to reconstruct the boundary between the two geologic facies in Fig. 3 from sets of randomly selected data points. A set of sampling densities, ranging from 0.25% (nine data points) to 20% (720 data points) was selected, and for each sampling density an ensemble of 20 randomly generated realizations of the sample locations was constructed. (To accentuate the sparsest datasets, the figures below only show the results up to the 10% sampling density.) Classification errors for each sampling density were computed as the fraction of misclassified grid points averaged over the classification results for each of the 20 realizations at that sampling density. Fig. 4 compares the performance of SVMs with the polynomial (PLM), exponential radial basis (ERB), Gaussian radial basis (GRB), and sigmoid (SIG) kernels in (30). One can see that the radial basis function kernels (GRB and ERB) provide the best performance, which is not surprising given the general popularity of these kernels in SVM applications [17].

Figs. 5 and 6 show the geologic facies reconstructed by an ERB SVM with $\sigma = 1.0$ from 9 and 180 sample points, respectively. The locations of sample points are indicated by the

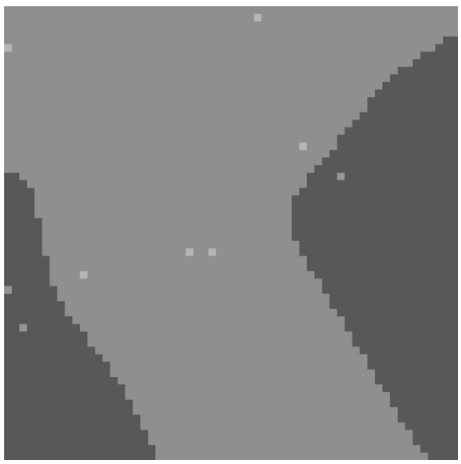

Fig. 5. Classification of data in Fig. 2, obtained by an ERB SVM ($C = 1.0$, $\sigma = 1.0$) using nine sample points (0.25% sampling density).

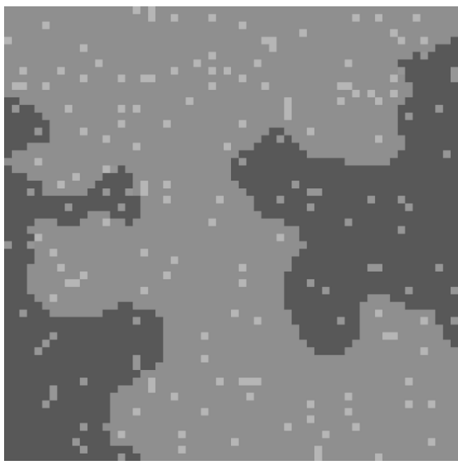

Fig. 6. Classification of data in Fig. 2, obtained by an ERB SVM ($C = 1.0$, $\sigma = 1.0$) using 180 sample points (5% sampling density).

lighter shades. The comparison of these reconstructions with the true field in Fig. 3 shows that even very sparse sampling might be sufficient for the SVMs to capture general trends in the spatial arrangement of geologic facies. However, the performance of SVMs on such sparse datasets is highly dependent on the actual locations of data points (i.e., highly variable from one realization to another). As the sampling density increases, the SVMs capture finer features of the spatial arrangement of geologic facies, and their performance is less dependent on a sampling realization.

A. Comparison With a Geostatistical Approach

We compare the accuracy of the facies reconstruction by means of the SVM with that obtained by a geostatistical approach (GSA) described in Appendix I. It is important to note that this and other geostatistical approaches to facies delineation assume that the relative volumes occupied by the two materials obtained from a sample are representative of the whole field. This assumption is usually difficult to validate *a priori*.

Fig. 7 shows the comparison of the performance of GSA and a SVM with $C = 10.0$ and the ERB kernel with $\sigma = 1.0$. When

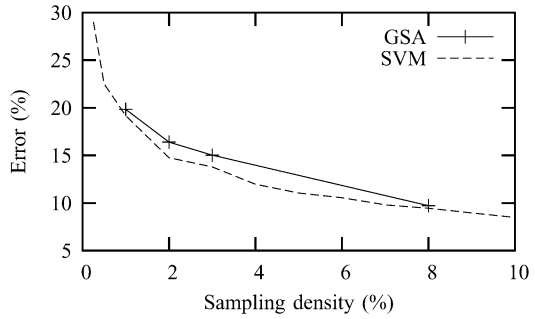

Fig. 7. Error rates corresponding to the GSA and SVM ($C = 10.0$, ERB kernel with $\sigma = 1.0$) approaches.

