Math Geosci (2010) 42: 433–446
DOI 10.1007/s11004-010-9268-7

CASE STUDY

# Applicability of Statistical Learning Algorithms for Spatial Variability of Rock Depth

**Pijush Samui · T.G. Sitharam**

Received: 1 December 2008 / Accepted: 26 January 2010 / Published online: 23 February 2010


**Abstract** Two algorithms are outlined, each of which has interesting features for modeling of spatial variability of rock depth. In this paper, reduced level of rock at Bangalore, India, is arrived from the 652 boreholes data in the area covering 220 sq·km. Support vector machine (SVM) and relevance vector machine (RVM) have been utilized to predict the reduced level of rock in the subsurface of Bangalore and to study the spatial variability of the rock depth. The support vector machine (SVM) that is firmly based on the theory of statistical learning theory uses regression technique by introducing $\varepsilon$-insensitive loss function has been adopted. RVM is a probabilistic model similar to the widespread SVM, but where the training takes place in a Bayesian framework. Prediction results show the ability of learning machine to build accurate models for spatial variability of rock depth with strong predictive capabilities. The paper also highlights the capability of RVM over the SVM model.

**Keywords** Support vector machine · Relevance vector machine · Rock depth · Spatial variability

## 1 Introduction

Rock depth in a site which is a very useful parameter to the geotechnical earthquake engineers to find their basic requirement of hard strata and ground motion at rock level. In most of the geotechnical investigations, knowledge of the hard strata or rock is essential to decide the type of foundations and design a suitable foundation for

P. Samui (✉)
Center for Disaster Mitigation and Management, VIT, Vellore, India
e-mail: pijush.phd@gmail.com

T.G. Sitharam
Department of Civil Engineering, Indian Institute of Science, Bangalore 560012, India
e-mail: sitharam@civil.iisc.ernet.in

a structure. In the ground response analysis, peak ground acceleration (PGA) and response spectrum for the particular site is evaluated at the rock depth levels and further on at the ground level considering local site effects. This is an essential step to evaluate site amplification and liquefaction hazards of a site and further to estimate induced forces on the structures. In ground response analysis, the response of the soil deposit is determined from the motion at the bed rock level. In all these problems, it is essential to evaluate the depth of the hard rock from the ground level. Thus, modeling of spatial variation of rock depth is an imperative task in geotechnical and geological engineering.

In this study, spatial variability of rock depth in Bangalore is modeled by statistical learning theory (Berger 1985) using a high-dimensional regression technique. This technique attempts to find a relationship between a training set of input vectors and corresponding outputs, also known as supervised learning (Tipping 2001). Once training has been done, the goal is to make predictions at unmeasured locations by incorporating any prior knowledge that may be available. Two types of statistical learning theory have been adopted. The first type uses SVM (Cristianini and Shawe-Taylor 2000) which is a relatively new type of learning algorithms, originally introduced by Vapnik and coworkers (Boser et al. 1992; Vapnik 1998). It achieves good generalization ability by adopting a structural risk minimization (SRM) induction principle that aims at minimizing a bound on the generalization error of a model rather than the minimizing the error on the training data only. It also 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. Recently, it has been used to solve nonlinear regression estimation and time series prediction by introducing $\varepsilon$-insensitive loss function (Mukherjee et al. 1997; Muller et al. 1997; Vapnik 1995; Vapnik et al. 1997). Application of SVM for geospatial data has been done by different researchers (Kanevski and Maignan 2004; Kanevski 2008). It is trained with optimization of a convex, quadratic cost function, which guarantees the uniqueness of the SVM solution. The recently introduced RVM by Tipping (2001) has been used as second type technique. RVM allows computation of the prediction taking into account uncertainties of both the parameters and the data. It provides much sparser regressors without compromising performance, and kernel bases give a small but worthwhile improvement in performance. A comparative study has also been done between the developed SVM and RVM model.

## 2 Subsurface of Bangalore and GIS Model Development

Bangalore covers an area of over 220 square kilometers and ground reduced levels (GRL) also vary a lot in the city. It varies from 810 m in the northeast part to 940 m in southwestern part of Bangalore. Ground reduced levels do not vary much in the central and northwestern parts of the city. There were more than 400 lakes at one time, and more than 340 lakes are dried up and have been encroached for development of residential and industrial layout. The population of the greater Bangalore region is over 6 million and it is the fifth largest city in India. It is situated on latitude of 1,208′ north and longitude of 77,037′ east. From geology, the subsurface of Bangalore region covers in Gneiss complexes, which is formed due to several tectonic-thermal

