A more accurate way to measure spatial correlation length from CPT data – ARMA models

Brigid Cami, Sina Javankhoshdel, & Thamer Yacoub
Rocscience, Toronto, ON, Canada

ABSTRACT
Spatial variability is one of the largest sources of uncertainty in geotechnical applications. This variability is primarily characterized by the spatial correlation length, a parameter that describes the distance over which the parameters of a material are similar. Spatial variability is generally described with traditional methods of time series analysis. In statistics, the Auto-Regressive Moving Average (ARMA) model is commonly used to describe the relationship between two points in time. Instead of assuming an autocorrelation model, the ARMA model calculates the necessary auto-regressive components (AR), as well as a decaying mean structure (MA). The advantage of this method is that it is calculated for each specific field study, so that the data is not forced to fit into a fixed autocorrelation model (e.g. Markovian, Gaussian, etc.). Additionally, a very simple and fast algorithm is needed to calculate the necessary AR, and MA estimates. In this study, the ARMA model is introduced as a means of measuring correlation length, and two case studies and a simulation are used to compare the correlation length values from the ARMA model to the other estimates. The ARMA model was able to find correlation length estimates that were very similar to other methods in the case study values, and much more accurate values in the simulation, compared to other methods.

1 INTRODUCTION

Spatial variability is one of the largest sources of uncertainty in geotechnical applications. In recent decades the necessity of considering spatial variability in geotechnical applications has been demonstrated in various studies (Cho 2010; Soubra and Massih 2010; Hicks and Spencer 2010; Huang et al. 2010; Stuedlein et al. 2012; Cassidy et al. 2013; Jha and Ching 2013; Jiang et al. 2014; Le 2014; Li et al. 2015; Xiao et al. 2016; Li et al. 2016; Luo et al. 2016; Javankhoshdel et al. 2017; Papaioannou and Straub 2017; Cami et al. 2018). This variability is primarily characterized by the spatial correlation length which describes the distance over which the parameters of a soil or rock are similar or correlated; soil properties sampled from adjacent locations in the soil profile tend to have similar values and as the sampling distance increases the correlation decreases. The correlation length parameter can be obtained in a variety of ways but is most commonly obtained from cone penetration test (CPT) measurements. Is required in order to characterize as well as simulate a spatially variable field. It should be noted that a different correlation length is defined for each material, so the CPT data considered here is material-specific.

Spatial variability is generally described by traditional methods of time series analysis in statistics, meaning that it constitutes of a trend component and a zero-mean spatial variability component (Equation 1). The reason for this is that as with measurements in time, soil property measurements that are closer together in space are more similar in value, as shown below:

$$X_i = X(s_i) = T(s_i) + \epsilon(s_i),\ i = 1, \ldots, k \qquad [1]$$

where X_i is the value of the soil property at location s_i, s_i is the vertical distance from the ground surface, for example, and k is the total number of measurements. $\epsilon(s_i)$ is the spatial variability component. The spatial correlation length describes the distance over which the spatial variability components $\epsilon(s_i)$ are correlated amongst themselves.

The commonly used methods of measuring correlation length in the geotechnical field assume an autocorrelation model. A method of moments can then be used to estimate the correlation length value, by minimizing the error between the theoretical autocorrelation model and the experimental one (Vanmarcke, 1977). An autocorrelation model describes the relationship between the distance separating two points and the correlation between them. Some typical autocorrelation models are shown in Table 1, where $\rho(\tau)$ is the correlation coefficient between two points separated by lag τ, and θ is the correlation length.

Table 1. Common autocorrelation models.

Autocorrelation Model	Relationship
Markovian	$\rho(\tau) = exp\{\frac{-2\|\tau\|}{\theta}\}$
Gaussian	$\rho(\tau) = exp\{-\pi(\frac{\|\tau\|}{\theta})^2\}$
Spherical	$\rho(\tau) = \{1 - 1.5\|\frac{\tau}{\theta}\| + 0.5\|\frac{\tau}{\theta}\|^3$ $if\ \|\tau\| \leq \theta;\ 0\ otherwise$

However, the autocorrelation model selected for a given set of CPT measurements is generally just assumed to be the one that describes the true structure of the data. Since no model can fit the data exactly, this makes the selection of an autocorrelation model difficult.