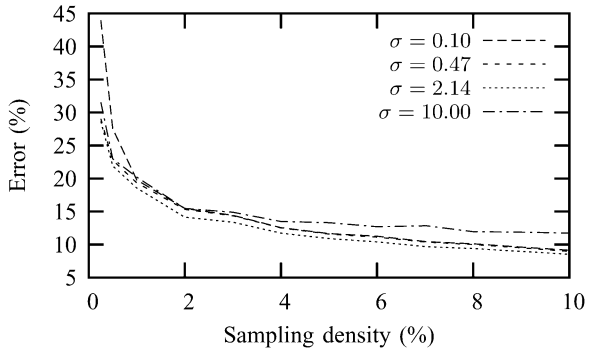

Fig. 8. Error rates corresponding to the SVMs with $C = 10.0$ and an ERB kernel with several values of σ.

enough measurements are available (i.e., when the sampling density is high enough), both methods have similar performance, with the SVM being slightly more accurate than GSA. Two factors, however, argue strongly in favor of SVMs. First, they perform relatively well even on highly sparse datasets (see the boundary reconstruction from nine sampling points in Fig. 5), on which GSA fails because the sample statistics (variograms in particular) become statistically meaningless. Second, SVMs are highly automated, while GSA generally requires manual data analysis to construct sample spatial variograms and to identify a proper interpretive theoretical model. As a result, GSAs are highly time consuming and depend on the subjective judgment of the practitioner.

B. SVM Parameter Sensitivity

The performance of the SVMs depends on the choice of C in (22) and the fitting parameters in the Mercer kernels (30). The optimal selection of these parameters is a nontrivial issue. The standard approach is via *cross validation*, which involves excluding a subset of the training data from the training step, and using it to evaluate the performance of the resulting classifier [12, Sec. 7.8.1]. This simplest form of cross validation is the *leave-one-out* (LOO) method [18], [19], described in some detail in Section IV-C. Alternatives to cross validation include those discussed in [17], [20]–[22].

We address this important issue in the context of subsurface characterization by performing a sensitivity analysis of the SVM performance with respect to the fitting parameters. Thereafter, as described in some detail in Section IV-C, we employ the LOO method to automatically select these parameters.

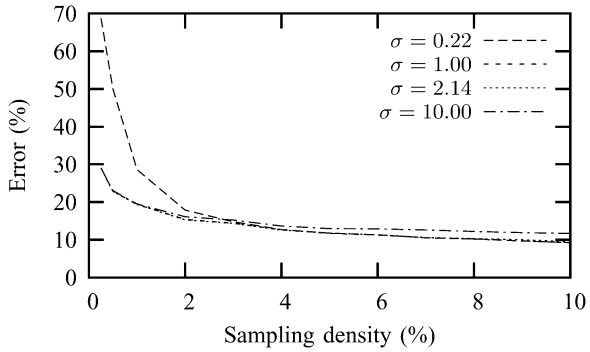

Fig. 9. Error rates corresponding to the SVMs with $C = 10.0$ and a GRB kernel with several values of σ.

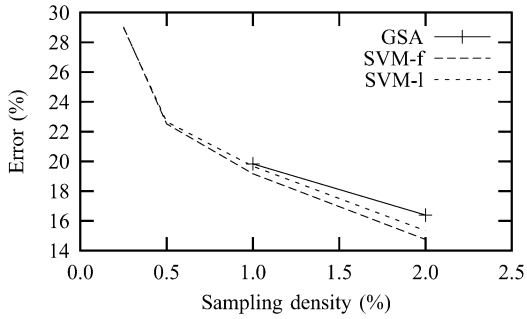

Fig. 10. Error rates corresponding to the GSA approach (GSA) and the SVM with the ERB kernel, with parameter σ selected either manually (SVM-f) or automatically by the leave-one-out method (SVM-l).

Figs. 8 and 9 demonstrate the sensitivity to a fitting parameter σ of the SVMs with the exponential radial basis (ERB) and Gaussian radial basis (GRB) kernels given by (30c) and (30d), respectively. One can see that when the sampling density exceeds 2%, the performance of both SVMs is relatively insensitive to the choice of σ (with its values varying over about two orders of magnitude). This finding is encouraging, since the optimal choice of σ is nontrivial, and computationally expensive when performed via LOO and related methods.

C. SVM Parameter Selection

The apparent lack of sensitivity of the SVM performance to the selection of kernel parameter σ in our application can be used to justify the manual selection of a set of SVM parameters. We used this approach in Section IV-A to compare the relative performance of the GSA and SVM methods. Such a procedure might bias this comparison in favor of the SVM method. To demonstrate that this bias is insignificant, we compare the performance of the SVM with manually selected parameters with the performance of an SVM, whose parameters are selected automatically by means of the standard LOO method. The LOO method consists of the following steps.