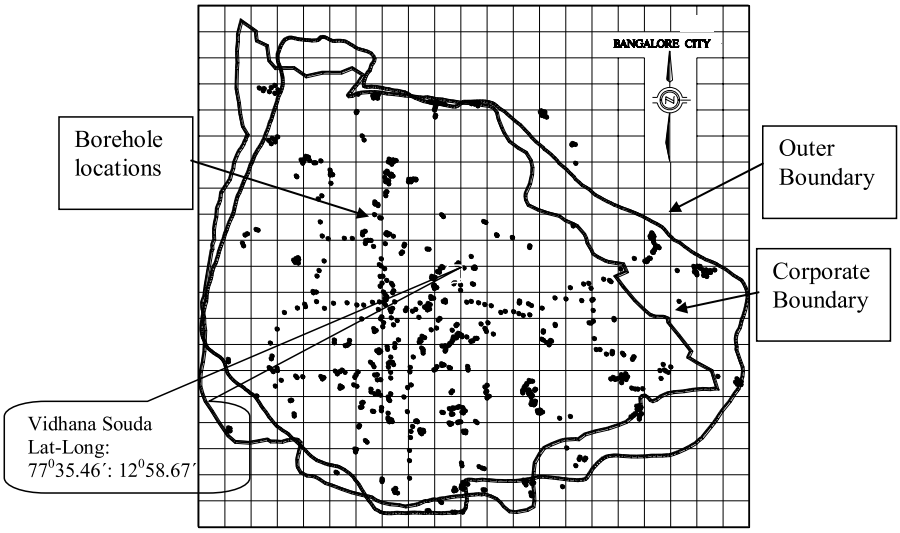


**Fig. 1** Borehole location in Bangalore Map (scale: 1:20000)

events with a large influx of sialic material, are believed to have occurred between 3,400 to 3,000 million years ago giving rise to an extensive group of gray gneisses designated as the older gneiss complex. These gneisses act as the basement for a widespread belt of schists. The younger group of gneissic rocks mostly of granodiomitic and granitia composition is found in the eastern part of the state, representing remobilized parts of an older crust with abundant additions of newer granite material, for which the name = younger gneiss complex has been given (Radhakrishna and Vaidyanadhan 1997). The soil is mostly a residual soil from granite gneiss due to weathering action. In the old tank beds, silty sand/clay is also found as overburden.

A geographic information system (GIS) model (see Fig. 1) of Bangalore with several layers on a scale of 1:20000 has been developed with a purpose of carrying out microzonation of Bangalore. The Bangalore map forms the base layer for GIS. The map entities have been developed for locating the boreholes to the utmost accuracy and at each location borelogs have been attached along with geotechnical data of each layer up to the hard rock. The digitized map has several layers of information. Some of the important layers considered are the boundaries (outer and administrative), highways, major roads, minor roads, Streets, railroads, water bodies, drains, ground contours and borehole locations. The locations of boreholes are shown in Fig. 1 along with ground reduced level with an interval of 10 m (see Fig. 2). Distribution of collected boreholes in Bangalore is shown in Fig. 3, indicating a very good distribution of the boreholes in each quadrant of Bangalore from the city center. Figure 1 also depicts grids of 1 km × 1 km along with the corporate boundary of Bangalore and outer boundary circumscribing the ring road. Figure 1 gives a clear view of the spatial distribution of boreholes in the Bangalore region. An average of about three boreholes data is available within the grid of 1 km × 1 km.

