

Statistical Estimation of Soil Parameters using CPT Data

Manish Kumar

Department of Civil, Structural and Environmental Engineering, 212 Ketter Hall, University at Buffalo, State University of New York, Buffalo, NY 14260, U.S.A.

ABSTRACT

Uncertainties in soil properties are characterized using probabilistic analysis. Soil models are required for reliability analysis of soil at a site. Parameters of a soil model, which includes mean, variance, and correlation structure, are estimated using experimental data. The obtained soil models can be used for reliability analysis at another site that has similar geological properties. The soil properties are known to be spatially correlated. The finite-scale model is one of the two model types used for probabilistic modeling of soil (other being the fractal models), and is used here. The assumptions used for application of finite-scale models are investigated. Seven analysis cases are explored by implementing two algorithms to determine the soil parameters. The difference in results obtained using different analysis cases and algorithm are investigated.

KEY WORDS: Random field, finite scale, first and second order, statistical, soil parameters

1. INTRODUCTION

The reliability analysis of geotechnical projects require random soil models to assess probabilities related to design. These soil models are required to predict the spatial variability of soil based on available field tests. Although spatial variability is observed in three-dimension, one-dimensional analysis is performed to obtain reasonable models for variation in soil properties along a line. The goal is to establish soil models using soil data at a site to make predictions about the variation in soil properties at a different site having similar geological formation.

Two types of stochastic models are used to estimate the spatially varying soil properties: a) finite-scale models, and b) fractal models. The finite scale models are also called short memory models, and are used for the soils where the correlation between soil properties at different locations dies out very rapidly over the distance. On the contrary, fractal models are used to model soils that have significant correlation over very large distances. Fenton [1] discusses different methodologies such as the periodogram, wavelet variance, and semivariogram plots to determine the soil model that should be used to model soil properties. Once the type of stochastic model has been determined, a correlation or covariance function should be selected to represent the stochastic model. For finite-scale models, usually one parameter, the scale of fluctuation known as correlation or covariance parameter, needs to be estimated from available data set to completely characterize a stationary stochastic model. The first order parameters of soil properties include mean and variance, and the second order parameter is the correlation parameter to characterize the correlation structure of soil properties. The estimation of these three parameters provides the sufficient information on first- and second-order parameters for random file modeling of soil properties. Soil properties such as cone penetration resistance, shear strength, pore pressure can be modeled using the procedure discussed here.

The cone tip resistance (q_c) data available on [2] were modeled using a finite scale random field stochastic process to estimate the first- and second-order parameters. The cone penetration test (CPT) soundings were obtained at 25 locations at the site, and at each location q_c values were obtained along the depth. Only the 18ft depth of cone penetration test (CPT) soundings were used. The CPT data can be represented by a 108×25 Q_c matrix, where each column represents the q_c observations at one of the 25 locations and rows represent q_c values at certain depth from ground surface. The Q_c matrix is presented in Appendix C for reference. The soil parameters are determined using a column vector that might be: 1) columns of Q_c matrix, or 2) a mean of 25 columns of Q_c matrix. Both options are explored and their effects on final results are discussed. The two algorithms discussed in Fenton [1] are implemented here for obtaining the soil parameters.

2. SPATIAL VARIATION

Consider a site where soil properties have been obtained at different locations over an area domain. Any observation, $S_u(z)$, along the depth would exhibit a deterministic trend, $m(z)$, and a random residual, $\varepsilon(z)$, as shown in Figure 1.

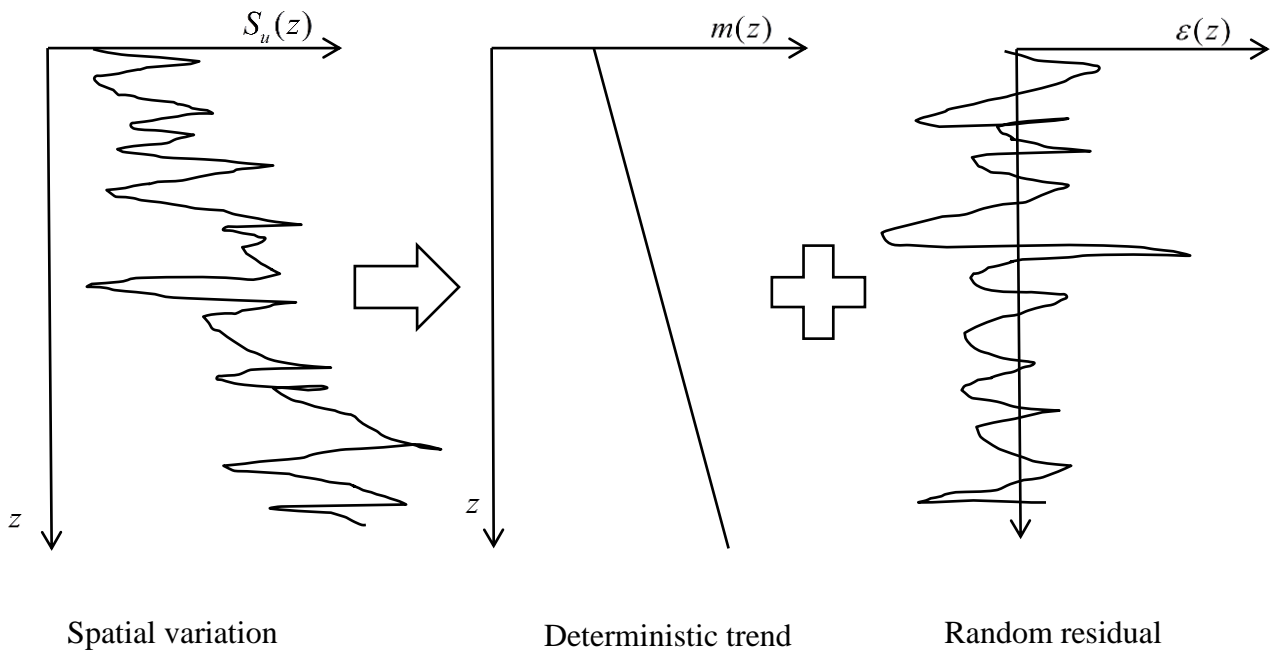


Figure 1: The spatial variation of soil properties

Random field modeling of a soil property can be performed using the $S_u(z)$ or the residual $\varepsilon(z)$ after extracting trend from $S_u(z)$. The trend should be removed from the spatial variability $S_u(z)$ only if a similar trend is expected on the target site; otherwise, the variability in trend should be included in uncertainty analysis. Removing the trend for analysis of spatial soil variability at a site where a similar trend is not expected, might lead to unconservative results. The main aim of detrending the data is to render the random residual spatially independent, which is desirable because all the statistical calculations are based on the assumption that samples consist of independent and identical observations.

3. CORRELATION FUNCTION

The mean, μ_x , and variance, σ_x^2 , at the target sites are obtained using preliminary information gathered for the design and then established soil models at a similar site is used to complete the uncertainty picture of a soil property. This information is then used in the reliability analysis of soil. An assumption of spatially statistically homogeneous soil, which means that mean, covariance, correlation structure, and higher-order moments are independent of position, is assumed for the statistical analysis.

Two types of stochastic soil models are used for the soil properties: a) finite-scale models, and b) fractal models. The finite scale models are also called short memory models, and are used for the soils where the correlation between soil properties at different locations dies out very rapidly over the distance. On the contrary, fractal models are used to model soils that have significant correlation over very large distances. A finite scale model called 1D Markov correlation model is used here. The 1D Markov model has exponentially decaying formulation given by:

$$\rho(z_i, z_j) = \exp\left(-\frac{|z_i - z_j|}{b}\right) \quad (1)$$

where z_i and z_j are location of two observations, and b is the scale of correlation beyond which soil properties are largely uncorrelated. Once the parameter b is obtained, the correlation structure of the soil property can be obtained.

4. IMPLEMENTATION

A statistical distribution for the observations $X(z)$ is assumed, which is used with chosen correlation function model (1D Markov model) is to estimate the unknown parameters (μ_x, σ_x^2, b) using maximum likelihood approach in the space domain.

