J. Intell. Syst. **20** (2011), 261–278
DOI 10.1515/JISYS.2011.014

Site Characterization Model Using Support Vector Machine and Ordinary Kriging

Pijush Samui and Sarat Das

Abstract. In the present study, ordinary kriging and support vector machine (SVM) have been used to develop three dimensional site characterization model of an alluvial site based on standard penetration test (SPT) results. The SVM is a novel type of learning machine based on statistical learning theory, uses regression technique by introducing ε-insensitive loss function has been adopted. The knowledge of the semivariogram of the SPT values (N) is used in the ordinary kriging method to predict the N values at any point in the subsurface of the site where field measurements are not available. The comparison between the SVM and ordinary kriging model demonstrates that the SVM model is superior to ordinary kriging model in predicting N values in the site.

Keywords. Ordinary kriging, semivariogram, support vector machine, standard penetration test, site characterization.

2010 Mathematics Subject Classification. 60-XX.

1 Introduction

Site characterization is an important task in geotechnical engineering practice and generally defined as the identification and description of the subsurface strata within the areas of influence of a project. The basic objective of site characterization is to provide sufficient and reliable information and data on the site condition to a level of compatible and consistent with the needs and requirements of the project. It is necessary to predict geotechnical properties at any point of a site based on a limited number of tests. Prediction of geotechnical properties of a site is a difficult task for uncertainty. Uncertainty comes from spatial variability, measurement noise, measurement and model bias, and statistical error due to limited measurement [1]. Case studies have shown that predictions solely based on engineering judgment fails in 70% of the cases. As such, need for characterization of soil strata based on more rational and scientific approach has been felt. The methods like wavelet based variance method, spectral density function, fractal model [2], Barlett's statistics method [3], kriging [1, 4–6] have been discussed to characterize soil stratification.

The ordinary kriging method was developed during the 1960s and 1970s in mining engineering and has been acknowledged as a good spatial interpolator [7–10]. This technique is widely used in the field of mining engineering, earth sciences, geochemistry, remote sensing, etc. A major advantage of kriging is that it is more flexible than other interpolation methods such as inverse-distance weighting, deterministic splines and Thiessen polygons. Soulie et al., (1990) [4] used the kriging method to describe the spatial variability of the undrained shear strength of clay deposits based on in-situ vane shear tests. kriging has also been used to study spatial variability of clay from both vane shear test and cone penetration test (CPT) data Chiasson et al. (1995) [5]. Sivakumar and Mukesh (2004) [11] have considered the effect of spatial soil variability on stability of slope. Elkateb et al. (2003a, 2003b) [6, 12] used kriging to study the spatial heterogeneity of soil both in horizontal and vertical direction based on CPT result. However, the literature on three dimensional site characterizations using geostatistics is not available [13]. This is quite likely due tom the following reasons:

(a) The physical difficulty in conducting soil investigation in different directions.

(b) High anisotropy generally observed within horizontal and vertical planes.

(c) Lack of proper model to describe the spatial variation of soil properties [13].

Recently, artificial neural network (ANN) has been also used for site characterization [11, 13, 14]. Itani and Najjar (2000) [14] used a back propagation neural network (BPNN) for three dimensional (3-D) characterization of a site based on SPT results. With five bore log data they found that predictions can be made with mean absolute relative error (MARE) of 35%. Juang et al. (2001) [13] used a radial basis neural network (RBNN) for 3-D site characterization of a site based on cone penetration test (CPT) data. They used the RBNN for predicting the cone resistance (q_c) value in 1-D, 2-D and 3-D characterizations and compared the result of 3-D characterization analysis with that obtained from kriging analysis, however, for the same they found that kriging is not suitable for 1-D characterization. Das and Basudhar (2004) [15] used BPNN to predict the SPT values for 1-D, 2-D and 3-D characterization. They observed that ANN could be applied successfully only to 1-D and 2-D problems. There are some limitations in using ANN; unlike other statistical models, ANN does not provide information about the relative importance of the various parameters [16] the knowledge acquired during the training of the model is stored in an implicit manner and very difficult to come up with reasonable interpretation of the overall structure of the network [17], it has some inherent drawbacks such as arriving at local minimum and less generalizing performance.

Support Vector Machine (SVM) developed by Vapnik (1995) [18], is a recently developed method based on statistical learning theory. It provides a new, efficient

novel approach to improve the generalization performance and can attain a global minimum. In general, SVM has been used for pattern recognition problems. But, recently it has been used to solve non-linear regression estimation and time series prediction by introducing ε-insensitive loss function [18, 19]. The details of SVM and its application to geotechnical engineering problems can be found in literatures [20–24].

The Standard Penetration Test (SPT) is one of the most widely used in-situ penetration test designed to provide information on the geotechnical properties of soils. The SPT values (N) can be empirically related to many engineering properties such as unit weight (γ), relative density (D_r), angle of internal friction (φ), and undrained compressive strength (q_u) and soil modulus (E_s). The paper has the following aims:

- To investigate the feasibility of the SVM and ordinary kriging model for site characterization.
- To develop a new type of cross-validation analysis for developed ordinary kriging model.
- To compare the performance of developed SVM and ordinary kriging model.

