Evaluation of Spatial Interpolation for RQD at Different Depths

Ambavarapu Rafi[1] and Saha Dauji[2][0000-0001-5766-3567] and Kapilesh Bhargava[3]

[1] SMF, BARC, Mysore
[2, 3] NRB, BARC, Mumbai
arafi@barc.gov.in

Abstract. Rock Quality Designation (RQD) of a site is a very important engineering property from geotechnical considerations. RQD is very useful in identifying potential problems related to bearing capacity, settlement, slope stability problems and also serves as warning indicator of low quality rock zones that requires greater scrutiny. However, due to time and cost considerations, geotechnical investigations are carried out at larger spacing which calls for appropriate techniques for estimation of data at intermediate locations for appropriate judgements. An earlier study reported IDW (Inverse distance weighing method) to be appropriate methodology for 20 m level below ground level at the site. It was deemed necessary to examine whether similar technique would be applicable for other layers also. This paper deals with estimation of RQD values at intermediate locations from an available coarse grid data using various spatial interpolation techniques for different depths. Geo-statistical methods selected for the present study are K nearest neighbour (KNN) mean method, Inverse Distance Weighing (IDW) method and Trend Surface Analysis (TSA). Spatial interpolation has performed at depths 20 m, 25 m and 29 m below ground level. IDW and KNN methods estimated the RQD values at all depths with good accuracy followed by TSA method. It is observed that site-specific parameters obtained for best performance are different for different depths. It is necessary to analyse the data at each level and perform spatial interpolation to obtain the site-specific parameters to obtain best results.

Keywords: RQD, geotechnical investigation, spatial interpolation.

1 Introduction

For analysis and design of important infrastructure projects, geotechnical parameters form an important input. Site investigations including bore-logs as well as laboratory tests on soil samples are carried out to arrive at the requisite geotechnical parameters. Locations of the bore-holes or sample collection for laboratory tests are fixed at the conceptual stage of the project based on the layout of the buildings at that stage. In the preliminary geotechnical investigations, which are targeted towards estimation of foundation design parameters, are sometimes carried out at grid pattern of 100 to 200m, because of time and cost considerations. Therefore, it becomes necessary to employ suitable techniques for preliminary estimation of properties of soils beneath and adjacent to the structures at a specific location, which would not, in most cases, be on the test grid locations. This preliminary estimation is of prime importance in terms of geotechnical considerations since behaviour of structures is strongly influenced by the response of soils, particularly in seismic conditions. Obviously, properties of the soils surrounding the structure also affect the bearing capacity. The preliminary estimates help for the analysis and design of the structures and subject to confirmatory soil investigations at the execution stage and thus would help shorten the total project time and cost.

Rock Quality Designation (RQD) of a site is a very important engineering property in from geotechnical considerations. RQD is very useful in identifying potential problems related to bearing capacity, settlement, slope stability problems and also serves as warning indicator of low quality rock zones that requires greater scrutiny.

The superiority of geostatistics in comparison with ordinary statistics is the inverse proportion between the strength of correlation among pre-defined parameters at two points under consideration and the distance between these points. Increasing the density of sampling points both in plan and depth is not cost effective. For this reason, usage of geostatistical methods to estimate the geotechnical parameters where test data is not available from the available data from the surroundings would be extremely useful for executing the activities of design and additional geotechnical investigations on parallel mode and thereby saving on the overall project time.

The geostatistics have found numerous applications in the domain of soil science and geotechnical engineering in the recent times. A number of studies have been carried out in the field of agriculture to model the soil properties like chemical content, pH, salinity etc., (Burgess & Webster 1980; Alexandra and Donald 1999; Robinson and Metternicht 2006; Binny et al. 2015). Studies have also been carried out to evaluate SPT variability to some extent (Sitharam and Samui 2007; Samui and Sitharam 2011; Dasaka and Zhang 2012; Selim et al., 2013; Masoud and Ahmed 2017). Annelies and Andre (2010) examined the applicability of ordinary kriging to interpolate the results of cone penetration tests. Prediction of Rock Mass rating (RMR) has been carried out by Chen et al. (2017). Soulie et al. (1990) demonstrated the applicability of geo-statistics in finding the structure of the spatial variability of the undrained shear strength. Evaluation of liquefaction potential using geo-statistics has been carried by Kevin and Laurie (2005).