Soil properties are strictly nonnegative, and are typically assumed to be of lognormal distribution. It is converted to normal distribution by simply taking the natural log of the soil data before any statistical analysis can be performed. Normal distribution is preferred because it can be completely characterized by information on only first two moments, mean and variation.

The space domain maximum likelihood estimation is obtained by maximizing the likelihood of observing the spatial data under the assumed distribution. The likelihood of observing the sequence of observations $x^T = \{x_1, x_2, \dots, x_n\}$ given the distribution parameters $\phi^T = \{\mu_x, \sigma_x^2, b\}$ is:

$$L(x|\phi) = \frac{1}{(2\pi)^{n/2} |C|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)^T C^{-1}(x-\mu)\right\} \quad (2)$$

where C is the covariance matrix between observations, and $|C|$ is the determinant of C . The data x is assumed to be stationary, i.e, the mean is spatially constant.

For spatially constant random field, the relation between covariance matrix and correlation matrix can be written as:

$$C = \sigma_x^2 \rho \quad (3)$$

The maximum likelihood function $L(x|\phi)$ is maximized with respect to ϕ to determine unknown parameters. Because $L(x|\phi)$ is nonnegative, taking natural log of Eqn (2), the log likelihood function is obtained as:

$$L(x|\phi) = -\frac{n}{2} \ln \sigma_x^2 - \frac{1}{2} \ln |\rho| - \frac{(x-\mu)^T \rho^{-1} (x-\mu)}{2\sigma_x^2} \quad (4)$$

Differentiating it with respect to the mean, μ_x , and equating to zero, value of μ_x is obtained as:

$$\mu_x = \frac{1^T \rho^{-1} x}{1^T \rho^{-1} 1}$$

Let us assume r and s are the solutions of:

$$\rho r = x$$

$$\rho s = 1$$

The mean, μ_x , is then given as:

$$\mu_x = \frac{1^T r}{1^T s}$$

Similarly equating partial derivative of $L(x|\phi)$ with respect to σ_x^2 equal to zero:

$$\sigma_x^2 = \frac{1}{n} (x - \mu_x 1)^T r$$

Once the mean and variance can be expressed in terms of the unknown correlation parameter b , substituting these values in $L(x|\phi)$, after ignoring constant terms, problem reduces to finding maximum of:

$$L(x|\phi) = -\frac{n}{2} \ln \sigma_x^2 - \frac{1}{2} \ln |\rho| \quad (5)$$

Fenton [1] discussed two algorithms to obtain soil parameters which are presented in the following sections.

4.1 Algorithm 1

The first algorithm uses an iterative algorithm to obtain the correlation parameter, along with mean and variance of soil data. The steps are:

1. Guess an initial value for b .
2. Obtain the correlation matrix ρ using relationship using $\rho(z_i, z_j) = \exp(-|z_i - z_j|/b)$

3. Obtain the value of vector r and s
4. Calculate the value of the determinant $|\rho|$
5. Compute the mean and variance
6. Compute the log-likelihood value $L(x|\phi)$
7. Guess a new value of b and repeat the steps 1 to 7 until a global maximum value of $L(x|\phi)$ is obtained.

An initial guess of the correlation length b can be obtained as a value at which $\rho(z_i, z_j)$ decays to $1/e$ of the initial value, or $b \sim |z_i - z_j|$.

A MATLAB [3] code developed to implement this algorithm is presented in Appendix A.

4.2 Algorithm 2

The second algorithm is applicable for a special case where observations are equispaced. If the correlation matrix elements are $q(z_i, z_j) = \exp(-|z_i - z_j|/b)$, or $q = \exp(-|\Delta z|/b)$ for equispaced observations, the determinant and inverse of correlation matrix has closed forms given by:

$$|\rho| = (1 - q^2)^{n-1} \quad (6)$$

$$\rho^{-1} = \left(\frac{1}{1 - q^2} \right) \begin{bmatrix} 1 & -q & 0 & 0 & \cdots & 0 & 0 \\ -q & 1 + q^2 & -q & -q & \cdots & 0 & 0 \\ 0 & -q & 1 + q^2 & -q & \cdots & 0 & 0 \\ 0 & 0 & -q & 1 + q^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 + q^2 & -q \\ 0 & 0 & 0 & 0 & \cdots & -q & 1 \end{bmatrix} \quad (7)$$

The maximum likelihood estimation of q using Eqn. (5) reduces to finding the root of the following equation on the interval $q \in (0, 1)$:

$$f(q) = b_0 + b_1 q + b_2 q^2 + b_3 q^3 = 0 \quad (8)$$

where the coefficients are:

$$b_0 = nR_1; b_1 = -(R_0 + nR_0'); b_2 = -(n-2)R_1 \quad (9)$$

$$b_3 = (n-1)R_0'; R_0 = \sum_{i=1}^n (x_i - \hat{\mu}_x)^2 \quad (10)$$

$$R_0' = R_0 - (x_1 - \hat{\mu}_x)^2 - (x_n - \hat{\mu}_x)^2 \quad (11)$$

$$R_1 = \sum_{i=1}^{n-1} (x_i - \hat{\mu}_x)(x_{i+1} - \hat{\mu}_x) \quad (12)$$

Solution of Eqn. (8) can be obtained using Newton-Raphson method. Initial estimate of q can be taken as 0.5. For an obtained value of q from Eqn. (8), the corresponding maximum likelihood estimates of μ_x and σ_x^2 are:

$$\hat{\mu}_x = \frac{Q_n - q(Q_n + Q_n') + q^2 Q_n'}{n - 2q(n-1) + q^2(n-2)} \quad (13)$$

$$\hat{\sigma}_x^2 = \frac{R_0 - 2qR_1 + q^2 R_0'}{n(1 - q^2)} \quad (14)$$

where

$$Q_n = \sum_{i=1}^n x_i; Q_n' = Q_n - x_1 - x_n \quad (15)$$

A MATLAB [3] code developed to implement this algorithm is presented in Appendix B.

5. RESULTS

The cone penetration resistance can be used as an indicator of soil strength. The strength of a soil increases with the depth due to confinement effects. It is expected that the trend of the cone penetration resistance should linearly increase with the depth, however if the soil strata consists layers of different soil types (ex., sand, clay), then increasing trend might not be obtained. The trend and residual obtained for the log of the cone penetration resistance (q_c) values at locations 1 and 25 are shown in Figure 2.

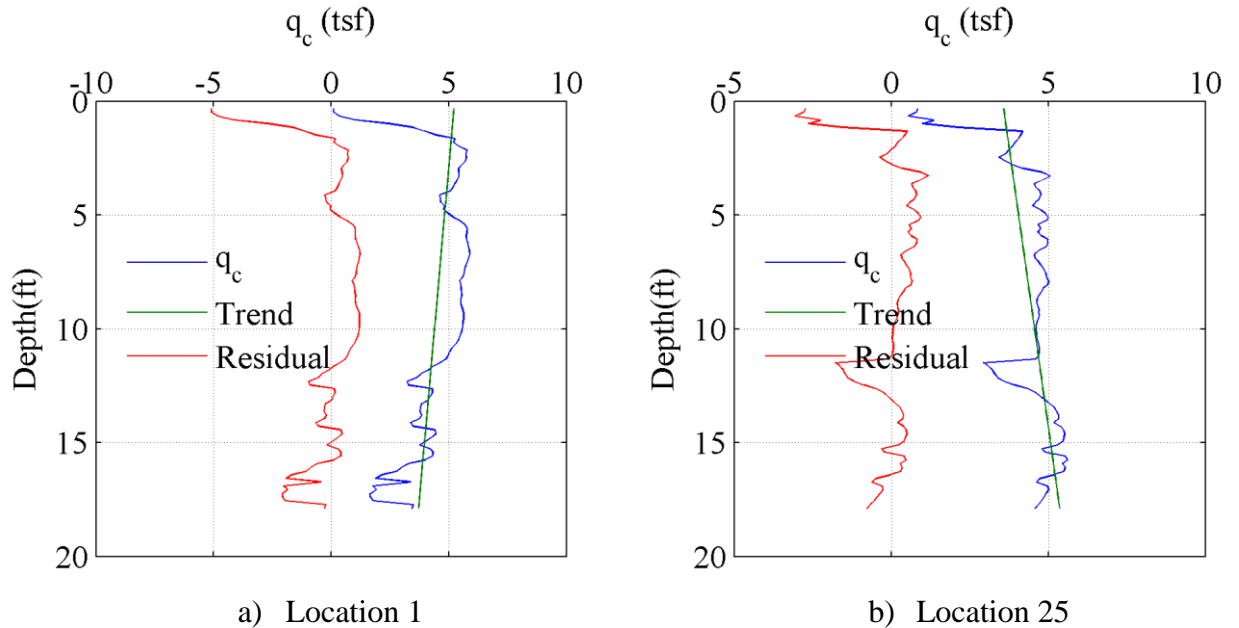


Figure 2: Spatial variation of cone penetration resistance

While the spatial trend of q_c at location 25 follows the expected behavior, it decreases with depth at location 1. An appropriate way to address such problem would be to divide the total depth in different layers of thickness corresponding to each soil type. A constant trend may also be used, however that would increase variance in the residuals. For the statistical analysis performed here, trend as obtained from linear regression analysis is used, even if the trend does not follow the expected behavior.

