Proceedings of the XVII ECSMGE-2019
Geotechnical Engineering foundation of the future
ISBN 978-9935-9436-1-3

doi: 10.32075/17ECSMGE-2019-0572

XVII European Conference
ECSMGE 2019 Reykjavik
on Soil Mechanics
and Geotechnical Engineering

Site spatial correlation estimation from CPT data using neural networks and random fields

Estimation de corrélation spatiale de site à partir de données CPT utilisant des réseaux de neurones et des champs aléatoires

J. D. Nuttall
Deltares, Delft, The Netherlands

ABSTRACT: Cone Penetration Test (CPT) data is commonly used to estimate the vertical and horizontal scales of fluctuation within site surveys. These measures of spatial correlation and variability are used in the modelling of soils especially in reliability methods, such as the Random Finite Element Method (RFEM). These parameters are conventionally estimated from CPT data, by fitting theoretical correlation functions to the site data. Presented is a new approach that trains a Convolutional Neural Network (CNN) with pseudo CPT data taken from generated 2D Random Field (RF) data with known scales of fluctuation. Once trained the network can predict these measures of spatial variability from real CPT data, with greater accuracy and with the need for fewer CPT measurements.

RÉSUMÉ: Les données issues de l'épreuve du pénétromètre statique (CPT) sont couramment utilisées pour estimer les échelles de fluctuations verticales et horizontales sur les sites étudiés. Ces mesures de corrélation et de variabilité spatiales sont utilisées dans la modélisation de sols, en particulier dans les méthodes de fiabilité, telles que la méthode des éléments finis aléatoires (RFEM). Ces paramètres sont conventionnellement estimés à partir des données d'une CPT en ajustant des fonctions de corrélation théoriques aux données du site étudié. Présentés sous forme d'une nouvelle approche, cette dernière entraine un réseau de neurones convolutifs (CNN) avec des pseudo-données d'une CPT produites à partir des données d'un champ bidimensionnel aléatoire (RF) pour lequel les échelles de fluctuation sont connues. Une fois entrainé, ce réseau peut prédire, avec une plus grande précision et avec un nombre restreint de CPT, ces mesures de variabilité spatiale contenue dans les données d'une vraie CPT.

Keywords: Scale of fluctuation; Spatial variability; Cone Penetration Test (CPT), Convolutional Neural Network (CNN); Random Fields;

1 INTRODUCTION

Soil is spatially varying, and as such it is important to consider this variability when modelling it's response. Methods such as the Random Finite Element Method (RFEM) (Fenton and Vanmarcke, 1999) and more recently the Random Material Point Method (RMPM) (Wang et al., 2016) predict the response of a geotechnical system incorporating this spatial variability. As such it is important to provide accurate measures of a sites spatial variability when undertaking a site survey for use in such a modelling approach.

The scale of fluctuation, θ, describes the distance over which a soil property remains significantly correlated (Vanmarcke, 1984). It is therefore an important parameter when describing the spatial variability of soils, and is used in the generation of random fields, which model them, such as in the Local Average Subdivision methods (LAS) (Vanmarke, 1977.)

Nuttall (2018) proposed a new method of estimating spatial correlation statistics from CPT data. This methodology involved training a 1D Convolutional Neural Network (CNN) using simulated CPT data, to estimate the vertical scale of fluctuation, θ_v. The simulated CPTs were generated using 1D Local Average Subdivision (LAS) random fields. The proposed method was shown to estimate θ_v, more accurately than more conventional methods, previously proposed, especially for larger scales of fluctuation.

This paper expands on this research by attempting to analyse a site survey using simulated CPT data generated using 2D LAS random fields, and shows that the proposed method can be beneficial, relative to traditional approaches, in particular when estimating the horizontal scale of a fluctuation, θ_h; which is considered difficult to estimate due to the available data from site surveys in the horizontal direction, and the need for appropriately spaced CPTs to capture the scale adequately.

Although in recent years several fitting algorithms have been proposed, especially for the horizontal scale of fluctuation, (Ching et al, 208, Lloret-Cabot et al, 2014), this paper will concentrate on the more traditional fitting approach for comparison, more specifically the fitting of an autocorrelation function.