2 General Site Description and Geotechnical Data

The considered site is located in the alluvial Gangetic plane (Sahajanpur of Uttar Pradesh, India). Site investigation was carried out to explore the soil strata. Out of these boreholes (Figure 1) two bore holes (Bore hole-1 and Bore hole-2) are of 27 m depth and the other bore holes are of 10.5 m depth. As the site is a seismic prone area SPT has been conducted at 1.0 m interval upto 10.5 m and then after at an interval of 1.5 m for Bore hole-1 (BH-1) and Bore hole-2 (BH-2). A typical bore log data for 6 bore holes are shown in Figure 2. The bore log data shows a top clay layer extending up to 2.0 m underlain by approximately 1.0 m of silty sand layer. Thereafter a sand layer with or without kanker extends till 10.0 m.

3 Methodology

In this study, two models (ordinary kriging and SVM) have been adopted for site characterization model. Brief description of two models developed for our study is given below.

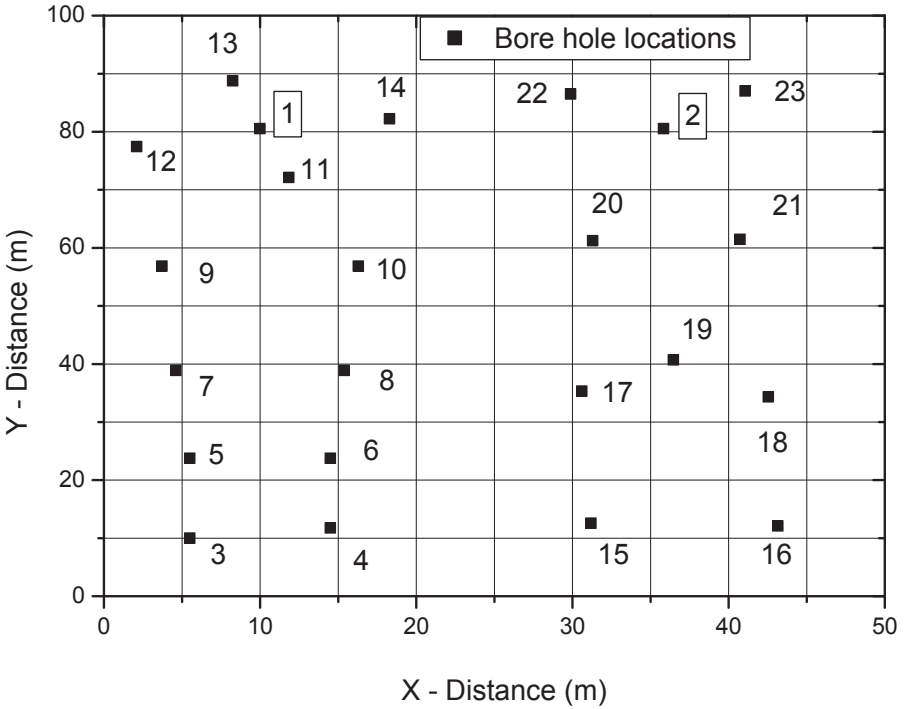

Figure 1. Location of boreholes.

3.1 Ordinary Kriging

Ordinary kriging is a linear geostatistical method which gives local estimation by interpolation. The covariance function between two points is defined as

$$C(h) = E[(N(\mathbf{x}) - m)(N(\mathbf{x}') - m)], \tag{1}$$

where m is the mean of $N(x)$ and $C(h)$ is the covariance function with lag h, with h being the distance between two samples x and x':

$$h = \|x - x'\| = \sqrt{(x - x')^2 + (y - y')^2 + (z - z')^2}. \tag{2}$$

The experimental semivariogram is defined as:

$$\gamma(h) = 0.5 * E[(N(\mathbf{x}) - N(\mathbf{x}'))^2]. \tag{3}$$

Three parameters that define the semivariograms are the nugget, range and sill. The nugget represents unresolved, sub-grid scale variation or measurement error

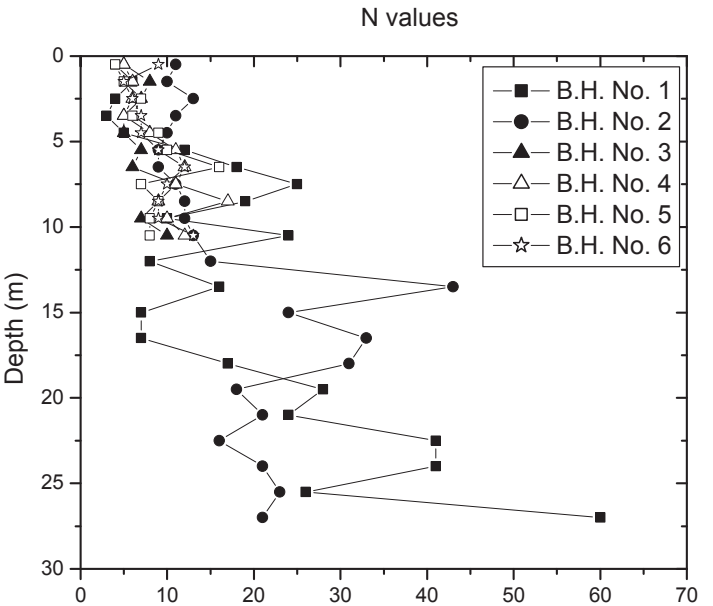

Figure 2. Bore log data showing the variation of SPT (N) value with depth for bore hole number 1 to 6.

and is seen on the semivariogram as the intercept of the semivariogram. Range is the scalar that controls the degree of correlation between data points, usually represented as a distance. Sill is the value of the semi-variance as the lag goes to infinity; it is equal to the total variance of the data set. The relation between the covariance function and the semivariogram is defined as

$$\gamma(h) = C(0) - C(h), \tag{4}$$

where $\gamma(h)$ is the semivariogram.