1) Identify the 2-D SVM parameter space (spanned by parameters C and σ), over which parameters C and σ are allowed to vary. In our computational example, we chose the ranges $C \in [0.10, 10.00]$ and $\sigma = [0.32, 10.00]$ to exclude parameters for which poor performance is expected *a priori* (see [23] for a discussion of asymptotic properties of GRB kernels with respect to these parameters).
2) Discretize the parameter space on a regular grid. In our example, we use $(C, \sigma) \in \{0.10, 0.32, 1.00, 3.16, 10.00\} \times \{0.32, 0.56, 1.00, 1.78, 3.16, 5.62, 10.00\}$.
3) Leave out the first of the data points $\{K_i\}_{i=1}^N$ and train the SVM on the remaining $N-1$ data points.
4) Compute the error in the SVM estimation of the first data point for each of the grid point of parameters C and σ.
5) Repeat this procedure by leaving out the second, third, and so forth data point, and training the SVM on the remaining $N-1$ data points.
6) Select the set of parameters to minimize the mean of these errors over all left-out samples.

Given the computational expense, we performed the LOO method for the four lowest sampling densities. The results of these simulations are presented in Fig. 10, which compares the boundary reconstruction errors resulting from the geostatistical approach (GSA), the SVM with manually selected parameters $C = 10.0$ and $\sigma = 1.0$ (SVM-f), and the SVM with parameters selected automatically using the LOO method (SVM-l). One can see that both SVM approaches outperform the GSA approach, with the difference between the SVM approaches being relatively small.

D. Identification of Hydraulic Parameters

In hydrologic applications, the delineation of geologic facies from parameter data is often not sufficient. Since many geologic facies are heterogeneous, it is also necessary to assign parameter values to locations (e.g., elements of a numerical grid) where data are not available. Geostatistical approaches achieve this goal through *data interpolation* algorithms, such as Kriging [15]. SVMs take an alternative route by employing *data regression* strategies as outlined in Appendix II.

In addition to the direct use of the nonlinear regression (41), we explore a two-step procedure. First, we use the SVM to delineate the geologic facies from a dataset $\{K(\mathbf{x}_i)\}_{i=1}^N$. Then we perform a separate nonlinear regression (41) regression on the data subsets within each facies. In the simulations presented here, we used the regression SVM [13] with the GBF kernel (30d) and $\sigma = 1$ for the direct data regression; and the classification SVM [13] with the GRB kernel (30c) and $\sigma = 1$ followed by the regression SVM [13] with the GBF kernel (30d) and $\sigma = 1$ for the two-step data regression.

Figs. 11 and 12 show the $K(\mathbf{x})$ fields reconstructed with these two regression strategies, from a sparse dataset $\{K_i\}_{i=1}^9$ (the sampling density of 0.25%) denoted in Fig. 5 by the lightly colored pixels. Figs. 13 and 14 do the same for a denser dataset (180 data points, corresponding to a sampling density of 5%) shown in Fig. 6. As one would expect, the visual comparison of these four figures with the "true" $K(\mathbf{x})$ field shown in Fig. 2 suggests that the quality of the reconstruction of the K field increases with the sampling density. The proposed two-step SVM regression (Figs. 12 and 14) outperforms the direct SVM regression (Figs. 11 and 13), capturing some of the main features of the K distribution even from an extremely

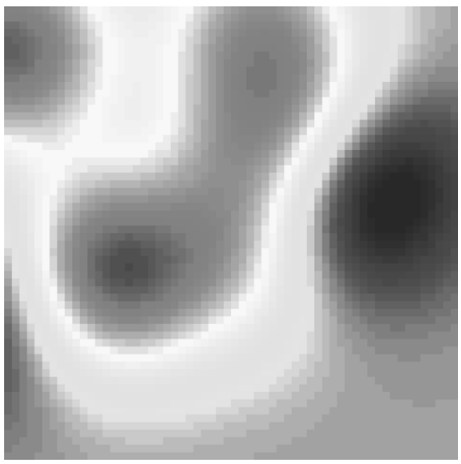

Fig. 11. Regression of nine sample points (0.25% sampling density) from data in Fig. 2, obtained by a GRB ($C = 10.0$, $\sigma = 10.0$) SVM regression.