The soil data is modeled using lognormal distribution, which is converted to Gaussian distribution by taking natural log of data before using it for first- and second-order parameter estimation. The assumption of Lognormal and Gaussian distribution is investigated here for few sample cases. The histograms and associated distributions are plotted for: 1) q_c observations at depths $z = 0.33$ ft and 17.88 ft obtained from all 25 site locations, 2) q_c observations along the depth at locations 1 and 25, and 3) q_c observations obtained for the whole site.

Histograms and lognormal fits to the CPT data at $z = 0.33$ ft and 17.88 ft are presented in Figure 3.

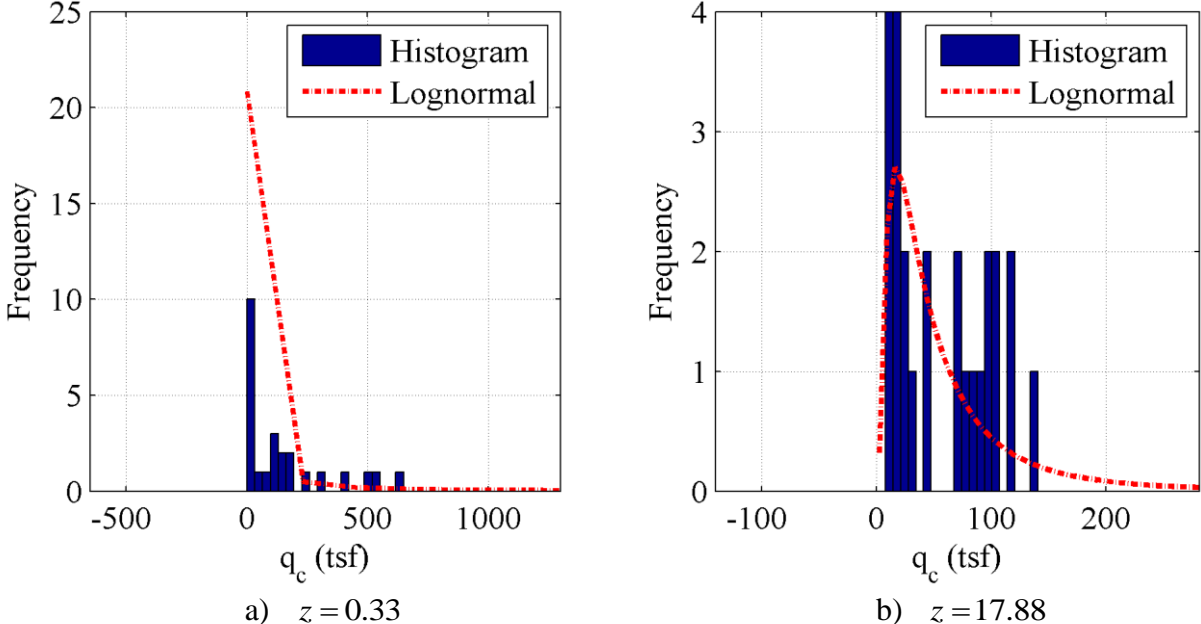


Figure 3: Histogram and Lognormal fit of CPT values

The corresponding histogram and normal distribution after taking natural log of q_c data at $z = 0.33$ ft and 17.88 ft are presented in Figure 4.

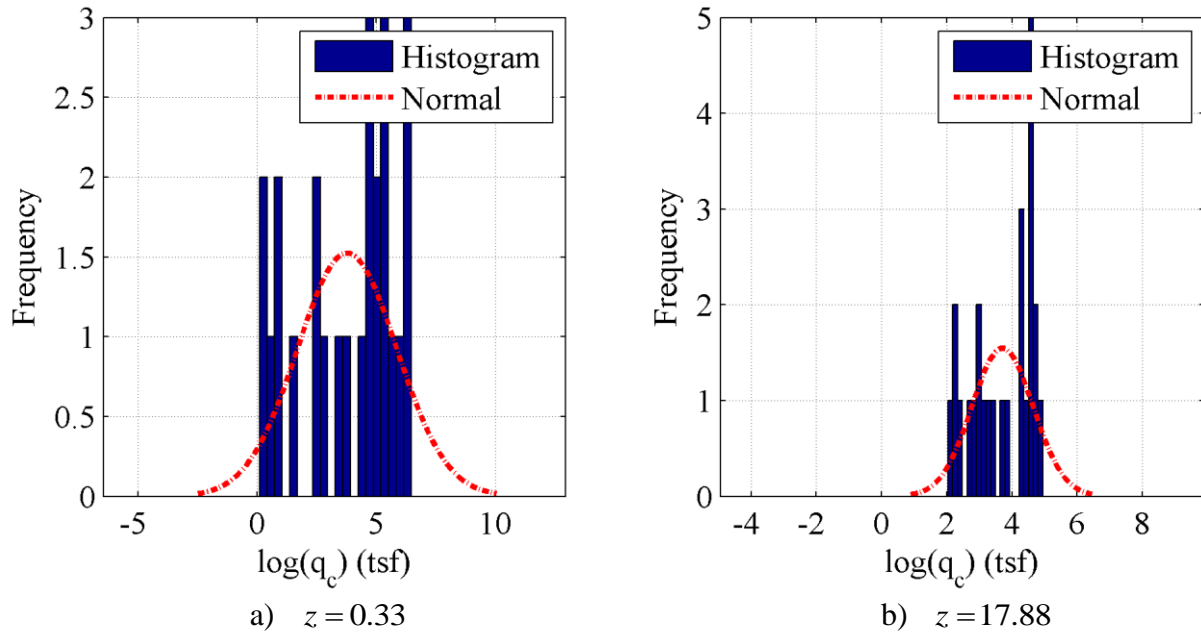


Figure 4: Histogram and Normal fit of log CPT values

The histograms and lognormal fits of CPT observations along the depth at locations 1 and 25 are presented in Figure 5.