The details about the kriging are available in literature [4, 7–10]. In the present analysis a linear model has been used.

In the soil strata exist as layers and formation of different layers in geologic age, variation of soil properties is always greater in vertical direction than in the horizontal direction. A vertical anisotropy factor has been taken into account to avoid the distortion of the semivariogram relationship due to the large fluctuation over small vertical distance. The range for horizontal to vertical anisotropy varies from 10–40 [25, 26] for the soil and in the present analysis a factor of 25 has been taken. A factor of 25 means that the known data values located vertically

from a prediction point influence the prediction of the same data points located 25 times the distance horizontally from the prediction point. Once the model of semivariogram is constructed, the weights are computed for ordinary kriging.

Every empirical model has to be tested before it is used for prediction which is known as cross-validation. In practice, model validation is made based on statistical tests involving the residuals. The detailed description of residuals in the case of kriging is given by Kitanidis (1991) [27]. It has been assumed that the n measurements are available at a time, in a given sequence. The kriging estimate of z at the second point, x_2 from the first measurement, x_1 is calculated. So, one can write $\hat{z}_2 = z(x_1)$ and $\sigma_2^2 = 2\gamma(x_1 - x_2)$. Where, $\hat{z}_2$ is the kriged value at the point x_2. The actual error $(\delta_2) = z(x_2) - \hat{z}_2$ is normalized by the standard error (σ_2) and this normalized value of the error is given by:

$$\varepsilon_2 = \frac{\delta_2}{\sigma_2}. \tag{5}$$

For the k-th measurement location, the actual error (δ_k) and normalized error (ε_k) can be written as respectively:

$$\delta_k = z(x_k) - \hat{z}_k, \quad \text{for } k = 2, \ldots, n, \tag{6}$$

$$\varepsilon_k = \frac{\delta_k}{\sigma_k}, \quad \text{for } k = 2, \ldots, n. \tag{7}$$

Cross-validation criteria Q_1 and Q_2 are used to check the statistical distribution of the residuals between the observed data and kriged values at the original observation location by using the same kriging parameters and semi-variogram model parameters. The Q_1 is the mean of the residual (ε_k) and it is written as:

$$Q_1 = \frac{1}{n-1} \sum_{k=2}^{n} \varepsilon_k. \tag{8}$$

Under the null hypothesis, Q_1 is normally distributed with mean 0 and variance $\frac{1}{n-1}$. The probability density function (PDF) of Q_1 is

$$f(Q_1) = \frac{1}{\sqrt{\frac{2\pi}{(n-1)}}} \exp\left(-\frac{Q_1^2}{\frac{2}{(n-1)}}\right), \tag{9}$$

where n is the number of data. If the experimental value of Q_1 turns out to be acceptable close to zero then this test gives no reason to question the validity of the model. Similarly the Q_2 is the variance of ε_k and it is written as:

$$Q_2 = \frac{1}{n-1} \sum_{k=2}^{n} \varepsilon_k^2. \tag{10}$$

$(Q_2) \cdot (n-1)$ approximately follows the chi-square distribution with parameter $(n-1)$. Where, n is the number of data points. The mean and variance of Q_2 are 1 and $\frac{2}{n-1}$ respectively. The PDF of Q_2 is given by the following equation:

$$f(Q_2) = \frac{(n-1)^{\frac{n-1}{2}} Q_2^{\frac{n-3}{2}} \exp\left(-\frac{(n-1)Q_2}{2}\right)}{2^{\frac{n-1}{2}} \Gamma\left(\frac{n-1}{2}\right)}, \tag{11}$$

where Γ is the gamma function. The experimental value of Q_2 should be close to one. In this work, the ordinary kriging equations and cross-validation methodology have been programmed using Matlab software.