Kriging is an advanced geospatial algorithm which requires computation intensive applications. Comparatively simpler geospatial methods such as the k-nearest neighbour means, inverse distance weighted schemes or trend surface analysis may serve the purpose equally well. Earlier studies shown that IDW method is suitable for estimation of standard penetration test value at 1.5 m and 3 m depths (Rafi, Saha and Kapilesh 2018) and RQD at 20 m (Rafi, Saha and Kapilesh 2019) below ground level at this site. Hence, in the present study, mapping of the RQD results at 20 m, 25 m and 29 m depth below ground level was attempted by use of the simpler spatial interpolation techniques mentioned earlier. RQD data from a large site was taken for the case study. The results are given in tabular form.

2 Data & Methodology

Data is obtained from field geotechnical investigations carried out at site located in northern Karnataka, approximately 200 km from Bangalore. The total area of the site is approximately 5.5 km^2. The layout of the site along with the borehole locations is shown in Fig. 1. The soil profile at site can be generalized as sandy gravel with thickness varying from 500 mm to 3 m on top followed by completely weathered rock of thickness 3 m to 5 m, highly weathered rock of thickness 5 m to 10 m, moderately weathered rock of thickness 3 m to 7 m, slightly weathered rock with thickness of 2 m to 5 m and fresh rock of thickness 1 m to 4 m along the depth of soil. Soil investigations have carried at an average spacing of 200 m in staggered grid pattern. Total of 126 boreholes were drilled for data collection.

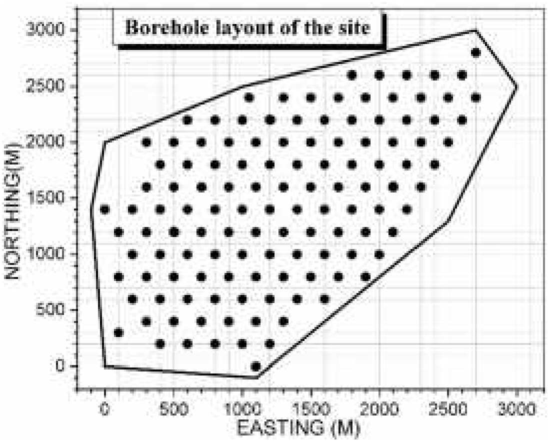


Fig. 1. Borehole layout of the investigation site

To arrive at suitable parameters of the methods for best performance, entire data available is divided into modelling and testing data sets in 3:1 proportion for individual depths. The modelling data set is used to formulate the model and estimate the model parameters. Using thus developed model, the RQD values were estimated at testing data locations. Subsequently, these estimated results were compared with testing data set to evaluate the suitability of the model. The descriptive statistics and of entire data available for 20 m, 25 m and 29 m is given in Table 1. The descriptive statistics and of modelling and testing data sets for 20 m, 25 m and 29 m is given in Table 2.

Table 1. Data statistics of RQD for different levels

Statistic	RQD 20 m	RQD 25 m	RQD 29 m
Number	70	87	109
Mean	45	44	47

Median	39	37	42
Mode	7	15	7
Std. dev	29	28	27
COV (%)	64.98	62.56	56.47
skewness	0.19	0.39	0.24
kurtosis	-1.38	-1.21	-1.01

Table 2. Data statistics of modelling and testing sets of RQD for different levels

Statistic	RQD					
	20 m		25 m		29 m	
	Modelling	Testing	Modelling	Testing	Modelling	Testing
Number	53	17	65	22	82	27
Mean	45	44	41	55	46	50
Median	44	34	31	60	41	50
Mode	7	8	15	68	7	54
Std. dev	29	31	27	28	26	27
COV (%)	63.89	70.45	66.47	50.02	57.74	53.35
skewness	0.20	0.19	0.61	-0.16	0.26	0.17
kurtosis	-1.30	-1.69	-0.95	-1.27	-1.02	-0.93

3 Geo-statistical techniques – An overview

An overview of the methods used for the present study is presented in this section.