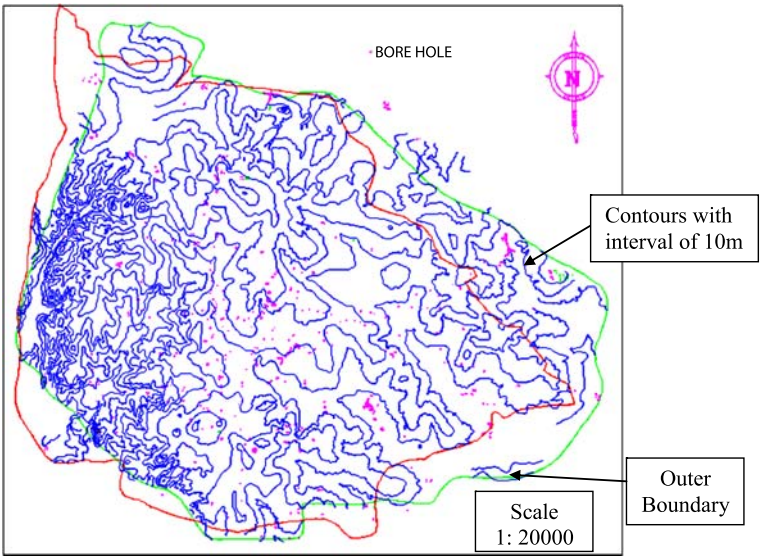


**Fig. 2** GIS model of borehole locations with respect to contours

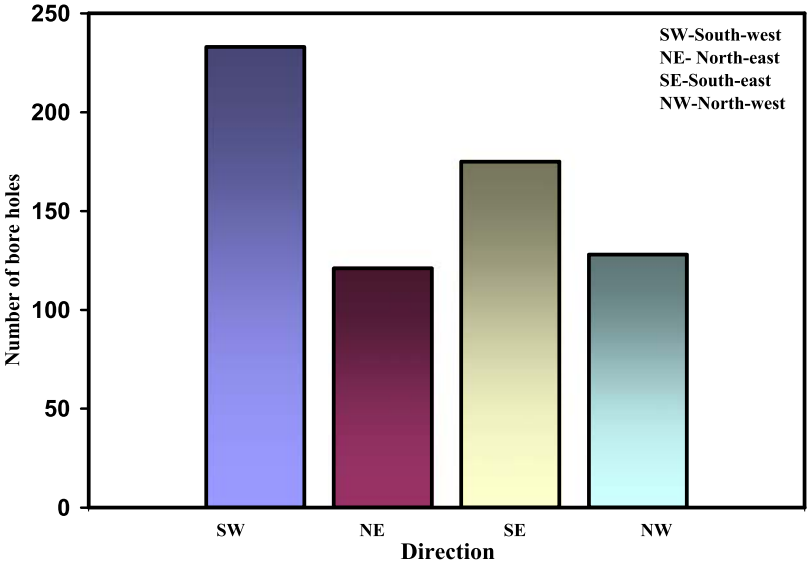


**Fig. 3** Distribution of boreholes in quadrants for Bangalore

Geotechnical data for 652 boreholes was collated from archives of only two organizations: Torsteel Research Foundation in India and Indian Institute of Science. This data was generated for geotechnical investigations carried out for several major projects in Bangalore including Bangalore metro project. The data collected is of very high quality and collected during the years 1995–2003. The data in the model

is on average to a depth of 30 m below the ground level. Each borelog contains information about depth, density of the soil, total stress, effective stress, fines content, and $N$ values, depth of ground water table, and reduced level rock depth ($d$). In GIS model, the boreholes are represented as a three-dimensional object spanning below the map layer. These three-dimensional boreholes are generated with several layers with a bore location in each layer overlapping one below the other and each layer representing 0.5 m interval of the subsurface. Each layer of this model is attached with borelog data at that depth. The data consists of visual soil classification, borehole location, ground water level, date, and time during which test has been carried out, other physical and engineering properties of soil, and rock depth. As such, when this model is viewed in a three-dimensional subsurface information on any borehole at any depth can obtained by clicking at that level. The hard rock has been identified by visual observation of the cores taken at these locations. Rock depth from ground level is the difference between the grounds reduced level at borehole location and reduced level of the hard rock at the same borehole location. The reduced level of the hard rock at borehole location is the difference between the ground reduced level at borehole location and depth of overburden thickness up to hard rock for the same borehole. The depth of overburden is estimated from the available borelogs.

## 3 Support Vector Machine Model

The support vector machine (SVM) based on statistical learning theory has been developed by Vapnik (1995). It represents a learning paradigm where prediction error and model complexity are simultaneously minimized. Unlike artificial neural network, the SVM structure is not fixed in advance with a specific number of adjustable parameters, but can adapt with data. In this study, SVM used regression technique by introducing $\varepsilon$-insensitive loss function. In support vector regression, the input $x$ is first mapped into a high dimensional feature space by the use of a kernel function, and then a linear model is constructed in this feature space. The kernel functions often used in SVM include linear, polynomial, radial basis function, and sigmoid function. The linear model $f(x; w)$ in the feature space is given by

$$f(x; w) = \sum_{j=1}^{m} w_j \times \Phi_j(x) + b, \tag{1}$$

where the functions $\{\Phi_j(x)\}_{j=1}^{m}$ are feature space representations of the input query $x$, $m$ is the number of patterns that contains all the information necessary to solve a given learning task, hereinafter, referred to as support vectors, and $w_j$ is the SVM weights. SVM regression uses a new type of loss function called $\varepsilon$-insensitive (see Fig. 4) loss function proposed by Vapnik (1998).

$$\begin{aligned} L_\varepsilon\big(y; f(x; w)\big) &= 0 \quad \text{for } \big|f(x; w) - y\big| < \varepsilon \quad \text{otherwise} \\ L_\varepsilon\big(y, f(x; w)\big) &= \big|f(x; w) - y\big| - \varepsilon. \end{aligned} \tag{2}$$