In statistics, the Auto-Regressive Moving Average (ARMA) model is commonly used to describe the relationship between two points in time. Instead of assuming an autocorrelation model, the ARMA model calculates the necessary auto-regressive components (AR), as well as a decaying mean structure or moving average (MA). The advantage of this method is that it is calculated for each specific field study, so that the data is not forced to fit into a fixed autocorrelation model. Additionally, a very simple and fast algorithm is needed to calculate the necessary AR, and MA coefficients.

In this study, the ARMA model is introduced as a means of measuring correlation length. Two case studies and a simulation are used to compare the correlation length values from the ARMA model to the method of moments estimates. There are no previous studies that use this method to measure correlation length.

2 ARMA

2.1 Stationary Time Series

As with the methods of moments, in order to measure the correlation length from CPT data, the data must first be stationary. A stationary time series has properties that do not depend on the time at which the series was observed. In the CPT realm, a stationary CPT is one whose properties do not depend on the depth.

Weakly stationarity is defined by a constant mean, variance, and covariance structure. This is necessary in order for the autocorrelation function to have meaning. While the constant variance and covariance must be assumed, the constant mean is the famous de-trending problem. This is analogous to removing the trend of a measurement and only looking at the spatial variability component ($\epsilon(s_i)$ in Equation 1). The readers are referred to the multitudes of literature on the subject some of which are included here (Ching et al. 2016, 2017; Ching & Phoon 2017). The data used in the remainder of the paper is assumed to satisfy weakly stationarity.

2.2 The ARMA Model

The auto-regressive (AR) component of the ARMA model, allows for current measurements in time to depend on a certain lag of past measurements. For example an AR(1) model indicates that the current measurement depends on the last. An AR(2) model indicates that the current measurement depends on the last and the one previous to that. This can be similarly applied to CPT measurements, such that for an AR(2) model, a measurement at a given location depends on the measurements at the two previous locations adjacent to it. An AR(p) model is expressed as shown in Equation 2 below, where α_i are the coefficients associated with each past measurement, and w_i are the random error components which are typically assumed to be independently and identically distributed white noise with some fixed variance σ_w^2. X_i is the value of the soil property at location s_i, X_{i-1} is the value at location s_{i-1}, and δ is the intercept.

$$X_i = \delta + \alpha_1 X_{i-1} + \cdots + \alpha_p X_{i-p} + w_i \qquad [2]$$

The moving average (MA) component indicates that the regression error is a linear combination of the error terms at the previous locations. Similarly to the AR, an MA(2) model indicates that the current error depends on the error at the previous *two* locations. An MA(q) model is expressed as shown in Equation 3 below, where θ_i are the coefficients associated with each past measurement error, w_{i-1} is the error associated with measurement X_{i-1}, and μ is the intercept.

$$X_i = \mu + \theta_1 w_{i-1} + \cdots + \theta_q w_{i-q} + w_i \qquad [3]$$

Therefore, for stationary data, an ARMA(p,q) model can be expressed as shown in Equation 4 below, where p is the order of the AR component, and q is the order of the MA component:

$$\begin{aligned} &X_i - \alpha_1 X_{i-1} - \cdots - \alpha_p X_{i-p} \\ &= w_i + \theta_1 w_{i-1} + \cdots + \theta_q w_{i-q} \end{aligned} \qquad [4]$$

In the equation above, X_i are the stationary measurements at each location, α_i are the coefficients of the AR components, θ_i are the coefficients of the MA components, and w_i are the errors associated with the MA model.

Once the coefficients α_i and θ_i are determined for the necessary number of p and q, then the autocorrelation function for the specific case is defined, and the correlation length can be calculated as simply the area under the correlation function.

It turns out that these coefficients and orders can be determined automatically and quickly with a simple algorithm.

2.3 Determining the ARMA coefficients

There are two ways to determine the ARMA coefficients. One is by visual inspection of the autocorrelation function and partial autocorrelation function plots. It is often evident from reviewing these plots what the values of p and q should be. An even simpler way is using the auto.arima algorithm from the forecast package in R (Hyndman and Khandakar, 2008; Hyndman et al. 2019). This code is open-source and available for implementation in other software.