3.1 'K' Nearest Neighbour (KNN) Mean Technique

It is one of the easiest method adopted for determining the properties at unsampled location based on the properties at the sampled location (1). In Fig. 2, Z_1, $Z_2 \ldots Z_k$ are the values at the sampled locations 1, 2…k respectively.

Let 'Z' be the value at unsampled location.

$$Z = \frac{\sum_{i=1}^{k} z_i}{k} \tag{1}$$

Where,

Z_i= value at sampled location

Z = value at unsampled location

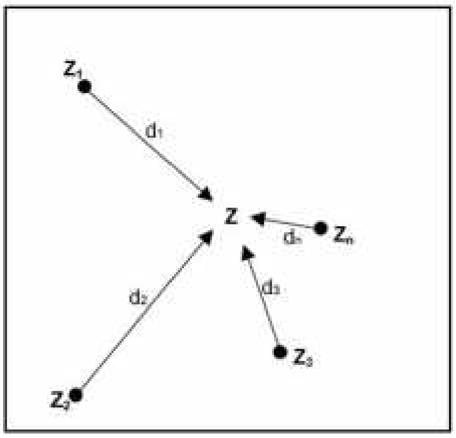


Fig. 2. Spatial arrangement of data points for interpolation

3.2 Inverse Distance Weighing (IDW) Technique

IDW method is another simple and readily available methods. It is local exact interpolation method and is based on an assumption that the value at an unsampled point can be approximated as a weighted average of values at points within a certain cut-off distance, or from a given number m of the closest points.

Weights are usually inversely proportional to a power of distance, which, at an unsampled location r, leads to an estimator (Fig. 2)

$$z(r) = \sum_{i=1}^{m} w_i z(r_i) = \frac{\sum_{i=1}^{m} \frac{z(r_i)}{d_i^e}}{\sum_{i=1}^{m} \frac{1}{d_i^e}} \tag{2}$$

Where, m (number of nearest neighbours) and e (exponent) are parameters of interpolation, which needs to be optimized for particular data set and site.

$d_i = |r-r_i|$ = distance from unsampled location to i^{th} location

r_i = position vector of i^{th} location

$z(r_i)$ = value at sampled location

$z(r)$ = value at unsampled location

For further details, texts of Marvasti (2001) and Baxter (2001) may be referred.

3.3 Trend Surface Analysis (TSA)

Trend surface analysis is a surface interpolation method that fits a polynomial surface by least-squares regression through the sample data points. This global inexact method results in a surface that minimizes the variance of the surface in relation to the input values. The resulting surface rarely goes through the sample data points. This is the simplest method for describing large variations. Trend surface analysis is used to find general tendencies of the sample data, rather than to model a surface precisely.

Let $\{x_i, y_i\}_{i=1}^{n}$ be a set of known data points with function value y_i at location x_i. The approximation function of degree 'm-1' can be write as

$$f(x) = \sum_{k=1}^{m} a_k x^{k-1} \quad (3)$$

Exact function can be write as

$$T(x) = f(x) + E(x) \quad (4)$$

Where E(x) is the residual.
Minimization of least squares error w.r.t. coefficients yield

$$\frac{\partial E}{\partial a_j} = 2\sum_{k=1}^{n}(f(x_k) - y_k)x_k^{j-1} = 0, 1 \leq j \leq m \quad (5)$$

The above equation implies a system of 'm' equations, and in the matrix form it can be written as

$$C_{kj}a_j = b_k \quad (6)$$

Where, Ckj and bk is written as

$$C_{kj} = \sum_{i=1}^{n} x_i^{k+j-2}, \quad 1 \leq k, \; j \leq m \quad (7)$$

$$b_k = \sum_{i=1}^{n} x_i^{k-1} y_i, \quad 1 \leq k \leq m \quad (8)$$

For further details, texts of Wren (1973) and Unwin (1975) may be referred.