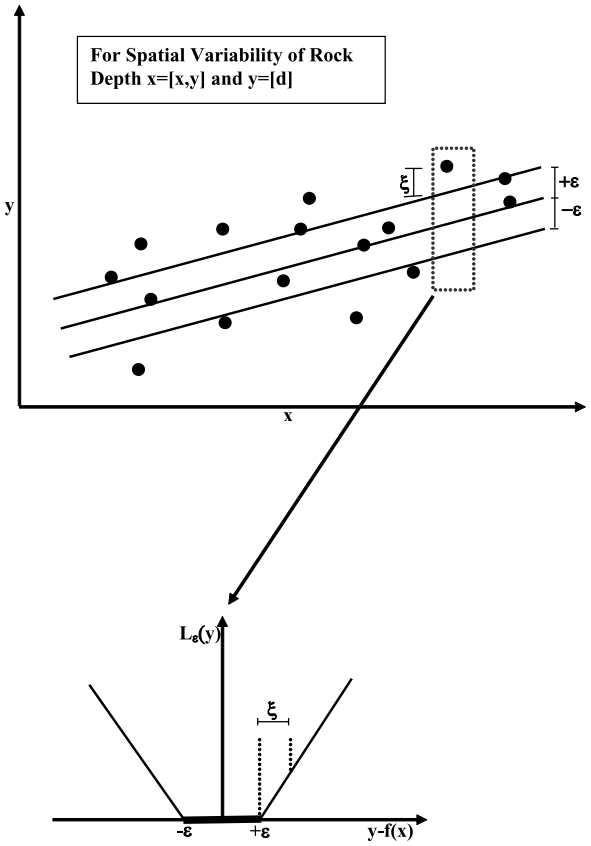


**Fig. 4** Prespecified accuracy $\varepsilon$ and slack variable $\xi$ in support vector regression (Scholkopf 1997)

The empirical risk is

$$R_{\text{emp}}(w) = \frac{1}{m}\sum_{i=1}^{m} L_e\big(y_i, f(x_i; w)\big). \tag{3}$$

SVM regression performs linear regression in the high dimension feature space using $\varepsilon$-insensitive loss and, at the same time, tries to reduce model complexity by minimizing $\|w\|^2$. Thus, SVM regression is formulated as a minimization of the following functional

$$\begin{aligned} &\text{Minimize} \quad \frac{C}{m}\sum_{i=1}^{m} L_e\big(y_i, f(x_i; w)\big) + \frac{1}{2}\|w\|^2 \\ &\text{Subjected to} \quad (2), \end{aligned} \tag{4}$$

where $C$ is referred to as the regularized constant and it determines the trade-off between the empirical risk and the regularization term. In practice, $C$ is selected by trial and error. To obtain the estimations of $w$ and $b$, (4) is transformed to the primal function given by (5) by introducing the positive slack variables and $\xi_i$ and $\xi_i^*$ as

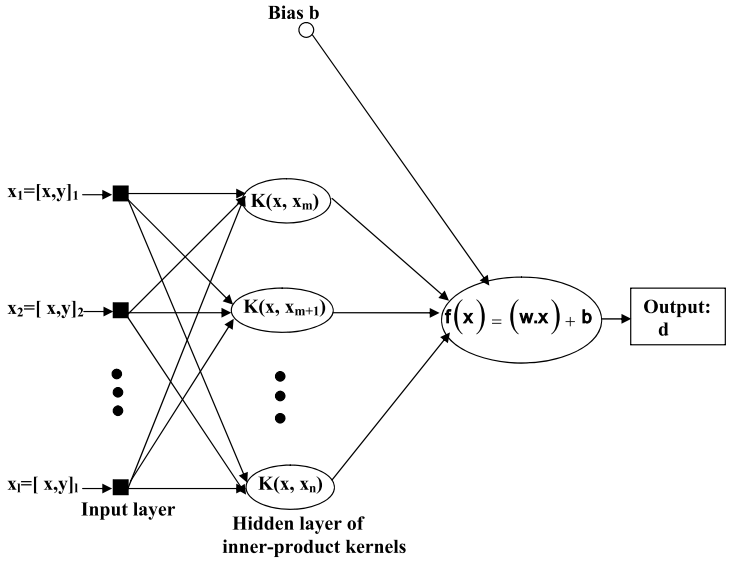


**Fig. 5** Architecture of SVM for rock depth prediction

follows

$$\begin{aligned} &\text{Minimize} && C\sum_{i=1}^{m}(\xi_i+\xi_i^*)+\frac{1}{2}\|w\|^2 \\ &\text{Subjected to} && y_i - f(x_i;w) \le \varepsilon+\xi_i \\ &&& f(x_i;w) - y_i \le \varepsilon+\xi_i^* \\ &&& \xi_i,\xi_i^* \ge 0, \quad i=1,\ldots,m. \end{aligned} \tag{5}$$

This optimization problem can be transformed into a quadratic programming problem (Vapnik 1998) and its solution is given by

$$\begin{aligned} &f(x)=\sum_{i=1}^{\text{nsv}}(\alpha_i-\alpha_i^*)K(x_i,x) \\ &\text{subjected to} \quad 0\le\alpha_i^*\le C,\ 0\le\alpha_i\le C, \end{aligned} \tag{6}$$

where nsv is the number of Support Vectors (SV) and the kernel function

$$K(x,x_i)=\sum_{i=1}^{m}\Phi_j(x)\Phi_j(x_i). \tag{7}$$

Figure 5 shows a typical architecture of SVM for d prediction.