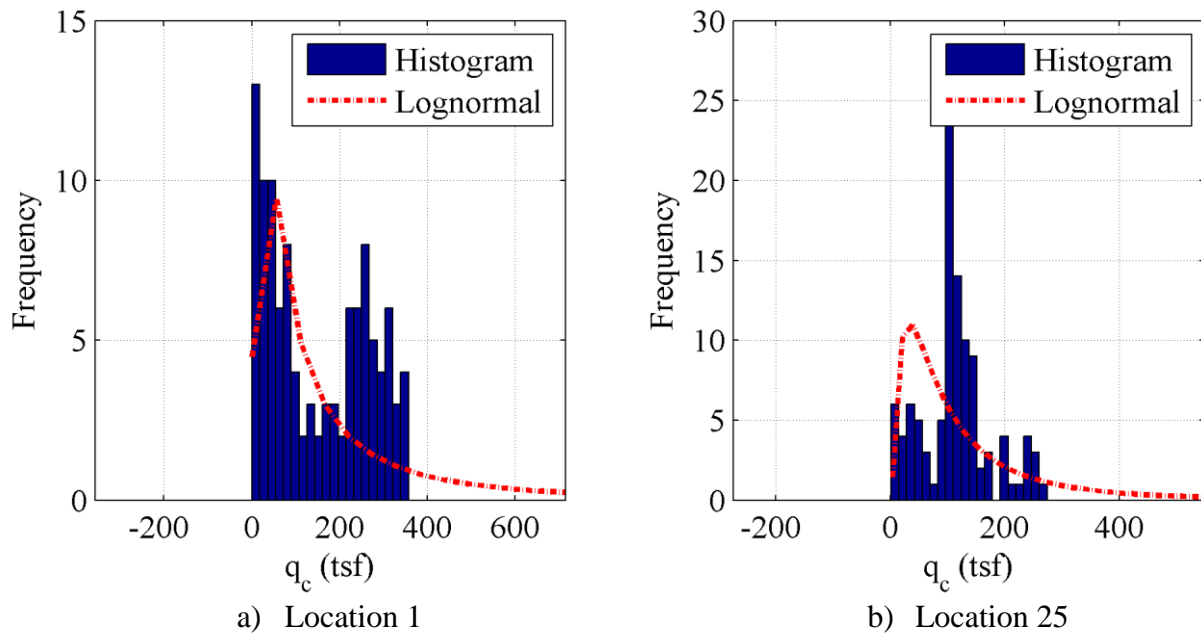
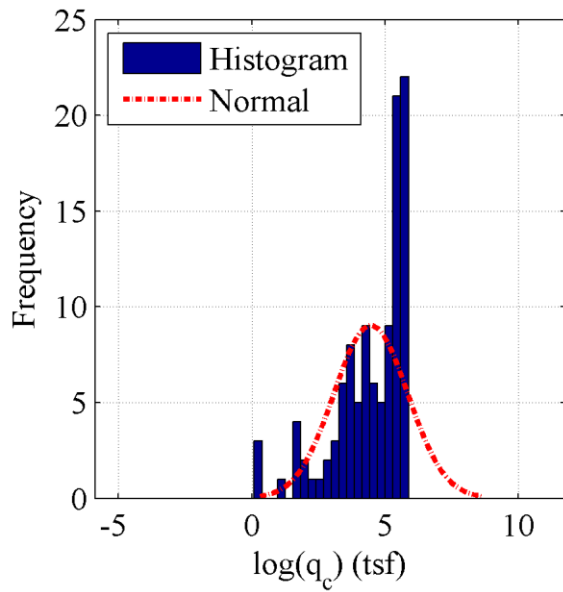
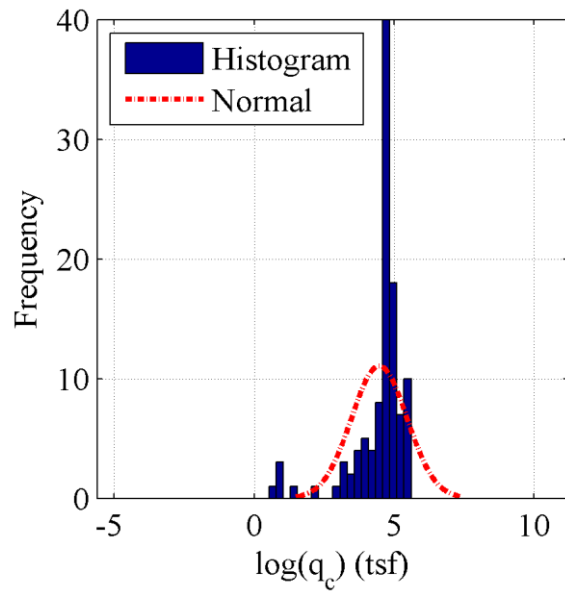


Figure 5: Histogram and Normal fit of CPT values

The histograms and lognormal fits of the log of CPT observations along the depth at locations 1 and 25 are presented in Figure 6.



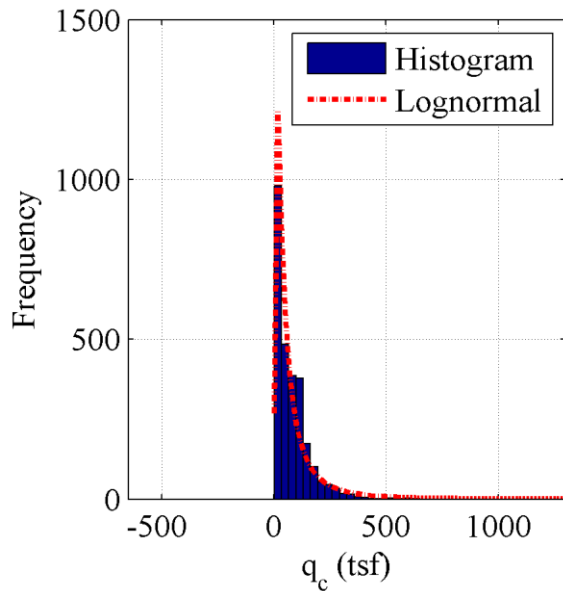
a) Location 1



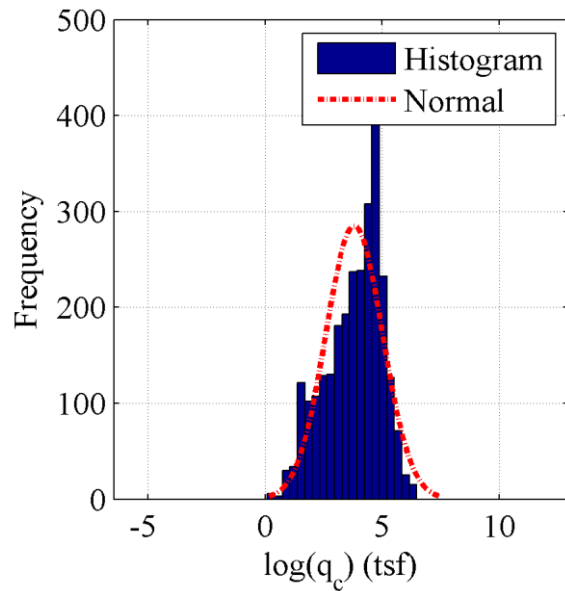
b) Location 25

Figure 6: Histogram and Normal fit of log CPT values

Histograms and Lognormal and Normal fit of CPT observations from the whole site is shown in Figure 7.



a) Cartesian space



b) Log-space

Figure 7: Histogram and statistical distribution fit of CPT data for whole site

The histograms and statistical fits developed for CPT data shows that lognormal fit is an appropriate distribution. The log of CPT data is used for further statistical analyses.

The CPT data were obtained at 25 locations, and each location had 108 observations along the depth were obtained. Although the soil properties vary in three-dimension, the goal here was to estimate first- and second-order soil parameters using a statistical analysis in one dimension along the depth. The statistical parameter estimation algorithms discussed in previous section require a x vector consisting of CPT observations along the depth. The vector x can be: 1) observation at each site location, or 2) mean of observations obtained from all the 25 locations. For each option statistical analysis can be performed using 1) spatial variability of CPT data, or 2) residual extracted from the spatial variability of the CPT data. If the residual of spatial variability is used, trend in CPT data along the depth needs to be calculated using linear regression analysis. For global averaged data, global trend is used, but for local data trend can be calculated as: 1) local trend using data at each location, or 2) a global trend using mean of the observations obtained from all 25 locations. Different analysis cases to determine soil parameters are presented in Figure 8. A special analysis case for CPT observations obtained at equidistant points was also considered for data at each location. The implementation of the special case is discussed in Algorithm 2 in the previous section. Although the CPT data observations were not obtained exactly at equispaced points, they were close enough, and a mean depth between the observations was used for analysis. Special case was implemented using only residual data for both local and global averaged data.