3.2 SVM Model

An interesting property of the SVM approach is that it is an approximate implementation of the Structural Risk Minimization (SRM) induction principle which suggests that the generalization ability of learning machines depends more on the capacity concept than merely the dimensionality of the space or the number of free parameters of the loss function. SRM has been shown to be superior to the empirical risk minimization (ERM) principle employed by conventional ANN [20]. In this section, a brief introduction on how to construct SVMs for regression problems is presented. There are three distinct characteristics when SVMs are used to estimate the regression function. First of all, SVMs estimate the regression using a set of linear functions that are defined in a high dimensional space. Secondly, SVMs carry out the regression estimation by risk minimization where the risk is measured using Vapnik's ε-insensitive loss function. Thirdly, SVMs use a risk function consisting of the empirical error and a regularization term which is derived from the SRM principle. In SVM, High generalization performance is achieved by minimizing the sum of the training set error and a term that depends on the Vapnik–Chervonenkis (VC) dimension. This study uses the SVM as a regression technique by introducing an ε-insensitive loss function. The ε-insensitive loss function $(L_\varepsilon(y))$ can be described in the following way

$$L_\varepsilon(y) = 0 \quad \text{for } |f(x) - y| < \varepsilon \quad \text{otherwise} \quad L_\varepsilon(y) = |f(x) - y| - \varepsilon. \tag{12}$$

This defines an ε tube so that if the predicted value is within the tube the loss is zero, while if the predicted point is outside the tube, the loss is the magnitude of the difference between the predicted value and the radius ε of the tube. Assume that the training dataset consists of l training samples $\{(x_1, y_1), \ldots, (x_l, y_l)\}$, where x is the input and y is the output. In this study, The inputs of SVM model are X, Y and Z, where X, Y, and Z are coordinates of borehole. So, $x = [X, Y, Z]$.

The output of SVM is N. Therefore, $y = [N]$. The problem of learning is that of choosing a function that predicts the actual response y as closely as possible, with a precision of ε. Let us assume a linear function

$$f(x) = (w.x) + b, \tag{13}$$

where $w \in \mathbb{R}^n$ and $b \in r$, where, $w =$ an adjustable weight vector, $b =$ the scalar threshold, $\mathbb{R}^n = n$-dimensional vector space and $r =$ one-dimensional vector space.

In the case of SVM, which is based on the SRM principle, the task of minimizing the risk functional with respect to both the empirical risk and the VC-dimension are carried out (Vapnik, 1998).The main aim in SVM is to find a function $f(x)$ that gives a deviation ε from the actual output(y) at the same time as flat as possible. Flatness in the case of (2) means that one seeks a small w. One way of obtaining this is by minimising the Euclidean norm $\|w\|^2$ [24]. The convex optimization problem thus involves:

$$\begin{aligned} &\text{Minimize:} \quad \frac{1}{2}\|w\|^2, \\ &\text{Subjected to:} \quad y_i - (\langle w.x_i \rangle + b) \leq \varepsilon, \quad i = 1, 2, \ldots, l, \\ &\qquad\qquad (\langle w.x_i \rangle + b) - y_i \leq \varepsilon, \quad i = 1, 2, \ldots, l. \end{aligned} \tag{14}$$

The best regression line is defined by minimizing the following cost function:

$$\begin{aligned} &\text{Minimize:} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*) \\ &\text{Subjected to:} \quad y_i - (\langle w.x_i \rangle + b) \leq \varepsilon + \xi_i, \quad i = 1, 2, \ldots, l, \\ &\qquad\qquad (\langle w.x_i \rangle + b) - y_i \leq \varepsilon + \xi_i^*, \quad i = 1, 2, \ldots, l, \\ &\qquad\qquad \xi_i \geq 0 \quad \text{and} \quad \xi_i^* \geq 0, \quad i = 1, 2, ..., l. \end{aligned} \tag{15}$$

The above optimization problem (15) consists of a 2-norm penalty on the regression coefficients, an error term multiplied by the error weight, C, and a set of constraints. Using this cost function, the goal is to simultaneously minimize both the coefficients size and the prediction errors. This is important because large coefficients might hamper generalization because these can cause excessive variance. This is an approach which is often used in multivariate calibration and that can deal with ill-posed problems. These can occur when performing spectral calibration (because the number of objects is smaller than the number of variables or the latter are correlated). In the error term, the prediction errors are penalized

linearly with the exception of a deviation of $\pm\varepsilon$. The predictions deviating more than $+\varepsilon$ or less than $-\varepsilon$, are taken into account, by the so-called slack variables, indicated by ξ_i and ξ_i^*, respectively. This value has to be optimized by the user. Note that the value of ε does not indicate the desired final prediction error of the model but is a characteristic of the prediction error penalty. In principle, it is possible to use an ε-value of zero. In this case, the error penalty comes down to a regular minimization of absolute values (minimizing the 1-norm of the error). C represents the penalty weight and in addition to the value of ε, this value also has to be optimized by the user. If its value is very high, deviations from $\pm\varepsilon$ count heavier in the cost function. For an infinite value of C, a solution is considered best if its error is minimal even though the regression coefficients size is very high. If an extremely low value of C is chosen, the best result is determined exclusively by the size of the regression weights. As stated above, the value of ε also has to be defined by the user and it is data and problem dependent. For example, if noise is present in the data, a larger ε can guide the solutions to be more independent of existing noise. On the other hand, too large values of ε lead to the situation in which no proper predictions can be made. This is caused by the fact that objects with prediction errors larger than $\pm\varepsilon$ are the so-called support vectors and only the support vectors determine the final prediction of the SVM model. This is explained later. ε is a trade of between the sparseness of the representation and closeness to the data. Hence the Lagrangian function is constructed from both the objective function (15) and corresponding constraints (15) as follows