The main scope of this work is to implement above methodology to study the spatial variability $d$ in the subsurface of Bangalore. For predicting $d$ in a given space, the two input variables ($x$, $y$; where $x$ and $y$ are the coordinates of borehole in Bangalore)

are used for the SVM model in this study. The only output is $d$. In SVM analysis, normalization of the data is very important. There are many ways of normalizing data, but the method used in this analysis is normalizing the data against their maximum values (Sincero 2003). In carrying out the formulation, 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, 457 out of the 652 $d$ values are considered as training dataset.
(b) A testing dataset: This is required to estimate the model performance. In this study, the remaining 195 data is considered as testing dataset.

The generalization performance of SVM depends on a good setting of parameters: $C$, $\varepsilon$, and the kernel type, and corresponding kernel parameters. The selection of the kernel function and corresponding parameters is very important because they define the distribution of the training set samples in the high dimensional feature space. To train the SVM model, the Gaussian kernel function has been used. In training process, $C$, $\varepsilon$, and other kernel specific parameters have been chosen by a trial-and-error approach. The other methods are also available in the literature (Kanevski and Maignan 2004). Different combinations of $C$ and $\varepsilon$ values are tried to yield the best performance on training data. Using a low value of $C$ could result in fewer penalties to errors. This allows the minimization of margin with errors, thus higher generalization ability. On the other hand, a large $C$ assigns higher penalties to errors so that the regression is trained to minimize error with lower generalization. If $C$ goes to infinitely large, SVM would not allow the occurrence of any error and result in a complex model, whereas when $C$ goes to zero; the result would tolerate a large amount of errors and the model would be less complex. In the present study, training and testing of SVM has been carried out by using MATLAB (MathWork Inc. 1999).

## 4 Relevance Vector Machine Model

Tipping (2000) proposed the RVM to recast the main ideas behind support vector machine (SVM) in a Bayesian context, and using mechanisms similar to Gaussian processes. A brief review of Tipping's paper is presented here for those unfamiliar with the work. The RVM model seeks to forecast $y$ for any quarry $x$ according to $y = f(x, w) + \varepsilon_n$, where the error term $\varepsilon_n = N(0, \sigma^2)$ is a zero-mean Gaussian process and $w$ is a vector of weights. The likelihood of the complete data set can be written as

$$p\left(y/w, \sigma^2\right) = \left(2\pi\sigma^2\right)^{-N/2} \exp\left\{-\frac{1}{2\sigma^2}\|y - \Phi w\|^2\right\}, \tag{8}$$

where $\Phi(x_i) = [1, K(x_i, x_1), K(x_i, x_2), \ldots, K(x_i, x_N)]^T$.

Without imposing the hyperparameters on the weights, w, the maximum likelihood of (1) will suffer from sever overfitting. Therefore, Tipping (2001) recommended imposition of some prior constraints, $w$, by adding a complexity penalty to the likelihood or the error function. An explicit zero-mean Gaussian prior probability distrib-

ution over the weights, $w$ with diagonal covariance of $\alpha$ is proposed as follows

$$p(w/\alpha) = \prod_{i=0}^{N} N\left(w_i/0, \alpha_i^{-1}\right), \tag{9}$$

with $\alpha$ a vector of $N+1$ hyperparameters. Consequently, using Bayes rule, the posterior over all unknowns could be computed given the defined noninformative prior distribution

$$p\left(w, \alpha, \sigma^2/y\right) = \frac{p(y/w, \alpha, \sigma^2)p(w, \alpha, \sigma)}{\int p(y/w, \alpha, \sigma^2)p(w, \alpha, \sigma^2)\,dw\,d\alpha\,d\sigma^2}. \tag{10}$$

Full analytical solution of this integral (3) is obdurate. Thus, decomposition of the posterior according to $p(w, \alpha, \sigma^2/y) = p(w/y, \alpha, \sigma^2)p(\alpha, \sigma^2/y)$ is used to facilitate the solution (Tipping 2001). The posterior distribution over the weights is thus given by

$$\begin{aligned} p\left(w/y, \alpha, \sigma^2\right) &= \frac{p(y/w, \sigma^2)p(w/\alpha)}{p(y/\alpha, \sigma^2)} \\ &= (2\pi)^{-(N+1)/2}|\Sigma|^{-1/2}\exp\left\{-\frac{1}{2}(w-\mu)^T\sum^{-1}(w-\mu)\right\}, \end{aligned} \tag{11}$$

where $\Sigma$ and $\mu$ are posterior covariance and mean, respectively, with $A = \mathrm{diag}(\alpha_0, \alpha_1, \ldots, \alpha_N)$. Therefore, learning becomes a search for the hyperparameter posterior most probable, i.e., the maximization of $p(\alpha, \sigma^2/y) \propto p(y/\alpha, \sigma^2)p(\alpha)p(\sigma^2)$ with respect to $\alpha$ and $\sigma^2$. For uniform hyperpriors over $\alpha$ and $\sigma^2$, one needs only maximize the term $p(y/\alpha, \sigma^2)$

$$\begin{aligned} p\left(y/\alpha, \sigma^2\right) &= \int p\left(y/w, \sigma^2\right)p(w/\alpha)\,dw \\ &= \left(\frac{(2\pi)^{\frac{-N}{2}}}{\sqrt{|\sigma^2 + \Phi A^{-1}\Phi^T|}}\right)\exp\left\{-\frac{1}{2}y^T\left(\sigma^2 + \Phi A^{-1}\Phi^T\right)^{-1}y\right\}. \end{aligned} \tag{12}$$