The auto.arima function takes as an input the CPT data in the format of measurement locations and measurements at each location. It outputs the necessary values for p and q and their respective coefficients. Once these coefficients are determined the correlation structure of the data is explained.

2.4 Determining the correlation length

Once the coefficients are determined, the autocorrelation function $\rho(\tau)$ can be defined and the corresponding correlation length, θ, is the area under this function, as shown in Equation 5 (Vanmarcke, 1984):

$$\theta = \int_{-\infty}^{\infty} \rho(\tau) d\tau = 2 \int_{0}^{\infty} \rho(\tau) d\tau \quad [5]$$

An important note is warranted here – the factor of 2 in the equation above is often omitted hence resulting in two definitions of correlation length. What is alternately referred to as *correlation length* or *scale of fluctuation* has been defined as both θ and $\theta/2$ in geotechnical literature, resulting in general confusion. In this study, the correlation length refers to θ as defined above.

This integral can be easily obtained with quadrature of the autocorrelation function.

3 VERIFICATION

Three examples are considered for verification of the ARMA method. The first two use CPT measurements from two studies, the correlation lengths of which were measured using a method of moments and an assumed autocorrelation model. These are used to verify that ARMA gives similar results to the classic methods. The third example is a simulated example where the correlation length is known, and ARMA as well as methods of moments are used to see how close they can get to the true measurement.

3.1 Example 1: Świebodzice

This example uses a CPT measurement from Świebodzice (Bagińska et al., 2012), the correlation length of which was measured by Pieczyńska-Kozłowska (2015). The Świebodzice CPT for q_c used in the study is shown in Figure 1.

Pieczyńska-Kozłowska (2015) used various autocorrelation models and de-trending methods and compared the resulting correlation lengths, measured using methods of moments. For comparison purposes, only the linearly de-trended measurements are used below. These results form Pieczyńska-Kozłowska (2015) are summarized in Table 2.

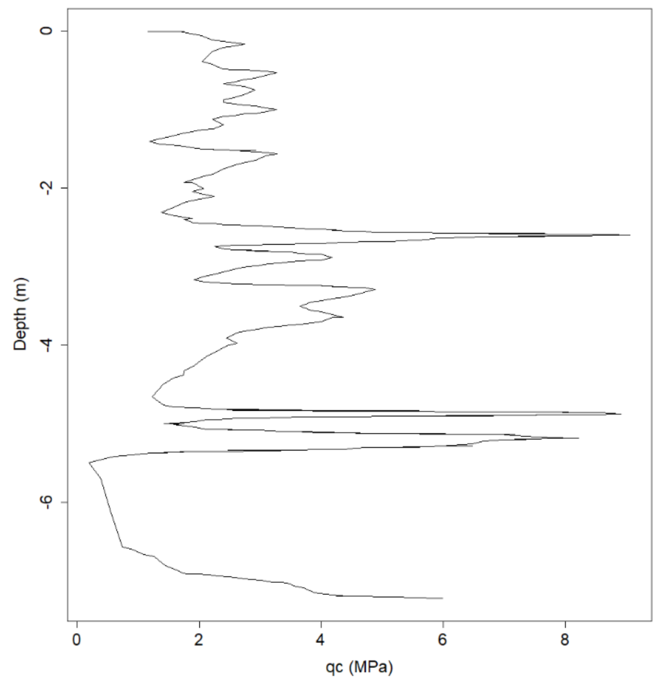

Figure 1 The Świebodzice CPT for q_c

Table 2. Pieczyńska-Kozłowska (2015) linearly de-trended correlation length results.

	Markov Autocorrelation	Gaussian Autocorrelation
Vanmarcke Method	0.28 m	0.22 m
Rice Method	0.23 m	0.29 m

The auto.arima function from the forecast package determined that an ARMA(4,4) model best described the correlation structure. That is, a model with 4 AR terms and 4 MA terms. The coefficients of this model are as shown in Table 3. Using these coefficients and quadrature of the resulting autocorrelation function, the estimated correlation length was found to be 0.26 m, which is in close agreement with the values found by Pieczyńska-Kozłowska (2015).