This paper does not concentrate on a particular piece of CPT data as it is applicable to all CPT data parameters, e.g. CPT tip resistance.

2 AUTOCORRELATION FITTING METHOD

Commonly the scale of fluctuation is estimated by fitting a suitable autocorrelation function to a measured experimental covariance function, taken from the CPT data (in this case the cone tip resistance), considering the correlation, ρ, and lag, τ (Vanmarcke, 1977). Typical correlation functions are shown in Table 1.

Table 1. Commonly used correlation functions

Correlation Model	**Function**
Gaussian	$\rho(\tau) = e^{-\pi\left(\frac{\lvert\tau\rvert}{\theta}\right)^2}$
Markov	$\rho(\tau) = e^{-\frac{2\lvert\tau\rvert}{\theta}}$

These correlation functions are typical used in the generation of the random fields to model soils in methods such as the RFEM and RMPM. In this paper the Markov model is used in the 2D generation of the simulated site soils, from which simulated CPTs are taken.

In this approach the layer to be analysed is extracted from the CPT data, any depth trend is removed from the data, before a normalized correlation function, ρ̂, is estimated from the data, using:

$$\hat{\rho}(\tau) = \frac{\hat{\gamma}(\tau)}{\hat{\gamma}(0)} \tag{1}$$

where,

$$\hat{\gamma}(\tau) = \frac{1}{(t-1)}\sum_{j=1}^{k}(x_j - \hat{\mu})(x_{j+\Delta\tau} - \hat{\mu}) \tag{2}$$

with,

- $\hat{\gamma}$ is the experimental covariance function,
- x_j is the value in the CPT data at obervation point, j,
- $\hat{\mu}$ is the mean of the CPT data,

- τ is the lag distance,
- k is the number of observations, and
- t is the number of pairs of data at lag.

The Markov correlation model, shown in Table 1, is then fitted to the resulting experimental function generated using Equation 1, in both the horizontal and vertical directions, thus obtaining estimates for the scales of fluctuation, θ, in both directions. This procedure is illustrated in Figure 1, which depicts the fitting of the Markov correlation function, $\rho(\tau)$ to the measured experimental correlation function, $\hat{\rho}(\tau)$, thus minimizing the error between them, to obtain the most suitable value of θ.

In this paper this fitting is carried out using the functionality provided by the SciPy library within the Python environment.

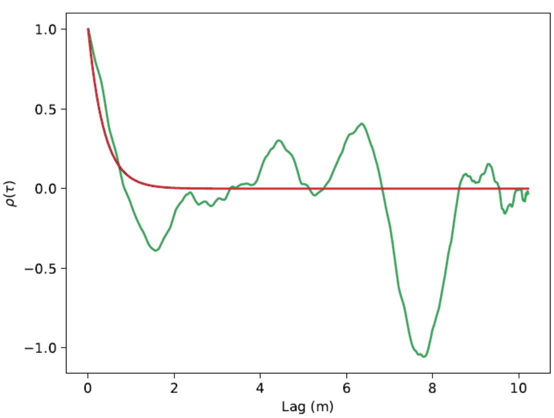


Figure 1. An illustration of the measured experimental correlation function (Green), with a fitted Markov correlation function.

3 PROPOSED CNN METHOD

The Convolutional Neural Network proposed is depicted in Figure 3. It is a series of filtering and perceptron layers, its structure and hyperparameters are beyond the scope of this paper. It is suffice to say that typically these types of neural network are used to recognise, and detect, features within, images. It is this functionality which the Author utilizes in the methodology to recognize and measure the features corresponding to the scale of fluctuation.

3.1 Training Data and Process

To train the CNN training data is required. This is generated using a 2D random field with known statistics, with no depth trend; from which regular columns of data are extracted to simulate a CPT site survey regime, as shown in Figure 2.

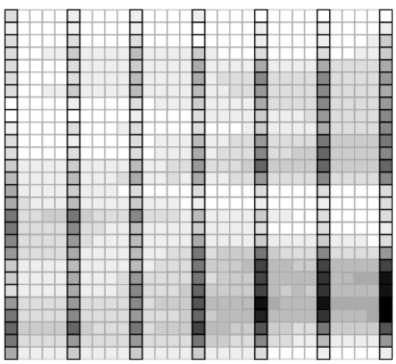


Figure 2. An illustration of the generation of a simulated CPT site survey for a single layer of soil, using a LAS generated 2D random field.

