Proceedings of the 21st ICSMGE, Vienna, Austria, 14 – 19 June 2026. Pistrol, Adam & Schweiger (eds.)
Published by: ÖGG, Austrian Society for Geomechanics, Salzburg, Austria, ISBN 978-3-9503898-4-5
https://doi.org/10.53243/ICSMGE2026-1697

Superposition of Random Fields to model multi-output CPTu spatial variability

Stefano Collico
Department of Civil and Environmental Engineering, Università degli studi di Firenze, Florence, Italy.

Lluís Monforte
Centre Internacional de Mètodes Numèrics en Enginyeria (CIMNE), Barcelona, Spain.

Dani Tarragó
Department of Civil and Environmental Engineering, UPC, Barcelona, Spain.

Amadeu Deu
Geociencias y Exploraciones Marítimas (GEM), Barcelona, Spain.

ABSTRACT: Accurately assessing and quantifying Cone Penetration Test (CPT) spatial variability is essential for reliability-based design. Among recent approaches, Gaussian Processes (GP) have emerged as an attractive method for random field modeling due to their intuitive application and workflow. Following previous works on the topic, this study introduces a multi-output extension of Gaussian Process to model the joint spatial variability of cone tip resistance, sleeve friction and pore pressure measurements. The random field is modeled as superposition of trend and stochastic components explicitly accounting for outputs interdependence through latent functions according to a Linear Model of Coregionalization. The methodology is first validated on a synthetic case study and subsequently applied to in-situ geotechnical data acquired at Barcelona harbor (Spain) highlighting the need to further implement approximate solutions for handling large datasets and multiple in-situ testing.

KEYWORDS: Random Fields, CPTu, spatial variability, Multi-output Gaussian Process.

1 INTRODUCTION

Site characterization from in-situ testing is a fundamental step at every project site. Among all, Cone Penetration Testing (CPTu) is one of the most informative tests to indirectly estimate geotechnical parameters and derive project-specific mechanical soil profiling, according to well-established Soil Behavior Types templates (e.g., Robertson 2016). Despite providing insightful vertical spatial information of soil response to the cone probe, the quantification of parameters to model horizontal spatial variability of CPTu measurements and corresponding uncertainties remains a key challenge for random field modeling (Vanmarcke 1983) and applicability of reliability-based design.

The variability of a soil parameter is usually modeled as a trend function and a zero-mean oscillating component about the trend (the so called-spatial variability). As highlighted by Cami et al. (2020), Ching and Phoon (2018) and Ching et al. (2019), the quantification of spatial variability is usually identifiable through two main parameters: the scale of fluctuation, to be used within an autocorrelation function (ACF) model, and a roughness parameter.

Several methods have been proposed for the computation of the scale of fluctuation from CPTu measurements: the method of moments (e.g., Uzielli et al. 2005, Lloret-Cabot et al. 2014), the maximum likelihood estimation method (e.g., DeGroot & Baecher, 1993) and Bayesian analysis (e.g., Tian et al. 2016). In this context, a key finding was illustrated by Ching et al. (2017a) by introducing the definition of an "identifiability problem" for scale of fluctuation estimation. Such identifiability issues arise in the general case—when both the trend function and the autocorrelation model are unknown—especially under sparse data conditions, and they only become identifiable when multiple CPT soundings are available.

However, even with additional soundings, the trend function selected to detrend the data and extract residuals for scale-of-fluctuation computation influences the results. Moreover, because the estimation procedures require the assumption of an autocorrelation model, the chosen model also impacts the estimates, with the degree of sensitivity depending on the estimation method used (see Cami et al. 2020 for an exhaustive overview of such concepts).

The identification of all random field parameters simultaneously without any assumptions is then a challenging task. Several studies have been proposed to tackle such problem complexity. Among all, Ching and Phoon (2017) initially proposed Bayesian learning for probabilistically modeling the spatial trend and further extending such concept to three-dimensional modeling (Ching et al., 2020, Ching et al. 2023a). Wang and Zhao (2016, 2017) proposed a compressive sampling technique, by equivalently comparing in-situ measurements to an incomplete signal, to quantify soil property variability profile from limited measurement data and extended this method for 2D sparse measurements.

In the same context, a different approach was proposed by Yoshida et al. (2021) who developed a method for simultaneously estimating the trend and random components using Gaussian Process (GP) regressions with the superimposition of Gaussian random fields. The trend component is then conceived as a random field with a large scale of fluctuation (SOF), whereas the stochastic component is a random field with a small SOF, that is, a rapidly changing distribution.

