Development of Optimal Geostatistical Model for Geotechnical Applications

Rojimol. J

A Dissertation Submitted to

Indian Institute of Technology Hyderabad

In Partial Fulfillment of the Requirements for

The Degree of Master of Technology

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Department of Civil Engineering

June, 2013

Declaration

I declare that this written submission represents my ideas in my own words, and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the Institute and can also evoke penal action from the sources that have thus not been properly cited, or from whom proper permission has not been taken when needed.

(Signature)

Rojimol. J
(Student name)

CE11M08
(Roll No)

Approval Sheet

This thesis entitled **'Development of Optimal Geostatistical Model for Geotechnical Applications'** by Rojimol J is approved for the degree of Master of Technology from IIT Hyderabad.

Dr. B. Umashankar
(Adviser)
Assistant Professor
Department of Civil Engineering
IIT Hyderabad

Dr. K. B. V. N. Phanindra
(Co-Adviser)
Assistant Professor
Department of Civil Engineering
IIT Hyderabad

Dr. K. Sri Rama Murty
(Examiner)
Assistant Professor
Department of Electrical Engineering
IIT Hyderabad

iii

Acknowledgements

First of all I would like to thank the Almighty God for giving me the knowledge and enthusiasm to carry out this research successfully. I wish to express my sincere thanks to my supervisor B. Umashankar for his constant encouragement and valuable support during my research time. Also, I would like to acknowledge my Co-supervisor K. B. V. N. Phanindra for his excellent guidance and support to complete all goals of my research.

I take this opportunity to thank my parents for their constant support. I am extremely grateful to all my friends who supported and encouraged me during my research period.

Dedicated to

My parents

Abstract

Evaluation and application of various geo-statistical interpolation techniques (including deterministic and probabilistic methods) to site characterization has received much attention in the recent years (Rouhani 1996, Fenton 1997, Asa *et al* 2012). However, the existing geo-statistical tools in their original form lack several inbuilt functionalities including hypothesis based normality check for the data; positional outlier separation; automated selection of base variogram and optimal kriging model; and elimination of negative kriging weights. This research addresses these issues, and aims at developing a generalized, public domain, open source and optimal linear geo-statistical model using MATLAB environment that best fits a given set of site specific parameters. The measured data at the random borehole locations were analyzed, and used to generate the prediction and error surfaces of the site parameters at user specified intervals. Normality of the data was statistically tested using Kolmogorov-Smirnov test at 5 and 10% significance levels. Positional outliers that may adversely affect the simulation were discarded from the analysis using the concept of point density. The best semi-variogram with optimum searching neighbourhood was automated using residual statistics. Negative kriging weights given at the known data locations were successively eliminated in the algorithm. A graphical user interface (GUI) in MATLAB for use with site managers / construction engineers of a region was developed in this work

Applicability of the developed code was tested for three cases. Case 1 considers the clay content values at 3 m depth for the Proposed refinery project region at Paradip, India. Case 2 considers the moisture content values at 1 m depth for the proposed power plant region in Kakinada, India, and Case 3 considers the sand content values at 1 m depth for the IIT Hyderabad campus, India. Results of the analysis were evaluated with ArcGIS based geo-statistical analyst® simulations and cross validated using residual statistical parameters. It was observed that spherical variogram model and the ordinary kriging methods were best suited for all the cases. Choice of lag distance, number of lags, and grid resolution have no effect on kriging predictions. Author recommends the need for additional borehole locations in the regions where prediction uncertainty is high. The developed algorithm has significantly improved the performance of the linear geo-statistical models over the conventional tools. A considerable decrease in RMSE from 40 to 76 % compared to ArcGIS was observed considering all the linear geostatistical models by using the modified algorithm.

Contents

Declaration .... **Error! Bookmark not defined.**
Approval Sheet.. **Error! Bookmark not defined.**
Acknowledgements iv
Abstract vi
1 Introduction……………………………………………………………………...1
1.1......... Reseach overview……………………………………………………………..1
1.2 Necessity of a geostatistical tool to site characterisation…………………………...1
1.3 Objectives of the study……………………………………………………………..2
1.4 Organization of the study……………………………………………………………2
2 Literature review………………………………………………………………………...4
2.1 Introduction……………………………………………………………………….4
2.2 Origin of Geostatistics……………………………………………………………….5
2.3 Geostatistics in geotechnical engineering…………………………………………...…6
2.3.1 Channel tunnel project…………………………………………………………………6
2.3.2 Spatial analysis of ground motion………………………………………………….7
2.3.3 Spatial analysis of ground water table elevation…………………………………….8
2.3.4 Spatial analysis of compacted earth materials……………………………………8
2.3.5 Spatial analysis of clay content……………………………………………......9
2.3.6 Spatial analysis of undrained shear strength……………………………………...9
2.3.7 Spatial analysis of SPT………………………………………………..................10
2.3.8 Spatial analysis of consolidation settlement……………………………………….11
2.3.9 Spatial analysis of Rock strength…………………………………………………..11
2.4 Tools available to apply geostatistics…………………………………………………11
2.5 Challenges in predicting soil parameters……………………………………......12
3 Geostatistical Modeling……………………………………………………………..13
3.1 Introduction…………………………………………………………………...13
3.2 Principles of Geostatistics………………………………………………………….13
3.3 Experimental semivariogram ……………………………………………………..…...14
3.4 Theoretical variogram models………………………………………………………..15
3.4.1 Gaussian model……………………………………………………………….16
3.4.2 Spherical model……………………………………………………………….16

3.4.3 Exponential model……………………………………………………………..…17
3.5 Model fitting……………………………………………………………...…17
3.6 Geostatistical interpolation……………………………………………….…18
3.7 Kriging techniques…………………………………………………………….…..19
3.7.1 Simple kriging (SK) ……………………………………………………....19
3.7.2 Ordinary kriging (OK) …………………………………………………..…20
3.7.3 Universal kriging (UK) ………………………………………......................21
3.8 Cross-validation……………………………………………………………………22
3.9 Prediction surfaces…………………………………………………………………23
4. Geostatistical algorithm development…………………………………………………25
4.1 Overview……………………………………………………………………...25
4.2 Geostatistical analyst in GIS………………………………………………………25
4.2.1 Data representation………………………………………………………….26
4.2.2 Exploring spatial data……………………………………………….……26
4.2.3 Evaluation of geostatistical techniques……………………………………...…27
4.3 Development of MATLAB based Geostatistical algorithm……………………………29
4.3.1 Conversion of co-ordinate system to projected system…………………………….29
4.3.1.1 Datum selection………………………………………………………………..29
4.3.1.2 Selection of projection system…………………………………………………30
4.3.2 Removal of positional outliers……………………………………….………30
4.3.3 Normality test of the data……………………………………………………….31.
4.3.4 Development of Experimental Semivariogram……………………………………31
4.3.5 Theoretical model fitting……………………………………………………...32
4.3.6 Simple kriging algorithm development…………………………………….…33
4.3.7 Ordinary kriging algorithm development……………………………………..33
4.3.8 Universal kriging algorithm development…………………………………………33
4.3.9 Cross validation……………………………………………………………………33
4.3.10 Prediction surfaces and Error surfaces………………………………………..34
5 Application of Geostatistical Algorithm……………………………………………..34
5.1 Overview……………………………………………………………………..35
5.2 Paradip Refinery project, Orissa (Case Study – 1)…………………………………….35
5.2.1 Site description……………………………………………………...……35
5.2.2 Spatial outlier removal……………………………………………………….36
5.2.3 Normality check…………………………………………………………....37

5.2.4 Semivariogram and model fitting……………………………………………………37
5.2.5 Evaluation of Kriging techniques……………………………………………………38
5.2.6 Comparison between conventional statistical tools and the developed tool………41
5.3 Kakinada Power Plant (Case study - 2)……………………………………………42
5.3.1 Site description……………………………………………………………………..43
5.3.2 Spatial outlier removal……………………………………………………………..43
5.3.3 Normality check…………………………………………………...……….43
5.3.4 Semivariogram and model fitting…………………………………………………….43
5.3.5 Evaluation of Kriging techniques……………………………………………………44
5.3.6 Comparison between conventional statistical tools and the developed tool………46
5.4 IIT Site, Hyderabad, Andhra Pradesh (Case study - 3)…………………………….47
5.4.1 Site description……………………………………………………………………..47
5.4.2 Spatial outlier removal………………………………………………………......48
5.4.3 Normality check……………………………………………………………………48
5.4.4 Semivariogram and model fitting…………………………………………………….48
5.4.5 Evaluation of Kriging techniques……………………………………………………49
5.4.6 Comparison between conventional statistical tools and the developed tool……...50
5.5 Summary……………………………………………………………………………….51
6. Summary and Conclusion………………………………………………………….52
6.1 Overview…………………………………………………………………………...52
6.2 Factors considered to improve linear kriging estimates ………………………...52
6.3 Limitations of the study……………………………………………………………54
6.4 Recommendations ……………………………………………………………………55
References…………………………………………………………………………….56

Chapter 1

Introduction

1.1 Research Overview

Site investigation studies are conducted to assess the unforeseen geological risks before executing any construction project. Even if proper site investigation have been done, it is common that most of construction projects experience delays due to unexpected failures and cost overruns due to overdesigning lesser risk areas. Generally, soil characterization is done by interpreting data from laboratory test and in-situ test using deterministic analysis methods. But these methods will not handle the uncertainties associated with the soil data. Geostatistical techniques have been used during the past few decades to help in improving the site characterization by handling the uncertainties methodically, thereby minimizing the unexpected failures after construction. These techniques have been applied to various fields such as mining, hydro-geology, water resource engineering, geotechnical engineering, etc. But most of the geotechnical engineers are still not much aware about the powerfulness of these techniques due to a lack of availability of handy geostatistical tools to meet their needs. Geostatistical techniques can be used to model the spatial variability of various soil parameters and generate the surface and error variance profiles for spatially varying soil parameters at a required depth with estimates of uncertainties. This research deals with the development of an automated, user-friendly and cost-effective tool to conduct a probabilistic linear geostatistical approach for evaluating the site characterization.

1.2 Necessity of a geostatistical tool to site characterisation

Detailed geotechnical site investigation is very expensive for any project. Hence, most of the engineers will assume that soil properties are similar throughout the region. In reality, in-situ soil parameters (Atterberg limits, soil type, Standard Penetration Resistance (SPT), Cone Penetration Resistance (CPT), undrained shear strength, bed rock depth, etc.) show high spatial variability. Geostatistical techniques can consider and model the underlying spatial variability (based on distance / direction) within the dataset for prediction at any unknown locations. These techniques can help in more accurate interpretation of ground conditions by creating sub surface profiles of any parameter of the study area. The technique can also be used in designing an optimal cost sampling program which can obtain the ground information thereby aiding construction managers for effective site characterization. Due to lack of a geostatistical tools specific to geotechnical applications, these principles are still unaware to many geotechnical engineers [Hammah and Curran, 2004]. The main objective of this research work is to develop a geostatistical tool specifically aimed for a geotechnical engineer to easily understand the principles of geostatistics and explore its best potential in geotechnics.

1.3 Objectives of the study

1. To develop an automated, public domain, open source, linear geostatistical model to apply for various geotechnical applications
2. To develop a cost-efficient geostatistical tool by reducing the computational time compared to conventional tool in choosing various factors affecting the best kriging method.
3. To apply the best semi-variogram and kriging algorithm for the parameter under consideration based on residual statistics.
4. To generate the prediction surface and error variance profiles for the spatially varying soil parameters at a required depth.
5. To achieve an improved site characterization using the best capability of tool developed.

1.4 Organization of the study

This thesis is organized into six main chapters.

Chapter1. INTRODUCTION Gives a brief overview of the thesis and Research objectives.

Chapter2. LITERATURE REVIEW

Reviews the origin of geostatistics and application of geostatistics in Civil engineering.

Chapter 3. GEOSTATISTICAL MODELLING

Focuses on the principles of geostatistics and formulation of linear geostatistical interpolation techniques.

Chapter 4. DEVELOPMENT OF GEOSTATISTICAL ALGORITHM

Explains the methodology adopted in developing the geostatistical algorithm. In addition, discusses the methodology adopted by conventional tools in evaluating geostatistical techniques.

Chapter 5. APPLICATION OF GEOSTATISTICAL ALGORITHM

Deals with the application of geostatistics to various case studies considered in the study. This chapter also provides a comparison of the effectiveness of the developed model to that of the conventional geostatistical tool results.

Chapter 6. CONCLUSIONS

Summarises the research methodology and discusses the advantages of the algorithm compared to conventional tools. It also provides findings from the study along with the recommendations for future research work.

Chapter 2

Literature review

1.1 Introduction

For the last few decades, geostatistical methods have been successfully applied to various fields of civil engineering viz. environmental science [Webster and Oliver, 2007], water resources engineering [Kumar, 2007; Gundogdu and Guney, 2007; Kambhamettu et al., 2011], hydro-geology [Kitanidis, 1997], mining [Matheron, 1976; Journel and Huijbregts, 1993], and geotechnical engineering [Rouhani, 1996; Exadaktylos, 2008; Samui and Sitharam, 2010; Asa *et al*. 2012]. Geostatistical techniques deal with the analysis of problems involving spatial variation of the data [Journel and Huijbregts, 1978]. Engineering properties of soil and rock are highly heterogeneous within a location. This inherent uncertainty of soil properties and assumption of averaging the properties in field can lead to uncertainty in engineering design. Application of principles of geostatistics to soil data can aid in obtaining the accurate interpretation of the ground parameters. Many researchers have applied geostatistics to the field of geotechnical engineering to determine the spatial variability of properties, such as, clay content, undrained shear strength of clay, SPT values, Atterberg limits, bed rock depth, assessment of liquefaction potential, ground motion, seismic hazard analysis, etc. Presently, ArcGIS® is widely used tool to analyze and model spatially varied data. Geostatistical analyst tool in GIS helps in exploring data variability, determining spatial relationships, examining trends and generating prediction and error surfaces. Given a sparse data, geostatistical techniques will help to create a statistically valid prediction surface, along with prediction error estimate. ArcGIS Geostatistical analyst tools have the capabilities of interactive graphical user interface (GUI) and web services. It has been found from the literature that lack of geostatistical software tools specific to geotechnical engineer has made it difficult for many to understand the principles of geostatistics and to apply the potential benefits of geostatistics in site exploration program. This review deals with the origin, principles and application of geostatistics to geotechnical engineering.

2.2 Origin of Geostatistics

Geostatistics is the collection of techniques to solve problems involving spatial variables [Matheron, 1971; Journel and Huijbregts, 1978]. These can be used for interpolation, integration, and differentiation on the data. It is assumed in the analysis that there is a strong relationship within the measured data that can be modeled using spatial variograms. Such an analysis can predict the spatial distributions of properties from known values at sampled points to extensive areas or volumes. Theory underlying the principles of geostatistics is well documented in literature [Isaaks and Srivastava, 1989; ASCE, 1990; Kitanidis, 1997; Deutsch and Journel, 1998; Chiles and Delfiner, 1999]. Geostatistical techniques (such as kriging, co-kriging and universal kriging) can generate a prediction surface, and also provide the accuracy of these predictions [Johnston et al., 2001]. Geostatistics ensures the accuracy by providing statistical tools for (1) calculating the most accurate predictions, based on measurements; (2) quantifying the accuracy of the same, and (3) selecting the parameters to be measured in the case of limited data points [ASCE, 1990]. Geo-statistical techniques fall in to two categories, viz., deterministic and probabilistic, based on the underlying functions. Probabilistic geostatistical techniques (commonly known as kriging techniques) have the capability of producing a prediction surface, along with prediction confidence / error variance [Johnston et al., 2001]. Kriging is an optimal linear geostatistical interpolation method which was pioneered by Krige (1951) and formalized into mathematical model by Matheron (1963b). Kriging was first applied in the field of mining [Journel and Huijbregts, 1978; Delhomme, 1979]. In kriging, prediction of spatial variability of a random variable is achieved by semivariogram functions [Gundogdu and Guney, 2007]. Kriging is used to construct a minimum error variance linear estimate at a location where the actual value is unknown [Deustch and Journel, 1998]. Over the past half century, kriging methods have been extensively applied to several disciplines such as engineering, earth, and environmental sciences [Goldberger, 1962; Matheron, 1976; Journel, 1989; Isaaks and Srivastava, 1989; Cressie, 1991; Goovaerts, 1997; Deustch and Journel, 1998; Journel and Huijbregts, 2003]. There are two broad categories of kriging methods: linear and nonlinear kriging. Commonly used linear kriging algorithms include: simple kriging, ordinary kriging and kriging with a trend (universal kriging). If measured data follows non-gaussian function, non-linear kriging techniques give more accurate estimates. Nonlinear kriging algorithms include lognormal kriging, multi-Gaussian kriging, disjunctive kriging, indicator kriging, probability kriging, and rank kriging. Various kriging algorithms can be compared with their statistical correctness parameters in order to determine the best method suited for the characterization and interpolation of soil data [Asa et al., 2012]. The kriging generated contour map and the error variance map can be used to infer on the spatial variation of the parameter under consideration [Kambhammettu et al., 2011]. Limited studies are available in the literature specific to factors affecting the kriging estimates, such as suitable kriging techniques, number of samples, grid size for prediction, and the quality of the data [Asa et al., 2012].