$$\begin{aligned} L(w,\xi,\xi^*,\alpha,\alpha^*,\gamma,\gamma^*) &= \frac{\|w\|^2}{2} + C\Big(\sum_{i=1}^{l}(\xi_i+\xi_i^*)\Big) - \sum_{i=1}^{l}\alpha_i[\varepsilon+\xi_i-y_i+\langle w.x_i\rangle+b] \\ &\quad - \sum_{i=1}^{l}\alpha_i^*[\varepsilon+\xi_i^*+y_i-\langle w.x_i\rangle-b] - \sum_{i=1}^{l}(\gamma_i\xi_i+\gamma_i^*\xi_i^*), \end{aligned} \tag{16}$$

where L is the Lagrangian and α, α^*, γ and γ^* are the Lagrangian multipliers. The partial derivatives of L with respect to ω, b, ξ and ξ^* have to be vanish to satisfy the saddle point condition.

$$\partial_w L = 0 \Rightarrow w = \sum_{i=1}^{l} x_i(\alpha_i - \alpha_i^*),$$

$$\partial_b L = 0 \Rightarrow \sum_{i=1}^{l}\alpha_i = \sum_{i=1}^{l}\alpha_i^*,$$

$$\partial_{\xi} L = 0 \Rightarrow \sum_{i=1}^{l} \gamma_i = \sum_{i=1}^{l} (C - \alpha_i),$$
$$\partial_{\xi^*} L = 0 \Rightarrow \sum_{i=1}^{l} \gamma_i^* = \sum_{i=1}^{l} (C - \alpha_i^*). \tag{17}$$

Substituting (17) into (16) yields the dual optimization problem

$$\text{Maximize:} \quad -\varepsilon \sum_{i=1}^{l} (\alpha_i^* + \alpha_i) + \sum_{i=1}^{l} y_i (\alpha_i^* - \alpha_i)$$
$$-\frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)(x_i . x_j),$$
$$\text{Subjected to:} \quad \sum_{i=1}^{l} \alpha_i = \sum_{i=1}^{l} \alpha_i^*; \quad 0 \le \alpha_i^* \le C \quad \text{and} \quad 0 \le \alpha_i \le C. \tag{18}$$

The coefficients α_i, α_i^* are determined by solving the above optimization problem (18). An important aspect is that some Lagrange multipliers (α_i, α_i^*) will be zero, implying that these training objects are considered to be irrelevant for the final solution (sparseness). The training objects with nonzero Lagrange multipliers are called support vectors. These are the objects with prediction errors larger than $\pm\varepsilon$. In this way, the value of ε determines the amount of support vectors. Obviously, if ε is too large, too few support vectors are selected which leads to a decrease of the final prediction performance. Furthermore, the Lagrange multipliers of the support vectors all have different values, which mean that one support vector is considered to be more important than another one. So equation (13) can be written as

$$f(x) = \sum_{\text{support vectors}} (\alpha_i - \alpha_i^*)(x_i . x) + b,$$

where

$$b = -\left(\frac{1}{2}\right) w.[x_r + x_s]. \tag{19}$$

From (18) it is clear that w has been completely described as a linear combination of training patterns. So, the complexity of a function representation by support vectors is independent of the dimensionality of input space and it depends only on the number of support vectors. The entry of the data in inner products is very

important because: (1) the dimension of the objects does not appear in the problem to be solved and (2) extension of this linear approach to nonlinear regression can be made easily.

When linear regression is not appropriate, then input data has to be mapped into a high dimensional feature space through some nonlinear mapping [28]. In optimization problem in (19), x has been replaced by feature space $\Phi(x)$. So, the optimization problem (19) can be written as

$$\begin{aligned}\text{Maximize:} \quad & -\varepsilon \sum_{i=1}^{l}(\alpha_i^* + \alpha_i) + \sum_{i=1}^{l} y_i(\alpha_i^* - \alpha_i) \\ & -\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)(\Phi(x_i).\Phi(x_j)), \\ \text{Subjected to:} \quad & \sum_{i=1}^{l}\alpha_i = \sum_{i=1}^{l}\alpha_i^*; \quad 0 \le \alpha_i^* \le C \quad \text{and} \quad 0 \le \alpha_i \le C. \end{aligned} \tag{20}$$

The concept of kernel function $[K(x_i, x_j) = \Phi(x_i).\Phi(x_j)]$ has been introduced to reduce the computational demand [29, 30]. So optimization problem can be written as:

$$\begin{aligned}\text{Maximize:} \quad & -\varepsilon \sum_{i=1}^{l}(\alpha_i^* + \alpha_i) + \sum_{i=1}^{l} y_i(\alpha_i^* - \alpha_i) \\ & -\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)(K(x_i.x_j)), \\ \text{Subjected to:} \quad & \sum_{i=1}^{l}\alpha_i = \sum_{i=1}^{l}\alpha_i^*; \quad 0 \le \alpha_i^* \le C \quad \text{and} \quad 0 \le \alpha_i \le C. \end{aligned} \tag{21}$$

The introduction of kernels according to Mercer's theorem [18] avoids an explicit formation of the nonlinear mapping, makes the dimension of feature space even infinite, and reduces the computation load greatly by enabling the operation in low dimensional input space instead of high dimensional feature space. Some common kernels have been used, such as polynomial (homogeneous), polynomial (non-homogeneous), radial basis function, Gaussian function, sigmoid for non-linear cases. The Kernel representation offers a powerful alternative for using linear machines in hypothesizing complex real world problems as opposed to ANN based learning paradigms, which use multiple layers of threshold linear functions [30].