Table 3. The ARMA coefficients determined for the linearly de-trended Świebodzice CPT.

AR Coefficients, α_i	MA Coefficients, θ_i
0.83	0.17
0.25	0.29
0.53	-0.34
-0.63	0.34

3.2 Example 2: Taranto Clay

The second example uses a CPT measurement from Taranto, Italy (Cafaro and Cherubini, 2002). The G1 borehole of the lower clay data is used for comparison purposes, as de-trended by Cafaro and Cherubini (2002). This de-trended data is shown in Figure 2.

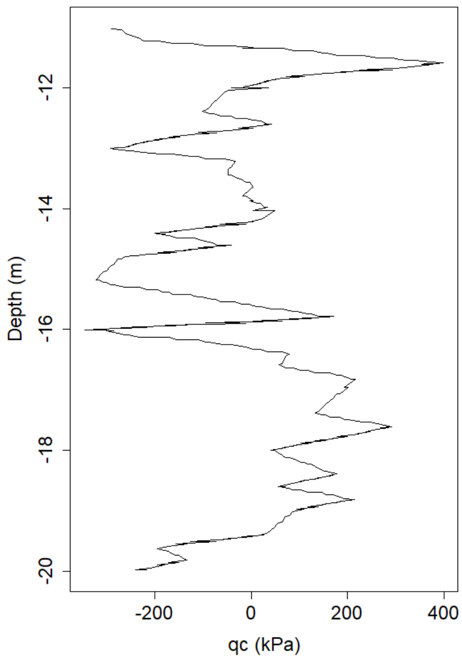

Figure 2 The de-trended CPT for q_c of borehole G1 per Cafaro and Cherubini (2002).

Cafaro and Cherubini (2002) used the variance function method to measure the correlation length and obtained a value of 0.536 m for the specific borehole, with an average measurement of 0.40 m over the five boreholes. The auto.arima function determined an ARMA(2,1) model to be the best fit for borehole G1, the coefficients of which are shown in Table 4. The estimated correlation length was found to be 0.40 m. This is in close agreement with the estimated measurement for the given borehole as well as the average over the five boreholes.

Table 4. The ARMA coefficients determined for the linearly de-trened Taranto CPT.

AR Coefficients, α_i	MA Coefficients, θ_i
1.98	-0.90
-0.98	-

3.3 Example 3: Simulated Data

Finally, the third example uses data which was simulated to have a correlation length of 5 m. This was done using the spatial variability field option in the Slide2 software (Rocscience, 2018), which uses Markovian and Gaussian autocorrelation functions together with a method known as Local Average Subdivision (LAS) (Fenton and Vanmarcke, 1990) to generate the field. The simulated field is a spatially variable cohesion parameter with a mean of 10 kPa, a standard deviation of 2 kPa, and a normal distribution. The spatial field with mesh size of 0.2 m in a typical slope with a unit weight of 19 kN/m^3 and a friction angle of 23 degrees is shown in Figure 3.

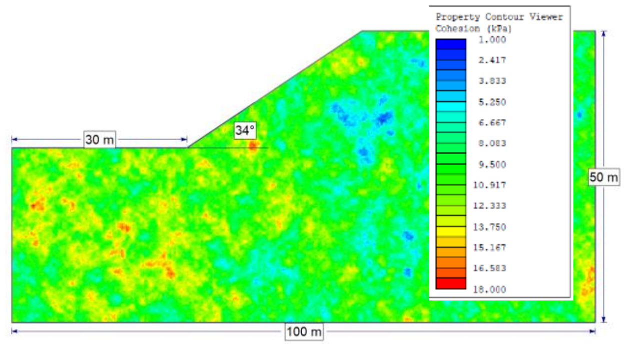

Figure 3. Random cohesion field generated with isotropic correlation length of 5 m.

Five relatively equi-spaced vertical samples were taken from the field, at x=1.1 m, x=20 m, x=50.1 m, x=75.1 m, and x=98.3 m. The correlation length was measured using both ARMA and an autocorrelation fitting method with a Markovian and Gaussian autocorrelation models. Since this data is simulated, de-trending was not necessary. The results are summarized in Table 5.