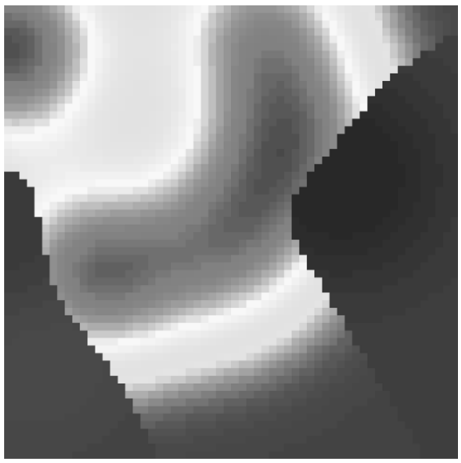

Fig. 12. Regression of nine sample points (0.25% sampling density) from data in Fig. 2, obtained by an ERB ($C = 1.0, \sigma = 1.0$) SVM classification followed by a GRB ($C = 10.0$, $\sigma = 10.0$) SVM regression.

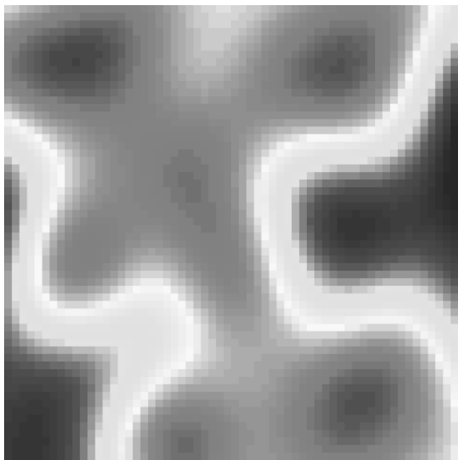

Fig. 13. Regression of 180 sample points (5% sampling density) from data in Fig. 2, obtained by a GRB ($C = 10.0$, $\sigma = 10.0$) SVM regression.

sparse (the sampling density of 0.25%) dataset. A quantitative comparison of the accuracy of these reconstructions is provided below.

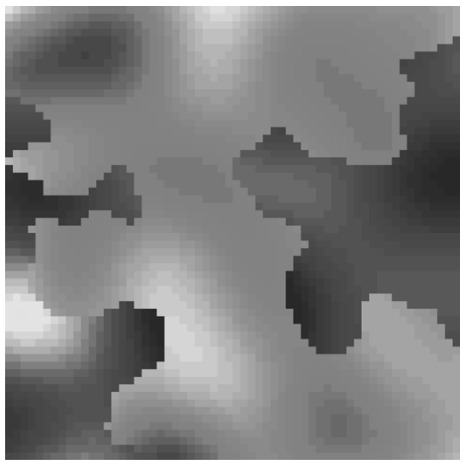

Fig. 14. Regression of 180 sample points (5% sampling density) from data in Fig. 2, obtained by an ERB ($C = 1.0$, $\sigma = 1.0$) SVM classification followed by a GRB ($C = 10.0$, $\sigma = 10.0$) SVM regression.

Finally, we compare the SVM regressions with two geostatistical approaches, which employ alternative interpolation strategies based on Kriging. The first approach uses Kriging to interpolate between the 180 data points, whose locations are shown in Fig. 6. An isotropic spherical variogram provides an appropriate interpretation of the data with parameters nugget 1.08, sill 12.01, and range 5.261. Note that the synthetic dataset in Figs. 2 and 3 is a realization of the random field with a Gaussian variogram. The discrepancy between the theoretical variogram model used to interpret the sample variogram inferred from the 180 data points and the variogram used to construct the underlying reference field is due to the finite number of samples and the aggregation into a single dataset of the data belonging originally to two different lithofacies.

The second approach, which we call Kriging with classification, consists of two steps. First, we use the geostatistical facies delineation procedure described in Appendix I. Second, we use simple Kriging on the two subsets of the 180 points shown in Fig. 6, each of which belongs to one of the two facies.This procedure results in an isotropic exponential variogram with nugget 0.0, sill 0.7 and range 3.357 for the high conductivity facies, and in an isotropic exponential variogram with nugget 0.035, sill 0.7 and range 1.768 for the low conductivity facies.

Figs. 15 and 16 provide the K fields reconstructed by Kriging, and Kriging with classification, respectively. Comparison of these fields with the "true" K field shown in Fig. 2 reveals that Kriging with classification outperforms simple Kriging, as is the case with their SVM counterparts. If an eyeball measure is used to compare the reconstructed K fields shown in Figs. 13–16 with their "true" counterpart in Fig. 2, one may conclude that the two Kriging approaches slightly outperform their respective SVM counterparts.