A different approach was also proposed by Zinas et al. (2025) who implemented a stochastic variational learning technique to approximate the posterior distribution of multi-output GP (MOGP), while handling heterotopic asymmetrical datasets to jointly predict categorical borehole information and CPT parameters. A key advantage of MOGP is to exploit parameters dependence such that the target outputs can leverage information from one another.

Following the same philosophy of Yoshida et al. (2021), this work applies a multi-output approach for CPTu parameters spatial variability and model their joint dependence through Linear Model of Coregionalization. The trend and stochastic components are treated as unknown random fields. The proposed MOGP aims to leverage on the dependence of CPTu parameters and therefore outperforming the standard independent analysis usually performed. The work is initially

tested on a simplified synthetic case study, where trend and stochastic components are known, and then applied to the highly heterogenous Muelle Prat site (Barcelona, Spain), exploiting limitation of proposed approach. Despite being of main importance, the statistical uncertainties of estimated random field parameters, are not assessed, and are left for future research.

2 METHODOLOGY

2.1 *Single-Output Gaussian Process*

Single-output Gaussian Process (GP) aims to approximate the target output $f(x)$, where $x \in \mathrm{R}^d$, by interpreting it as a probability distribution in function space as (Liu et al. 2018):

$$f(x) \sim \mathrm{GP}(m(x), k(x,x')) \quad (1)$$

$$\mathrm{y(x)} = f(x) + \varepsilon \quad (2)$$

The target output $f(x)$, that approximates observations y at known location x with identically distributed (i.i.d.) noise $\varepsilon \sim \mathrm{N}(0, \sigma_\varepsilon^2)$, is defined by a mean function $m(x)$, usually taken as zero without loss of generality, and the covariance function $k(x,x')$, which can assume different forms depending on assumed autocorrelation function model (i.e., squared exponential, Markovian, binary noise, Matern function). The covariance function should be then constructed according to some unknown parameters θ (e.g., scale of fluctuation and variance for squared exponential function) by fitting the function model to data. Given a set of observations $X = [x_1, x_t, \ldots, x_n] \in \mathrm{R}^{nxd}$, and $y = [y_1, \ldots, y_n]^T \in \mathrm{R}^n$, the posterior distribution at a new location x_*, $f(x_*)$ can be analytically derived as (Liu et al. 2018):

$$f(x_*)|X, y, x_* \sim \mathrm{N}(\hat{f}(x_*), \sigma^2(x_*)) \quad (3)$$

with $\hat{f}(x_*)$, predictive mean and $\sigma^2(x_*)$ predictive variance at a new location x_* that read:

$$\hat{f}(x_*) = k_*^T [K(X,X) + \sigma_\varepsilon^2 I]^{-1} y \quad (4a)$$

$$\sigma^2(x_*) = k(x_*, x_*) - k_*^T [K(X,X) + \sigma_\varepsilon^2 I]^{-1} k_* \quad (4b)$$

where $k_* = k(X, x_*) \in \mathrm{R}^{nx1}$ denotes the covariance between the new location x_* and the n training datapoints and $K(X,X) \in \mathrm{R}^{nxn}$ covariance matrix.

2.2 *Multi-output Gaussian Process*

A CPTu allows to record three independent measurements (i.e. cone tip resistance q_t, sleeve friction f_s, and pore pressure u_2). Therefore, for a single CPTu performed, three outputs will be available. Liu et al. (2018, 2023) defines such problem as a Multi-Output Gaussian Process (MOGP), which intends to approximate T outputs simultaneously. Since only CPTu are analyzed herein, the problem is simplified and categorized as symmetric, isotopic (i.e. T outputs share the same training set) scenario within Gaussian Process regression notation. Nevertheless, the approach can be extended for a heterotopic and asymmetric training set condition to include different in-situ tests and/or geological information (see Zinas et al. 2025 as an example of application to in-situ testing).