2.3 Geostatistics in geotechnical engineering

To determine the geotechnical or geological condition of a site, boreholes are drilled at some specified random locations. It is reasonable to assume that observation from nearby boreholes will have similar values and far boreholes have different values. This observation satisfies the basic assumption of geostatistics. But, the volume of total samples extracted for characterizing soil masses constitutes only a minute fraction compared to total volume of material that defines engineering behaviour. This is because engineering properties of soil masses are heterogeneous in nature. But geotechnical engineers assume properties are same throughout the space which can be entirely different from actual behaviour. Thus more accurate knowledge of the spatial distribution of material properties promoting safe and economic design is required. Many researchers have applied kriging algorithms to solve for geotechnical applications (Soulie et al., 1990; Jaksa, 1993; Rouhani, 1996; Fenton, 1997; Robinson and Metternicht, 2006; Lenz and Baise, 2007; Exadaktylos, 2008; Samui et al, 2010; Asa et al., 2012). Geo-statistical techniques are widely used to generate the continuous profile of various soil parameters from the measurements made at random bore hole locations [Kulatilake, 1989, Fenton, 1997, Samui, 2010]. Geostatistcs aids a geotechnical engineer to estimate the engineering properties of soil at any location with minimum estimation error. In addition, decisions for optimizing the borehole locations can be made based on the error contours. Important applications of geostatistics to site characterization in the past few years are presented in the following sections.

2.3.1 Channel tunnel project

The channel tunnel project, between France and Britain was one among the successful application of principles of geostatistics to geotechnical engineering. The principle was applied to optimize the alignment of the tunnel by considering various factors of geological risks [Blanchin and Chilès, 1993]. Chalk Marl is a soft, impermeable and homogeneous rock, which is an ideal medium for tunneling and is overlain by grey chalk, a highly porous layer of fractured and altered rocks, and underlain by gault clay, which makes penetration difficult. Typical geological cross section is shown in Fig. 2.1. It was ensured that the tunnel was bored only within the Chalk Marl, avoiding the Gault Clay material. Kriging was used to determine the boundary between the Chalk Marl and the Gault Clay, based on data available prior to construction. Kriging provided both estimates and its confidence level with its standard deviation. Contours of the standard deviations of predicted depths of this boundary were generated. These helped in identifying tunnel sections for which improved precision was required, which in turn enabled design of successful complementary geophysical surveys of the seafloor. As further information was collected, geostatistics was applied to improve the spatial model of the Chalk Marl–Gault Clay interface. This helped the tunnel engineers to maintain the penetration of Gault clay within acceptable levels.

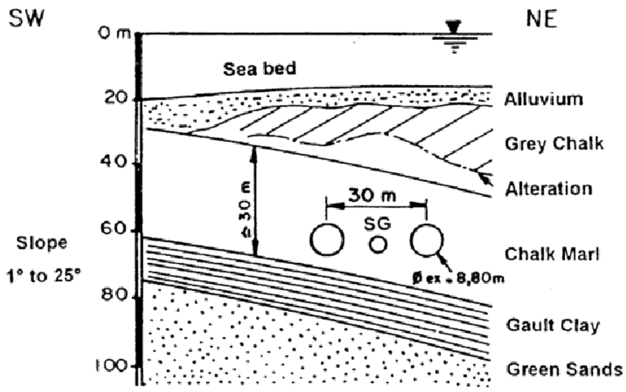


Figure 2: Geologic cross section of the tunnel [Blanchin and Chiles, 1993]

At the end of the project, actual locations of chalk Marl- Gault clay boundary were found to compare well with the predictions from the geostatistical model, thus validating the effectiveness of geostatistical modelling.

2.3.2 Spatial analysis of ground motion

Earthquake ground motion studies require accurate replication of actual ground motion. Carr and Glass (1985) applied kriging techniques to estimate earthquake ground motion. Ground motion data from four California earthquakes were analyzed. They demonstrated that disjunctive kriging was accurate for modelling ground motion. In 1989, Carr and Roberts compared universal kriging with ordinary kriging for estimation of earthquake ground motion. Spherical model was used based on the assumption of quasi stationary behavior of the ground motion data. Hypothesis testing for cross validation results of kriging techniques showed that the accuracy obtained using universal kriging is similar to accuracy obtained using ordinary kriging. In 1993, Carr and Mao compared disjunctive kriging to generalized probability kriging and found that both are giving similar results for real data possessing normal distribution, and differ significantly for non-normal distribution.

2.3.4 Spatial analysis of ground water table elevation

Kumar (2007) has applied universal kriging to generate the contour and to estimate the variance maps of groundwater level of a command area in north western India using water table elevations from the 143 observation wells during September 1990. Experimental semi-variograms was generated using a FORTRAN code. Kumar found that the predicted groundwater levels compared well with the observed levels at the monitoring well locations. Kambhamettu et al. (2010) has evaluated universal kriging for optimal contouring of groundwater levels from their measurements at random locations. Geographic system tools were used to evaluate kriging techniques and generate water table elevations in 38 monitoring wells of the Carlsbad area alluvial aquifer located south east of New Mexico, USA. A generalized MATLAB code was written to generate omni and directional semi-variograms. Different theoretical semivariogram models were tried to select the base semi-variogram to evaluate kriging. Several combinations of neighbourhood size, polynomial trend and semi-variogram models for the residuals were tried to obtain the optimal residual statistics for Universal kriging. The contour maps of ground water level obtained shows significant decrease in the water table from 1996 to 2003 and their statistical analysis revealed that the decrease in water table is between 0.6 m and 4.5 m at 90% confidence. The estimation variance contours showed that the error in estimation was more than 8 m^2 in the west and south-west portions of the aquifer due to the absence of monitoring wells.

2.3.5 Spatial analysis of compacted earth materials

Geostatistical analysis was used to quantify non-uniformity of compacted earth materials using spatially referenced roller-integrated compaction measurements [Vennapusa et al., 2009]. Roller-integrated compaction measurements obtained from two case studies viz. 1) A test area selected to include three different layered subsurface conditions: (a) compacted sandy lean clay subgrade (CL), (b) compacted gravelly sand subbase material (SW-SM) underlain by the sandy lean clay subgrade, and (c) scarified/uncompacted gravelly sand subbase material underlain by the sandy lean clay subgrade and 2) TH64 reconstruction project located to south of Akeley, Minnesto, USA were analysed using geostatistical methods. Exponential models were found to fit well with most of the experimental semivariogram, while spherical models fit less frequently. Models were checked for “goodness” using the modified Cressie goodness of fit method suggested by Clark and Harper (2002), and a cross-validation process. The nugget effect was modelled using the variance of the measured value from the nearest neighbour statistics as the upper bound of the nugget value. The geostatistical analysis as a paradigm shift helped contractor in identifying localized, poorly compacted areas or areas with highly non-uniform condition that needed additional compaction. It is suggested that if automated technique of geostatistics considering various hurdles in semivariogram modelling could also be used, the results of analysis could be effectively used to target quality assurance testing by field engineers.

2.3.6 Spatial analysis of clay content

Three linear (simple, universal, and ordinary) and three nonlinear (indicator, probability, and disjunctive) kriging methods were applied to characterize and interpolate the clay content of soil data for the Williston Department of Transportation district of North Dakota [Asa et al., 2012]. The data was analysed in both vector and raster formats using GIS. Experiments were run using the geostatistical analyst tool in ArcGIS software. Exploratory data analysis was used to determine the statistical properties of the data. The data which were log normally distributed was transformed before variogram modelling and applying kriging algorithm. Spherical variogram model was arbitrarily chosen and employed for the analysis. The characterization results were cross-validated to assess their validity and correctness. Best method for the data was chosen using a robust approach.The method which results in root mean- squared error closer to average error was chosen as the final model. It was observed that different interpolation methods chosen led to different results of the spatial interpolation. The best results were obtained using indicator and probability kriging with the vector data set and raster data set.

2.3.7 Spatial analysis of undrained shear strength

Undrained shear strength, S_u, was modelled by Soule et al. (1990). S_u was measured at regular depth intervals for a number of borings in B-6 clay in Quebec. Soulie developed horizontal and vertical variograms and had formed a vertical grid that represents a cross section of S_u values.The geostatistical technique was applied on stiff, over consolidated clay known as the Keswick Clay of the city of Adelaide [Jaksa 1993]. Jaksa has analysed 3D spatial variability of the clay deposit. He used twenty semi-variograms and found that spherical model has best fitted the experimental semi-variograms. It has been found to have a range of influence between 600 mm and 1,750 mm in the vertical direction and so samples should not be treated as uncorrelated random variables. Jaksa suggested that, to obtain the range of influence of the lateral variability of the Keswick Clay, it is necessary to sample at spacing less than one metre. Also while evaluating semivariogram, clay exhibited nugget effect, small scale random behaviour. Jaksa found that semivariogram is a useful technique in the assessment of the range of correlation of the undrained shear strength of clays compared to time series analysis and random field theory.

2.3.8 Spatial analysis of SPT

The knowledge of semivariogram was used to analyse ordinary kriging method to predict SPT values at any point in the subsurface of Bangalore [Samui and Sitharam, 2012]. More than 2,700 field standard penetration test (SPT) values were collected from 766 boreholes spread over an area of 220-km^2 area in Bangalore, India, and geostatistical analysis was done. Samui and Sitharam (2012) used spherical semivariogram for the analysis without mentioning the best fitted model for experimental semivariogram. A vertical anisotropy factor was introduced to the semi-variogram model for vertical dimension. Because of layers and formation of different layers in geologic age, variation of soil properties is always greater in vertical direction compared to the horizontal direction. Ordinary kriging model and artificial neural network model (ANN) was applied for predicting SPT values in the 3D subsurface of Bangalore. In their study, MATLAB software was used to model ordinary kriging model and tested using cross-validation. The training and testing of back propagation model was carried out using neural network tool box in MATLAB. They have adopted a new type of cross validation technique based on residual analysis. A comparative study was made between the ordinary kriging technique and developed ANN model. ANN model was found to give good results compared to kriging while testing with the data known in prior.

2.3.9 Spatial analysis of consolidation settlement

Fenton (1997) demonstrated an example to calculate consolidation settlement of an unknown location using kriging technique. Initially, factors affecting consolidation settlement at various locations were considered as the spatial variables. Fenton ignored the source of uncertainty in field to be unknown. All random fields were assumed to be stationary, with spatially constant mean and variance. Covariance structure for the field was established to obtain a best linear estimate. Settlement coefficient of variation was estimated. It was observed that reduction in variance can be found before performing the sampling since the estimator variance depends totally on the covariance structure and the assumed functional form for the mean. Fenton observed that kriging technique can also be used to plan an optimal sampling scheme to minimize the estimator error. Fenton also demonstrated in his study that nature of permeability of soils is based on the principle of geostatistics.

2.3.10 Spatial analysis of Rock strength

The prediction of spatial distribution of rock strength over the tunnel length by using the limited number of samples from boreholes is one among the challenging task in rock excavation engineering. Exadaktylos (2008) has applied geostatistical technique to predict spatial analysis of rock strength. He suggested that mechanical behaviour of rock can be derived prior to excavation, if most probable values of penetration rate and wear of a tunnel boring machine (TBM) can be predicted. Initially the rock mass classification indices such as rock mass rating or rock quality index (RMR or Q) from borehole data is estimated and kriging is applied to interpolate the rock mass classification data and TBM data using semivariogram. A computational code was written in FORTRAN to perform kriging predictions in a regular or irregular grid in 1D, 2D or 3D space based on sampled data. The code developed has the capability (1) to establish a correlation between SE and rock mass rating (RMR or Q) along the chainage of the tunnel, (2) to predict RMR, Q or SE along the chainage of the tunnel from boreholes at the exploration phase and design stage of the tunnel, and (3) to make predictions of SE and RMR or Q ahead of the tunnel's face during excavation of the tunnel based on SE estimations during excavation. Also it was possible to continuously update the geotechnical model of the rock mass based on logged TBM data. Case studies considered for the proposed methodology includes: (a) data from a system of twin tunnels in Hong Kong, (b) data from three tunnels excavated in Northern Italy, and (c) data from the section Singuerlin - Esglesias of the Metro L9 tunnel in Barcelona. Results proved that there is a good agreement of RMR predictions from the borehole with the RMR prediction estimated from the SE of TBM in the L9 case study.

2.4 Tools available to apply geostatistics

Commercially available geostatistical tools include ArcGIS, Surfer, Grass, Gstat, mGstat etc. ArcGIS is one among the widely used commercial package that consists of variety of tools (Spatial Analyst® and Geostatistical Analyst®) in order to explore spatially distributed data, to evaluate the prediction uncertainty and to create surfaces for efficient decision making in various fields (Johnston et al., 2001). Though ArcGIS is a powerful tool for geo-statistics, it has some limitations such as (a) expensive, (b) lacking automation in selecting the best theoretical semi-variogram model for the experimental data, (c) lacking automation in selecting the best kriging technique for the parameter, and (d) lacking identification and separation of the positional outliers that take part in simulation. Dace (2002) and Van Beers el al. (2003) have developed geostatistical kriging tool (mGstat) in MATLAB. This tool being a user friendly tool has most of the limitations similar to ArcGIS for application to geotechnical engineering. Recently Exadaktylos (2008) has developed a geostatistical analysis code KRIGSTAT using FORTRAN 77 for considering of normality check, outlier removal, selecting best fit semivariogram and selecting the best model to apply for rock strength analysis. Even though there is no automatic consideration of outlier removal, normality test, optimized model and optimized kriging technique for the given parameter. Also ArcGIS require more user interaction to choose the best model for the data given.

2.5 Challenges in predicting soil parameters

There are many practical challenges in predicting the soil parameters. Few of them include inaccuracy of data, unaffordable cost for drilling borehole and conducting laboratory tests on samples collected in the field. In addition, the problem of data collection is aggravated as the owner of a given project aims for a fast pace construction. The parameter showing heavy nugget effect and not satisfying the semivariogram nature cannot be used in prediction using probabilistic spatial interpolation techniques. There are various tools that are commercially available for extensive application of kriging, but no particular tool meets the specific needs of a geotechnical engineer [Hammah and Curran, 2004]. This is the reason that best potential of geostatistics methods finds under utilization in the field of geotechnical engineering. If results of geostatistical methods are analyzed and interpreted properly, various inferences can be made to improve estimates and thus help to judiciously plan the sampling programs, thereby build realistic models of soil and rock of the site.

Chapter 3

Geostatistical Modeling

3.1 Introduction

Recent studies show that knowledge of spatial distribution and variation of in-situ soil parameters will help a geotechnical engineer in characterizing the mechanical properties of the in- situ rock or soil masses [Rouhani, 1996; Fenton, 1997; Phoon and Kulhawy, 1999; Uzielli *et al*. 2005 and Asa *et al.* 2012]. Geo-statistics is the branch of statistics dealing with spatial (or) spatio-temporal datasets. Geostatistical analysis can be used to predict spatial distributions of soil properties across medium to large areas or volumes**.** This method uses the parameter values at random locations and generate the prediction surface using mathematical/statistical functions. In addition**,** geo-statistics can help in integration, and differentiation of hydro-geologic and geotechnical parameters. Kriging is one of the geo-statistical algorithms introduced by Krige and is based on the theory of regionalized variables. Kriging interpolation methods are applicable when estimates with prediction uncertainty are required. There are mainly three types of linear kriging techniques: Simple kriging, Ordinary kriging and Universal kriging (linear/ quadratic/ cubic trend). This chapter provides principles of geostatistics and various geostatistical interpolation techniques available in the literature.

3.2 Principles of Geostatistics

Geostatistical methods offer a systematic approach for obtaining inferences about the quantities that vary in space. Corner stones of geostatistical modeling include- description, analysis and interpretation of spatial variability of a given parameter [Srivastava, 1996]. A parameter, z, is generally defined as z(x), where x is vector of spatial co-ordinates (i.e., in 1D, 2D and 3D) [ASCE, 1990].

The technique assumes expectation value of a soil property at location ‘x’ represented in the form:

$$Z(x) = m(x) + \xi(x) \tag{3.1}$$

where,

Z(x) is random variable representing z(x), z(x) is the parameter of interest

m(x) is the mean or expected value of variable Z(x), also popularly known as trend

ξ (z) is the residual component representing stochastic variation

Mean value being deterministic, is generally represented as a smooth function of space. The residual value is assumed to oscillate about zero. Also, the residual component shows a statistical dependence. This dependence is based on Tobler’s (1970) first law of geography, i.e., things that are closer in space are more related than those that are farther apart. This is the underlying feature in geostatistics. It is assumed in the modelling that deviations from mean value is to be normally distributed with a mean of zero and finite variance σ^2 [Sokal and Rohlf, 1969].

There are two major classes of probabilistic geostatistics: linear and non-linear geostatistics. The main features of linear geostatistics include: (1) usage of spatial correlation structure of spatial functions; (2) prediction of estimates based on weights (subjected to unbiased constraints) obtained by minimization of mean square error; and (3) capability to average measurements over different volumes [ASCE, 1990].

3.3 Experimental semivariogram

Semi-variogram is the statistical tool to model spatial variability within the data set for any parameter. There are certain weaker stationarity assumptions for generalisation of variogram model, known by intrinsic hypothesis, viz., (1) the mean is constant throughout space; and (2) the variance of $[Z(x) - Z(x + h)]^2$ is defined as a function of h [ASCE, 1990].