The random field is generated with a mean, $\mu = 0.0$, and standard deviation, $\sigma = 1.0$, and known scales of fluctuation; it is also assumed to have no vertical trend. Furthermore the field is rescaled to fall between 0 and 1. As such the point statistics should not influence the results of the data; thus Site Surveys with a variety of point statistics can be analysed. It is generated using a Markov correlation function, as shown in Table 1, and with a lognormal distribution.

Each random field is 512×512 cells, with a cell width of 0.01 m, thus representing a layer of 5.12 m in depth and width. From this layer 10 equidistant columns are taken, to represent the CPTs, thus each CPT is 1.0m apart and has a depth of 5.12 m. It should be noted however that

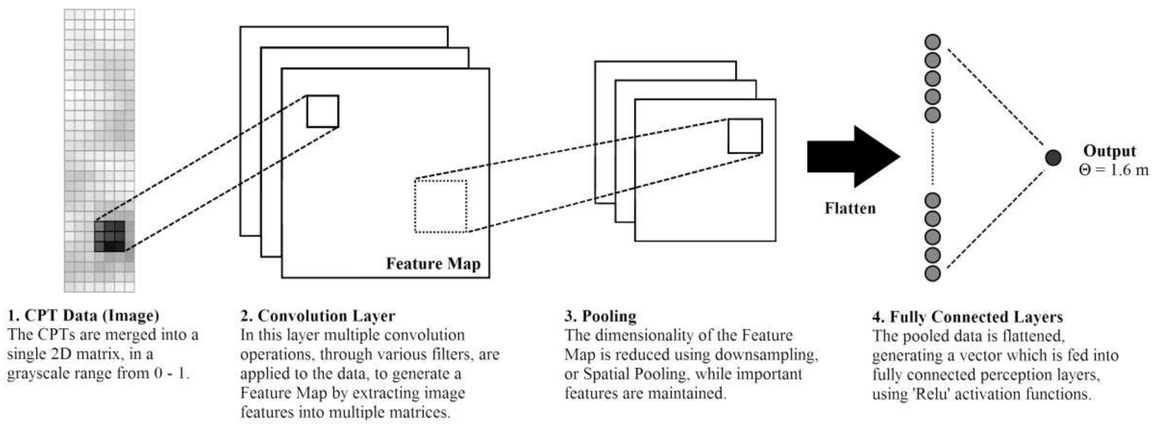


Figure 3. The structure of the Convolutional Neural Network (CNN) used to predict the coefficients of variation from CPT data in the proposed methodology.

the CPT would only have 512 sampling points in the vertical direction (i.e. around every 1cm), which is comparable to the readings from a typical CPT.

Two CNNs are trained, one for each plane, to predict θ_h and θ_v. Each CNN is trained with up to 200 epochs of 100,000 simulated CPT fields per epoch, differing at each epoch, i.e. 20,000,000 random fields, with the best results saved for use. In all cases the input was the simulated CPT values, scaled between 0 – 1, in matrix form, and the corresponding output as the relevant known statistic, i.e. θ_h or θ_v, also scaled.

As to utilize the mechanism efficiently the Author has trained the neural networks with θ_h and θ_v from 0.1 – 5.1 m, in this way it will be possible to scale the methodology based on the depth of layer and CPT readings and the spacing between reading horizontally when making the predictions. Thus the analysis is not restrained to a field with a 1m CPT spacing, or a depth of 5.12 m.

4 RESULTS AND COMPARISON

The two CNN's were trained and the method was used to estimate the scale of fluctuation, from 100 simulate site surverys; the results were compared with the known generated statistics, and the autofitting methodology.

4.1 Vertical scale of fluctuation θ_v

Figure 4 shows the comparison for the vertical scale of fluctuation. All the comparison figures show the range of the results, using ± 1 standard deviation of the collated results. The method was tested over the range $0.1m \leq \theta_v \leq 2.8m$, this the range of typical indicated by the literature review by Phoon and Kulhawy (1999).