Tables I and II provide more rigorous quantitative comparisons of the SVM and GSA approaches without and with classification, respectively. These comparisons are given in terms of: 1) the L^1 error computed as the mean of the absolute value of the error at each element (pixel); 2) the L^2 error computed as the square root of the mean of the square of the error at each element; and 3) the fractional error (FE) computed as

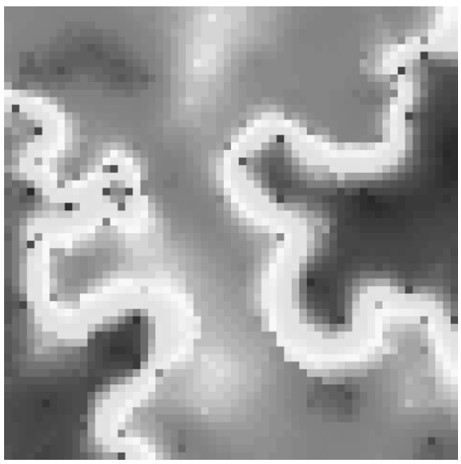

Fig. 15. Interpolation of 180 sample points (5% sampling density) from data in Fig. 2, obtained by Kriging.

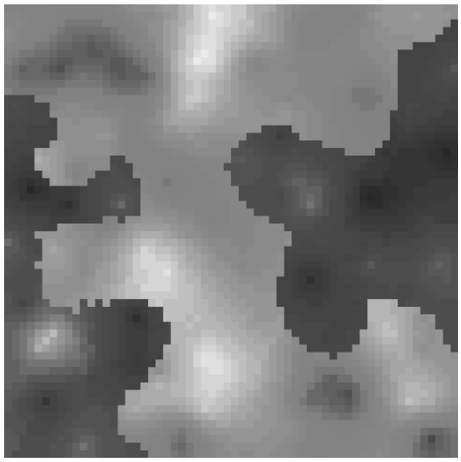

Fig. 16. Interpolation of 180 sample points (5% sampling density) from data in Fig. 2, obtained by Kriging with IK classification.

the fraction of the elements in the total number of elements (3600) for which the relative error exceeds a given threshold specified in parenthesis.

One can see that in every norm but L^2 the use of classification prior to regression improves the quality of the reconstructed K fields. The L^2 norm is an exception, because it unduly accentuates the errors introduced by the misclassification of the elements close to the boundary between two materials. The corresponding SVM and GSA approaches introduce comparable errors, but the Kriging approaches (especially Kriging with classification) are prone to subjective modeling choices and much more labor intensive than their SVM counterparts.

V. Conclusion

We explored the potential of support vector machines for the delineation of geologic facies from limited data. This was accomplished: 1) by reconstructing, from a few data points, a synthetic randomly generated porous medium consisting of two heterogeneous materials and 2) by comparing the performance of SVMs with that of the geostatistical approach [24].

Key differences between SVMs and geostatistics, first pointed out in [7], are as follows.

TABLE I
Several Measures of the Reconstruction Errors Introduced by the SVM and GSA Approaches Without Classification

	L^1	L^2	FE (0.1)	FE (0.2)	FE (0.3)
SVM	1.51	2.39	0.57	0.44	0.39
GSA	1.28	2.11	0.53	0.44	0.40

TABLE II
Several Measures of the Reconstruction Errors Introduced by the SVM and GSA Approaches With Classification

	L^1	L^2	FE (0.1)	FE (0.2)	FE (0.3)
SVM	1.22	2.53	0.45	0.36	0.33
GSA	1.13	2.53	0.42	0.35	0.33

- Since SVMs do not treat the subsurface environment as random, they do not require ergodicity and other statistical assumptions that lie at the heart of geostatistics.
- While geostatistics provides a set of interpolation tools, SVMs use regression.

Our analysis leads to the following major conclusions.

- For any sampling densities, SVMs slightly outperform the geostatistical approach in reconstructing the boundaries between two geologic facies, while being significantly less labor intensive.
- For very low sampling densities (e.g., 0.25%), which make the inference of statistical parameters meaningless, the geostatistical approach fails, while SVMs still do a reasonably good job in reconstructing the boundaries.

We also employed SVMs and geostatistics to infer parameter values at spatial locations where parameter data are not available. This was accomplished by two alternative approaches. The first employed direct regression SVMs (or Kriging). The second consisted of a sequential two-step regression SVMs (or geostatistical approach), in which the use of SVMs (or indicator Kriging) to delineate geologic facies was followed by the use of SVM regression (or Kriging interpolation) to infer parameter values. We found the following.

- Sequential approaches are more accurate than their direct counterparts by all measures but L^2.
- The reconstructions of parameter fields obtained with the SVM and geostatistical approaches are comparable at medium to high sampling densities.
- Regression SVMs are highly automated, while their geostatistical counterparts typically require manual estimation of variograms. Consequently, they require less user effort and are less prone to subjective interpretive errors.
- Regression SVMs can be used at low sampling densities (e.g., for the 0.25% sampling density or nine data points) where geostatistical inference becomes meaningless.