Experimental semivariogram (generally known as variogram) from the measured data points is defined as [Matheron, 1972]:

$$\gamma = \frac{1}{2N\,|\,h\,|}\sum_{i=1}^{N}[z(x_i + h) - z(x_i)]^2 \tag{3.2}$$

where,

Z(x) is the measured value of the parameter at x_i

Z (x_i+h) is the measured value at (x_i+h)

|h| is the average distance between the pairs of data points

N (|h|) is the number of pairs of data points that belongs to the distance interval, h

A semi-variogram in general increases non-linearly with distance, and levels off at certain distance, beyond which, distance has no effect on the variability in the parameter (Fig. 3.1).

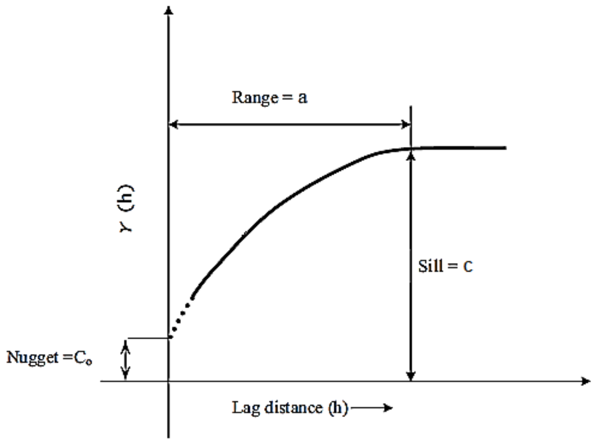


Figure 3.1: Typical empirical semivariogram

The three main characteristics of a semivariogram plot include the following [Issaks and Srivastava, 1989]:

Range (a): Semi variance value will generally increase as the separation distance between pairs increase. However at some point, an increase in the distance no longer causes a corresponding increase in the semi variance and the semivariogram reaches a plateau. The distance at which the semivariogram reaches this plateau is called the *range*. Large *range* values indicate greater spatial continuity.

Sill (c)*:* Vertical limit of the levelling of semivariogram is called *sill* [Goovaerts, 1997]. A semivariogram generally has a sill that is approximately equal to variance of data (Srivastava, 1996).

Nugget Effect (c_0):Though the value of the semivariogram at h = 0 is strictly zero, various inherent factors, such as sampling error and short scale variability, may cause sample values separated by extremely short distances to be quite dissimilar. This causes a discontinuity of the semivariogram at the origin called the *nugget effect*.

Semivariogram model is stable only if the measured values are stationary over an aerial extent. If the data values are non-stationary, spatial variability should be modeled only after appropriate transformation of the data [Clark and Harper, 2002].

3.4 Theoretical variogram models

Estimated semi-variogram can be modelled by considering various theoretical semi-variogram models to give an algebraic expression for the relationship between values at specified distances. Various theoretical models are available in the literature that include

Gaussian, spherical and exponential [Isaaks and Srivastava, 1989; Clark and Harper, 2000; Goovaerts, 1997; Deutsch and Journel, 1998]. Each model varies depending upon range, sill and nugget component(Fig. 3.2-3.4).

3.4.1 Gaussian model

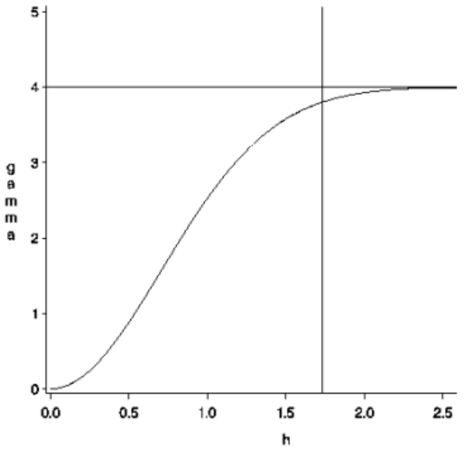


Figure 3.2: Gaussian model

Gaussian model is defined by:

$$\gamma_z(h) = c_0 \left[1 - \exp\left(-\frac{h^2}{{a_0}^2}\right)\right] \tag{3.3}$$

Effective/ practical range defined by Deutsch and Journel (1992) or range ε for Gaussian model as defined by Christakos (1992) is represented using a vertical line at $h=r_e=\sqrt{3}a_0$. For example the sill value for Gaussian model shown in Fig. 3.2 is about four variance units represented by a horizontal line that is asymptotic in nature.

3.4.2 Spherical model

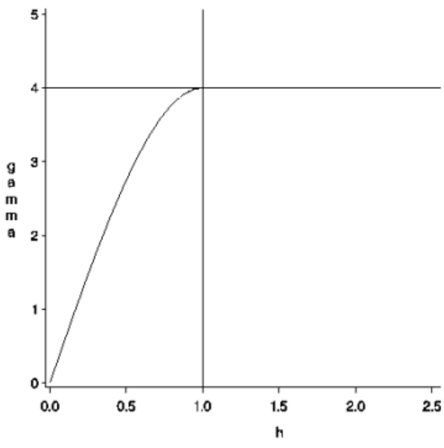


Figure 3.3: Spherical model

Spherical model is defined by:

$$\gamma_z(h) = \begin{cases} c_0 \left[\frac{3}{2}\frac{h}{a_0} - \frac{1}{2}\left(\frac{h}{a_0}\right)^3\right], \ for\ h < a_0 \\ c_0, \ for\ h < a_0 \end{cases} \quad (3.4)$$

This model has a definite range, hence a unique sill value. Range of the model is indicated by the vertical line at h=1 and sill using horizontal line at 4.0 variance units (as shown in Fig. 3.5).

3.4.3 Exponential model

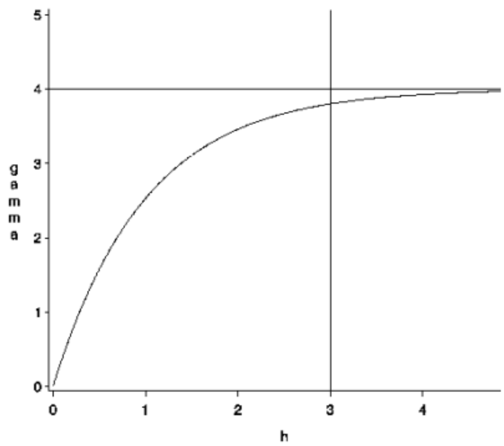


Figure 3.4: Exponential model

Exponential model is defined by:

$$\gamma_z(h) = c_0 \left[1 - \exp\left(-\frac{h}{a_0}\right)\right] \quad (3.5)$$

Effective/ practical range defined by Deutsch and Journel (1992) or range ε for exponential model by Christakos (1992) is indicated using a vertical line drawn at $h=r_e=3a_0$. For exponential model shown in Fig 3.6, the Sill value equal to 4 represented by a horizontal line and it is asymptotic in nature.

3.5 Model fitting

In order to select the best fitting theoretical model for experimental semivariogram to a given study parameter, each theoretical model is to be optimized for parameters such as sill and range of the experimental semivariogram. Most of the models available in the literature randomly choose a theoretical model without considering the best-fitted theoretical model. Generally, there are two methods adopted to select the best fit semi-variogram. The first method is based on visual inspection, and second method uses an automated approach using methods, such as, least squares, maximum likelihood, and various other robust parameters [Cressie, 1993 and Goovaerts, 1997: 98]. Goovaerts (1997) says ‘the *goodness* behavior of a fitted model is one among the toughest task and cannot be simply inferred based on rigorous tests; hence there is no best semivariogram model’. He has suggested to adopt a combination of visual assessment and statistical methods. Also care must be taken to ensure that the parametric assumptions for these methods are reasonable [Cressie, 1991].

3.6 Geostatistical interpolation

After obtaining the best theoretical semivariogram for the parameter, next step is to apply various geostatistical interpolation methods. Linear geostatistics predicts estimate of $z^*(x_0)$ at a location x_0, as weighted sum of the measured data $z(x_1)$, $z(x_2)$ … $z(x_n)$, at *n* locations x_1, x_2 …. x_n [Rouhani, 1989], such that:

$$z^*(x_0) = \sum_{i=1}^{N} |\lambda_i z(x_i)| \qquad (3.6)$$

Where, λ_is are the weights estimated for each random data locations to satisfy following statistical conditions.

The prediction estimator is subjected to two set of conditions.

(1) First condition is that the prediction estimator $z^*(x_0)$should be unbiased, i.e.,

$$E[z^*(x_0) - z(x_0)] = 0 \qquad (3.7)$$

where,

$z(x_0)$ is the value of the random function z at x_0. Substituting Eq. 3.8 in Eq. 3.7 yields to

$$E\{[\sum_{i=1}^{N} \lambda_i z(x_i)] - z(x_0)]\} = 0 \qquad (3.8)$$

Taking the expectation of each value and equating it to the mean, *m*, which is assumed to be constant, leads to:

$$\sum_{i=1}^{n} \lambda_i E(z(x_i))] - E(z(x_0)] = \sum_{i=1}^{N} \lambda_i m - m = 0 \qquad (3.9)$$

yielding to unbiased condition:

$$\sum_{i=1}^{N} \lambda_i = 1 \qquad (3.10)$$

(2) The second condition is that the estimator $z^*(x_0)$ should have minimum variance of estimation as follows:

$$V[z*(x_0) - z(x_0)] = V(z^*) - 2V(z^*, z_0) + V(z_0)$$

$$= \sum_{i=1}^{N} \sum_{I=1}^{N} \lambda_i \lambda_j \gamma_{ij} - 2 \sum_{I=1}^{N} \lambda_i \gamma_{i0} + C(0) \qquad (3.11)$$

Where γ_{ij} are semivarance values between points x_i and x_j defined as $\gamma(|x_i - x_j|)$

The minimization of Eq. 3.11 subjected to Eq. 3.10 is easily achieved by using Lagrangian method by determining λ_is, and this form of linear geostatistical interpolation is known as kriging.

3.7. Kriging techniques

Kriging was pioneered by Krige (1951) and was validated as a mathematical model by Matheron (1963b). There are two major classes of kriging methods: linear and nonlinear kriging. Commonly used linear kriging algorithms include: simple kriging, ordinary kriging and kriging with a trend model (universal kriging). Non-linear kriging techniques include lognormal kriging, multi-Gaussian kriging, disjunctive kriging, indicator kriging, probability kriging, and rank kriging.

This research work deals with the models that use linear kriging techniques. These techniques consider the minimization of the variance of estimation based on unbiasedness constraints. The estimation values are weighted averaged input point values, similar to the moving average technique. Various linear kriging methods used in this research are ordinary, simple, and universal kriging.

3.7.1. Simple kriging (SK)

Simple kriging technique assumes mean value, *m*, of the stationary random parameter is constant, also known prior to kriging [Deutsch and Journel, 1992]. The prediction estimate Z^*_{SK} at location 'x' is given by:

$$Z^*_{SK}(x) = m + \sum_{i=1}^{n} \lambda_i [Z(x_i) - m] \qquad (3.12)$$

where,

$Z(x_i)$ is the value of random variable at i^{th} location

n is the total number of data locations

λ_i is the kriging weight for measured value to be determined

m is the mean value of stationary variable

The simple-kriging estimator is exact interpolator and considers that kriging weights λ_is are unbiased. This can be proved by substituting Eq. 3.12 in the unbiasedness condition of Eq. 3.10, and equating the prediction estimates to m. The minimization of the estimation variance (given in Eq. 3.11), can be done by taking its partial derivatives with respect to each λ_i, and setting them equal to zero. This yields n simultaneous equations as follows:

$$\begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} & \cdots & \gamma_{1n} \\ \gamma_{21} & \gamma_{21} & \gamma_{23} & & \gamma_{2n} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} & \ddots & \gamma_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ \gamma_{n1} & \gamma_{n2} & \gamma_{n3} & \cdots & \gamma_{nn} \end{bmatrix} \times \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \vdots \\ \lambda_n \end{bmatrix} = \begin{bmatrix} \gamma_{01} \\ \gamma_{02} \\ \gamma_{03} \\ \vdots \\ \gamma_{0n} \end{bmatrix} \quad (3.13)$$

The minimum estimation variance for simple kriging is given by:

$$\sigma^2 sk = \sum_{i=1}^{n} \lambda_i \gamma_{oi} \quad (3.14)$$

The only solution to obtain an estimate at a point where a sample at known location, is λ_j, = 1 for the j that corresponds to the same data location of the point being estimated, making all other weights equal to zero. This demonstrates that simple kriging is an exact interpolator. The only disadvantage of simple kriging is that the constant mean value, m, must be known prior to interpolation.

3.7.2 Ordinary kriging (OK)

The ordinary kriging technique is a non-stationary algorithm that involves estimating the mean value, which is constant. A location-dependent estimate of the mean is used to replace the constant mean of the simple kriging technique. Mean value estimation is done by moving search neighbourhoods. Ordinary kriging is defined as [Deutsch and Journel, 1992]:

$$Z_{OK}^*(x) = \sum_{i=1}^{n} \lambda_i(x).[Z(x_i) - m(x)] \quad (3.15)$$

where,

m(x) =E {Z(x)}, location-dependent expected value of Z(x)

$\lambda_i(x)$ = estimated kriging weights

In this technique, unbiasedness condition is written in such a manner that the mean becomes part of the solution. Minimization of the variance of estimation, Eq. 3.11, subject to Eq.3.10 is done by introducing a Lagrange multiplier, μ. This minimization process can be written in the form

$$\frac{\partial \mathrm{Var}[Z(x_0) - Z^*(x_0)]}{\partial \lambda_i} - 2\mu = 0 \quad \text{For i=1, 2, 3….. n} \quad (3.16)$$

The equations considering the minimization of variance and unbiased condition forms system of linear equations as follows

$$\begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} & \cdots & \gamma_{1n} & 1 \\ \gamma_{21} & \gamma_{21} & \gamma_{23} & & \gamma_{2n} & 1 \\ \gamma_{31} & \gamma_{32} & \gamma_{33} & \vdots & \gamma_{3n} & 1 \\ \vdots & \vdots & & & \vdots & \vdots \\ \gamma_{n1} & \gamma_{n2} & \gamma_{n3} & \cdots & \gamma_{nn} & 1 \\ 1 & 1 & 1 & & 1 & 0 \end{bmatrix} \times \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \vdots \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} \gamma_{01} \\ \gamma_{02} \\ \gamma_{03} \\ \vdots \\ \gamma_{0n} \\ 1 \end{bmatrix} \quad (3.17)$$

The resulting estimation variance for ordinary kriging is

$$\sigma^2 sk = \sum_{i=1}^{n} \lambda_i \gamma_{0i} + \mu - \lambda_{00} \quad (3.18)$$

3.7.3 Universal kriging (UK)

In certain cases, the study parameter may exhibit drift or trend (eg. northeast or southwest drift) leading to non-stationary behaviour of the mean. In such a case, sampling domain can be limited such that very few locations will be influenced for the prediction, thereby preserving the local stationarity of mean. This approach is called as search neighbourhood method. Kriging with trend is known as universal kriging [Goovaerts 199]. This modeling approach requires more set of unbiasedness conditions. In UK, random function Z(x) is combination of trend component with a deterministic variation, m(x), and a residual component, R(x) as given by:

$$Z(x) = m(x) + R(x)$$

where, mean $m(x) = E\{Z(x)\} = \sum_{i=0}^{n} a_p f_p(x)$ (3.19)

a_p= p^{th}coefficient

f_p = p^{th}basic function that describes the drift

l = number of functions used in modelling the mean.

In a 2-D space with Cartesian coordinates (x,y), if the drift component is modelled as a first- order polynomial, the basic functions to be chosen are 1,x,y,while for second-order polynomial the functions are: 1, x, y, x^2, y^2, and xy. All these functions belong to unbiasedness conditions. Generalisation of modelling with ‘l’ polynomial function can be written as

$$\sum_{i=1}^{n} \lambda_i f_p(x_i) = f_p(x_0) \qquad p = 1, 2, 3 \ldots l \quad (3.20)$$

The universal kriging system with ‘l’ unbiasedness conditions is then found by minimizing the variance of estimation, given in Eq. 3.11, subject to the conditions of Eq. 3.20 using the Lagrange multiplier technique, which results in system of linear equations of the form

$$\begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} & f_1^1 & \cdots & f_l^1 \\ \gamma_{21} & \gamma_{22} & \gamma_{23} & f_1^2 & & f_l^2 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ \gamma_{n1} & \gamma_{n2} & \gamma_{n3} & f_1^n & \dots & f_l^n \\ f_1^1 & f_1^2 & f_l^n & 0 & \cdots & 0 \\ \vdots & & \vdots & & & \\ f_l^n & f_l^n & f_l^n & 0 & \dots & 0 \end{bmatrix} \times \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \mu_1 \\ \vdots \\ \mu_l \end{bmatrix} = \begin{bmatrix} \gamma_{01} \\ \gamma_{02} \\ \vdots \\ \gamma_{0n} \\ f_1^0 \\ \vdots \\ f_l^0 \end{bmatrix} \quad (3.21)$$

where:

γ_{ij} = semivariogram values between points x_i and x_j

$f^l{}_p = f_p(x_j)$, p^{th} basis function

μ_p is the Lagrangian multiplier associated with the p^{th} unbiasedness condition

The estimation variance of universal kriging is

$$\sigma^2_{UK} = \sum_{i=1}^{n} \lambda_i \gamma_{0i} + \sum_{p=1}^{l} \mu_p f_p(x_0) - \gamma_{00} \quad (3.22)$$

To evaluate the drift in UK, semivariogram should be known, and to evaluate semivariogram the drift must be known (as the semivariogram is found from the residuals, i.e., [Z(x) — m(x)]. One solution to such a circular nature of UK is the residual approach [Matheron, 1969 and Olea, 1975], where polynomials are used to model the drift within local neighbourhood. The residuals are then calculated by direct subtraction of the polynomial trend model from the initial values. Several combinations of neighbourhood size, polynomial degree, and semivariogram model for the residuals should be experimented before deciding on the selection that best matches the actual trend and the corresponding semivariogram. The second transformation approach is based on using only increments of data that do not depend on the drift [Matheron 1973]. In this study, the drift was modelled as a polynomial of order k, with a corresponding underlying polynomial covariance of order 2k + 1. The random function is defined as an intrinsic random function of order, k, and the measure of spatial correlation as the generalized covariance.