Given a set of isotopic training set $X = X_1, X_2 = \ldots = X_T \in \mathrm{R}^{nxd}$, with $n = n_t$ and $x_{t,i} = x_i$ for t = 1, ..., T the latent MOGP $f(x)$ reads (Liu et al. 2018):

$$f(\mathrm{x}) = \begin{bmatrix} f_1(x) \\ \cdot \\ \cdot \\ \cdot \\ f_T(x) \end{bmatrix} \sim \mathrm{GP}(0, k_M(\mathrm{x}, \mathrm{x}')) \quad (5a)$$

$$k_M(\mathrm{x}, \mathrm{x}') = \begin{bmatrix} k_{11}(x,x') & \ldots & k_{1T}(x,x') \\ \vdots & \ddots & \vdots \\ k_{T1}(x,x') & \ldots & k_{TT}(x,x') \end{bmatrix} \quad (5b)$$

with $k_M(x,x') \in R^{TxT}$ multi-output covariance matrix with elements $k_{tt'}(x,x')$ that quantify the degree of similarity between target outputs $f_t(x)$ and $f_{t'}(x')$.

The t-th output observations are then sampled as:

$$y_t(\mathrm{x}) = f_t(x) + \varepsilon_t \quad (6)$$

with ε_t i.i.d. noise $\sim \mathrm{N}(0, \sigma_t^2)$ associated with the t-th output. According to such structure, the predictive mean and variance at a new location x_* are analytically derived following the analogous procedure of single output GP, such that:

$$\mathrm{f}(x_*)|\,\mathrm{X}, \mathrm{y}, x_* \sim \mathrm{N}(\hat{f}(x_*), \Sigma_*) \quad (7)$$

$$\hat{f}(x_*) = K_M^T [K_M(X,X) + \Sigma_M]^{-1} y \quad (8)$$

$$\Sigma_* = K_M(x_*, x_*) - K_{M*}^T [K_M(X,X) + \Sigma_M]^{-1} K_{M*} \quad (9)$$

where $K_M(X,X) \in \mathrm{R}^{nTxnT}$ is the full muti-output covariance matrix, $K_{M*} = K_M(X, x_*) \in \mathrm{R}^{nTxT}$, has blocks $k_{tt'}(X, x_*)$ and $K_M(x_*, x_*) \in \mathrm{R}^{TxT}$ has elements $k_{tt'}(x_*, x_*)$.

It is then intuitive to observe that the effectiveness of MOGP relies on selecting an admissible multi-output covariance function $k_M(\mathrm{x}, \mathrm{x}')$ defines a positive semi-definite covariance matrix $K_M(X,X)$ and enables the transfer of information among T outputs.

Let now assume that the target outputs $f(x)$, is composed of superposition of N independent random fields (Yoshida et al. 2021). In this study a trend component $f_{tr}(\mathrm{x}) \sim \mathrm{GP}(0, K_M^{tr}(\mathrm{x}, \mathrm{x}'))$, and stochastic component $f_{st}(\mathrm{x}) \sim \mathrm{GP}(0, K_M^{st}(\mathrm{x}, \mathrm{x}'))$ are assumed such that N=2.

The total multi-output kernel of Equation (5b) and consequently the kernel terms appearing in Equation (8, 9) can be expressed as the sum of the component kernels:

$$k_M(\mathrm{x}, \mathrm{x}') = \sum_{i=1}^N K_M^i(\mathrm{x}, \mathrm{x}') \quad (10a)$$

$$K_{M*} = \sum_{i=1}^N K_M^i(X, x_*) \quad (10b)$$

$$K_M(x_*, x_*) = \sum_{i=1}^N K_M^i(x_*, x_*) \quad (10c)$$

2.3 *Linear Model of coregionalization*

A widely used approach for constructing admissible multivariate covariance functions $k_M(\mathrm{x}, \mathrm{x}')$ is known as linear model of coregionalization (LMC). When the outputs are represented as the superposition of N random fields, the LMC expresses each output as a linear combination of Q latent Gaussian processes ($Q \leq T$), (Figure 1):

$$f_t(\mathrm{x}) = \sum_{i=1}^{N} \sum_{q=1}^{Q} a_{tq}^{(i)} u_q^{(i)}(x) \quad (11a)$$

where $u_q^{(i)}(x)$ is the q-th latent function of the i-th random field and a_{tq}^i are the mixing coefficients linking latent function q of component i to output t. Each latent function is modeled as a Gaussian process $u_q^i(x) \sim \mathrm{GP}(0, k_q^{(i)}(x,x'))$. Using the LMC construction, the multi-output covariance function of Equation (5b) becomes:

$$k_M(\mathrm{x}, \mathrm{x}') = \sum_{i=1}^N \sum_{q=1}^Q A_q^{(i)} k_q^{(i)}(x,x') \quad (12)$$

with $A_q^{(i)} \in R^{TxT}$ coregionalization matrix symmetric and PSD, wherein the (t, t') element $\left[A_q^{(i)}\right]_{tt'} = a_{tq}^{(i)} a_{t'q}^{(i)}$. For simplicity, the parameters defining the MOGP ($A_q^{(i)}, \sigma_t^2$ and parameters that define $k_q^{(i)}(x, x')$) will be referred to as *hyperparameters* and collectively denoted by θ.