3.4 Criteria For Comparison

To compare different interpolation techniques, we examined the difference between the observed data and the estimated data using the coefficient of correlation (R) (9), mean absolute error (MAE) (10) and root mean squared error (RMSE) (11).

$$R = \frac{N\sum(y_i\hat{y}_i) - (\sum y_i)(\sum \hat{y}_i)}{\sqrt{[N\sum y_i^2 - (\sum y_i)^2}\sqrt{[N\sum \hat{y}_i^2 - (\sum \hat{y}_i)^2}} \quad (9)$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i) \quad (10)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \quad (11)$$

Where,
y_i = actual value
$\hat{y}_i$ = estimated value

N = number of data pairs

Whilst such validation techniques are not confirmatory tools, as exploratory tools they greatly assist in choosing appropriate interpolation procedures and their associated parameters.

4 Results and Discussion

Modelling data set is used to find the parameters of the methods KNN, IDW and TSA for best performance at 20 m, 25 m and 29 m levels. To decide the best parameters statistical indicators such as coefficient of correlation (R) between actual RQD value and estimated RQD value, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are computed. Table 3 shows the performance metrics obtained with best parameters.

Table 3. Summary of RQD interpolation using KNN, IDW and TSA methods

Perfo-mance metric	RQD at different depths below ground level								
	20 m			25 m			29 m		
	KNN (K=8)	IDW (e=0.2, K=8)	TSA (degree =4)	KNN (K=5)	IDW (e=0.2, K=5)	TSA (degree =5)	KNN (K=4)	IDW (e=0.2, K=4)	TSA (degree =5)
R	0.84	0.82	0.71	0.8	0.82	0.65	0.67	0.67	0.51
MAE	20	20	18	20	20	20	18	18	20
RMSE	23	23	23	24	24	23	22	22	23

Amongst the three methods studied, from comparison of the best performance from each interpolation technique, KNN and IDW methods performance is similar and better than TSA method for all three levels considered.

The low value of the exponent (e=0.2) has been found suitable for interpolation using IDW method is possibly due to high coefficient of variation of data. (56 % to 65 % for 20, 25 and 29 m levels)

For deeper layers, the number of points for interpolation reduces. One possible reason could be that due to more number of points for which data is available for lower layers (70 for 20 m; 87 for 25 m; 109 for 29 m), less number of points are yielding better results. Another reason can be that due to more heterogeneity of strata at higher depths compared to that of strata at shallow depths gives better results with less number of neighbours at higher depths.

RQD estimations are carried out at locations of testing data set locations using the parameters obtained above to assess the performance of the model. Performance of the models (KNN and IDW) is given below in the form of scatter plots with one standard deviation bound on either side are shown in Fig. 3 through Fig. 5.

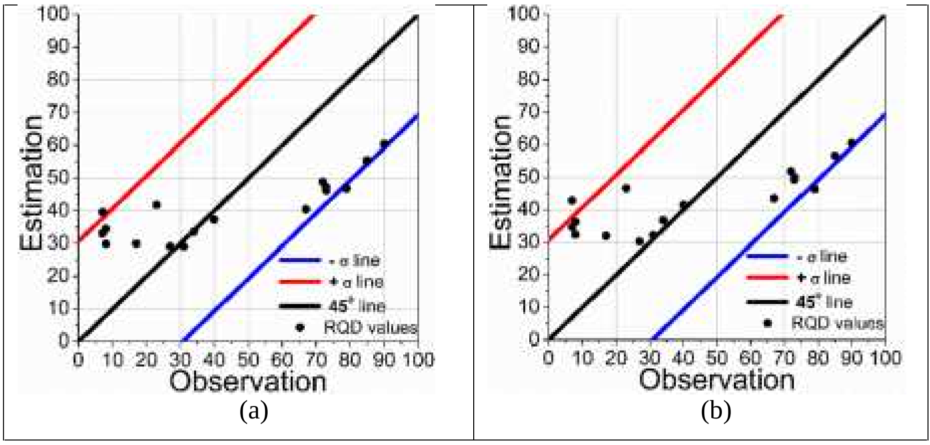


Fig. 3. Scatter plot of estimated and observed RQD values at 20 m depth using (a) KNN method (K=8) (b) IDW method with parameters (K=8, e=0.2)