3.8 Cross-validation

According to Cressie (1993), cross validation is the common means to assess statistical correctness of prediction at measured location. Cross-validation omits a known (measured) point and calculates the value of the parameter at the same location using the predicted model parameters and neighbourhood type.

Optimal kriging technique is determined based on method which gives the minimum root mean squared error (RMSE) value, mean error estimate near to zero. Measure of kriging accuracy is in the form of estimation variances. These variances are used to design sampling plans as each estimate has an estimation variance and they do not depend on individual observations [Rouhani, 1996]. Rouhani and Hall (1988) have found that it is not sufficient to consider only the estimation variance, but some more factors are required for expanding the sampling plan. A common practice is to assume that at any location errors are normally distributed with a mean zero and a standard deviation equal to square root of estimation variance [Journel and Hujbregts, 1978].

3.9 Prediction surfaces

Prediction surfaces show the continuous variation of the parameter considered within the study area. Prediction value can be obtained as:

$$Z^* = \sum_{i=1}^{N} w_i * z_i \qquad (3.23)$$

where,

Z^* is the predicted value at unknown location

w_i is the estimated weight for i^{th} known location

z_i is the input value at i^{th} known location.

In addition, the quality of the predictions can be examined by generating a prediction standard error surface which quantifies the uncertainty for each location on the prediction surface. Prediction error estimate can be obtained as:

$$\sigma = \sqrt{\left(\sum_{i=1}^{N}(w_i * \gamma(h_{pi})) + \lambda\right)} \qquad (3.24)$$

where,

h_{pi} is the distance between the prediction p and input location i

$\gamma(h_{pi})$is the value of the semi-variogram model for the distance h_{pi}

λ is a Lagrange multiplier minimizing error.

If the data is normally distributed, a simple rule of thumb followed is that 95 percent of the time, the absolute value of the surface will be within the interval formed by the predicted value, plus or minus two times the prediction standard error. Also, it can be observed in the prediction error surfaces that locations near sample point will show less error. Fenton (1997) suggested that prediction errors are useful to suggest optimal sampling locations.

Chapter 4

Geostatistical algorithm development

4.1 Overview

Conventional geostatistical tools are oriented mainly towards large-scale projects, for example, mining projects (mine scheduling, pit design, etc.), transportation planning projects (optimal network route detection, identifying noice regulation violations, maintaining transportation systems etc.), construction projects (construction planning, construction scheduling, construction designing etc), and other major projects (environmental management, fire mapping, weather warnings, flood management, etc.). Hence, geostatistical tools may have limitations when applied to problems in other fields, like geotechnical engineering. The main objective of this research is to develop a MATLAB based efficient, automated, public domain, and cost effective tool that can generate prediction surface of any spatially varied parameter from the observations made at random locations (with small samples) using linear kriging principles. This chapter discusses the various steps followed to develop an automated geostatistical tool (using MATLAB) to apply specifically to problems in the field of geotechnical engineering. The advantages of developed tool over the conventional geostatistical analyst tool available in ArcGIS is discussed at the end.

4.2 Geostatistical analyst in GIS

Geostatistical analyst helps in exploring spatial variability of data, examining global and local trends, developing normal probability plots (known as quantile plots), optimizing model parameters based on cross validation and, producing reliable maps, viz., predictions, prediction errors, etc. Figure 4.1 is a screenshot showing various sub-components of ArcGIS Geostatistical analyst.

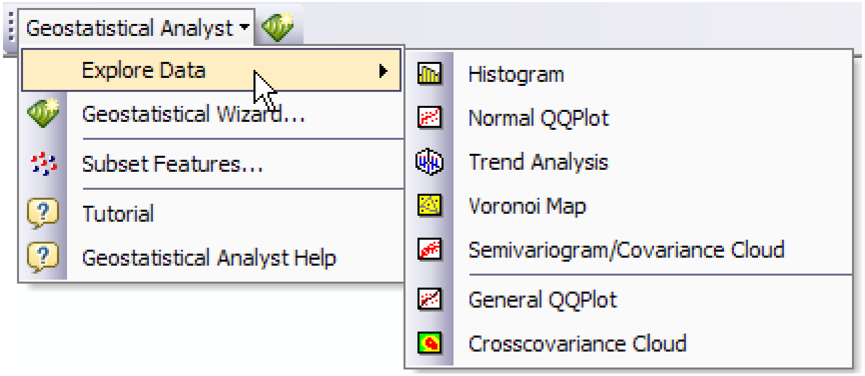


Figure. 4.1: Geostatistical analyst in ArcGIS

Various steps adopted for interpolation of a spatially-varying parameter using ArcGIS are detailed in the following sections.

4.2.1 Data representation

In order to represent borehole information in ArcGIS Environment, Universal Transverse Mercator (UTM) 44 North zone projections with World Geographic System (WGS) datum were considered. An accurate spatial representation of various features of the region, that include point features such as boreholes and wells; line features such as roads and streams; and polygon features such as buildings, land zoning, and study area boundary was done in GIS. This overall process is called *digitization* of given study area. Each borehole location contains attribute data on Atterberg limits, soil type, insitu test parameters such as Standard Penetration Test (SPT) values, Cone Penetration Test (CPT) values, etc., at various depths. The base map of the study area with various features overlaid along with the attribute information is developed in ArcGIS for further analysis. Geo-referencing of the image files was performed to rectify the feature set towards spatially correct location and is achieved by selecting at least four ground control points (GCPs) that have accurate latitude and longitude information. Further information of soil properties were prepared in comma separated value (CSV) files, exported to GIS, and later converted into shapefiles.

4.2.2 Exploring spatial data

Data exploration components in geostatistical analyst (as shown in Fig 4.2) involves developing (1) Histogram – to analyse summary statistics, (2) Normal and General QQ Plots – to check if the data is normally distributed and if the two data sets have similar distributions, (3) Trend Analysis – to detect visually spatial trends, (4) Semivariogram/ Covariance cloud – to evaluate spatial dependence of data, and (5) Crosscovariance Cloud – to examine spatial dependence between two data.

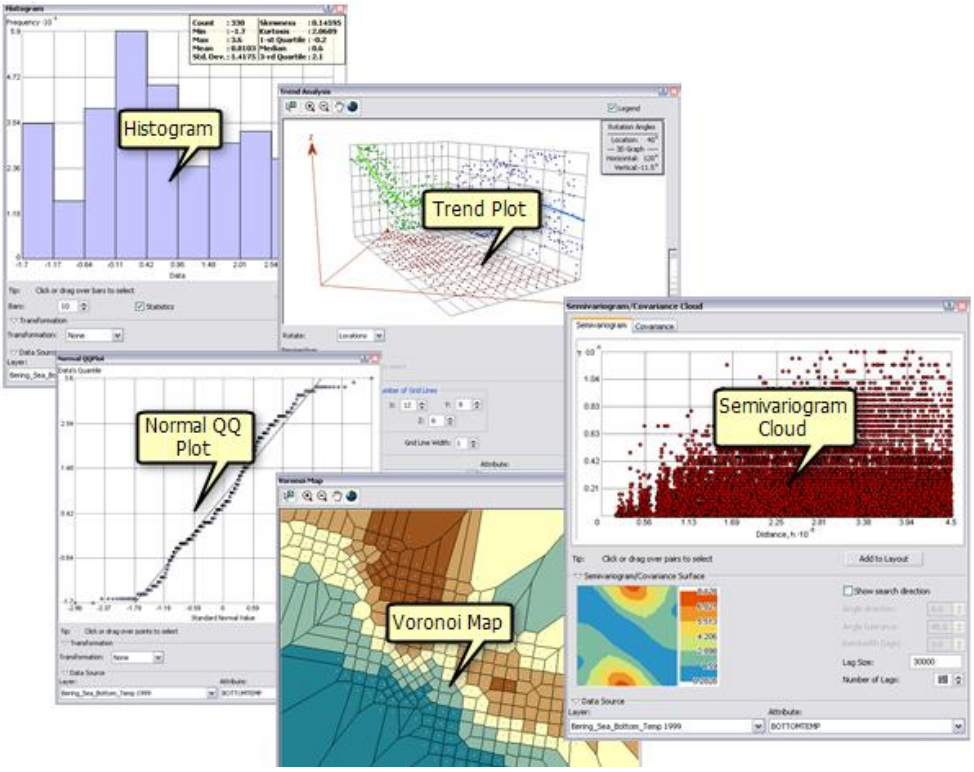


Figure 4.2: Graphical illustration of various tools to explore data in GIS (ArcGIS tutorial)

As per ASTM D- 5923, linear geostatistical methods are applied only when the data follows normal distribution. Semivariogram modelling is the next step to assess the evaluation of spatial dependence. Trend analysis is used to check the presence and direction of trend for the soil parameter.

4.2.3 Evaluation of geostatistical techniques

Once the data exploration is achieved, the next step is to evaluate various geostatistical techniques based on the normal distribution of data. Geostatistical analysis is done using geostatistical wizard of ArcGIS. Various steps to conduct geostatistical analysis are shown below graphically in Fig 4.3:

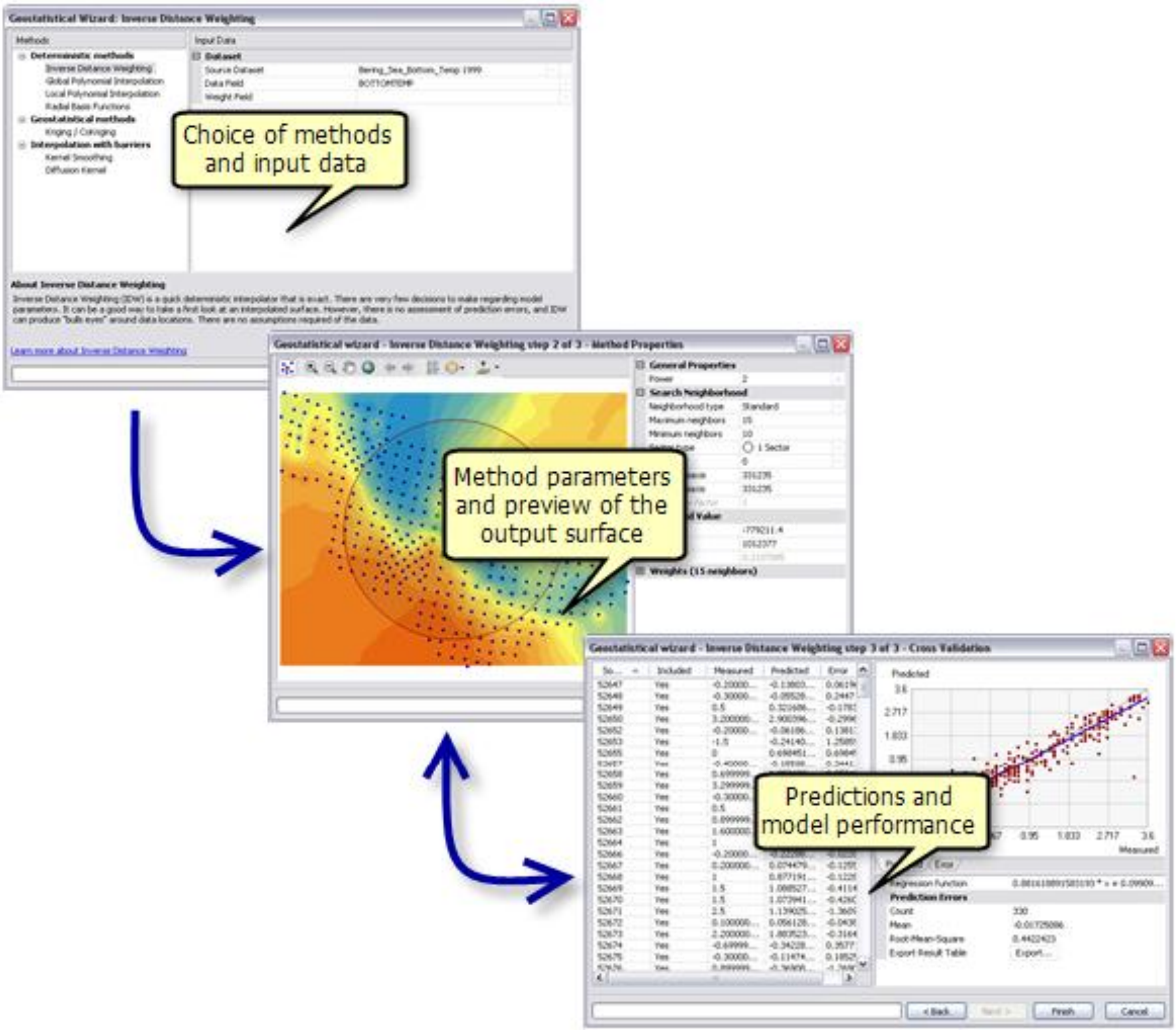


Figure 4.3: User interactive steps for spatial interpolation (ArcGIS tutorial)

Geostatistical Wizard can aid the user to select the interpolation method, select the parameter to be interpolated, vary lag divisions of semivariogram, select various theoretical variogram models and select its parameter and vary the neighbourhood factors in order to conduct cross validation of the model and obtain minimal residual statistics. Certain procedures, shown in Fig 4.3, are repeated back and forth until model parameters satisfying optimal residual statistical parameters are obtained. Geostatistical Wizard offers tools to analyse deterministic geostatistics, viz., inverse distance weighed method and radial basis functions and probabilistic geostatistics (simple kriging, ordinary kriging, universal kriging, indicator kriging, probability kriging, and disjunctive kriging). Geostatistical Wizard considers various theoretical semivariogram models, viz., spherical, Gaussian, exponential, circular, pentaspherical, tetraspherical, etc. Various components of the semivariogram such as nugget, sill and range are fixed based on the cross validation results. Geostatistical wizard helps the user to vary neighbourhood factors, such as nearby minimum number of points, nearby maximum number of points, radial distance, direction of ellipse, etc.

Cross validation process helps to assess the quality of the output map by comparing real values at unknown locations and predicted values. Due to lot of practical difficulties to collect real values to a great extend from the field, some values among random data set are used to model and generate the surface. The remaining part is used to compare and validate the output surface. The Subset Features tool in Geostatistical analyst (refer Fig 4.1) helps in dividing the data set into two parts. Residual statistics used in cross validation includes Mean Error (ME), Standard Error (SE), Root Mean Squared Error (RMSE), Kriged Root Mean Square Error (KRMSE), etc. Once the geostatistical method, variogram and neighbourhood factors are fixed, prediction and error maps can be developed for further decision.

4.3.1 Development of MATLAB based Geostatistical algorithm

Various steps involved in the development of automated geostatistical algorithm are as follows:

4.3.2 Conversion of co-ordinate system to projected system

The borehole information collected from the agencies will have co-ordinates either in GPS co-ordinates in spherical system (latitude / longitude) with respect to an assumed datum and spheroid or localized co-ordinate system (based on triangulation / total station surveying). For smaller areas (such as in small-scale construction project sites) representation in planar co-ordinate systems is convenient and appropriate [Canters, 2002]. Hence, if given borehole information have global co-ordinate system, it is necessary to convert to planar co-ordinate system. The developed code considers the datum and projections that is appropriate to the geographic location of the study area. Transformation to projected co-ordinate system comprises of the following steps:

4.3.2.1 Datum selection

In order to get accurate results of the distance between two points on earth, a spheroid (or ellipsoid) which best fits the shape of earth between two points should be assumed. This spheroid is characterized by semi-major axis distance ‘a’, semi-minor axis distance ‘b’, flattening ‘(a-b)/a’ values with which it is possible to calculate distance between points.WGS-84 coordinate System is being adopted universally as the standard form of Geographical Coordinates System as it has a common origin and a common spheroid. This particular datum is a geocentric geodetic datum, established through space geodetic observations, which is earth-centered, earth-fixed (ECEF). Global Positioning System (GPS) and other similar modern aids needed for Global Navigation Satellite System (GNSS) use this co-ordinates system. Everest-1830 data is the oldest datum (specific to India) which is used for preparing maps by Survey of India and other agencies. Various types of datum chosen for the study and their characteristics are shown in Table 4.1.