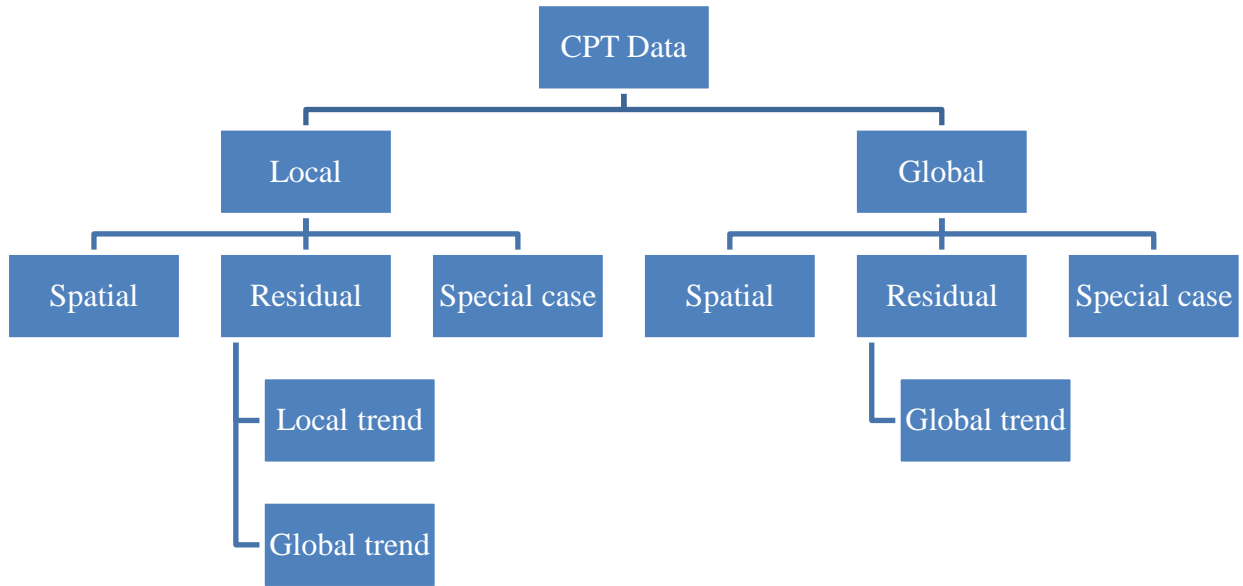


Figure 8: Options available for statistical analysis

Once the statistical soil parameters are available from analysis for each of the 25 site-locations, their global mean are calculated and they are transferred back from log-space to cartesian-space using transformation relation between normal and lognormal distribution:

$$\mu_x = e^{\left(\mu_{\ln x} + \frac{\sigma_{\ln x}^2}{2} \right)} \quad (16)$$

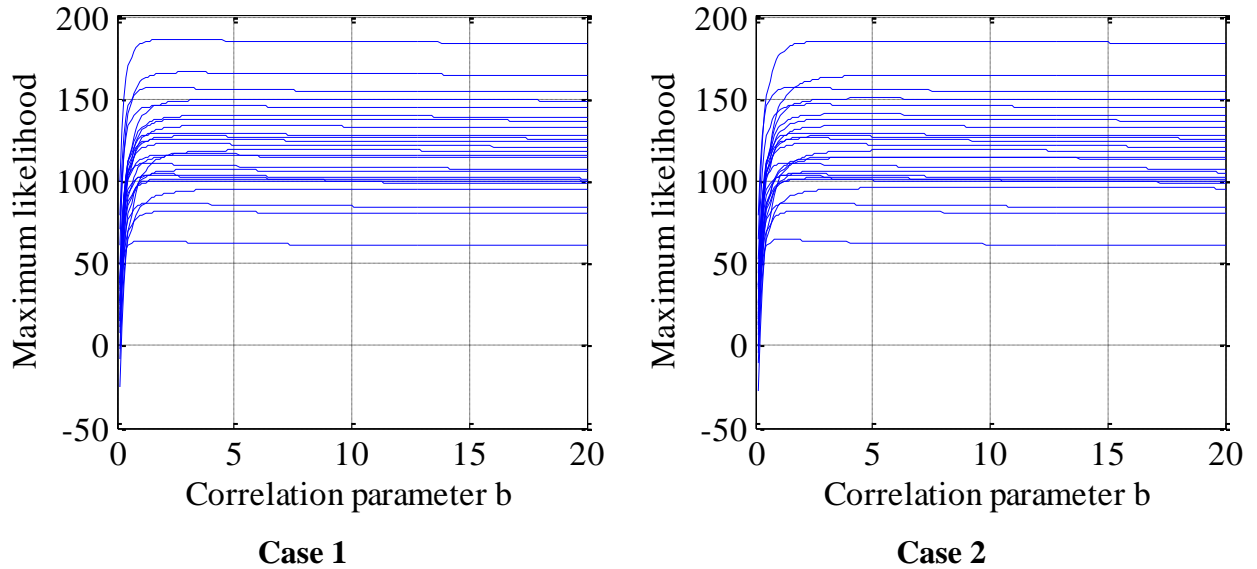
$$\sigma_x^2 = e^{\left(\mu_{\ln x} + \frac{\sigma_{\ln x}^2}{2}\right)} \left(e^{\sigma_{\ln x}^2} - 1\right) \quad (17)$$

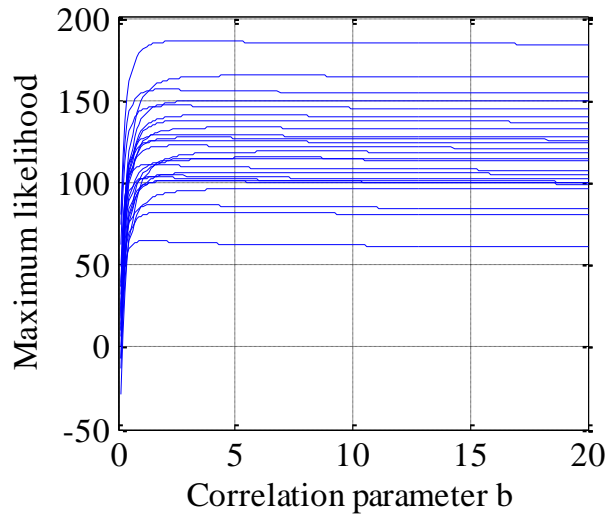
The results obtained from different analysis cases are presented in Table 1.

Table 1: First- and second-order parameters for the soil data

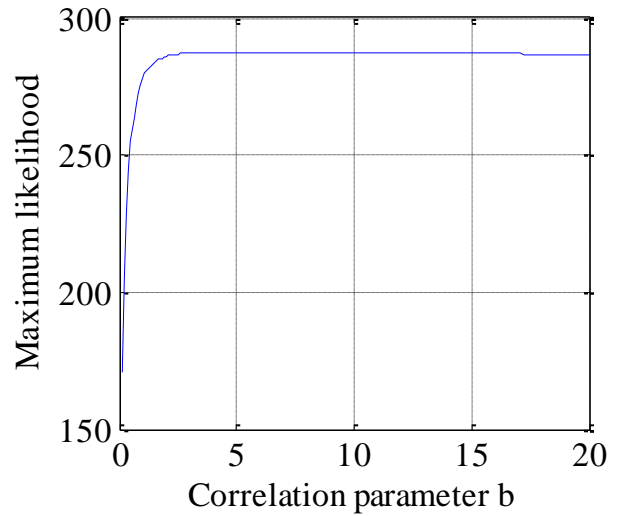
No.	Data source	Data type	Trend	Mean (μ_x)	Variance (σ_x^2)	Correlation parameter (b)
1	Local	Residual	Local	78.17	13448	3.22
2	Local	Residual	Global	80.51	15898	3.62
3	Local	Spatial	n.a.	81.10	16105	3.54
4	Local	Special case	n.a.	77.52	12103	6.78
5	Global	Residual	Global	84.12	516	4.70
6	Global	Spatial	n.a.	91.53	1392	10.20
7	Global	Special case	n.a.	79.08	663	13.25

The variation of log maximum likelihood function with correlation parameter b for five of the seven analysis cases in Table 1 is presented in Figure 9. For special cases 4 and 6 using algorithm 2, the correlation parameter is calculated directly using maximization of log maximum likelihood function and not by iteration. Hence, variation of maximum likelihood function with correlation parameter b is not available. For the cases where local data were used for statistical analysis (case 1, 2, and 3), 25 plots are obtained for each site location. For analysis done on global mean data (case 5 and 6), only one plot is obtained.