The regression function (13) has been obtained by applying same procedure as in linear case. An important characteristic of this optimization problem is that the solution is global and deterministic (i.e., given the same training set and values of ε and C, the same solution is always found, that is, no stochastic events are present during the building of the model), which is in contrast with ANNs. a regression technique by introducing a ε-insensitive loss function. The data has been divided into two sub-sets: such as

(a) A training dataset: This is required to construct the model. In this study, 300 out of the 334 N values are considered for training the dataset.

(b) A testing dataset: This is required to estimate the model performance. In this study, the remaining 34 N values are considered as testing dataset.

The data has been scaled (normalized) between 0 and 1 before being presented to the model. To train the SVM model, radial basis function has been used as kernel functions. While applying SVM, in addition to the specific kernel parameters, the optimum values of the capacity factor C and the size of the error-insensitive zone ε should be determined during the modeling experiment. In training process, C, ε and width (σ) of radial basis function have been chosen by trial-and-error approach. The above formulation is implemented using Matlab.

4 Results and Discussion

In case of ordinary kriging, the semivariogram of SPT values N obtained from the experimental values and the linear model has been plotted in Figure 3. In the semivariogram, "relative to the full length scale" means normalized lag distance ($\frac{h}{a}$). It can be seen that it gives a reasonable fit to the values obtained. The range and nugget of the semivariogram are found to be 0.95 and 0.025 respectively. One of the most important findings of this study is that the semivariogram is free from white noise or a pure nugget effect that corresponds to the total absence of auto-correlation. A weighted nonlinear least squares method has been used to fit semivariogram model. The point closer to the origin are given higher weights than points further away, because they are inherently more accurate, as they are calculated using more data pairs. The developed kriging model gives a unique solution. The reason for unique solutions is explained below.

Considering a case of two different measurements obtained at the same location, if the semivariogram for both models is continuous (with zero nugget) the function $N(x)$ will be continuous. This means that one of the two measurements is redundant. Thus, one must be discarded; otherwise, a unique solution cannot be obtained because the determinant of the matrix coefficients of the kriging vanishes.

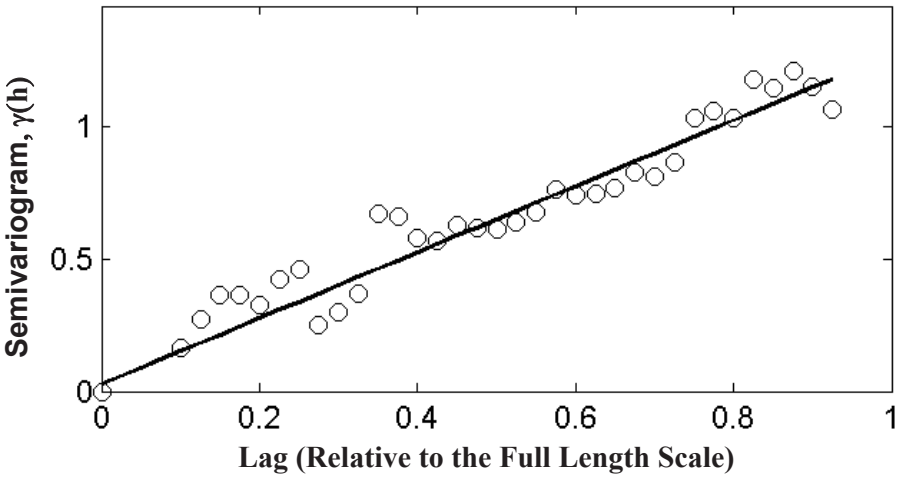

Figure 3. Semivariogram model for N values using ordinary kriging.

This problem is solved by adding a nugget term to the semivariogram. As a result, the linear model adopted shows the nugget effect. The nugget effect decreases the condition number of the kriging matrix. In practical sense, nugget effect gives the kriging equations a stability and robustness. Without a nugget effect, inverting the kriging matrices may lead to computational round-off error. As the sampling distance decreases, it is possible to obtain a better estimate of nugget effect. But if the sampling interval is reduced to a very small distance, then the cost of the exploration program increases enormously. So it is undesirable to reduce the sampling spacing below some nominal minimum value. The value of Q_1 (0.028) close to 0 and Q_2 (1.003) close to 1.0 indicates that the developed ordinary kriging model is good for the estimation of N value in the site.

In this study, coefficient of correlations (R) is used to evaluate the performance of the developed SVM models. In SVM training consisted in solving a-uniquely solvable-quadratic optimization problem, unlike the ANN training, which requires complex non-linear optimization with the danger of getting stuck in local minima. The design value of C, ε and σ is 6, 0.03 and 0.4 respectively. The number of support vector is 238. Figure 4 illustrates the performance of the training dataset ($R = 0.930$) and the result are almost identical to the original data. Figure 5 shows the performance of the SVM model for testing dataset ($R = 0.889$).