Maximization of this quantity is known as the type II maximum likelihood method (Berger 1985; Wahba 1985) or the evidence for hyperparameter—(MacKay 1992). Hyperparameter estimation is carried out in iterative formulae, e.g., gradient descent on the objective function (Tipping 2001). The outcome of this optimization is that many elements of $\alpha$ go to infinity such that w will have only a few nonzero weights that will be considered as relevant vectors. The relevance vector can be viewed as counterparts to support vectors in SVM. Therefore, the resulting model enjoys the properties of SVM.

The plausibility of the above proposed models is evaluated using $d$ in the subsurface of Bangalore. In RVM model, the same training dataset, testing dataset, kernel, and normalization technique have been used as used in SVM model. The kernel size value has been chosen by trial and error approach for RVM. In the present study, training and testing of RVM has been carried out by using MATLAB (MathWork Inc. 1999).

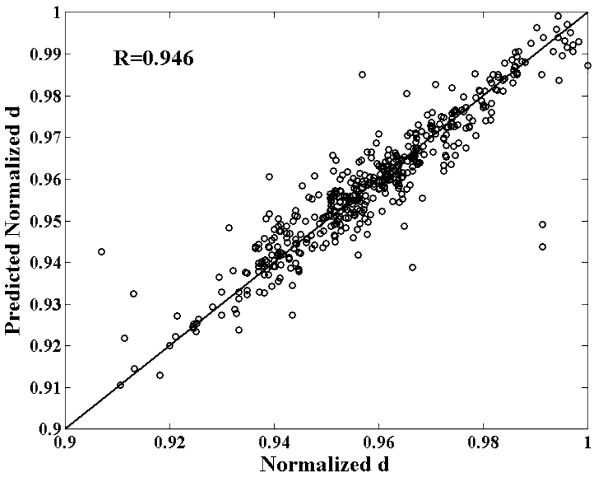


**Fig. 6** Performance of SVM model for training dataset using Gaussian kernel

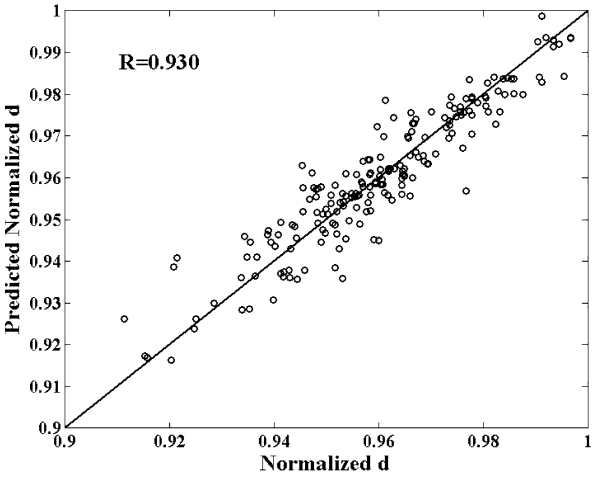


**Fig. 7** Performance of SVM model for testing dataset using Gaussian kernel

## 5 Results and Discussion

The coefficient of correlations ($R$) is the main criteria that are used to evaluate the performance of the SVM and RVM models developed in this work. Different combinations of $C$ and $\varepsilon$ values are tried to yield the best performance on training data for SVM model. The design value of $C$, $\varepsilon$ and width of Gaussian kernel ($\sigma$) is 100, 0.001 and 0.09 respectively. Figure 6 shows the performance of the SVM model for training dataset using Gaussian kernel. In order to evaluate the capabilities of the SVM model, the model is tested with new d values that are not part of the training dataset. Figure 7 shows the performance of the SVM model for testing dataset using Gaussian kernel. The number of support vector is 390.