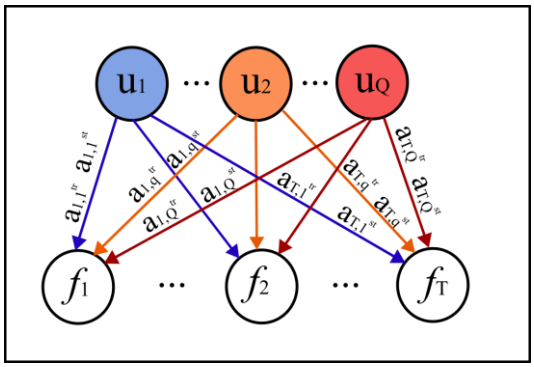

Figure 1. Graphical model for superposition MOGP.

3 SYNTHETIC CASE SCENARIO

Let's assume six sounding records along a 2D section for each parameter. Therefore, for each test, three different measurements are available (i.e., T=3, as for real CPTu), characterized by different magnitude, trend and stochastic components. Each parameter (i.e., T_1, T_2, T_3) assumes a quadratic vertical trend and linear horizontal trend, with different coefficients (see Equation 13a-c). Although the Whittle–Matérn (WM) model generally provides a more flexible and realistic representation of spatial correlation of geotechnical parameter (Ching et al. 2023), the squared exponential autocorrelation function is employed in this study to reduce computational cost, while still allowing for different horizontal and vertical scales of fluctuation for each parameter (Equation 13d):

$$T_1(x, z) = 7 + 4z + 4z^2 + 0.5x \quad (13a)$$

$$T_2(x, z) = 2.5 + 2z + z^2 + 0.1x \quad (13b)$$

$$T_3(x, z) = 7 - 0.4z - 2z^2 + 2x \quad (13c)$$

$$k_i(x, x') = \exp\left[-\pi\left(\sqrt{\frac{\Delta_x^2 + \Delta_y^2}{\delta_{h-i}^2}} + \frac{|\Delta_z|}{\delta_{v-i}}\right)\right] \text{ with i=1,..,T} \quad (13d)$$

An overview of trend, stochastic components and of the total random field of each parameter is reported in Figure 2. For such scenario low noise values are considered for each output (i.e. σ_t^2 1e-4). Out of six generated sounding records, five were taken for MOGP training and reported in Figure 3a. The overall true spatial variability of each parameter is reported in Figure 3b-d. True *hyperparameters* to be identified are listed in Table 1. Noting that a normalization of coordinates has been assumed, typical of Gaussian Process workflow (see Zinas et al. 2025).

A total of Q = T latent functions are employed since only three CPTu parameters are considered as outputs. The maximum likelihood estimation method is employed for the estimation of trend and stochastic *hyperparameters*. To assess the varying magnitudes of three parameters, which may degrade the computation of latent coefficients A_q^i across outputs, a well-known cumulative distribution function (CDF) transform is applied during MOGP training (see Betz et al. 2016, Collico et al. 2025 for application examples). Results of total (i.e., sum of stochastic and trend components) and the estimated trend function at the validation location are reported in Figure 4 for each parameter.