From the results, it is clear that the SVM model has predicted the actual value N very well and it can be used as a practical tool for the determination of N. The loss of performance with respect to the testing set addresses SVM's susceptibility

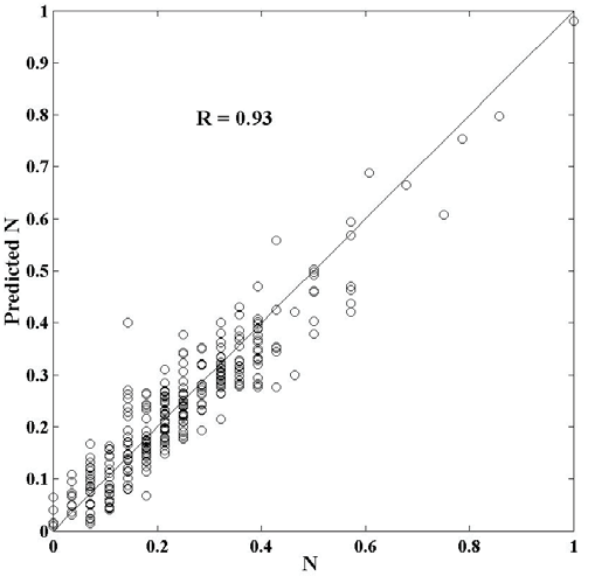

Figure 4. Performance of SVM model for training dataset using radial basis function.

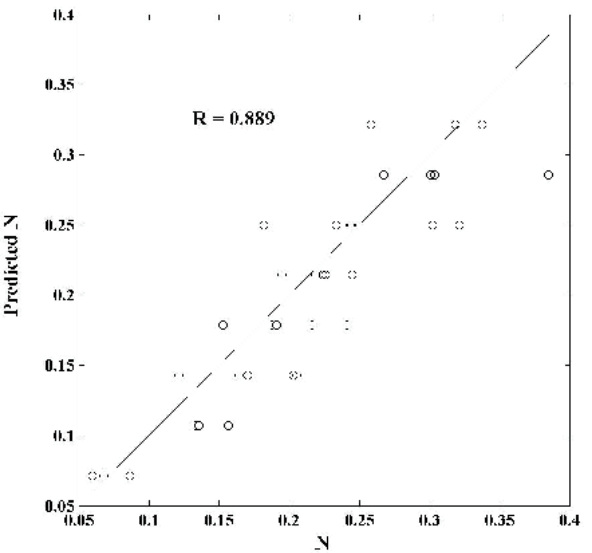

Figure 5. Performance of SVM model for testing dataset using radial basis function.

$X(m)$	$Y(m)$	$Z(m)$	Actual N	Predicted N by ordinary kriging	Predicted N by SVM
12.35	1.3	9.9	11	14	11
18.15	7.12	5.1	5	3	6
16.28	68.97	6.5	9	6	10
5.87	68.74	4	11	14	9
10.84	75.5	0.9	10	13	10

Table 1. Comparison between SVM and Ordinary Kriging model.

to overtraining. There is a marginal reduction in performance on the testing dataset (i.e., there is a difference between SVM performance on training and testing) for the SVM model. So, SVMs have the ability to avoid overtraining, and hence it has good generalization capability. In this study, SVM model employs approximately 79 to 86 percent (for radial basis function = 79.33%, polynomial kernel = 85.33% and spline kernel = 83.33%) of the training data as support vectors. It is worth mentioning here that the support vectors in the SVM model represent prototypical examples. The prototypical examples exhibit the essential features of the information content of the data, and thus are able to transform the input data into the specified targets. So, there is real advantage gained in terms of sparsity. Sparseness means that a significant number of the weights are zero (or effectively zero), which has the consequence of producing compact, computationally efficient models, which in addition are simple and therefore produce smooth functions.

In order to compare between the ordinary kriging and SVM models, five points have been chosen randomly from known N values of 334 points in the subsurface of the site. The predicted values of these points are shown in Table 1. It can be seen from the table that the SVM model gives better prediction than ordinary kriging model. Similar random pick up of points are made several times and it was observed that SVM model prediction was better than ordinary kriging model.

5 Conclusion

Ordinary kriging model and SVM model has been developed for predicting N values in the three dimensional subsurface of a site. In case of ordinary kriging model, the procedure to determine the semivariogram model that quantifies the spatial variability of N values in the three dimensional subsurface of a site has been discussed. By the use of the semivariogram, it is possible to make an estimation of the N value at points of the site where N value measurements were not known.

A new type of cross-validation analysis ($Q_1 = 0.028$ and $Q_2 = 1.003$) proves the robustness of the developed Ordinary Kriging model presented in this study. For SVM model, the result indicated that SVM model have the ability to predict N values in three dimensional subsurface with an acceptable degree of accuracy. Based on statistical performances and predicted SPT values at random points SVM model is found to better than ordinary kriging model.

Bibliography

[1] G. B. Baecher, Geotechnical error analysis. Trans. Research Record, 1105 (1986), 23–31.

[2] G. A. Fenton, Estimation for stochastic soil model., J. Geo. and Geoenv. Engg. ASCE, 125(6) (1999), 470–485.