Table 4.1: Various geodetic datum and their characteristics

Datum	Equatorial Radius, meters(a)	Polar Radius, meters (b)	Flattening, (a-b)/a
WGS 84	6378137	6356752.31	1/298.257223563
Everest-1830	6377276.34	6356075.4	1/300.8017

4.3.2.2 Selection of projection system

Modern mapping systems uses a transverse mercator (or close variant) to preserve conformality, thereby generating less distorted maps. In order to project the locations in global co-ordinates to planar co-ordinates, Universal Transverse coordinate Mercator (UTM) grids are used. These are projections created by laying a square grid on the Earth.

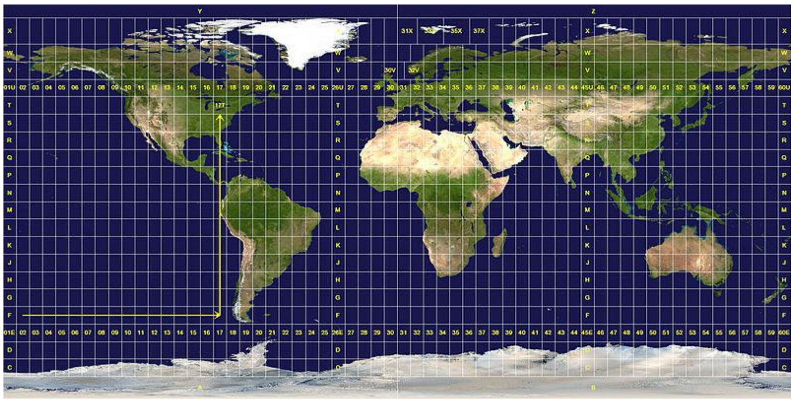


Figure 4.4: UTM System

UTM System divides the Earth into sixty zones, each at a six-degree band longitude (Figure 4.4). Based on the selection of datum, different maps will have different grids. Army Technical Manual TM 5-241-8 formulas were used to convert Spherical co-ordinate system to Cartesian system.

Code was developed to consider the datum specific to India and calculate northing, easting and zone appropriate to the geographic location of the study area.

4.3.3 Removal of positional outliers

Study parameter is checked for the presence of any value which positionally deviates by excessive amount from other observations as to cause suspicious nature is termed as positional outlier (Hawkins 1980). These may be due to erroneous entry during data collection or due to sudden variation of property. Outlier detections are often based on distance measures, clustering and spatial methods. In the present work, a robust approach using point density was employed for identifying spatial outlier.

The developed code can automatically detect the outermost data point using point density (number of data points per the rectangular area outlined by the four extreme directional points) approach, by suppressing each borehole location, one at a time and comparing the point density for each altered rectangular area. A borehole location is considered as a positional outlier if the point density is rapidly increased because of its absence. An increase in point density by 10-15% due to the absence of a bore hole location was considered as threshold for separating out the given data point.

4.3.3 Normality test of the data

As per ASTM D-5923- 96 (2010), it is necessary that the data to be normally distributed to evaluate linear kriging techniques. One of the available methods to evaluate normal distribution is to use normal score transformation [Journel and Hujbergts, 1978; Isaaks and Srivastava, 1989; Cressie, 1993]. The conventional geostatistical tools use graphical methods (Q-Q plots) to test for the normality of the data. Normal QQ plot is created by plotting data values versus value of a standard normal where the cumulative distributions are equal. Visual assessment of normality is done based on the closeness of various data points to form a straight line. Such techniques are not suitable for small samples (as in case of site parameters), due to difficulty in comparison. The Code developed in MATLAB conducts the Statistical based Kolmogorov-Smirnov test to check the normality of the data at 5% and 10% significance levels. Null hypothesis assumed considers that the data follows the normal distribution. The developed code proceeds further only when the null hypothesis is accepted.

4.3.4 Development of Experimental Semivariogram

The variation of the semi-variance in the measurement of a given parameter with distance is grouped into bins (based on the distance pairs) and represented as the empirical semivariogram [Matheron 1972] of the data set, and is given by (Eq. 4.1):

$$\gamma(h) = \frac{1}{2N|h|}\sum_{i=1}^{N}[z(x_i + h) - z(x_i)]^2 \qquad (4.1)$$

where, z (x_i) is the measured value at x_i

z (x_i+h) is the measured value at neighbour point at an average distance |h|

N (|h|) is the number of pairs of data points that belongs to the distance interval represented by h.

A code was developed to calculate the distance between pairs and group them into bins based on the given lag divisions and also calculate the variance of values of the parameter based on binning, thereby preparing the experimental semi-variogram for the data.

4.3.5 Theoretical model fitting

In order to select the best fitting theoretical models for various study parameters, explained in section 3.5, each model is to be optimized for constants such as sill and range. It may not be always easy for a user to visually assess the model fitting parameters, hence optimization using the least square technique has been adopted to obtain the best-fitted model. Model fitting parameters were obtained by non-linear least square fitting method using optimization tool in MATLAB. The model having minimal residual, i.e., Root mean squared error (RMSE) in semi-variance values of theoretical and experimental semivariogram was chosen as the best-fitted model. Best-fitting model and its parameters (a, c, c_0) were then extracted.

4.3.6 Simple kriging algorithm development

After calculating the distance between various known locations and semivariogram values using fitted semivariogram model, linear kriging techniques using Eq. 3.13 were evaluated. Mostly, kriging techniques are written of the form

$$[C]*\{w\} = \{D\} \tag{4.2}$$

where, C is the matrix that contains all the semivariogram values for the distance between the values at known locations, w is the vector that contains weights estimated for the unknown location, and D is the vector that contains all the semivariogram values for the distance between the values of known locations and unknown locations

Initially, distance between two known locations were estimated and semivariogram values determined. This process was repeated for all input points and semivariogram values, and filled in C matrix. Unknown locations were selected by gridding the entire region by dividing the maximum x and y distances with an integer factor (minimum of 10). Then the distance between the various known locations and first output grid location was estimated and semivariogram values were filled in D vector. Various kriging weights of input values for the first output location were determined using Eq. 4.2 by finding the inverse of the C matrix and multiplying the inverse with the D vector. Since the matrix C might be badly scaled, decomposition techniques and least square techniques were used in solving for unknown weights. Prediction estimates were calculated for each output locations as the sum of the products of the weight factors and the input values. Estimation variance is calculated using Eq. 3.14 as summation of products of kriging weights and corresponding variance. Standard error for each predicted estimate can be obtained as the square root of estimation variance.

4.3.7 Ordinary kriging algorithm development

Ordinary kriging algorithm development, based on Eq. 3.17, is similar to simple kriging technique. In this case, only certain input points which will contribute in improving the accuracy of the output value are selected depending on factors such as specified limiting radius, and minimum and maximum values of neighbour points. Farther points are ignored based on Tobler's law of geography which says that as the distance between the points increases, properties are less co-related in space. This procedure is called searching neighbourhood. The developed MATLAB code will calculate the best suitable neighbourhood combination factors for each grid location. Calculation of kriging weights, prediction estimates, estimation variance, and standard error are similar to simple kriging algorithm development.

While estimating the kriging weights, some values are observed to be negative as some points are "shadowed" by closer points. Negative weights can affect the accuracy of prediction by increasing or decreasing the prediction estimate/ estimation variance. Program will eliminate the points with the most negative weight, and recompute the weights and repeat the process until the value becomes positive (satisfying unbiased condition). Thus, the algorithm resolves the issue by converting negative weights to positive weights, thereby improving the performance of predictions.

4.3.8 Universal kriging algorithm development

Universal kriging algorithm is developed based on Eq. 3.21. This algorithm is similar to ordinary kriging, except that the equations concerned with the local trend (which considers the various degree of spatial co-ordinates) have to be solved. Procedures of searching neighbourhood points and limiting radius, negative kriging weight elimination, prediction estimate, standard variation estimate and error estimate is similar to that of ordinary kriging.

4.3.9 Cross validation

Cross validation technique improves the accuracy of the prediction estimates. Cross validation method developed in MATLAB is such that after estimating the values at all unknown locations, random location values are suppressed and re-estimated. The method which gives the least RMSE value among the cross validation results of various methods is automatically chosen by code as the most suitable method for the parameter.

4.3.10 Prediction surfaces and Error surfaces

Prediction contours, prediction surfaces and error surfaces are automatically created using contouring tool available in MATLAB from various methods based on Equations 3.23 and 3.24.

4.3.11 Development of GUI

A Graphical User Interface (GUI) tool was developed in MATLAB to aid the user to perform tasks interactively, such as plotting, fitting curves and surfaces, developing contour profiles, etc. Fig 4.5 shows the screen shot of GUI screen developed using MATLAB to execute the developed automated program for any user to interact and apply the principles of geostatistics to geotechnical applications in India.

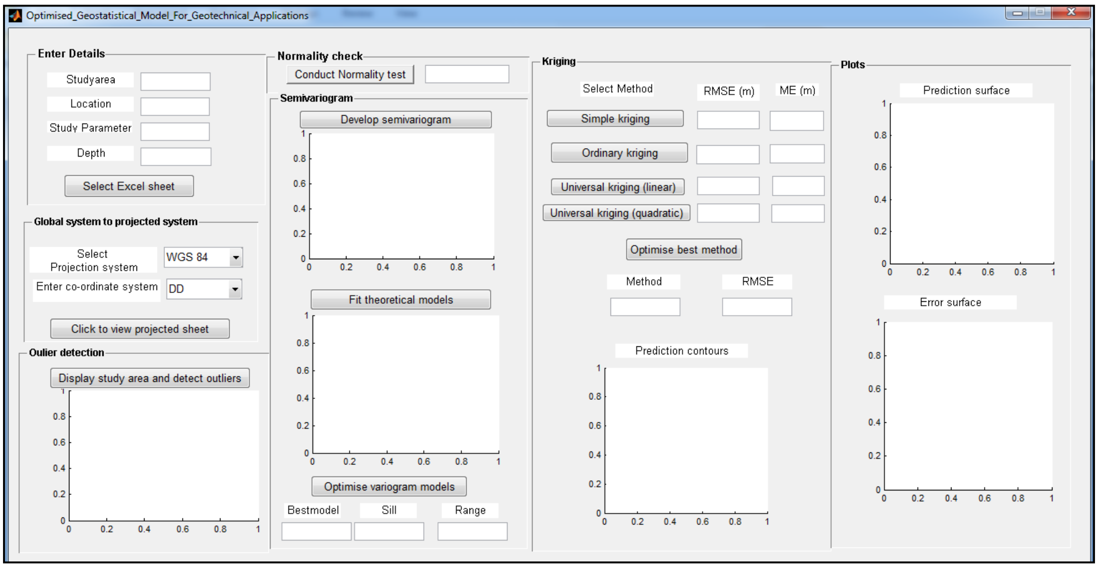


Figure.4.5: Developed Kriging GUI

Figure 4.6 provides the flow chart detailing various processes involved in the generation of prediction and error surfaces of a parameter.

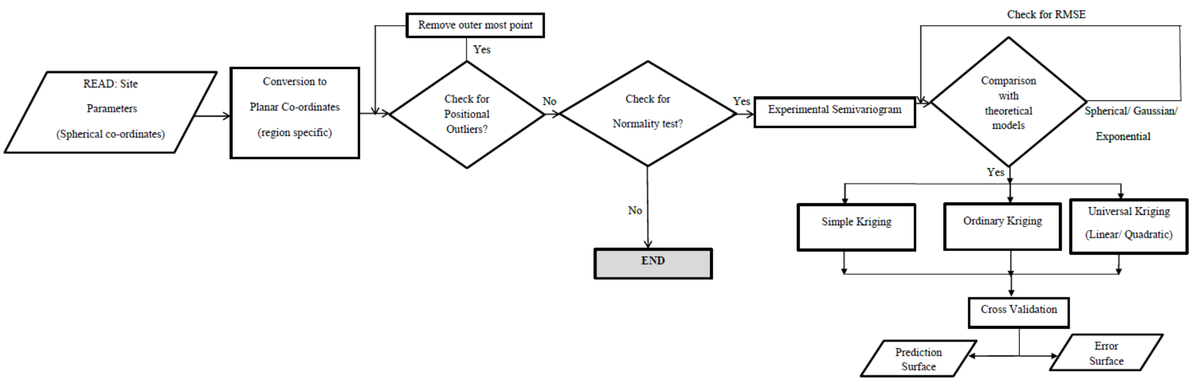


Figure 4.6: Workflow of developed kriging algorithm

Chapter 5

Application of Geostatistical Algorithm

5.1 Overview

The robustness of geostatistical methods in the field of geotechnical engineering is assessed through application of automated geostatistical algorithm for various problems in geotechnics. This chapter deals with the application of the developed algorithm to generate prediction surfaces for various soil parameters. The application of developed tool to various case studies helps to measure the effectiveness of the developed algorithm. For all the case studies, Universal Transverse Mercator (UTM) system with 44 North zone were considered for projecting the geographic locations from spherical to planar system. Based on the site exploration record, various parameters were chosen for a particular depth based on the maximum number of data points, and the corresponding prediction surfaces were generated. Results of the developed algorithm were compared with results obtained from the linear geo-statistical models using conventional tools.

5.2 Paradip Refinery project, Orissa (Case Study – 1)

5.2.1 Site description

The proposed refinery site (Fig. 5.1) is located approximately 7 km South West of Paradip Port on the North bank of the River, Kansarbatia, and is located near Paradip port in Jagatsingpur district of Orissa, India. This region frequently experiences cyclones. Geographic location of the site is 21^{o} 07’11.17” N latitude, and 90^{O} 18’ 20.28” E. longitude. Geographical coverage area of the region is about 3549 acres (14.96 km^2). There were total fifty seven (57) boreholes drilled to conduct extensive site investigation. It was found that no rock was encountered even after exploring up to 100m depth below existing ground level. Ground surface was slightly uneven as bore holes drilled in the area under study differed by 0.63 m to 4.78 m, due to part of the area having been filled. In order to conduct the site investigations, the entire site was classified as filled and unfilled areas. During the investigation, it was observed that the filled up area constitutes yellowish brown fine to medium sand to a depth of about 3.0m, followed by a layer of soft to firm clay followed by sand strata which is loose at the top, becoming medium dense to occasionally dense. These are underlain by alternate layers of (medium dense to very dense) sand and (stiff to hard) clay up to the maximum depth 100m. The variations of static ground water levels were found to be in the range of 0.4m to 3.8m below existing ground levels.

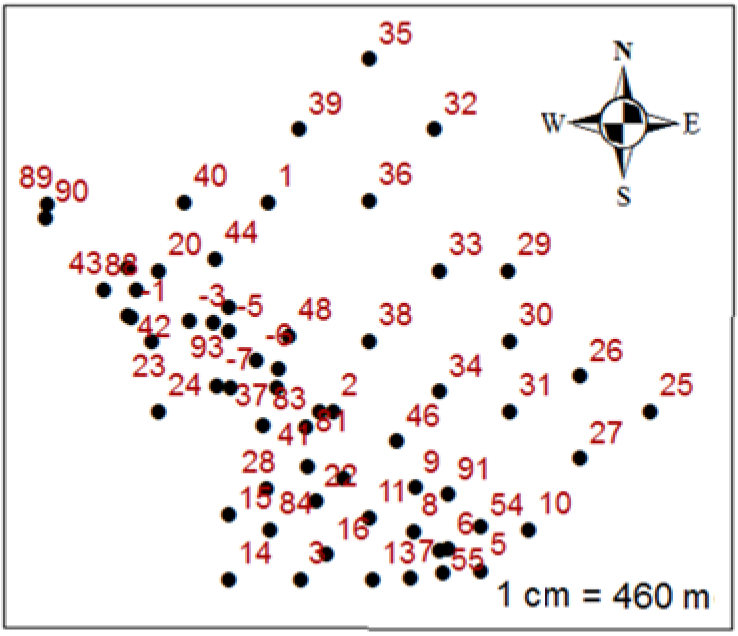


Figure 5.1: Schematic of Paradip refinery project site with boreholes

5.2.2 Spatial outlier removal

Study parameter taken for the case was clay content at 3m depth and had values ranging from 1 to 47. There were a total 35 data values for the study parameter with a point density of 1.21 /km^2 of borehole coverage area. The first objective was to separate spatial outliers present in the data. The algorithm estimates initial point density (with *n* data points), and then iteratively compares with point density obtained after removing one point at a time (with *n-1* data points). Point density is estimated by dividing the number of points under consideration with the rectangular area formed by considering outer most data points in the *x-y* plane. A borehole location is considered as an outlier when there is an increase in point density by more than a threshold value of 15 % of the initial point density. One such dominant positional outlier which has increased point density from 1.21 /km^2 to 2.36 /km^2 was observed during the process (Fig. 5.2), and eliminated from the analysis.