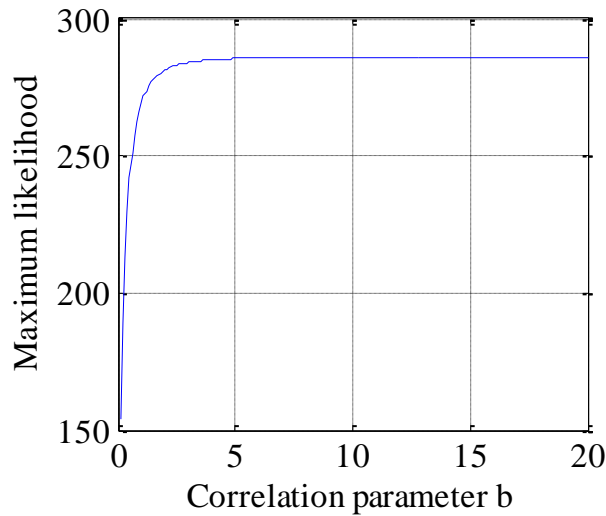




Case 3



Case 5



Case 6

Figure 9: Variation of log maximum likelihood function with correlation parameter

6. CONCLUSIONS

Random field modeling of cone penetration resistance of soil at a site was performed using finite-scale stochastic model. First- and second-order parameters were estimated using different analysis cases. The key conclusions of the study are:

1. Different types of soils are encountered along the depth. A single trend obtained for the whole depth, irrespective soil layer, might not follow an expected behavior. Dividing the soil strata in different layers based on the soil type and then performing the statistical modeling is a more appropriate method to address variability in soil type.
2. A Lognormal distribution in Cartesian-space or Gaussian distribution in log-space might not be appropriate for single site location; however the data obtained from the whole site

suggest that these distributions are appropriate for statistical analysis of a large sample space.

3. Using local or global trend to determine residual at a site location does not significantly affect the final values of soil parameters.
4. For the case (case 1, 2, and 3) where the soil parameters are obtained as global mean of parameters from each of 25 locations, using either residual or spatial variation does not affect the parameter values.
5. If the parameters obtained from global mean data, then using residual and spatial data gives substantially different values correlation parameter.
6. The variation in soil properties for local cases is greater than the global case.
7. Special cases implemented using algorithm 2, produced values of correlation parameter that were much different than the ones obtained using Algorithm 1.

REFERENCES

1. Fenton, G. (1999). "Estimation for Stochastic Soil Models." *Journal of Geotechnical and Geoenvironmental Engineering*, **125**(6), 470-485.
2. (2014). "CIE 512: Reliability Methods course materials: cone penetration test (CPT) soundings." <<http://ekalavya.eng.buffalo.edu/~kallol/512/SoilProperties>>. (04/09/2014, 2014).
3. Mathworks (2011). Computer Program Matlab R2011b, The mathworks Inc., Natick, MA.

Appendix A: Algorithm 1

```
%Implementation of algorithm 1 proposed by Fenton (1999)
%Program written to estimate second order parameter of random fields
%Written by Manish Kumar, 03/25/2014
clc, clear;
load data.txt;
load depth.txt;
[m,n]=size(data);
Corr=zeros(m);
one=ones(m,1);

% set variables for graph plotting
% set size of plots in inches
plot_height = 3.25;
plot_width = 3.25;
% set plot resolution in dpi
resolution = '-r300'; % default: 150
% set plot line width l = 1/72 inch
line_width = 0.5; % default: 0.5
% change default font
set(0,'DefaultAxesFontName', 'Times New Roman')
set(0,'DefaultAxesFontSize', 12)
set(0,'DefaultTextFontname', 'Times New Roman')
set(0,'DefaultTextFontSize', 12)
% set file path of graphs
filepath=[cd, '\Graphs\'];

%Method 1 (identified by m1)
%Compute trend and residual for each location
%Average the second order parameters obtained from all locations

param_m1=[];
trend_m1=[];
for i=1:n
    x_m1=log(data(:,i));
    p_m1 = polyfit(depth,x_m1,1);
    xfit_m1=polyval(p_m1,depth);
    %xfit_m1=mean(x_m1)*ones(m,1);
    xresid_m1=(x_m1-xfit_m1);
    fig_plot = figure('visible','off');
    plot(x_m1,depth, xfit_m1,depth, xresid_m1, depth,'LineWidth',line_width);
xlabel('q_c (tsf)');ylabel('Depth(ft)');
    set(gcf, 'PaperUnits', 'inches', 'PaperPosition', [0 0 plot_width
plot_height]);
    set(gca, 'XAxisLocation', 'top', 'YAxisLocation', 'left', 'ydir', 'reverse');
    legend('q_c', 'Trend', 'Residual');
    legend('Location', 'Best');
    legend('boxoff'); grid on;
    print(fig_plot, '-dpng', ['Spatial', num2str(i), '.png'], resolution);

param=[];
b=0.1:0.1:20;
for j=b
    for k=1:m
        for l=1:m
```

```

        Corr(k,1)=exp(-abs(depth(k,1)-depth(1,1))/j);
    end
end
r=Corr\xresid_m1;
s=Corr\one;
mu_resid=one'*r/(one'*s); %Mean of the residual
var=(xresid_m1-mu_resid*one)'*r/m;
v = logdet(Corr);
L=-m*log(var)/2-v/2;
param=[param;L,mu_resid,var,j];
end
trend_m1=[trend_m1; mean(xfit_m1)]; %Mean of the trend at each location
Lmax=max(param(:,1));
[a1,a2] = find(param(:,1)==Lmax);
param_m1=[param_m1;i,param(a1,:)];

plot(b,param(:,1));
xlabel('Correlation parameter b');ylabel('Maximum likelihood');
grid on;
hold on
end

parameters_m1=mean(param_m1(:,2:end))
[mu_m1,var_m1]=lognstat(parameters_m1(1,2)+mean(trend_m1),sqrt(parameters_m1(
1,3)))

%Method 2 (identified by m2)
%Compute trend from global and residual for each location
%Average the second order parameters obtained from all locations

param_m2=[];
x_m2_global=log(mean(data,2));
p_m2 = polyfit(depth,x_m2_global,1);
xfit_m2=polyval(p_m2,depth);
trend_m2=xfit_m2;
hold off;
for i=1:n
    x_m2=log(data(:,i));
    xresid_m2=x_m2-xfit_m2;
    param=[];
    b=0.1:0.1:20;
    for j=b
        for k=1:m
            for l=1:m
                Corr(k,1)=exp(-abs(depth(k,1)-depth(1,1))/j);
            end
        end
        r=Corr\xresid_m2;
        s=Corr\one;
        mu_resid=one'*r/(one'*s);
        var=(xresid_m2-mu_resid*one)'*r/m;
        v = logdet(Corr);
        L=-m*log(var)/2-v/2;
        param=[param;L,mu_resid,var,j];
    end
end
Lmax=max(param(:,1));

```



```

[a1,a2] = find(param(:,1)==Lmax);
param_m2=[param_m2;i,param(a1,:)];

plot(b,param(:,1));
xlabel('Correlation parameter b');ylabel('Maximum likelihood');
grid on;
hold on
end

parameters_m2=mean(param_m2(:,2:end))
[mu_m2,var_m2]=lognstat(parameters_m2(1,2)+mean(trend_m2),sqrt(parameters_m2(
1,3)))

%Method 3 (indentified by m3)
%Compute trend and residual each from global average

param_m3=[];
x_m3=log(mean(data,2));
p_m3 = polyfit(depth,x_m3,1);
xfit_m3=polyval(p_m3,depth);
xresid_m3=x_m3-xfit_m3;
trend_m3=xfit_m3;

param=[];
b=0.1:0.1:20;
for j=b
    for k=1:m
        for l=1:m
            Corr(k,l)=exp(-abs(depth(k,1)-depth(l,1))/j);
        end
    end
    r=Corr\xresid_m3;
    s=Corr\one;
    mu_resid=one'*r/(one'*s);
    var=(xresid_m3-mu_resid*one)'*r/m;
    v = logdet(Corr);
    L=-m*log(var)/2-v/2;
    param=[param;L,mu_resid,var,j];
end

hold off
plot(b,param(:,1));
xlabel('Correlation parameter b');ylabel('Maximum likelihood'); grid on;

Lmax=max(param(:,1));
[a1,a2] = find(param(:,1)==Lmax);
param_m3=[1,param(a1,:)];

parameters_m3=param_m3(:,2:end)
[mu_m3,var_m3]=lognstat(parameters_m3(1,2)+mean(trend_m3),sqrt(parameters_m3(
1,3)))

% Method 4 (indentified by m4)
% No separation of trend or residual
% Average the parameters obtained from all locations