Table 5. Correlation length measurements for simulated data.

Measurement Location	Autocorrelation Fitting with Markovian Model	Autocorrelation Fitting with Gaussian Model	ARMA
1.1	3.32 m	3.36 m	5.58 m
20	1.77 m	1.36 m	1.65 m
50.1	5.32 m	6.41 m	6.60 m
75.1	3.51 m	3.92 m	6.27 m
98.3	2.26 m	2.47 m	3.58 m
Average	**3.24 m**	**3.51 m**	**4.73 m**

This simulated example has attempted to replicate what might happen in the field, where only a handful of boreholes are taken and must be used in order to characterize the field. It is seen that although all methods in Table 5 tend to deviate from the true value at specific locations, when averaged the ARMA model gives a value that is much closer to the 5 m measurement. This is due to the fact that ARMA defines an autocorrelation model for a each of the five locations exactly, instead of assuming the Markovian or Gaussian autocorrelation model.

These three average correlation lengths were input into a spatial variability analysis for the slope in Figure 3 using 500 Latin-Hypercube samples and Morgenstern-Price limit equilibrium method in order to get a rough idea of the expected difference in probability of failure when the correlation length is misrepresented.

Table 6. Probability of failure values for the slope in Figure 3 using the three correlation lengths in Table 5.

Markovian Model 3.24 m	Gaussian Model 3.51 m	ARMA 4.73 m
16.8%	17.6%	18.4%

It can be seen in the table that the correlation length parameter has a considerable effect on probability of failure.

4 DISCUSSION

In this study, the ARMA model is introduced as a means of measuring correlation length. The advantage of this method is that it allows the autocorrelation model to be defined exactly, instead of forcing the data to fit into a pre-defined model such as Gaussian or Markovian. Additionally, an open-source algorithm is available for finding the coefficients of the model quickly and easily.

Two case studies and a simulation are used to compare the correlation length values from the ARMA model to the method of moments estimates. The two case studies were found to be in good agreement with the ARMA measurement. The simulated study showed that the ARMA model got much closer to the true correlation length than the methods of moments. This has a considerable effect on the computed probability of failure.

5 REFERENCES

Bagińska I., Kupis R., Pochrań Z., Badania onda statyczna CPTU gruntu nasypowego oraz rodzimego celem analizy stanu i odkształacalności nasypu, Politechnika Wrocławska, Instytut Geotechniki i Hydroechniki, Raport Serii U 18/12, 2012.

Cafaro, Francesco, and Claudio Cherubini. "Large sample spacing in evaluation of vertical strength variability of clayey soil." *Journal of geotechnical and geoenvironmental engineering* 128.7 (2002): 558-568.

Cami, B., Javankhoshdel, S., Yacoub, T., and Bathurst, R.J. 2018. 2D Spatial Variability Analysis of Sugar Creek Embankment: Comparative Study. Proceedings of the 2nd GeoMEast International Congress and Exhibition on Sustainable Civil Infrastructures, Egypt 2018.

Cassidy, M. J., M. Uzielli, and Y. Tian. 2013. Probabilistic combined loading failure envelopes of a strip footing on spatially variable soil. Comput. Geotech. 49 (Apr): 191–205.

Ching, J., Lee, S. W. & Phoon, K. K. 2016. Undrained strength for a 3D spatially variable clay column subjected to compression or shear. Probabilistic Engineering Mechanics, 45, 127–139.

Ching, J. Y., Phoon, K. K. & Pan, Y. K. 2017. On characterizing spatially variable soil Young's Modulus using spatial average", Structural Safety, 66, May 2017, 106–117.

Ching, J. Y., Sung, S. P. & Phoon, K. K. 2017. Worst Case Scale of Fluctuation in Basal Heave Analysis Involving Spatially Variable Clays. *Structural Safety*, 68, pp.28-42.

Cho, S. E. 2010. Probabilistic Assessment of Slope Stability That Considers the Spatial Variability of Soil Properties. ASCE Journal of Geotechnical and Geoenvironmental Engineering 136 (7): 975–984.