[3] K. K. Phoon, Q. Kwang, S. Tong and P. An, Identification of statically homogeneous soil layers using modified Barlett statistics, Journal of Geotech. and Geoenv. Engineering ASCE, 129(7) (2003), 649–659.

[4] M. Soulie, P. Montes and V. Silvestri, Modeling spatial variability of soil parameters. Canadian Geotechnical Journal, 27(5) (1990), 617–630.

[5] P. Chiasson, J. Lafleur, M. Soulie and K. T. Law, Characterizing spatial variability of clay by geostatistics, Canadian Geotechnical Journal, 32(1) (1995), 1–10.

[6] T. Elkateb, R. Chalaturnyk and P. K. Robertson, Simplified geostatistical analysis of earthquake-induced ground response at the Wildlife Site, California, U.S.A., Canadian Geotechnical Journal, 40(1) (2003b), 16–35.

[7] G. Matheron, Principles of geostatistics, Economic Geology, 58 (1963), 1246–1266.

[8] E. H. Isaaks and R. M. Srivastava, an Introduction to Applied Geostatics, Oxford University Press, New York, 1989.

[9] J. C. Davis, Statistics and Data Analysis in Geology, Wiley, New York, 2002.

[10] A. G. Journel and C. J. Huijbregts, Mining Geostatistics. New York Academic Press, NewYork, 1978.

[11] G. L. Sivakumar Babu and M. D. Mukesh, Effect of soil variability on reliability of soil slopes, Geotechnique, 54(5) (2004), 335–337.

[12] T. Elkateb, R. Chalaturnyk and P. K. Robertson, An overview of soil heterogeneity: quantification and implications on geotechnical field problems, Canadian Geotechnical Journal, 40(1) (2003a) 1–15.

[13] C. H. Juang, T. Jiang and R. A. Christopher, Three-dimensional site characterization: neural network approach, Geotechnique, 51 9 (2001), 799–809.

[14] O. M. Itani and Y. M. Najjar, Three-dimensional modeling of spatial soil properties via artificial neural network, Transportation research record, 1709 (2000), 50–59.

[15] S. K. Das and P. K. Basudhar, Site characterization of an alluvial site using artificial neural network, Proceedings of the International workshop on Risk Assessment in Site Characterization and Geotechnical Design, Bangalore, 2004.

[16] D. Park and L. R. Rilett, Forecasting freeway link ravel times with a multi-layer feed forward neural network, Computer Aided Civil and Infrastructure Engineering. 14 (1999), 358–367.

[17] V. Kecman, Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models, The MIT press, London, 2001.

[18] V. N. Vapnik, 1995. The nature of statistical learning theory, Springer, New York, 1995.

[19] S. Mukherjee, E. Osuna and F. Girosi, Nonlinear prediction of chaotic time series using support vector machine, Proc., IEEE Workshop on Neural Networks for Signal Processing 7, Institute of Electrical and Electronics Engineers, New York, 1997, 511–519.

[20] V. N. Vapnik, Statistical learning theory, Wiley, New York, 1998.

[21] A. T. C. Goh and S. H. Goh, Support vector machines: Their use in geotechnical engineering as illustrated using seismic liquefaction data, Computers and Geotechnics, 34(5) (2007), 410–421.

[22] P. Samui, K. Pradeep and T. G. Sitharam, OCR prediction using support vector machine based on piezocone data, Journal of Geotechnical and GeoEnvironmental engineering, 134(6) (2008), 894–898.

[23] P. Samui, Support vector machine applied to settlement of shallow foundations on cohesionless soils, Computers and Goetechnics, 35(3) (2008), 419–427.

[24] A. J. Smola and B. Scholkopf, “A tutorial on support vector regression,” Statistics and Computing, 14 (2004), 199–222.

[25] D. J. Degroot, Analyzing spatial variability of in situ soil properties. Proceedings of uncertainty’96, uncertainty in the geologic environment: from theory to practice, ASCE geotechnical special publications, 58 (1996), 210–238.

[26] K. K. Phoon and F. H. Kulhawy, On quantifying inherent soil variability, Proceeding of uncertainty ’96, uncertainty in the geologic environment: from theory to practice, ASCE geotechnical special publication, 58 (1996), 326–340.

[27] P. K. Kitanidis, Orthonormal residuals in geostatistics: Model criticism and parameter estimation, Math. Geology 23(5) (1991), 741–758.

[28] B. E. Boser, I. M. Guyon, and V. N. Vapnik, A training algorithm for optimal margin classifiers, 5th Annual ACM Workshop on COLT, 1992, 144–152.

[29] N. Cristianini and J. Shawe-Taylor, An introduction to Support vector machine, Cambridge University press, London, 2000.

[30] C. Cortes and V. N. Vapnik, Support vector networks, MachineLearning, 20 (1995), 273–297.

Received July 18, 2011.

Author information

Pijush Samui, Associate Professor, Centre for Disaster Mitigation and Management,
VIT University, Vellore-632014, India.
E-mail: pijush.phd@gmail.com

Sarat Das, Associate Professor, Department of Civil Engineering,
National Institute of Technology Rourkela, 769 008, India.
E-mail: saratdas@rediffmail.com