```

```

param_m4=[];
hold off
for i=1:n
    x_m4=log(data(:,i));
    param=[];
    b=0.1:0.1:20;
    for j=b
        for k=1:m
            for l=1:m
                Corr(k,l)=exp(-abs(depth(k,l)-depth(l,l))/j);
            end
        end
        r=Corr\x_m4;
        s=Corr\one;
        mu=one'*r/(one'*s);
        var=(x_m4-mu*one)'*r/m;
        v = log(det(Corr));
        L=-m*log(var)/2-v/2;
        param=[param;L,mu,var,j];
    end

    plot(b,param(:,1));
    xlabel('Correlation parameter b');ylabel('Maximum likelihood');
    grid on;
    hold on

    Lmax=max(param(:,1));
    [a1,a2] = find(param(:,1)==Lmax);
    param_m4=[param_m4;i,param(a1,:)];
end

parameters_m4=mean(param_m4(:,2:end))
[mu_m4,var_m4]=lognstat(parameters_m4(1,2),sqrt(parameters_m4(1,3)))

%Method 5 (indentified by m3)
%No separation of trend and residual
%Compute parameters from global averaged data

param_m5=[];
x_m5=log(mean(data,2));
param=[];
b=0.1:0.1:20;
for j=b
    for k=1:m
        for l=1:m
            Corr(k,l)=exp(-abs(depth(k,l)-depth(l,l))/j);
        end
    end
    r=Corr\x_m5;
    s=Corr\one;
    mu=one'*r/(one'*s);
    var=(x_m5-mu*one)'*r/m;
    v = logdet(Corr);
    L=-m*log(var)/2-v/2;
    param=[param;L,mu,var,j];
end

```

```
end
hold off
plot(b,param(:,1));
xlabel('Correlation parameter b');ylabel('Maximum likelihood');
grid on;
Lmax=max(param(:,1));
[a1,a2] = find(param(:,1)==Lmax);
param_m5=[1,param(a1,:)];

parameters_m5=param_m5(:,2:end)
[mu_m5,var_m5]=lognstat(parameters_m5(1,2),sqrt(parameters_m5(1,3)))
```

Appendix B: Algorithm 2

```
%Implementation of algorithm 2 proposed by Fenton (1999)
%Program written to estimate second order parameter of random fields
%Written by Manish Kumar, 03/25/2014
clc, clear;
load data.txt;
load depth.txt;
[m,n]=size(data);
Corr_inv=zeros(m);
one=ones(m,1);
dz=0.160;

%Method 1 (identified by m1)
%Compute soil parameters for each location
%Average the second order parameters obtained from all locations

param_m1=[];
for i=1:n
    x_m1=log(data(:,i));
    p_m1 = polyfit(depth,x_m1,1);
    xfit_m1=polyval(p_m1,depth);
    %xfit_m4=mean(x_m4)*ones(m,1);
    xresid_m1=x_m1-xfit_m1;
    %figure;
    %plot(depth,x_m1,depth, xfit_m1)

    R0=0;
    R1=0;
    for k=1:m-1
        R0=R0+xresid_m1(k,1)^2;
        R1=R1+xresid_m1(k,1)*xresid_m1(k+1,1);
    end
    R0=R0+xresid_m1(m,1)^2;
    R0_dash=R0-xresid_m1(1,1)^2-xresid_m1(m,1)^2;
    b0=m*R1;
    b1=-(R0+m*R0_dash);
    b2=-(n-2)*R1;
    b3=(n-1)*R0_dash;

    q = 0.5;
    q_old = 1;
    iter = 0;
    while abs(q_old-q) > 10^-6 && q ~= 0
        q_old = q;
        q = q - (b0+b1*q+b2*q^2+b3*q^3) / (b1+2*b2*q+3*b3*q^2);
        iter = iter + 1;
    end

    Qn=sum(xresid_m1);
    Qn_dash=Qn-xresid_m1(1,1)-xresid_m1(m,1);
    mu=(Qn-q*(Qn+Qn_dash)+q^2*Qn_dash)/(n-2*q*(m-1)+q^2*(m-2));
    var=(R0-2*q*R1+q^2*R0_dash)/(m*(1-q^2));
    b=-2*dz/log(q);
    Lmax=-m*log(var)/2-(1/2)*(m-1)*log(1-q^2);
    mu=mu+mean(x_m1); %Adding trend mean to the residual mean
```

```

        param_m1=[param_m1;Lmax,mu,var,b];
end

parameters_m1=mean(param_m1(:,2:end))
[mu_m1,var_m1]=lognstat(parameters_m1(1,1),sqrt(parameters_m1(1,2)))

%Method 2 (identified by m1)
%Compute soil parameters for mean global data

x_m2=log(mean(data,2));
p_m2 = polyfit(depth,x_m2,1);
xfit_m2=polyval(p_m2,depth);
xresid_m2=x_m2-xfit_m2;
%figure;
%plot(depth,x_m1,depth, xfit_m1)

R0=0;
R1=0;

for k=1:m-1
    R0=R0+xresid_m2(k,1)^2;
    R1=R1+xresid_m2(k,1)*xresid_m2(k+1,1);
end
R0=R0+xresid_m2(m,1)^2;
R0_dash=R0-xresid_m2(1,1)^2-xresid_m2(m,1)^2;
b0=m*R1;
b1=-(R0+m*R0_dash);
b2=-(n-2)*R1;
b3=(n-1)*R0_dash;

q = 0.5;
q_old = 1;
iter = 0;
while abs(q_old-q) > 10^-6 && q ~= 0
    q_old = q;
    q = q - (b0+b1*q+b2*q^2+b3*q^3)/(b1+2*b2*q+3*b3*q^2);
    iter = iter + 1;
end

Qn=sum(xresid_m2);
Qn_dash=Qn-xresid_m2(1,1)-xresid_m2(m,1);
mu=(Qn-q*(Qn+Qn_dash)+q^2*Qn_dash)/(n-2*q*(m-1)+q^2*(m-2));
var=(R0-2*q*R1+q^2*R0_dash)/(m*(1-q^2));
b=-2*dz/log(q);
mu=mu+mean(x_m2); %Adding trend mean to the residual mean
param_m2=[mu,var,b];

parameters_m2=param_m2
[mu_m2,var_m2]=lognstat(parameters_m2(1,1),sqrt(parameters_m2(1,2)))