For RVM model, the design value of $\sigma$ is 0.15. Figure 8 illustrates the performance of the training dataset using Gaussian kernel, and the results are almost identical to the original data. In order to evaluate the capabilities of the RVM model, the model

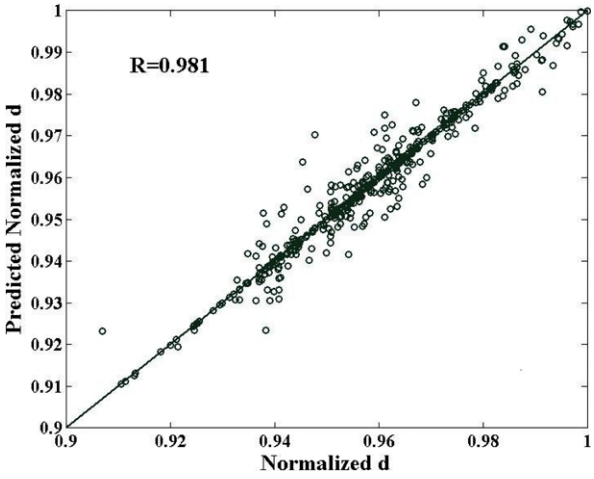


**Fig. 8** Performance of RVM model for training dataset using Gaussian kernel

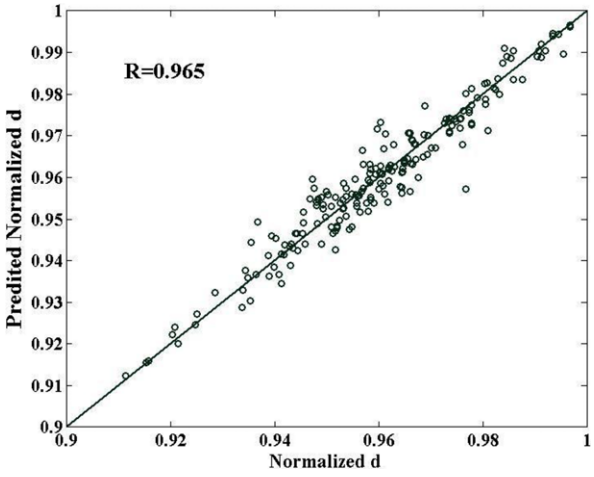


**Fig. 9** Performance of RVM model for testing dataset using Gaussian kernel

is tested with new $d$ data that are not part of the training dataset. Figure 9 shows the performance of the RVM model for testing dataset using Gaussian kernel. The number of relevance vectors is 130. The performance of RVM model is slightly better than SVM model.

There is a noticeable difference between SVM and RVM in the number of vectors employed in the machine structure. RVM uses lesser relevance vector than the support vectors in SVM. Generally, good performance in the testing phase is considered to be evidence of an algorithm's practical plausibility and provides an evaluation of the model's predictive abilities. Both SVM and RVM have better performance in the training phase than in the testing phase. The loss of performance with respect to the testing set addresses a machine's susceptibility to overtraining. There is a very marginal reduction in performance on the testing dataset (i.e., there is a difference between machine performance on training and testing) for the SVM as well as RVM model. This relatively small decline of performance of the SVM and RVM model

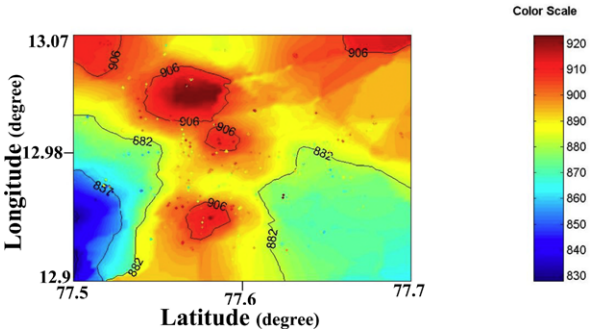


**Fig. 10** Spatial variability rock depth using SVM model

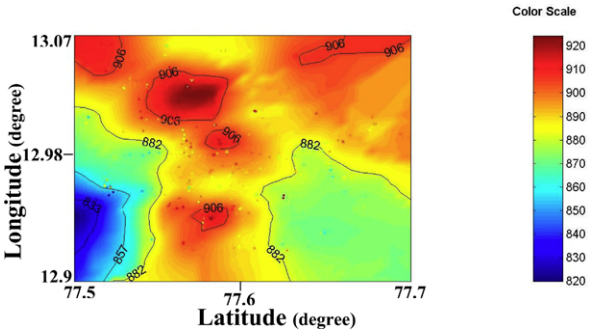


**Fig. 11** Spatial variability rock depth using RVM model

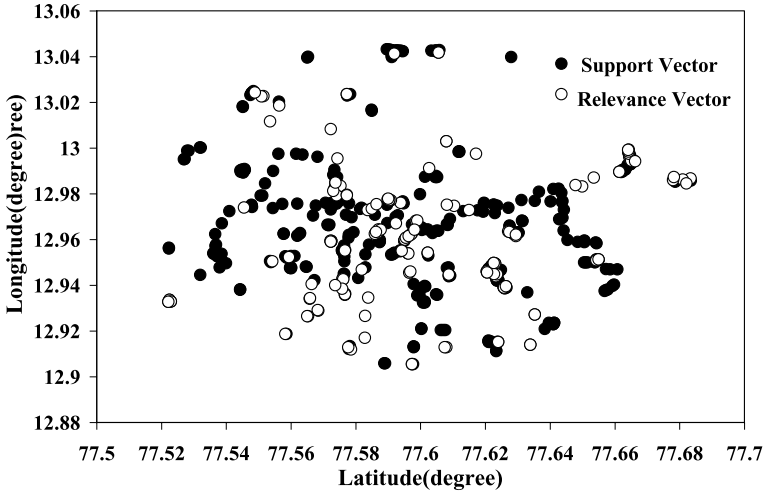


**Fig. 12** Support vector and relevance vector in Bangalore City

indicates its ability to avoid overtraining, and hence it can be expected to generalize better. SVM and RVM provide functional formulations that produce a high degree of generalization without resorting to the use of a large number of parameters (i.e., degrees of freedom). The SVM formulation leads to a sparse model dependent only