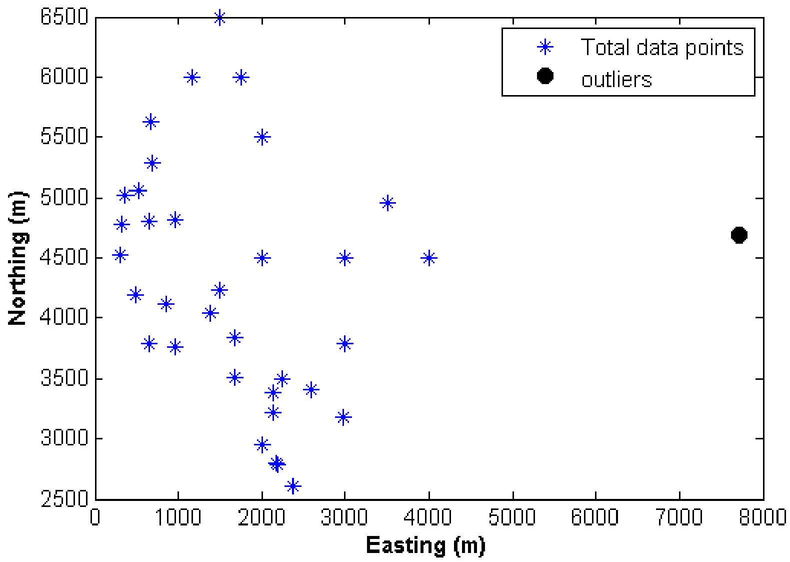


Figure 5.2: Spatial distribution of data considered in the analysis

5.2.3 Normality check

According to ASTM D- 5923, method of linear geostatistics is applicable only when data follows normal distribution. Since, visual inspection of the normality of the data using Q-Q plots cannot accurately check for the normality, a hypothesis based Kolmogorov-Smirnov test that suits for smaller samples was applied with the algorithm. Test was conducted in MATLAB and the results shows that distribution follows the normal distribution at a significance level of 5%.

5.2.4 Semivariogram and model fitting

First step before applying linear geostatistics was to develop the experimental semivariogram model. Algorithm is designed so as to consider optimal lag divisions of 10 and generate the semivariogram. Figure 5.3(a) shows that experimental semivariogram has a nugget effect initially, followed by a gradual non-linear increase indicating that there is a strong influence of distance on the study parameter and then a sudden decrease and increase of the values. This is because, certain points have failed in satisfying the basic assumption of correlation of parameter with distance. This observation led to the development of outlier separation study for the data. As the point causing semivariance value less than 150 m^2 has been separated as outlier, the semivariogram has a gradual increasing nature (Fig. 5.3(b)), closely following the ideal nature. The decreasing trend observed in the semivariogram is mainly due to either positional outlier or inaccuracy in data. Hence, accuracy in data collection is an important factor before the application of kriging technique.

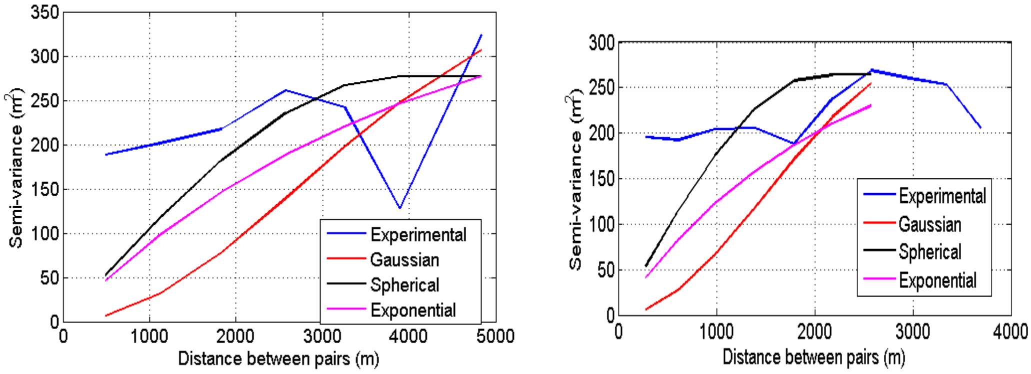


Figure 5.3: Comparison of experimental variogram with theoretical models before (left) and after (right) outlier separation

Next step in the analysis was to select the best fitting theoretical model for the empirical model. Various theoretical semivariograms were fitted to the experimental variogram (Fig. 5.3(b)) using the developed algorithm based on the optimization of the parameters (such as range and sill). Final value of the theoretical semivariogram was taken as initial guess for theoretical model parameters, and optimization was done by giving upper bound and lower bound between 0.8 to 1.2 times the initial guess values. The optimal theoretical model is selected based on the minimum residual (RMSE) values for the semivariance obtained from theoretical and experimental semivariogram. Best fitted model to the data was spherical model with sill and range values 263.8 m and range 2059 m respectively.

5.2.5 Evaluation of Kriging techniques

Once the theoretical model is fixed, various kriging techniques such as simple kriging, ordinary kriging and universal kriging (linear and quadratic trend) are evaluated using the algorithm developed. Simple kriging requires a known value for mean (or, mean surface for local search) as input to the model; ordinary kriging assumes mean to be constant, unknown and estimated in the searching neighbourhood; and universal kriging models local mean as low order polynomial functions of the spatial coordinates. Unknown locations were specified by gridding entire location with a division factor of 10 for the largest dimensions across the region. The optimal search neighborhood factors to evaluate ordinary and universal kriging techniques are obtained using algorithm by varying model sensitive parameters including −minimum and maximum number of neighborhood points and the searching radius. It was observed that an increase in searching radius has an effect in simulation accuracy upto certain extent, beyond which, there is no further reduction in RMSE (Table 5.1).

Table 5.1: Selection of optimal neighborhood parameters

	2 points(Min)			3 points(Min.)		
Limiting Radius (%max. distance between pairs)	3 points (Max)	4 points (Max)	5 points (Max)	4 points (Max)	5 points (Max)	6 points (Max)
15%	7.68	8	8.11	9.15	9.24	9.39
20%	7.6	7.95	8.11	8.12	8.28	8.46
25%	7.59	7.98	8.15	7.98	8.15	8.58
30%	7.59	7.98	8.14	7.98	8.14	8.57
35%	7.59	7.98	8.15	7.98	8.15	8.57
40%	7.59	7.98	8.14	7.98	8.14	8.57

Negative weights computed were converted to positive weights. The method which gives the optimal residual statistics (least RMSE and close to zero ME) was chosen as the best method for the parameter. Ordinary kriging with a minimum and maximum neghborhoods of 2 and 3 was obtained as the best kriging method to generate the prediction and error variance surfaces after the removal of outliers (as Table 5.1 and 5.2). It can be clearly seen from Table 5.2 that outlier has a significant effect in minimizing the residual statistics, there by increasing the model performance.

Table 5.2: Effect of outliers on Kriging simulations

Kriging Algorithm	RMSE (m)		Mean Error (m)	
	Before outlier separation	After outlier separation	Before outlier separation	After outlier separation
Simple kriging	10.63	9	-0.01	-0.53
Ordinary kriging	9.1	7.59	-0.77	0.24
Universal kriging (Linear)	11.08	8.28	-1.89	0.33
Universal kriging (Quadratic)	14.72	10.48	0.47	-0.87

Cross validation was performed by suppressing values at each known location, and re-computing the value using the model parameters. Figure 5.4 shows the prediction surfaces and error surfaces for the optimal kriging technique created using the algorithm. Clay content values are very low along the north east region of study area and higher in western region. The gaps in the prediction surface show the inability to interpolate for the unknown with the given model and neighborhood parameters. The conventional tools at such locations will execute extrapolation techniques to predict the unknown values. Further sampling locations are suggested for the portion of the prediction map where data is not available indicated by white spaces.

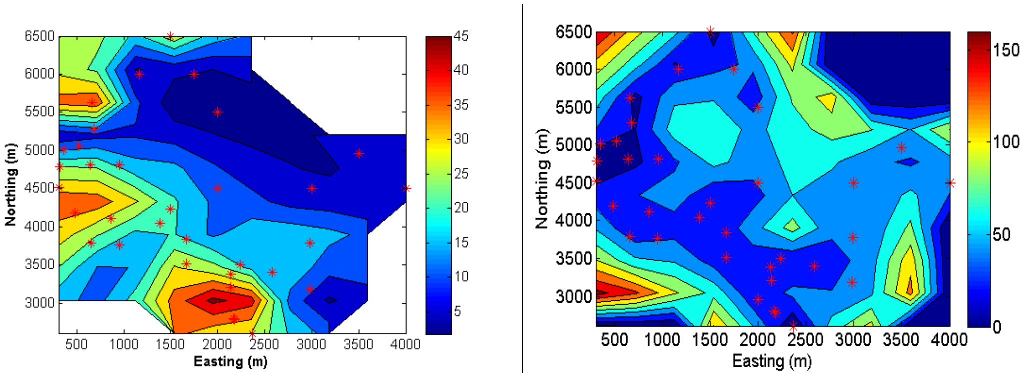


Figure 5.4: Model generated prediction (left) and error variances (right) surfaces

It was also observed that Lag distance / lag number and grid divisions have neglgible effect on the choice of best semi variogram (as Table 5.3 and 5.4). Hence an optimal lag division of 10 and grid division of 10 was taken to reduce the computational time in each analysis.

Table 5.3: Selection of optimal lag divisions

Lag divisions	RMSE(m)
10	7.60
15	7.59
20	7.60

Table 5.4: Selection of optimal grid divisions

Grid Divisions	RMSE (m)
5 * 5	12.49
10 * 10	7.59
15 * 15	7.59

Results of cross validation (Fig. 5.5) suggest that the model predicted clay content values are in convergence with the observed data at the known locations.

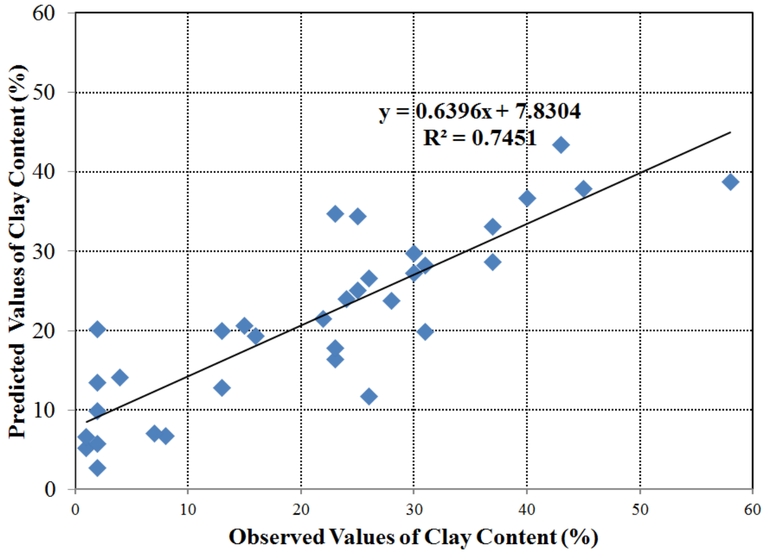
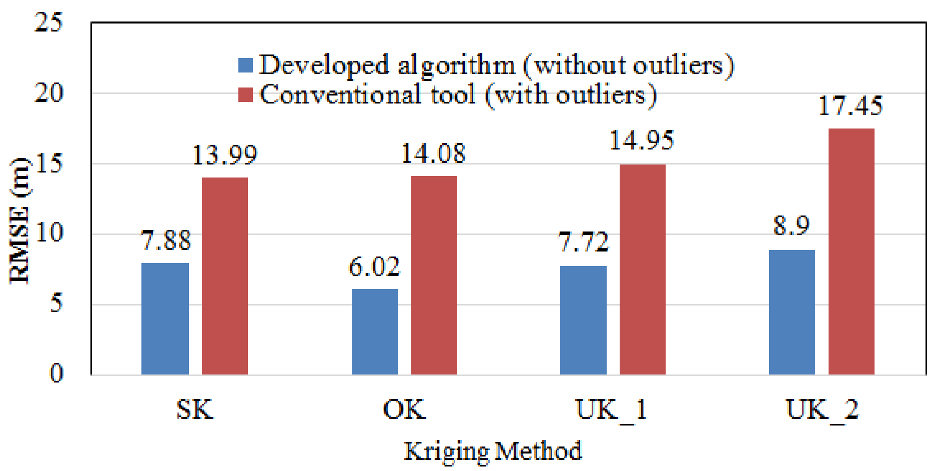


Figure 5.5: Comparison of observed and predicted moisture content values during cross validation

5.2.6 Comparison between conventional statistical tools and the developed tool

Geostatistical analysis was performed in ArcGIS. Comparative study of residual statistical parameters (as Fig. 5.6) shows that the developed tool has improved the prediction accuracy (in terms of RMSE) by *38.8 - 48.4%.* Factors considered by developed algorithm to improve the estimates are the outlier separation process, best theoretical model, elimination of negative weights, providing the optimum grid intervals for interpolation, etc.



SK-Simple Kriging, OK- Ordinary Kriging, UK_1-Universal Kriging Linear, UK_2-Universal Kriging Quadratic

Figure 5.6: Comparison of Model performance over conventional algorithm

5.5 Kakinada Power Plant (Case study - 2)

5.3.1 Site description

Kakinada region is located about 465 kilometers (289 mi) east of capital city of Andhra Pradesh. The study area (Fig. 5.7) is situated at16.93°N latitude 82.22°E longitude, and has an area of about 0.56 km^2. The average elevation across the study area is about 2 meters above mean sea level (AMSL). The region has approximately a north-south orientation and is confined to a long narrow strip parallel to the sea coast. The maximum temperatures in this region area about 38-to-42 °C (100-to-108 °F) and the minimum temperatures are about18-to-20 °C (64-to-68 °F). The region experiences an average annual rainfall between 110 and 115 centimetres.

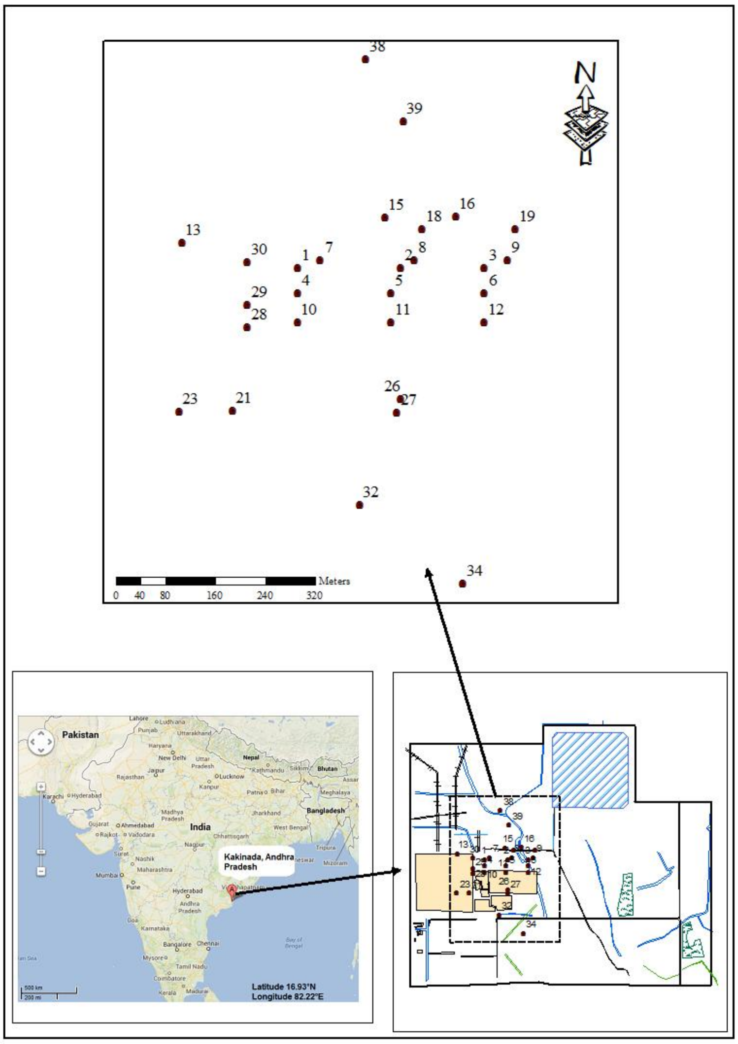


Figure 5.7: Schematic of Kakinada Power Plant with boreholes

A number of private power plants that supply electricity to the State's transmission utility were situated near to the study area. Bore log data of a major power plant project in Kakinada region of Andhra Pradesh was collected. A total of 39 borehole locations were made across the study area to conduct an extensive site investigation. However, data from only 28 boreholes contains information on the study parameter, and hence considered for the analysis.

5.2.2 Spatial outlier removal

For the present study, moisture content at 1m depth from borehole information was taken as the study parameter. There were about 28 moisture content data values ranging from 16-85% with a point density. About four spatial outliers (Fig. 5.8) were identified with a threshold value of 10% and were discarded from the analysis. A considerable increase in the point density from 61.96 /km^2 to 142.54 /km^2 was observed due to outlier separation.

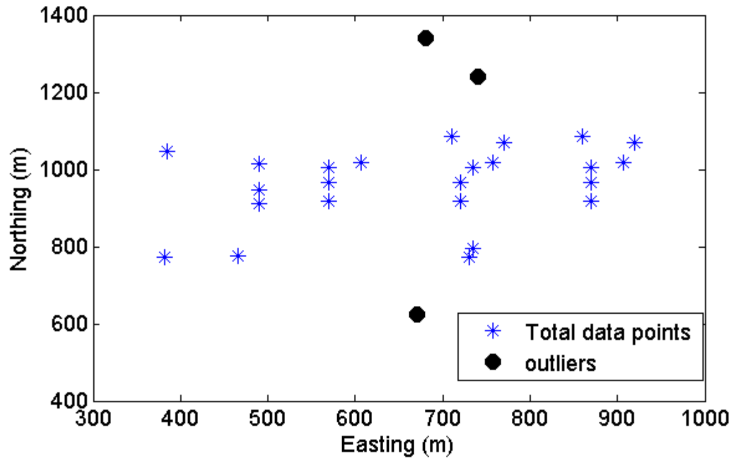


Figure 5.8: Spatial distribution of data considered in the analysis

5.2.3 Normality check

Results of the Kolmogorov-Smirnov test show that the data follows normal distribution at a significance level (α) of 5%.