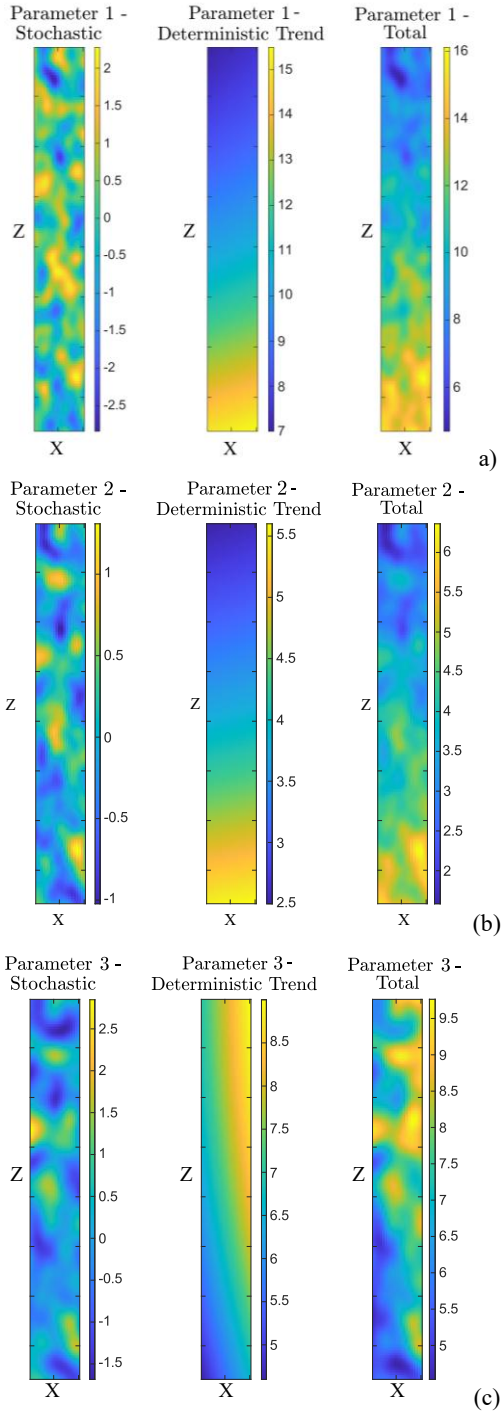

Figure 2. Stochastic, Trend components and Total Random Field: a) Parameter 1. b) Parameter 2. c) Parameter 3.

The estimated *hyperparameters* are reported in Table 1. The predicted trend profile along depth (red line) at validation location remarkably identify the true trend (blue line), with a clear identification of the reverse trend that characterizes the T_3 parameter. By considering the total random field, a good resemblance is obtained with respect to referenced validation sounding record. For the stochastic component, estimated *hyperparameters* (Table 1), well agree with true values, particularly for vertical scale fluctuation. A minor difference can be observed for the estimated horizontal scale of fluctuation, whose accuracy and precision are strongly

dependent on number of CPTu profiles employed within MOGP training. A slight difference between true and predicted diagonal coefficients of matrix A_{stoch} without altering the general good prediction capability of the proposed method.

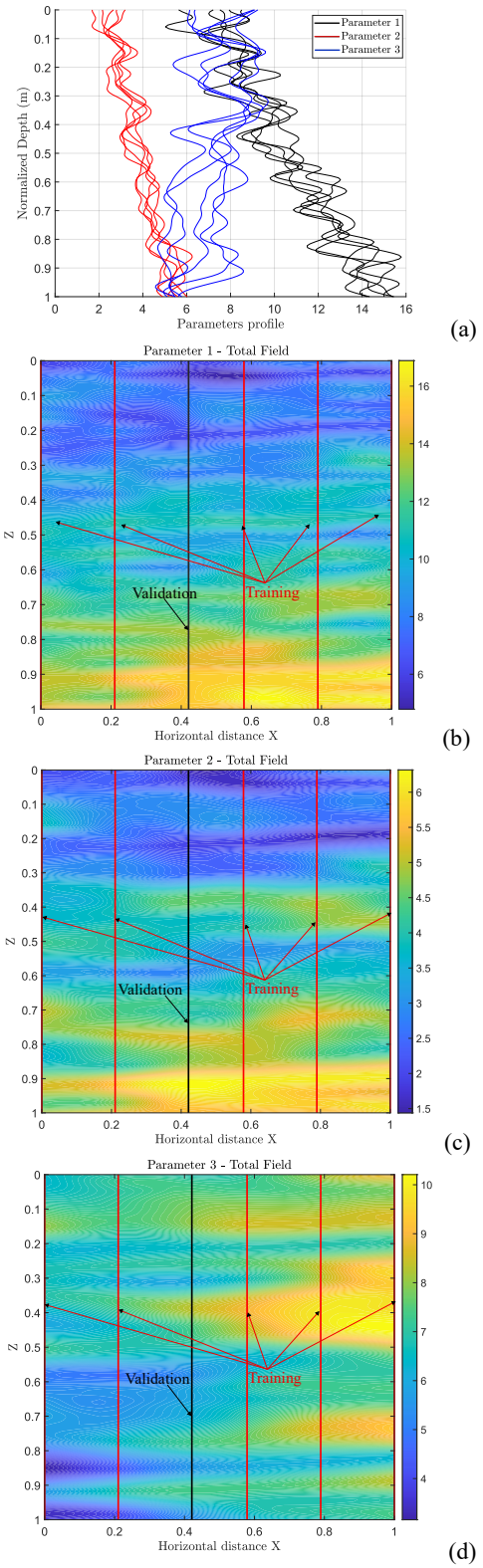

Figure 3. a) Synthetic CPTu profiles employed for MOGP training. b) Parameter 1. c) Parameter 2. d) Parameter 3.