Appendix I
Geostatistical Approach

The geostatistical approach due to [24] was used for performance comparisons with SVMs. This approach consists of the following steps. First, we use Kriging [15] to construct a map of the ensemble average of the indicator function $\langle I(\mathbf{x})\rangle$ from the data $\{I(\mathbf{x}_i)\}_{i=1}^N$. The ensemble mean $I(\mathbf{x})$ is the probability

that a point $\mathbf{x}$ lies in Material 1, $\langle I(\mathbf{x})\rangle = P[\mathbf{x} \in M_1]$. Then we define a boundary between the two materials as an isoline $P[\mathbf{x} \in M_1] = c$, where c is a number of data points in Material 1 (or 2) relative to the total number of data points, after accounting for data clustering.

In some cases, this value does not guarantee that the Kriging estimate of the fraction of the total area covered by the low-conductivity material equals the declustered global mean of the original indicator data, resulting from the raw data. In such cases, c is set to a value of the Kriged indicator field which allows one to recover a reconstruction that honors the empirical relative volumetric fractions of the two materials.

Appendix II
Support Vector Regression

SVM regression may be viewed as a generalization of the SVM classification introduced in Section II to delineate the boundaries between geologic facies. Let $\hat{K}(\mathbf{x}, \boldsymbol{\alpha})$ be an estimate of a "true" parameter field $K(\mathbf{x})$, whose adjustable parameters $\boldsymbol{\alpha}$ are determined from the available parameter data $\{K_i = K(\mathbf{x}_i)\}_{i=1}^N$. SVM regression aims to minimize the difference between the two, while providing a probabilistic bound on the accuracy of the estimator $\hat{K}$ at a randomly drawn point $\mathbf{x}$ [11, Sec. 4.5].

Similar to SVM classification in Section III, SVM regression is first introduced for linear regression, and is then generalized to nonlinear regression via the kernel technique. For linear regression, we seek to approximate a dataset $\{K_i = K(\mathbf{x}_i)\}_{i=1}^N$ with a linear function

$$\hat{K}(\mathbf{x}) = \mathbf{a} \cdot \mathbf{x} + b. \tag{31}$$

The regression equivalent of the classification functional (12) is [11, Sec. 6.2]

$$\frac{1}{2}\|\mathbf{a}\|^2 + C\sum_{i=1}^{N} L_\epsilon(\mathbf{x}_i, K_i, \hat{K}) \tag{32}$$

where the pointwise sum of classification errors $\sum_{i=1}^N \xi_i$ in (12) is replaced by pointwise sum $\sum_{i=1}^N L_\epsilon(\mathbf{x}_i, K_i, \hat{K})$ of a *loss function* $L(\mathbf{x}, K, \hat{K})$ which measures the error in approximating K_i at $\mathbf{x}_i$ by $\hat{K}(\mathbf{x}_i)$. While it is not the only choice, we utilize the ϵ-insensitive loss function

$$L_\epsilon(\mathbf{x}, K, \hat{K}) = \begin{cases} 0, & \text{for } |\hat{K}(\mathbf{x}) - K| < \epsilon, \\ |\hat{K}(\mathbf{x}) - K| - \epsilon & \text{otherwise} \end{cases} \tag{33}$$

first used in SVM regression. With this loss function, the primal optimization problem for SVM regression is the minimization of

$$\frac{1}{2}\|\mathbf{a}\|^2 + C\sum_{i=1}^{N}(\xi_i + \hat{\xi}_i) \tag{34}$$

subject to the constraints

$$(\mathbf{a} \cdot \mathbf{x}_i + b) - K(\mathbf{x}_i) \le \epsilon + \xi_i \tag{35a}$$

$$K(\mathbf{x}_i) - (\mathbf{a} \cdot \mathbf{x}_i + b) \le \epsilon + \hat{\xi}_i \tag{35b}$$

where $i \in \{1, \ldots, N\}$, and $\xi_i \ge 0$ and $\hat{\xi}_i \ge 0$ are the slack variables similar to those first introduced in (12).