5.2.4 Semivariogram and model fitting

It was observed that Exponential model best fits to the data before outlier separation, and spherical model fits after the outlier separation. The final spherical model has a sill and range of 856.57 m and 690.43 m respectively. The shape of the experimental semivariogram (Figure 5.9 (b)) followed the assumption in geostatistics defined by Tobler and is close to the ideal shape of semivariogram after outliers have been separated. No correlation between the lag distance / number of bins and the choice of best semivariogram was observed in the analysis.

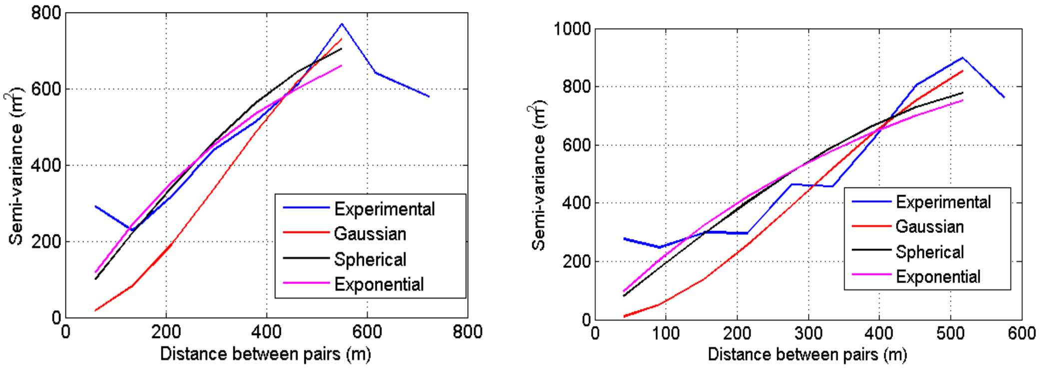


Figure 5.9: Comparison of experimental variogram with theoretical models before (left) and after (right) outlier separation

Figure 5.9.b shows the fitting of various theoretical models to the experimental model. The best semivariogram fitted with minimum MSE (as given in Table 5.5) was spherical model with sill and range values of 810.55 m and 620.14 m, respectively.

5.2.5 Evaluation of Kriging techniques

As the theoretical model was fixed, various linear kriging techniques were evaluated. Neighborhood parameters were varied to obtain the best neighborhood combination to predict the estimate at various grid locations (Table 5.5).

Table 5.5: Selection of optimal neighborhood parameters

	2 points(Min)			3 points(Min.)		
Limiting Radius (%max. distance between pairs)	3 points (Max)	4 points (Max)	5 points (Max)	4 points (Max)	5 points (Max)	6 points (Max)
15%	5.05	6.67	6.69	6.65	6.76	6.81
20%	4.28	6.22	6.32	7.19	7.26	7.49
25%	4.31	6.22	6.32	6.29	6.38	6.83
30%	4.31	6.3	6.44	6.3	6.44	6.89
35%	4.31	6.3	6.42	6.3	6.42	6.87
40%	4.31	6.3	6.42	6.3	6.42	6.86

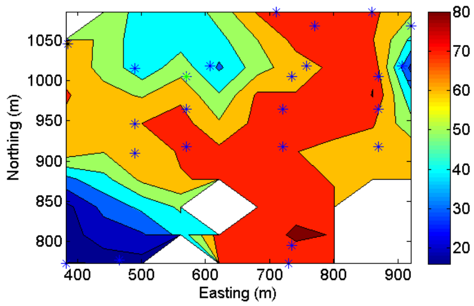
Table 5.6 shows that ordinary kriging is the most suitable kriging method for the parameter considered in the given case study with optimal neighbor combination of

minimum 2 and maximum 3 points. Four major outlier separation has reduced the residual statistics to minimum.

Table 5.6: Effect of outliers on Kriging simulations

Kriging Algorithm	RMSE (m)		Mean Error (m)	
	Before outlier separation	After outlier separation	Before outlier separation	After outlier separation
Simple kriging	10.51	7.97	-1	-0.48
Ordinary kriging	9.16	4.28	1.45	-0.06
Universal kriging (Linear)	10.56	5.05	1.25	-0.23
Universal kriging (Quadratic)	8.88	6.42	1.1	0.42

Prediction surfaces and error surfaces (Fig. 5.10) are generated for the moisture content across the entire region. A north-east downward trend in the moisture content was observed across the region.



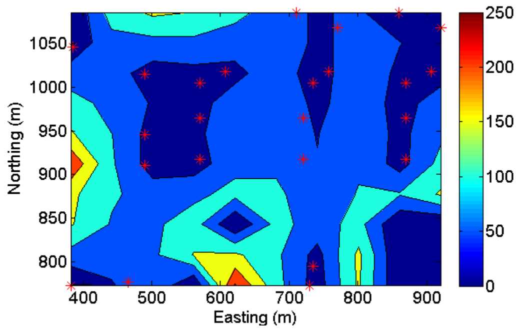


Figure 5.10: Model generated prediction (left) and error variances (right) surfaces

Predicted moisture content values are in convergence with the observed data at the known locations (Fig. 5.11).

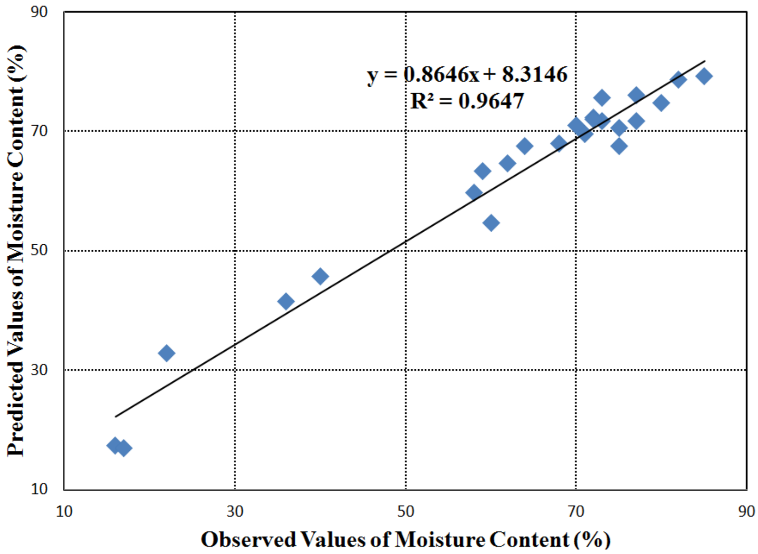
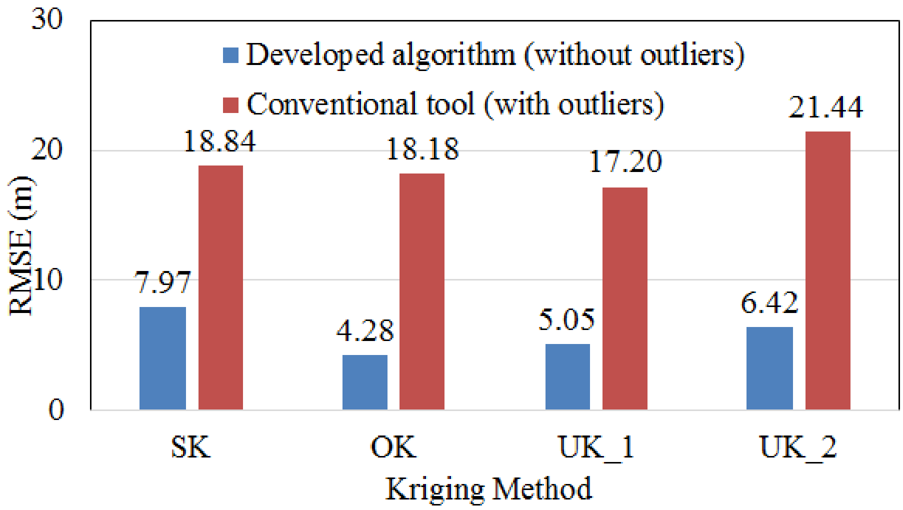


Figure 5.11: Comparison of observed and predicted moisture content values during cross validation

5.2.6 Comparison between conventional statistical tools and the developed tool

It was found from the results that the developed tool has improved the performance of prediction (evident from low RMSE values) by *57 – 76 %* compared to conventional tools (Fig. 5.12).

.



SK-Simple Kriging, OK- Ordinary Kriging, UK_1-Universal Kriging Linear, UK_2-Universal Kriging Quadratic

Figure 5.12: Comparison of Model performance over conventional algorithm

5.4 IIT Site, Hyderabad, Andhra Pradesh (Case study – 3)

5.4.1 Site description

Site for proposed campus of IIT Hyderabad (Fig. 5.13) is located about 61 km from Hyderabad on Hyderabad- Mumbai National Highway (NH 9). The site is located in Medak District of Andhra Pradesh, which is under Seismic Zone II as per IS 1893 (Part 1 -2002). Geographical area is about 543 acres(2.19 km^2). Total difference in level of about 10m exists within the site. Around 39 boreholes were constructed to have an extensive site investigation programme. Refusal strata (SPT value, N>50) was found at depth varying between 0.5 to 3m.Various soil stratum found were - 1) Reddish/ Whitish / Brownish/ Dark Grey Clayey Sandy Silt with/ without Gravels 2) Whitish /Reddish/ Brownish/ Dark Grey Clayey Silty Sand with/ without Gravels/ Intrusion of Lime 3) Whitish /Yellowish/ Greyish/ Reddish/ Brownish Grey Granite Based Weathered Rock 4) Whitish/ Greyish Brown Fissured and Fractured Granite Rock. Water table below the existing ground level varied from 14 to 17m.

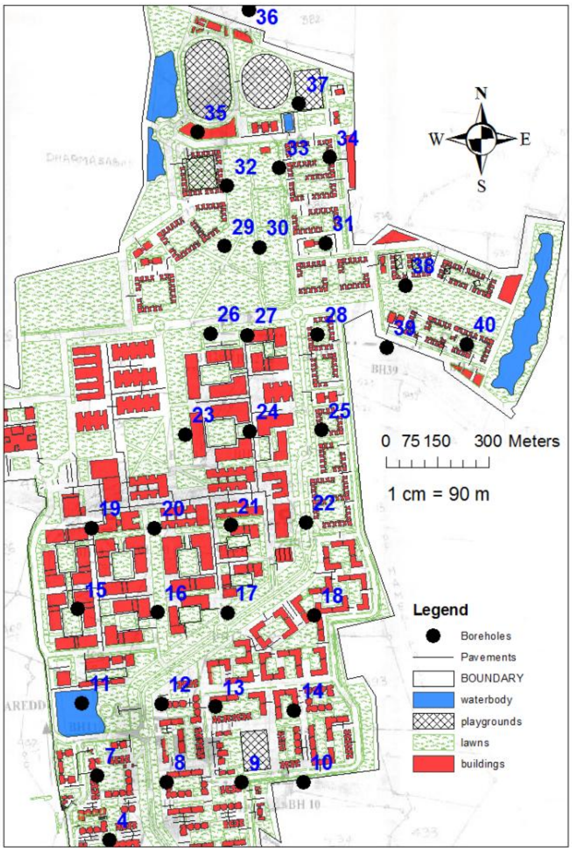


Figure 5.13: Schematic of study area with boreholes

5.4.2 Spatial outlier removal

Study parameter taken for the case study was sand content at 1m depth with values ranging from 0 to 72%. There were total 39 data values for the study parameter with a point density of 15.02 per km^2 of borehole coverage area. There was no outlier found during the outlier separation process (Fig. 5.14) .

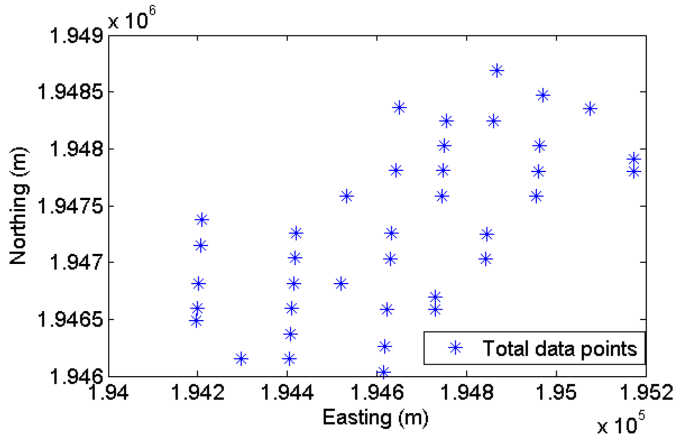


Figure 5.14: Spatial distribution of data considered in the analysis

5.4.3 Normality check

Kolmogorov-Smirnov test resulted that data was following normal distribution.

5.4.4 Semivariogram and model fitting

Experimental semivariogram was developed giving a lag divisions of 15 to display the increasing behavior of nature of semivariogram. Spherical model was fitted with a sill and range as 211.46 m^2, 1764.67 m respectively (Fig. 5.15).

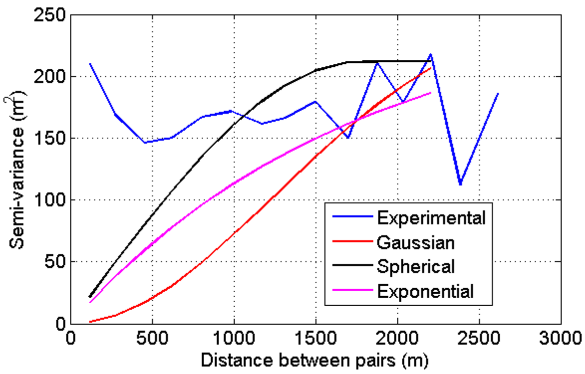


Figure: 5.15 Experimental variogram with theoretical models

5.4.5 Evaluation of kriging technique

The optimal method with minimum residual statistics obtained using algorithm (as tabulated in Table.5.7 and Table 5.8) was ordinary kriging technique with minimum neighbor 2 points and maximum neighbor 3 points.

Table 5.7: Selection of optimal neighborhood parameters

	2 points (Min)			3 points (Min.)		
Limiting Radius (%max. distance between pairs)	3 points (Max)	4 points (Max)	5 points (Max)	4 points (Max)	5 points (Max)	6 points (Max)
15%	6.97	7.59	8.05	7.59	8.07	8.7
20%	6.97	7.61	8.06	7.61	8.06	8.67
25%	6.97	7.61	8.06	7.61	8.06	8.67
30%	6.97	7.61	8.06	7.61	8.06	8.67
35%	6.97	7.61	8.06	7.61	8.06	8.67
40%	6.97	7.61	8.06	7.61	8.06	8.67

Table 5.8: Residual statistics obtained using developed algorithm

METHOD	RMSE (m)	ME(m)
Simple kriging	8.85	-0.14
Ordinary kriging	6.97	-0.44
Universal kriging(Linear)	8.53	-0.03
Universal kriging (Quadratic)	9.26	0.72

Prediction and Contour surfaces (Fig. 5.16) for the optimal method were generated to infer about the quality of prediction and infer about the additional sampling locations to improve the prediction. Sand content values are found to be highly concentrated on some portions of the study area.

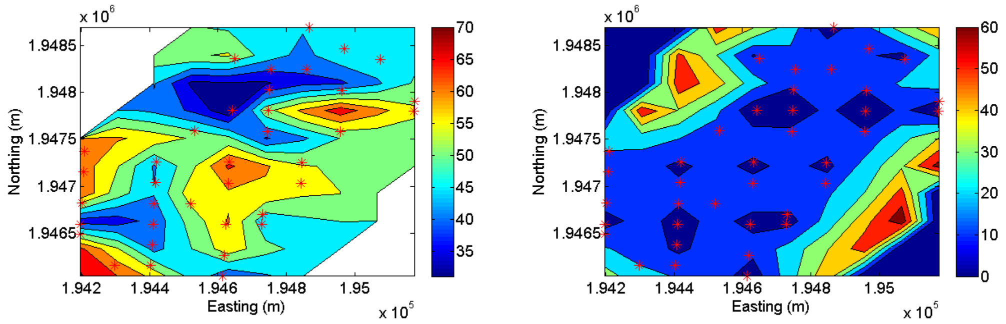


Figure: 5.16 Model generated prediction (left) and error variances (right) surfaces

Convergence of predicted sand content values with the observed data at the known locations can be seen in cross validation plot (Fig. 5.17).