Fenton, G.A. and Vanmarcke, E.H. 1990. Simulation of random fields via local average subdivision. Journal of Engineering Mechanics, 116(8): 1733-1749.

Hicks, M.A., and W.A. Spencer. 2010. Influence of heterogeneity on the reliability and failure of a long 3D slope. Comput. Geotech. 37 (7): 948–955.

Huang, J. S., Griffiths, D.V., and Fenton, G.A. 2010. System reliability of slopes by RFEM. Soils Found. 50 (3): 343–353.

Hyndman R, Athanasopoulos G, Bergmeir C, Caceres G, Chhay L, O'Hara-Wild M, Petropoulos F, Razbash S, Wang E, Yasmeen F (2019). *forecast: Forecasting functions for time series and linear models*. R package version 8.6, <URL: http://pkg.robjhyndman.com/forecast>.

Hyndman RJ, Khandakar Y (2008). "Automatic time series forecasting: the forecast package for R." *Journal of Statistical Software*, *26*(3), 1-22. <URL: http://www.jstatsoft.org/article/view/v027i03>.

Javankhoshdel, S., Luo, N., and Bathurst, R.J. 2017. Probabilistic analysis of simple slopes with cohesive soil strength using RLEM and RFEM. Georisk 11 (3): 231–246.

Jha, S.K., and Ching, J. 2013. Simplified reliability method for spatially variable undrained engineered slopes. Soils Found. 53 (5): 708–719.

Jiang, S. H., Li, D.Q., Zhang, L.M., and Zhou, C.B. 2014. Slope reliability analysis considering spatially variable shear strength parameters using a non-intrusive stochastic finite element method. Eng. Geol. 168 (Jan): 120–128.

Le, T.M.H. 2014. Reliability of Heterogeneous Slopes with Cross-Correlated Shear Strength Parameters. Georisk 8 (4): 250–257.

Li, D. Q., Jiang, S.H., Cao, Z. J., Zhou, W., Zhou, C.B., and Zhang, L.M. 2015. A multiple response-surface method for slope reliability analysis considering spatial variability of soil properties. Eng. Geol. 187 (Mar): 60–72.

Li, J.H., Zhou, Y., Zhang, L.L., Tian, Y., Cassidy, M.J., and Zhang, L.M. 2016. Random finite element method for spudcan foundations in spatially variable soils. Eng. Geol. 205 (Apr): 146–155.

Luo, N., Bathurst, R.J. and Javankhoshdel, S. 2016. Probabilistic stability analysis of simple reinforced slopes by finite element method. Computers and Geotechnics 77: 45-55.

Papaioannou, I., and Straub, D. 2017. Learning soil parameters and updating geotechnical reliability estimates under spatial variability Theory and application to shallow foundation. Georisk 11 (1): 116–128.

Pieczyńska-Kozłowska, Joanna M. "Comparison between two methods for estimating the vertical scale of fluctuation for modeling random geotechnical problems." *Studia Geotechnica et Mechanica* 37.4 (2015): 95-103.

Rocscience Inc. 2018. Slide2 Version 2018 – 2D Limit Equilibrium Slope Stability Analysis. www.rocscience.com, Toronto, Ontario, Canada.

Soubra, A.H., and Massih, D.Y.A. 2010. Probabilistic analysis and design at the ultimate limit state of obliquely loaded strip footings. Géotechnique 60 (4): 275–285.

Stuedlein, A.W., Kramer, S.L., Arduino, P., and Holtz, R.D. 2012. Reliability of spread footing performance in desiccated clay. J. Geotech. Geoenviron. Eng. 138 (11): 1314–1325.

Vanmarcke, E. H. 1977. “Probabilistic Modeling of Soil Profiles.” ASCE Journal of the Geotechnical Engineering 103 (11): 1227–1246.

Vanmarcke, E.H. (1984). *Random Fields: Analysis and Synthesis,* MIT Press, Cambridge, Massachusetts.

Xiao, T., Li, D.Q., Cao, Z.J., Au, S.K. and Phoon, K.K. 2016. Three-dimensional slope reliability and risk assessment using auxiliary random finite element method. Comput. Geotech. 79 (Oct): 146–158.