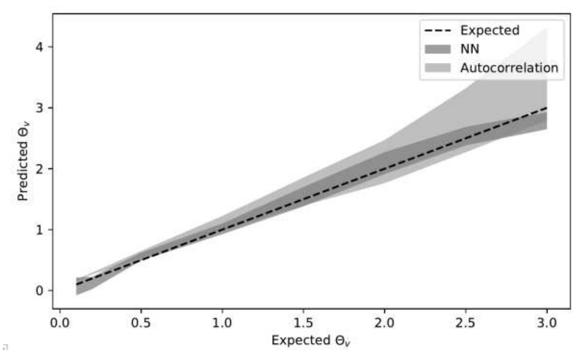
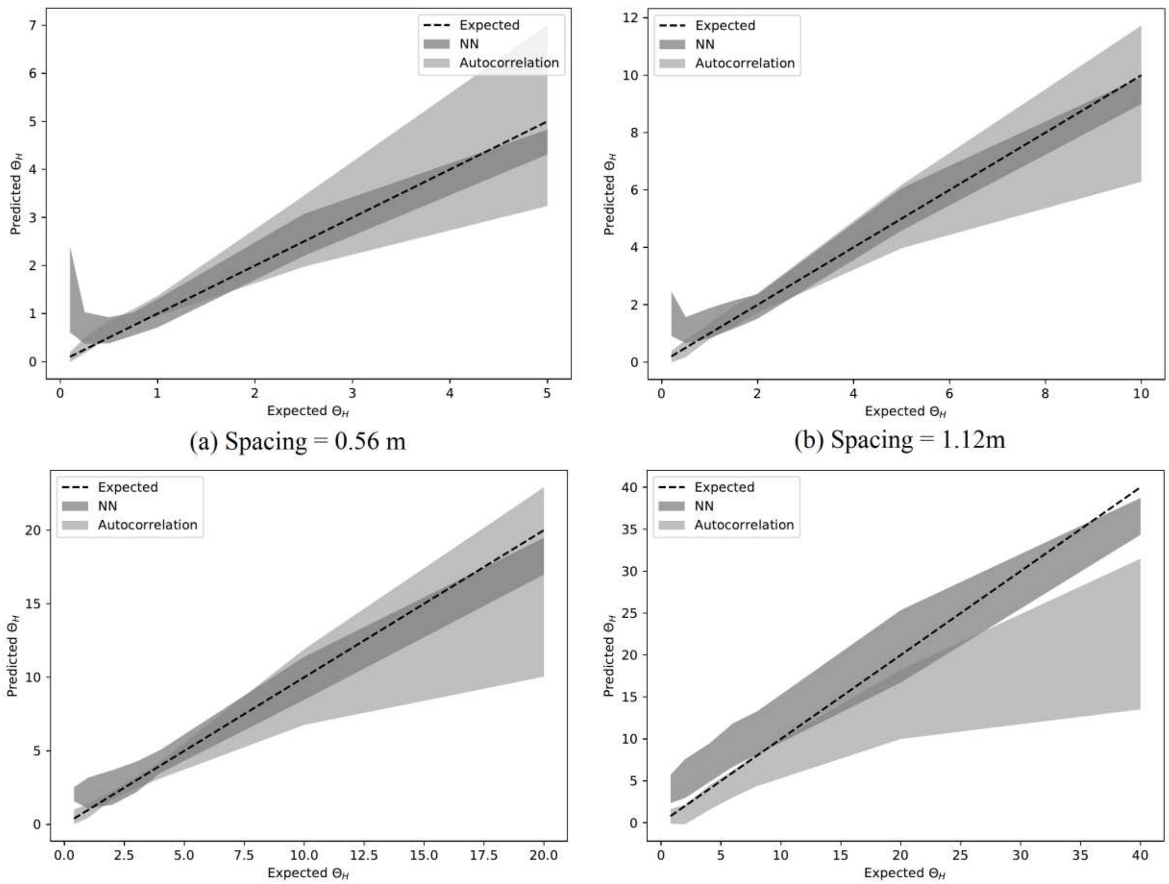


Figure 4. A comparison of the results from the proposed method and the autocorrelation fitting method over 100 simulated per tested θ_v.