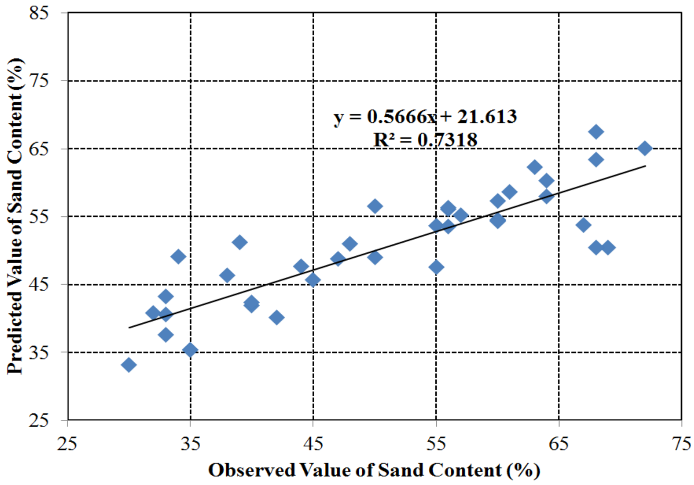
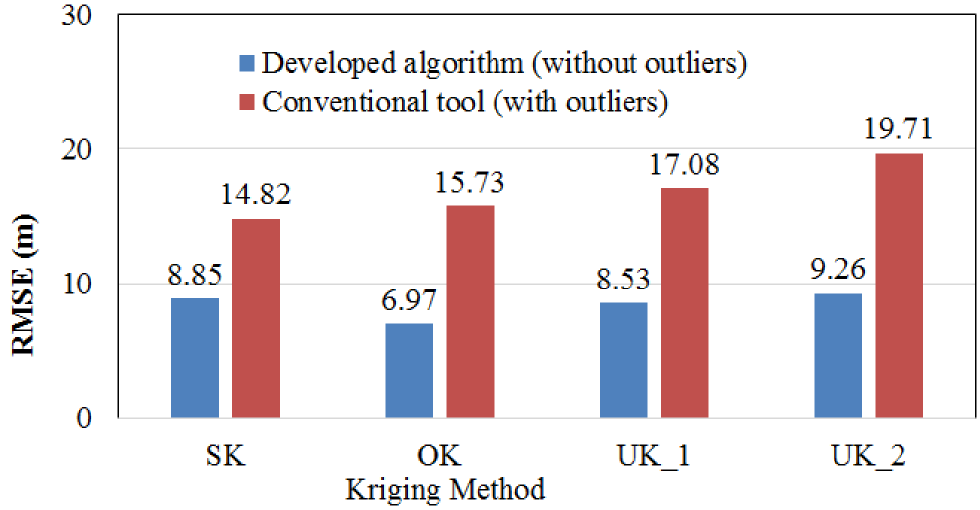


Figure 5.17. Comparison of observed and predicted moisture content values during cross validation

5.4.6 Comparison between conventional statistical tools and the developed tool

Comparison of results with conventional tool and developed algorithm is graphically represented in Fig 5.18.

Results prove that for sand content at 1m in IITH campus, the developed algorithm predicts more accurately with *40-50%* minimum residual statistics.



SK-Simple Kriging, OK- Ordinary Kriging, UK_1-Universal KrigingLinear, UK_2-Universal Kriging Quadratic

Figure 5.18. Comparison of Model performance over conventional algorithm

5.5 Summary

Three case studies were considered to evaluate the applicability of developed geostatistical algorithm. First case study taken was clay content at 3m depth in Paradip refinery project site, second case study was moisture content at 1m at the proposed Kakinada power plant and third case study was sand content at 1m depth in IIT Hyderabad permanent campus, In first two case studies it was found out that outlier separation is an important factor to be considered in the geostatistical analysis. One outlier was found in first case study and 4 outliers were found in second case study. Third case study doesn’t had any outlier. Normality test considered in the developed algorithm had overcome the difficulty in visual assessing normality behavior of parameters using Normal QQ plots by conventional tools. Automated selection of best fitted model to evaluate any kriging technique is not considered by conventional tools. For all three case studies, spherical model was found to be the best fitting model after outlier removal. As the program considers optimization of various factors and automation, huge amount of computational time was saved for finding out the best theoretical model, best neighborhood factors and best kriging methods based on cross validation results. Ordinary kriging was found to be the most suited kriging technique for three case studies with optimal neighborhood combination of 2 and 3 (minimum and maximum points). Residual statistics (RMSE) results has reduced about 38.8-48.4% in case study 1, 57-76% in case study 2 and 40-50% in case study 3 compared to conventional geostatistical tools. It was also concluded that the outlier separation has a significant effect in minimizing the RMSE values.

Chapter 6

Summary and Conclusions

6.1 Overview

The research was aimed at developing an automated, cost-efficient, generalized, user-friendly, public domain, and accurate linear geostatistical tool to apply in geotechnics. All the important linear geostatistical methods were considered in the algorithm. Most of the limitations of the conventional tools viz. hypothesis based normality check, separation of positional outliers, automated selection of base variogram and kriging method, and successive elimination of negative weights were overcome by the developed algorithm. Code also considers appropriate datum to geographic location of study area, and project it to Cartesian system by improving the accuracy of predictions. The developed code was tested for parameters collected from three regions in India, and evaluated using cross-validation and residual statistics. Hypothesis based normality test was done to select the most suitable parameter for each locations to evaluate linear geostatistics. The parameters showing strong semivariogram nature were- 1) Clay content at 3m depth of Paradip site; 2) Soil moisture content at 1m depth of Proposed Kakinada Power Plant; and 3) Sand content at 1m depth of IIT Hyderabad Kandi campus. After selecting the appropriate semivariogram model, kriging techniques are evaluated. Surface profiles and error surfaces were generated using the most suited kriging technique for these parameters.

6.2 Factors considered to improve the linear kriging estimates

1. *Spatial outlier separation based on simplified point density approach*

Conventional tools mostly do not consider the outlier separation as significant, though it can affect the predictions. Here, the simplified point density approach has been adopted to separate the outliers, thereby minimizing the residuals. Also, it was found from the Case Studies considered in the study that model fitting and cross validation are influenced by outliers.

2. Normality test based on hypothesis tests

Hypothesis test considered in the algorithm ensured normality of the data overcoming the difficulty of visual inspection of data as is done in conventional ArcGIS tools using smaller sample.

3. Automation in selection of best fit theoretical semi-variogram model

Conventional tools do not consider the optimization of various theoretical semi-variogram models for best-fitted model. The developed code considers least-squared nonlinear optimization of models reducing the effort and time of the user.

4. Automation in optimization of neighborhood factors

In conventional tools, various neighborhood factors have to be varied to select the best combination of neighborhood factors. Most important factors governing the searching neighborhood criterion, such as limiting radius of the ellipse, and minimum and maximum number of neighbors have been varied iteratively in the developed code resulting in best neighborhood factors.

5. Automation in optimization of best linear kriging technique

In conventional geostatistical tools selection of best linear kriging technique is circular in nature and the user spends most of the time in varying the factors, such as, the best theoretical semivariogram selection, varying neighborhood factors for testing the method which gives the best minimal residual statistical parameters. In the developed algorithm, this tedious procedure has been reduced and the program will provide the user with best linear kriging technique.

6. Oriented towards geotechnical applications as per ASTM requirements

Most of the conventional tools are oriented towards large mining programs [Hammah, 2004]. For geotechnical engineer to have an accurate site characterization, necessary steps to generate surface profiles based on principle of geostatistics are sufficient. The tool developed in this research was based on this objective. ASTM D 5923-96 suggests various important factors to apply kriging techniques. Code recommends that linear geostatistical techniques should be applied only when the soil data passes normality. In other cases, nonlinear geostatistical techniques should be applied. Code suggests that if very few spatial outliers are present, then one can go with ordinary kriging technique. If large amount of spatial outliers are present, then nonlinear kriging techniques are to be adopted.

Ordinary kriging is the appropriate estimation method, if the mean is assumed to be constant but is unknown. Simple kriging is the appropriate estimation method, if the mean is presumed to be known. If kriging variance is used to quantify uncertainty then ordinary or simple kriging can be chosen as appropriate estimation methods. Ordinary kriging is default kriging method suitable for the soil parameters. The developed tool resulted with the best kriging method for all case studies as the ordinary kriging technique. If a drift is present in the data, then universal kriging is an appropriate estimation method. ASTM also recommends that if drift can be accommodated by modifying the search neighborhood configuration, then ordinary kriging is the best method. As evaluation results of all case studies resulted in best kriging technique as ordinary kriging, statement is agreed.

7. *Fixing the influencing factors such as lag divisions, grid size, etc.*

It was found from the case studies that lag divisions have no effect on prediction estimates. This factor has only differences when plotting the semivariogram. Considering this factor lag divisions has been chosen to be equal to 15. Another minor factor concerning the prediction estimate is the grid size. It was found that higher the grid divisions, the higher is the accuracy of the predictions. The code has considered an optimal number of divisions equal to 10, reducing the computational time. Hence, these two factors have been chosen to be optimal reducing the effort for user and hence computational time compared to conventional tool approach.

The developed algorithm was found to be effective in generating the prediction and error variance surfaces with a decrease in RMSE values of kriging methods about *38.8-48.4%* in case study 1, *57-76%* in case study 2 and *40-50%* in case study 3 over conventional ArcGIS tools.

6.3 Limitations of the study

During the study it was very hard to get enough number of data values at various depths. It would have been more useful when enough data was obtained. The study was limited only to linear geostatistics as the modelling of nonlinear geostatistics is more involved. Only very few parameters of various case studies including soil type, Atterberg limits has shown the normality behaviour. For site characterisation, many other important soil parameters are required, especially the soil strength parameters. These parameters were not enough to consider for a linear geostatistics. Nugget effect has not considered for the fitting of semivariogram. The solutions of kriging weights were based on the least square solution as the study was aimed for an optimised geostatistical development.

6.4 Recommendations

This study was limited only to linear geostatistics. As most of the parameters governing soil strength have exhibited non-normal distributions, it will be better to evaluate nonlinear geostatistics. Case studies having data values at a minimum of 30 locations should be used to have an accurate check for normality. Also the study considered only lateral correlation of the soil parameter. Vertical correlation studies has to be done to model the 3D variability of the soil parameter. A web based GUI should be created so that user across the globe can easily access the application at any place.

References

[1] ASCE Task Committee. Review of geostatistics in geohydrology, 1: Basic concepts. J. Hydraul. Eng. ASCE 116 (5),(1990) 612–632.

[2] ASTM D 5923-96. Standard guide for selection of Kriging Methods in Geostatistical Site Investigations. ASTM International 4 (9), (2010) 1-4.

[3] E. Asa, M. Saafi, J. Membah, and A. Billa. Comparison of Linear and Nonlinear Kriging Methods for Characterization and Interpolation of Soil Data. J. Comput. Civ. Eng., 26 (1), (2012) 11-18.

[4] R. Blanchin and J. P Chiles.The channel tunnel: Geostatistical of the Geological Conditions and its Validation by the reality. Mathematical geology 25 (7), (1993) 963-974

[5] F. Canters. Small scale map projection design. Taylor and Francis, London, (2002) 343.

[6] J. R. Carr and C. E. Glass Treatment of earthquake ground motion using regionalized variables. Math. Geol. 1 , (1985) 221–241.

[7] J. R. Carr and N. Mao, Comparison of disjunctive kriging to generalised probabibility kriging in application to the estimation of simulated and real data.Quantitative Geology and Geostatistics 5. Geostatistics Troia. (1992) 13-24.

[8] J. R. Carr and K. P. Roberts. Application of universal kriging for estimation of earthquake ground motion: Statistical significance of results. Mathematical Geology ,21(2), (1989), 255-265.

[9] J. P. Chiles and P. Delfiner. Geostatistics: Modeling spatial uncertainty. Wiley series in probability and statistics. Wiley. New York, (1999) 687.

[10] Clark, I., and Harper, W. Practical geostatistics 2000. Ecosse. North America Llc, (2002)

[11] N. A. C. Cressie. Statistics for spatial data, Wiley, New York, (1991) 900.

[12] J. P. Delhomme. Spatial variability and uncertainty in groundwater flow parameters: A geostatistical approach. Water Resources Research, 15 (2), (1979) 269–280.

[13] C. V. Deutsch and A. G. Journel. GSLIB. Geostatistical software library and user’s guide. Oxford University Press, New York, (1998) 369.

[14] G. Exadaktylos, M. Stavropoulou, G. Xiroudakis, M. de Broissia, and H. Schwarz. A spatial estimation model for continuous rock mass characterization from the specific energy of a TBM, Rock Mech. Rock Eng. 41 (6), (2008) 797-834.

[15] G. A. Fenton. Data analysis/geostatistics. Probabilistic methods in geotechnical engineering. ASCE Geotechnical Safety and Reliability Committee, (1997).

[16] A. S. Goldberger. Best linear unbiased prediction in the generalized linear regression model. J. Am. Stat. Assoc. 57(298), (1962) 369–375.

[17] P. Goovaerts. Geostatistics for natural resources evaluation. Oxford University Press. New York, (1997) 483 .

[18] K. S. Gundogdu. and I Guney. Spatial analyses of groundwater levels using universal kriging. J. Earth Syst. Sci., 116 (1), (2007) 49–55.

[19] T. M. Hansen. mGstat, Ed. 0.97, A geostatistical Matlab toolbox, Lyngby, Denmark, (2010).

[20] R. E. Hammah and J.H. Curran. Geostatistics in Geotechnical Engineering: Fad or Empowering?. GeoCongress.ASCE, (2006) 1-5.

[21] D. M. Hawkins. Identification of Outliers. Biometrical Journal, 29 (2), (1980) 198.

[22] Y. Honjo. and K. Kuroda. A new look at fluctuating geotechnical data for reliability design, soils and foundations. Jpn. Soc. Soil Mech. Found. Eng 31 (1), (1991) 110–120.

[23] E. H. Isaaks and R. M. Srivastava. An introduction to applied geostatistics, Oxford University Press, New York, (1989) 561.

[24] M. B. Jaksa, P. I. Brooker and W. S. Kaggwa. Modelling the Spatial Variability of the Undrained Shear Strength of Clay Soils Using Geostatistics. Proceedings of 5th Int. Geostatistics Congress. Wollongong, (1997) 1284-1295.

[25] K. Johnston, J. M. Ver Hoef, K. Krivoruchko and N. Lucas. Using ArcGIS-geostatistical analyst. GIS by ESRI. Redlands. USA, (2001).

[26] A. G. Journel. Fundamentals of geostatistics in five lessons, American Geophysical Union, Washington, DC, (1989).

[27] A. G. Journel and C. J. Huijbregts. Mining geostatistics. Academic. New York, (1978).

[28] P. B. V. N. Kambhammettu, P. Allena, and J. P. King. Application and evaluation of Universal Kriging for Optimal Contouring of Groundwater levels. J. Earth Syst. Sci. 120 (3), (2011) 413-422.

[29] P. K. Kitanidis. Introduction to geostatistics: Applications in hydrogeology. Cambridge University Press. New York, (1997) 249.

[30] D. G. Krige. A statistical approach to some basic mine valuation problems on the witwatersrand. J. Chem. Metall. Min. Soc. South Africa, 52 (6), (1951) 119–139.

[31] P. H. S. W. Kulatilake, and K. M. Miller. A scheme for estimating the spatial variation of soil properties in three dimensions. Proc. 5th Int. Conf. on Application of Statistics and Probabilities in Soil and Struct. Eng., (1987) 669–677.

[32] V. Kumar. Optimal contour mapping of groundwater levels using universal kriging – a case study. J. Hydrol. Sci. 52 (5), (2007) 1039–1049.

[33] J. A. Lenz, and L. G. Baise. Spatial variability of liquefaction potential in regional mapping using CPT and SPT data. Soil Dyn. Earthquake Eng. 27 (7), (2007) 690–702.

[34] G. Matheron. Principles of geostatistics. Econ. Geol. 58 (8), (1963) 1246–1266.

[35] K.K Phoon. and F.H. Kulhawy. Characterization of geotechnical variability. Can. Geotech. J. 36 (4), (1999) 612–624.

[36] T. P. Robinson and G. Metternicht. Testing the performance of spatial interpolation techniques for mapping soil properties. Comput. Electron. Agric. 50 (2), (2006). 97–108.

[37] S. Rouhani, R.M. Srivastava, A.J. Desbarats, M.V. Cromer, and A.I. Johnson (eds.). STP 1283 Geostatistics for Environmental and Geotechnical Applications. ASTM. Philadelphia, (1996) 3-31;102-116

[38] P. Samui and T. Sitharam. Site Characterization Model Using Artficial Neural Neural Network and Kriging. Int. J. Geomech.10 (5), (2010) 171-180.

[39] R. R. Sokal, and J. F. Rohlf. Biometry: The principals and practice of statistics in biological research. W. H. Freeman and Co. San Francisco. California, (1969) 776.

[40] M. Soulie, P. Montes, and V. Sivestri. Modelling spatial variability of soil parameters. Can. Geotech. J. 27, (1990) 617–630.

[41] W. Tobler. A computer movie simulating urban growth in the Detroit region. Economic Geography 46, (1970) 234-240.

[42] M. Uzielli, G. Vannucchi, and K.K. Phoon. Random field characterization of stress-normalised cone penetration testing parameters. Geotechnique. 55 (1), (2005) 3–20.

[43] P. Vennapusa, D. J. White, and M. Morris. Geostatistical analysis for spatially referenced roller-integrated compaction measurements. J. of Geotech. and Geoenv. Eng. 136 (6). ASCE, (2010) 813-822.
[44] R. Webster, and M. A. Oliver. Geostatistics for environmental scientists. 2nd Ed. Wiley. West Sussex. England, (2007) 295.