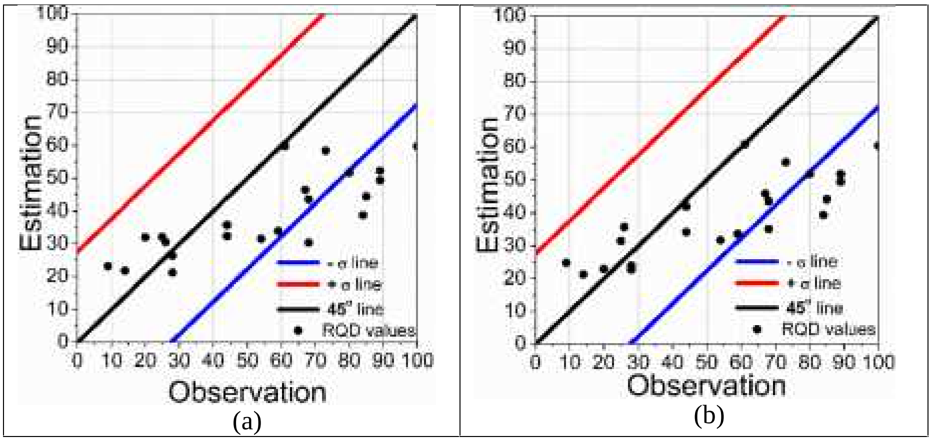
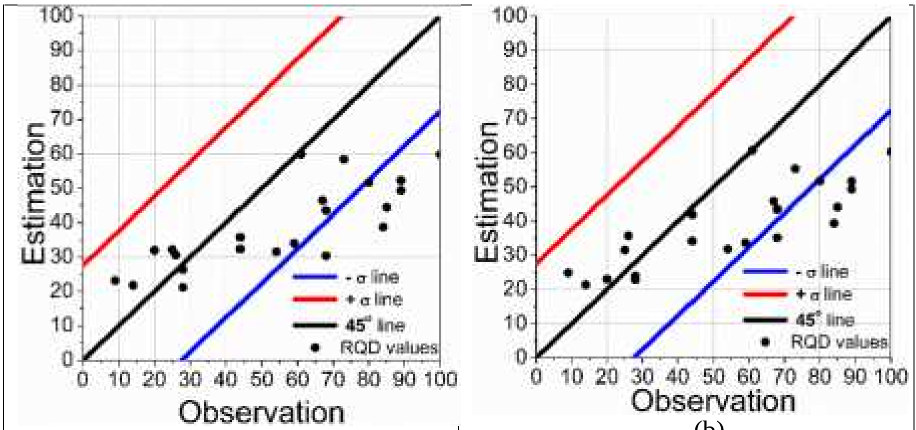


Fig. 4. Scatter plot of estimated and observed RQD values at 25 m depth using (a) KNN method (K=5) (b) IDW method with parameters (K=5, e=0.2)



(a) (b)

Fig. 5. Scatter plot of estimated and observed RQD values at 29 m depth using (a) KNN method (K=4) (b) IDW method with parameters (K=4, e=0.2)

5 Conclusions

This study has shown that, out of the three spatial interpolation methods used, KNN and IDW methods perform equally well (correlation coefficient of around 0.8 for 20m, 25m and 0.67 for 29m depth; RMSE and MAE between 18 and 24) and better than TSA (maximum of correlation coefficient of 0.7) for all the levels considered. The exponent remains same for all depths equal to 0.2 for IDW method to obtain best performance. Number of points of interpolation varies from one level to another level (number of neighbours of 8, 5 and 4 for 20m, 25m and 29m depths respectively) due to different spatial dependencies at different depths. Therefore, it is necessary to evaluate the specific parameters to estimate RQD for each level. These parameters may be different for another site. It is useful to analyse the available data to estimate the site-specific parameters of the model to estimate RQD values to use in planning, design and execution.