```


11.48	100.2	2.1	105.3	116.6	64.2	98	88.6	78.2	171	62.2	5.5	58.1	160.3	116.8	72	16.7	11.2	5.2	3.8	4.2	6	4.5	4.8	4.1
11.65	73.6	2.1	118.5	111.9	73	89.6	93.7	84.8	167.5	69.5	8	55.7	155.8	139	80.1	31.2	12	4.8	4	3.8	4.8	4	4.9	4.8
11.81	61.6	2.3	125.3	111.9	80.4	100.5	97.7	91.7	140	71.8	8.3	48.8	160	171.6	93.1	51.8	17.8	4.6	4.4	3.8	3.8	4.3	6.2	5.2
11.98	44.8	3.1	115.5	104.8	80.7	95.7	90.9	113.6	126.2	71.6	5.6	39.6	158.5	194.8	104.2	66.2	9.8	4.5	4.4	4	3.5	4.4	6.9	4.9
12.14	37.7	2.2	117.8	111	84.8	94.7	81.8	121.3	119.8	83.8	6.7	27	153.4	204.9	107.9	80.6	13.1	4.9	4.3	5.1	3.5	4.5	6.3	5
12.3	24.9	2.2	122	113.6	87	95.9	84.7	121.2	114.5	94.8	13.3	24.6	144.5	203	111.9	87.5	92.4	5	4.3	6.4	3.5	4.3	7.5	5.2
12.47	26.2	2.4	125.6	111.9	83.7	100.2	77.5	125.7	109.2	99	13.9	20.8	135	192.7	114.7	84.8	164.9	5.3	4.4	6.1	4.3	4.7	22.6	4.7
12.63	73	2.5	123.1	111.3	78.5	116.4	81.4	123.2	105.5	99.7	15	12.3	128	186.7	124.3	79.2	239.4	5	4.9	10.2	5	5.8	15.8	4.5
12.8	73.2	2.8	124.4	113.3	78.7	131.8	89.4	112.3	103.1	102.7	32.5	20.6	122.4	182.1	139.8	64.8	185.7	4.8	4.8	19.9	5	6.6	8.5	4.9
12.96	66	5	123.1	120.6	77.2	138.8	85.7	105.6	100	101.2	59.7	36.5	122.9	179.6	147.8	55.9	230.8	4.2	4.5	20.3	6.1	7.5	5.6	4.7
13.12	61.5	29.5	122.5	129.3	80.9	146.4	84.6	104.7	97	99.9	79.4	40.3	116.5	182.5	154.4	46.9	201.9	4.8	5.4	17.4	6.8	7.3	5.1	4.8
13.29	46	45.4	125.7	137.4	82.7	143.8	83.1	108.5	93.9	103.3	96.8	42.9	106	196.7	161.8	42.4	189.2	5	5.8	16.3	9.9	11.4	4.9	4.8
13.45	44.1	53	130.1	147.7	76.7	137.5	82	117.9	90.6	104.1	105.7	48.5	97.3	212.6	169.5	47.3	171	6.4	7.8	14.1	9.3	21.8	6.8	4.8
13.62	42.8	56.5	136.3	134.8	71.8	125.9	77.4	123.7	84.2	114.2	104.8	59.4	83.3	217	174.6	47.6	146.8	11.8	10.6	10.8	10.9	35.3	13.3	4.8
13.78	47.1	67.1	143	135.3	82.1	116.5	74	118.2	72.6	122.6	103.8	80.2	64.5	215	179.9	26	107.9	12.1	12	7.1	20.4	41.7	18.8	4.8
13.94	42.2	76	146.6	138.2	90	114	75.3	103.6	66	129.8	113.6	86	28.6	211.4	187.4	14.3	84.4	9	7.1	6.4	45.3	43.4	23.6	5.1
14.11	29	77.1	152.2	142.9	101	111.9	74.2	101	67.7	129.9	116.4	94.2	16.8	203.2	189.7	8.6	86.6	26.9	19.8	7.4	71.5	40.6	16.1	5
14.27	31.9	75	163.8	147.5	101.6	103.4	73.6	107.1	79.5	125.9	113.8	105.2	8.5	193.6	192.9	6.3	65.7	33.6	44	6.7	89.2	38	11.6	5
14.44	82.6	79.1	167.1	136.6	98.7	94.1	64.3	107.2	87.7	121.5	108	115.6	5.3	185.1	185.3	6.8	43.7	40.5	46	7.3	106.4	40.8	8	4.8
14.6	85.4	80.6	160.1	124.9	86	84.2	60.9	103.1	96.7	111.7	93.6	115.5	4.2	175.6	165.2	6.8	46.7	48.9	41.9	8	115.2	45.5	8.7	4.8
14.76	71.2	76.1	165	110.7	76.8	78.2	59.7	107.3	109.3	99.8	78.9	109.9	4.2	172.9	160.8	7	47.3	45.8	25.3	8.5	116.3	58.7	11.1	4.9
14.93	54.9	68.7	173	110.5	73.6	76.2	60.4	103.6	127.4	99.9	80.1	107.2	3.5	180	142	7.1	37.4	27.2	13.9	7.8	100.7	56.1	16.2	5.4
15.09	42.6	66.9	160.5	122.2	66.3	81.4	63.8	97.8	118	98.6	85.7	100.7	2.9	183.3	136	7.4	27.8	14.1	11.5	8.6	88.4	56.7	7.9	8.5
15.26	66.7	69.7	163.1	129.5	59.7	92.3	67.4	84.2	118.8	100.7	83.4	98	2.7	174.7	130.3	7.5	25.8	13.6	18.3	12.8	84.4	60.9	6.3	24.2
15.42	75.2	78.8	166.8	135.5	59	97.5	68.7	88.3	116.3	101.5	81.3	93	2.9	157.5	119.9	7.3	25.4	10.4	26	17.7	67.9	65	9.5	47.7
15.58	74.4	91.5	164.8	131.2	61.3	102.2	66.5	91.1	118.8	103.6	88.4	94	5	149	101.7	6.8	75	23	43.3	26.2	51.3	70.4	7.8	53.6
15.75	54.3	100.7	152.9	123.9	60.5	100.3	61.5	93.1	117.1	108.9	97.1	94.2	5.9	133.1	71.1	6.6	93.2	52.4	46	32.7	39.2	65.5	14.4	36.1
15.91	26.4	101.8	143.4	118.6	58.8	94.2	56	96.9	110.8	113.5	101.3	86	3.2	108.6	40.2	6.4	81.8	95.9	51.4	34.5	33.8	49.5	28.8	22.9
16.08	19.8	110.6	138.3	117.8	59	81.3	58.3	98	105.9	119.7	97.6	71.7	2.6	62.6	22	6.6	68.1	103.4	67.5	32.3	29.8	31.8	39.1	14.4
16.24	16.1	108.9	139.4	112.5	64.8	86.2	61.4	97.6	111	130.6	91.1	57.2	2.7	34.8	20.1	6.6	65.7	95.4	77.5	24	19.7	37.1	8.4	
16.4	8.5	107.6	143.1	110.3	72.3	78	61	98.7	104.4	132.9	93.6	44.7	2.9	18	14	7.8	69.7	82.3	60.7	18.9	25.3	12.5	39.7	6.1
16.57	6.6	113.3	141	117.5	75.2	72.8	64.4	93	114.5	126.2	95.7	32.2	2.6	12.6	9.2	8.3	65.3	74.2	30.5	14.5	19.8	8.3	42.7	5.7
16.73	29.4	111.3	151.6	118.2	76.6	69.8	69	94.8	126.1	124.8	95.7	26.5	2.9	9.1	6.4	9.8	65.8	45.4	14.1	13	14.2	8.2	44.2	6.9
16.9	5.8	110.9	159.7	117.2	83.4	67.9	78.3	103.4	123.6	127.4	92.7	27.8	13.6	8.4	6	18.7	59.3	19.3	12.6	15.4	16	13.6	46.1	9.6
17.06	6.7	110.2	163.1	113.3	87.8	70.9	92.7	112.6	113.3	122	89.6	29.8	5.3	7	5.8	31.2	51.9	11.7	11.8	22.2	36	13.8	46.8	15.7
17.23	5.3	107.3	158.1	125	88.4	76.1	98.7	114.4	120.1	121	85.6	30.7	7.7	5.8	10.5	45	46.8	11.6	11	25.1	31.1	12.8	46.3	13.5
17.39	5.3	103.2	149.4	139.3	87.4	76	96.7	115.4	133.9	127.5	81	30.8	8.7	7.4	10.6	56.5	36.4	8.6	21.3	35.4	24.9	12.3	45.8	9.9
17.55	5.9	98.5	146.1	144.3	85.8	75.1	86.2	108.9	130.5	128.7	77.3	31.4	10.7	17	8.7	67.4	21.9	15.6	27	54.4	21.9	12.1	45	6.8
17.72	32.8	91	135.1	148.6	85.4	74	76.7	108.6	110.2	119	73.1	27.2	11.4	23.7	13.1	82.6	13.2	23.9	23.9	50.9	13.1	10.7	44.2	6.8
17.86	30.4	102.7	116.8	140.7	83.4	71.8	75.6	102.7	94.9	118.7	71	20.9	10	25.3	15.2	91.4	16.2	21.9	20.6	42.6	11.4	9.5	45.4	7.7