on a subset of training examples and their associated kernel functions (Vapnik 1995). SVM suffers the necessity to manually tune some parameters and from the selection of kernel function parameters which also must satisfy Mercer's condition (Vapnik 1995; Tipping 2001). The results proved that RVM is remarkable in producing an excellent generalization level while maintaining the sparsest structure. For example, whereas the RVM model employs 28.44% of the training dataset as relevance vectors; the SVM model utilized 85.33% of the training dataset as support vectors. It is worth mentioning here that the support vectors in the SVM model represent decision boundaries, while the RVM relevance vectors represent prototypical examples (Li et al. 2002). 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. Figures 10 and 11 shows the spatial variability of reduced level of rock in Bangalore by SVM and RVM model, respectively. Figure 12 shows the support vector and relevance vector in Bangalore city.

## 6 Conclusion

The machine learning induction techniques examined here have shown the ability to build accurate models with high predictive capabilities for modeling of spatial variability of reduced levels of rock in the subsurface of Bangalore. On the basis of the evidence of this experimental research, learning machines appear to be highly effective. The results of the analyses presented here show distinct performances of each machine in a supervised-learning task.

In the development of the machine models discussed here, significant effort is required to build the machine architecture. However, once developed and trained, the transpired models performed the simulations in a small fraction of the time required by the physically based model. SVM suffers from as many difficulties as RVM in finding the optimum solution when the size of the dataset and/or the dimension of the input vector is large. When SVM is applied for solving large-size problems, the computation time is prohibitively high. Both SVM and RVM exploit only the set of observations that contains all the information necessary for defining the final decision surface. In summary, this paper has surveyed two learning machines that could be viewed as powerful alternative approaches to a physically based model for the modeling of spatial variability of reduced level of rock.

## References

Berger JO (1985) Statistical decision theory and Bayesian analysis. Springer, New York

Boser BE, Guyon IM, Vapnik VN (1992) A training algorithm for optimal margin classifiers. In: Haussler D (ed) 5th annual ACM workshop on COLT. ACM Press, Pittsburgh, pp 144–152

Cristianini N, Shawe-Taylor J (2000) An introduction to support vector machine. Cambridge University Press, Cambridge. www.support-vector.net

Kanevski, Maignan (2004) Analysis and modelling of spatial environmental data. PPUR

Kanevski M (ed) (2008) Advanced mapping of environmental data. Geostatistics, machine learning and Bayesian maximum entropy. iSTE and Wiley, New York

Li Y, Campbell C, Tipping M (2002) Bayesian automatic relevance determination algorithms for classifying gene expression data. Bioinformatics 18(10):1332–1339

MacKay DJ (1992) Bayesian methods for adaptive models. PhD thesis, Dep of Comput and Neural Syst. Calif Inst of Technol, Pasadena, CA

MathWork, Inc. (1999) Matlab user's manual, Version 5.3. Natick, The MathWorks, Inc

Mukherjee S, Osuna E, Girosi F (1997) Nonlinear prediction of chaotic time series using support vector machine. In: Proc IEEE workshop on neural networks for signal processing, vol 7. Institute of Electrical and Electronics Engineers, New York, pp 511–519

Muller KR, Smola A, Ratsch G, Scholkopf B, Kohlmorgen J, Vapnik V (1997) Predicting time series with support vector machines. In: Proc int conf on artificial neural networks. Springer, Berlin, p 999

Radhakrishna BP, Vaidyanadhan R (1997) Geology of Karnataka. Geological Society of India, Bangalore

Scholkopf B (1997) Support vector learning. R. Oldenbourg, Munich

Sincero AP (2003) Predicting mixing power using artificial neural network. EWRI world water and environmental

Tipping M (2000) The relevance vector machine. Adv Neural Inf Process Syst 12:652–658

Tipping ME (2001) Sparse Bayesian learning and the relevance vector machine. J Mach Learn 1:211–244

Vapnik V (1995) The nature of statistical learning theory. Springer, New York

Vapnik V (1998) The nature of statistical learning theory. Wiley, New York

Vapnik V, Golowich S, Smola A (1997) Support method for function approximation regression estimation and signal processing. In: Mozer M, Petsch T (eds) Advance in neural information processing system, vol 9. MIT Press, Cambridge

Wahba G (1985) A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline-smoothing problem. Ann Stat 4:1378–1402