It is important to note that RQD values estimated using the geo-spatial interpolation techniques are useful as design input in preliminary design stages where data is not available at desired locations. However, one cannot completely rely on these estimates, and it is always necessary to perform the confirmatory geotechnical investigations to validate the design before final execution. Further studies may be directed towards development of suitable interpolation strategies for other variables, such as hydraulic conductivity, rock mass rating, permeability, etc. would be useful for estimation (interpolation) of these variables in fine grid resolution from coarse grid data.

References

1. Alexandra, K., Donald, B. G.: A Comparative Study of Interpolation Methods for Mapping Soil Properties. Agronomy journal, 393-400 (1999)
2. Annelies, G., Andre, V.: Geostatistical Interpolation of soil properties in boom clay in Flanders. geoENV VII — Geostatistics for Environmental Applications, 219-230 (2010)
3. Baxter, V. E.: Distributed Hydrologic Modeling Using GIS. Springer Netherlands (2001)
4. Binny, G., Shetty, A., Jayaprakash, and Chaya, D. Y.: Spatial variability of topsoil chemical properties. Indian Journal of Agricultural Research, 49(2), 134-141 (2015)
5. Burgess, T. M., Webster, R.: Optimal interpolation and isarithmic mapping of soil properties ii, block kriging. Europian Journal of Soil Science, 333-341 (1980)
6. Burgess, T. M., Webster, R.: Optimal interpolation and isarithmic mapping of soil properties. I. The semivariogram and punctual kriging . Europian Journal of Soil Science, 315--331 (1980)
7. Chen, J., Li, X., Zhu, H.: Geostatistical method for predicting RMR ahead of tunnel face excavation using dynamically exposed geological information. Engineering Geology, 214-223 (2017)
8. Dasaka, S. M., Zhang, L. M.: Spatial variability of in situ weathered soil. Geotechnique, 375-384 (2012)
9. Kevin, D. M., Laurie, B. G.: Three-dimensional liquefaction potential analysis using geostatistical interpolation. Soil Dynamics and Earthquake Engineering, 369–381 (2005)
10. Marvasti, F.: Nonuniform Sampling: Theory and Practice, Volume 1. Springer US (2001)
11. Masoud, A. A., Ahmed, K. A.: Three-dimensional geotechnical modeling of the soils in Riyadh city, KSA. Bulletin of Engineering Geology and the Environment, 1-17 (2017)
12. Rafi, A., Saha Dauji & Kapilesh Bhargava.: Estimation of SPT from coarse grid data by spatial interpolation technique. In: Proceedings of Indian Geotechnical Conference, TH-10-66. Indian Institute of Science, Bengaluru, India (2018)
13. Rafi, A., Saha Dauji & Kapilesh Bhargava.: Spatial interpolation methods for development of fine resolution RQD map from coarse grid data. In: Proceedings of Indian Conference on Geotechnical and Geo-Environmental Engineering, Paper No. 29. Motilal Nehru National Institute of Technology, Allahabad, India (2019)
14. Robinson, T. P., Metternicht, G.: Testing the performance of spatial interpolation techniques for mapping soil properties. Computers and Electronics in Agriculture, 97-108 (2006)
15. Samui, P., Sitharam, T. G.: Application of Geostatistical Models for Estimating Spatial Variability of Rock Depth. Scientific Research, Engineering, 886-894 (2011)
16. Selim, A., Alper, S., Gokpete, B. A.: Geostatistical interpolation for modelling SPT data in northern Izmir. Sadhana, 38(6), 1451-1468 (2013)
17. Sitharam, T. G., Samui, P.: Geostatistical modelling of spatial and depth variability of SPT data for Bangalore. Geomechanics and Geoengineering: An International, 307-316 (2007)
18. Soulie, M., Montes, P., Silvestri, V.: Modelling spatial variability of soil parameters. Canadian Geotechnical Journal, 27, 617-630 (1990)
19. Unwin, D. J.: AN INTRODUCTION TO TREND SURFACE ANALYSIS. Norwich: Geo Abstracts, University of East Anglia (1975)
20. Wren, E. A.: TREND SURFACE ANALYSIS - A REVIEW. Canadian Society of Exploration Geophysicists, vol 9, 39-44 (1973)