Introducing Lagrange multipliers $\gamma_i, \hat{\gamma}_i, \delta_i, \hat{\delta}_i \ge 0$ $(i \in \{1, \ldots, N\})$ gives the objective function

$$\begin{aligned} L(\mathbf{a}, b, \boldsymbol{\xi}, \hat{\boldsymbol{\xi}}, \boldsymbol{\gamma}, \hat{\boldsymbol{\gamma}}, \boldsymbol{\delta}, \hat{\boldsymbol{\delta}}) = {} & \frac{1}{2}\|\mathbf{a}\|^2 + C\sum_{i=1}^{N}(\xi_i + \hat{\xi}_i) \\ & - \sum_{i=1}^{N}\gamma_i(\epsilon + \xi_i + K_i - \mathbf{a}\cdot\mathbf{x}_i - b) \\ & - \sum_{i=1}^{N}\hat{\gamma}_i(\epsilon + \hat{\xi}_i - K_i + \mathbf{a}\cdot\mathbf{x}_i + b) \\ & - \sum_{i=1}^{N}\delta_i\xi_i - \sum_{i=1}^{N}\hat{\delta}_i\hat{\xi}_i. \end{aligned} \tag{36}$$

In analogy with (16), the dual optimization problem is

$$\begin{aligned} \max_{\boldsymbol{\gamma}, \hat{\boldsymbol{\gamma}}} \Bigg\{ & \sum_{i=1}^{N}(\gamma_i - \hat{\gamma}_i)K_i - \epsilon\sum_{i=1}^{N}(\gamma_i + \hat{\gamma}_i) \\ & - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\gamma_i - \hat{\gamma}_i)(\gamma_j - \hat{\gamma}_j)\mathbf{x}_i \cdot \mathbf{x}_j \Bigg\} \end{aligned} \tag{37}$$

subject to the constraints

$$0 \le \gamma_i \le C,\ 0 \le \hat{\gamma}_i \le C, \qquad \sum_{i=1}^{N}(\gamma_i - \hat{\gamma}_i) = 0. \tag{38}$$

Let $\gamma_i^\star$ and $\hat{\gamma}_i^\star$ $(i \in \{1, \ldots, N\})$ denote a solution of (37) and (38). Then the optimal parameters $\mathbf{a}$ and b are given by [12, Sec. 9.2]

$$\mathbf{a}^\star = \sum_{i=1}^{N}(\gamma_i^\star - \hat{\gamma}_i^\star)\mathbf{x}_i \tag{39}$$

and

$$b^\star = K_j - \mathbf{a}^\star \cdot \mathbf{x}_j - \epsilon \tag{40}$$

respectively, where K_j and $\mathbf{x}_j$ in (40) are chosen such that $0 < \gamma_j^\star < C$.

For a nonlinear regression, we once again replace $\mathbf{x}$ by $\mathcal{F}(\mathbf{x})$ to give the regression

$$\hat{K}(\mathbf{x}) = \mathbf{a} \cdot \mathcal{F}(\mathbf{x}) + b$$

and utilize the dual form of the optimization to express the problem in terms of the kernel $\mathcal{K}$ associated with mapping $\mathcal{F}$, giving the regression

$$\hat{K}(\mathbf{x}) = \sum_{i=1}^{N}(\gamma_i - \hat{\gamma}_i)\mathcal{K}(\mathbf{x}_i, \mathbf{x}) + b. \tag{41}$$

See [25] for a discussion of methods for choosing C, ϵ, and the kernel parameter(s) in SVM regression.

ACKNOWLEDGMENT

The authors thank the anonymous reviewers for their insightful comments.