The results show that although both methodologies perform well at lower levels of θ_v, the new analysis performs better at higher



(a) Spacing = 0.56 m (b) Spacing = 1.12m

(c) Spacing = 2.24m (d) Spacing = 4.48m

Figure 5. Graphs showing the comparison of θ_h predictions for different CPT spacing, over 100 simulated CPT sites.

levels, both in terms of accuracy and consistency, and are in general agreement with those in Nuttall (2018).

4.2 Horizontal Scale of Fluctuation θ_h

Figure 5 shows the results, as previously described, but in the horizontal direction. Here the results are based on the sampling of the 10 CPTs in the simulated site survey, at four different spacings. In the proposed method described, the CNN was trained using 0.56m spacings, illustrated in Figure 5(a), whereas the results in the other graphs are taken by scaling the results accordingly based on the distance.

The results show that autocorrelation method perform better at lower values of θ_h, while again the performance of the new methodology shows improved estimates in terms of consistency and accuracy at higher values.

The poorer CNN results at lower values in each range, can be explained by the fact that the θ_h value is less than the CPT spacings, as such the CNN is unable to detect the changing features in this period adequately to make a prediction.

An explanation for the poorer performance of the autocorrelation fitting method is due to the amount of data available at greater lags distances, for instance, in the vertical direction, there is only 1 sampling point per CPT with a distance of 5.12m compared with 512 per CPT at 0m. There is less data available to fit the correlation model sufficiently. It is a rule that the distance over which a variogram can be considered reliable is D/2, where D is the distance over which the CPT data is available. i.e. $\tau \leq D/2$. As such the curve fitting for $\theta > D/2$, would be outside the reliable data zone.

Furthermore in the horizontal direction there is only 10 sampling points per row of data, as such lag distances which are larger, again have fewer members, but the large sampling intervals means that it is more difficult to fit the data, particularly at greater distances.

The new method however differs in that it detects features within the data (image) that correspond to the correlation length, and as such can provide a more accurate estimation across a fuller range of data.

5 CONCLUSIONS

This paper describes the continued development of a new method for the estimation of the scales of fluctuation in both the horizontal and vertical directions, θ_h and θ_v. The results have shown that compared with the traditional method of fitting the autocorrelation function, the new methodology can provide a consistently more accurate estimate of the statistics, particularly for higher values of scales of fluctuation.

6 REFERENCES

Ching, J., Wu T.-J., Stuedlein, A.W., B. Taeho, 2018, Estimating horizontal scale of fluctuation with limited CPT soundings, *Geoscience Frontiers*, **9:6**, 1597-1608.

Fenton, G.A, Vanmarcke, E.H., 1990, Simulation of Random Fields Vai Local Average Subdivision. *ASCE Journal of Engineering Mechanics,* **116:8**, 1773-1749

Lloret-Cabot, M., Fenton, G.A., Hicks, M.A. 2014, On the estimation of scale of fluctuation in geostatistics, *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, **8:2**, 129-140.

Nuttall, J.D., 2018, Estimating spatial correlations from CPT data using neural networks and random fields, *Numerical Methods in Geotechnical Engineering: Proceedings of the 9th European Conference on Numerical Methods in Geotechnical Engineering* (Eds: Cardoso, A.S., Borges, J.L, Costa, P.A., Gomes, A.T., Marques, J.C., Vieira C.S.), 713-718, CRC Press, Leiden, the Netherlands.

Phoon, K.-K. and Kulhawy, F.H., 1999, Characterization of geotechnical variability. *Canadian Geotechnical Journal,* **36:4**, 612-624

Vanmarcke, E.H., 1977, Probablistic modelling of soil properties. *ASCE Journal of the Geotechnical Engineering,* **103:11**, 1227-1246.

Vanmarcke, E.H. 1984, *Random Fields: Analysis and Synthesis.* Cambridge, MA: The MIT Press.

Wang, B., Hicks, M.A., Vardon, P.J., 2016, Slope failure analysis using the random material point method. *Géotechnique Letters*, **6:2**, 113-118.