REFERENCES

[1] Y. Rubin, *Applied Stochastic Hydrogeology*. New York: Oxford Univ. Press, 2003.
[2] A. G. Journel, "Nonparametric estimation of spatial distribution," *Math. Geol.*, vol. 15, no. 3, pp. 445–468, 1983.
[3] A. G. Journel and E. K. Isaaks, "Conditional indicator simulation: Application to a Saskatcheouan uranium deposit," *Math. Geol.*, vol. 16, no. 7, pp. 685–718, 1984.
[4] A. G. Journel and J. J. Gomez-Hernandez, "Stochastic imaging of the Wilmington clastic sequence," *SPE Form. Eval.*, vol. 8, no. 1, pp. 33–40, Mar. 1993.
[5] W. Oh and W. B. Lindquist, "Image thresholding by indicator Kriging," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 7, pp. 590–602, Jul. 1999.
[6] S. Moysey, J. Caers, R. Knight, and R. M. Allen-King, "Stochastic estimation of facies using ground penetrating radar data," *Stoch. Environ. Res. Risk Assess.*, vol. 17, pp. 306–318, 2003.
[7] D. M. Tartakovsky and B. E. Wohlberg, "Delineation of geologic facies with statistical learning theory," *Geophys. Res. Lett.*, vol. 31, no. 18, L18502, 2004.
[8] V. N. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.
[9] L. Guadagnini, A. Guadagnini, and D. M. Tartakovsky, "Probabilistic reconstruction of geologic facies," *J. Hydrol.*, vol. 294, pp. 57–67, 2004.
[10] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining Knowl. Discov.*, vol. 2, no. 2, pp. 121–167, 1998.
[11] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, U.K.: Cambridge Univ. Press, 2000.
[12] B. Schölkopf and A. J. Smola, *Learning With Kernels*. Cambridge, MA: The MIT Press, 2002.
[13] S. R. Gunn. (1998) Support vector machines for classification and regression. Univ. Southampton, Southampton, U.K., Tech. Rep., School of Electron. Comput. Sci., Southampton, U.K. [Online]. Available: http://www.ecs.soton.ac.uk/~srg/publications/pdf/SVM.pdf.
[14] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. New York: Springer-Verlag, 2001.
[15] C. V. Deutsch and A. G. Journel, *Geostatistical Software Library and User's Guide*. New York: Oxford Univ. Press, 1992.
[16] S. R. Gunn. (2001) Matlab support vector machine toolbox. Univ. of Southampton, Southampton, U.K.. [Online]. Available: http://www.isis.ecs.soton.ac.uk/isystems/kernel/svm.zip.
[17] W. J. Wang, Z. B. Xu, W. Z. Lu, and X. Y. Zhang, "Determination of the spread parameter in the Gaussian kernel for classification and regression," *Neurocomputing*, vol. 55, no. 3–4, pp. 643–663, Oct. 2003.
[18] O. Chapelle and V. Vapnik, "Model selection for support vector machines," *Adv. Neural Inf. Process. Syst.*, vol. 12, pp. 230–236, 2000.
[19] J.-H. Lee and C.-J. Lin. (2000) Automatic model selection for support vector machines. Dept. Comput. Sci. Inf. Eng., Nat. Taiwan Univ., Taipei, Taiwan, R.O.C.. [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/papers/modelselect.ps.gz.
[20] K. Duan, S. S. Keerthi, and A. N. Poo, "Evaluation of simple performance measures for tuning SVM hyperparameters," *Neurocomputing*, vol. 51, pp. 41–59, Apr. 2003.
[21] M. Cassabaum, D. Waagen, J. J. Rodriguez, and H. Schmitt, "Unsupervised optimization of support vector machine parameters," in *Proc. SPIE Conf Automatic Target Recognition XIV*, vol. 5426, F. Sadjadi, Ed., Orlando, FL, Apr. 2004, pp. 316–325.
[22] C. Soares, P. B. Brazdil, and P. Kuba, "A meta-learning method to select the kernel width in support vector regression," *Mach. Learning*, vol. 54, no. 3, pp. 195–209, Mar. 2004.
[23] S. S. Keerthi and C.-J. Lin, "Asymptotic behaviors of support vector machines with Gaussian kernel," *Neural Comput.*, vol. 15, no. 7, pp. 1667–1689, 2003.
[24] R. W. Ritzi, D. F. Jayne, A. J. Zahradnik, Jr, A. A. Field, and G. E. Fogg, "Geostatistical modeling of heterogeneity in glaciofluvial, buried-valley aquifer," *Ground Water*, vol. 32, no. 4, pp. 666–674, Aug. 1994.
[25] V. Cherkassky and Y. Q. Ma, "Practical selection of SVM parameters and noise estimation for SVM regression," *Neural Netw.*, vol. 17, no. 1, pp. 113–26, Jan. 2004.

Brendt Wohlberg received the B.Sc. (Hons.) degree in applied mathematics, the M.Sc. degree in applied science, and the Ph.D. degree in electrical engineering from the University of Cape Town, Cape Town, South Africa, in 1990, 1993, and 1996, respectively.

He is currently a Technical Staff Member in the Mathematical Modeling and Analysis Group (T-7) at Los Alamos National Laboratory, Los Alamos, NM. His research interests include image coding, pattern recognition, wavelets and adaptive signal decompositions, and inverse problems in signal and image processing.

Daniel M. Tartakovsky received the M.Sc. (with honors) degree in applied mathematics and fluid mechanics from Kazan State University, Kazan, Russia, and the Ph.D. degree in hydrology from the University of Arizona, Tucson, in 1991 and 1996, respectively.

He is currently a Team Leader and a Technical Staff Member in the Mathematical Modeling and Analysis Group (T-7) at Los Alamos National Laboratory, Los Alamos, NM. His research interests include uncertainty quantification, stochastic partial differential equations, and inverse modeling.

Alberto Guadagnini received the laurea and doctoral degrees in hydraulic engineering from the Politecnico di Milano, Milan, Italy, in 1989 and 1992, respectively.

He is currently a Full Professor of hydraulics at the Department of Hydraulic, Environmental, Infrastructure, and Surveying Engineering, Politecnico di Milano. His research interests include hydrogeology, fluid mechanics, modeling of contaminant fluxes in porous formations, inverse modeling, and stochastic partial differential equations.