Such simplified case highlights the potential applicability of random field superposition to model multiple CPTu data spatial variability. Nevertheless, the synthetic case scenario provides a rather simplistic example, coherent with assumptions that the trend is smooth and characterized by long scale of fluctuation and analogous polynomial form.

Table 1. Summary of estimated vs True *hyperparameters.*

Hyperparameter	True	Estimated
$\delta_{h1}\ \delta_{h2}\ \delta_{h3}$	4,6,10	3.87, 5.83, 10.4
$\delta_{v1}\ \delta_{v2}\ \delta_{v3}$	0.5, 0.8, 0.9	0.5, 0.78, 0.89
A_{stoch}	$\begin{bmatrix} 0.8 & 0.2 & 0.1 \\ 0.2 & 0.3 & 0.4 \\ 0.1 & 0.4 & 0.8 \end{bmatrix}$	$\begin{bmatrix} 0.14 & 0.19 & 0.15 \\ 0.19 & 0.28 & 0.36 \\ 0.15 & 0.36 & 0.69 \end{bmatrix}$
$\delta_{h\,tr1}\ \delta_{h\,tr\,2}\ \delta_{h\_tr\,3}$	- (imposed trend)	21, 26, 27
$\delta_{v\_tr\,1}\ \delta_{v\_tr\,2}\ \delta_{v\_tr\,3}$	- (imposed trend)	18.6, 10.6, 9.9
A_{tr}	−(imposed trend)	$\begin{bmatrix} 1 & 0.99 & -0.45 \\ 0.99 & 0.96 & -0.51 \\ -0.45 & -0.51 & 0.67 \end{bmatrix}$

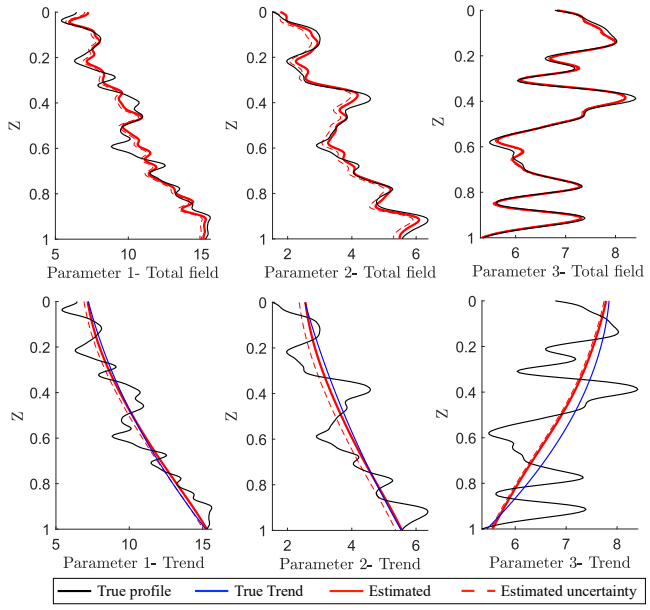

Figure 4. Validation and generated profiles at validation location. a) Parameter 1. b) Parameter 2. c) Parameter 3.

4 BARCELONA HARBOR

Within the project for the expansion of Barcelona harbor, a large offshore test campaign was carried out for soil characterization in 2018. The geotechnical campaign comprises laboratory test, CPTu and Dilatometer Marchetti Tests (DMT) and boreholes. The Muelle Prat site is investigated since several studies at such site were published (Deu et al. 2021, Peña et al. 2021, Peña and Asanza 2021, Collico & Arroyo 2023, Collico et al. 2024). A simplified overview of the geotechnical campaign is provided in Figure 5. The relative distance between CPTu varies between 30-80m, while an average penetrated depth of about 38m is considered. Regarding the study, the section AA' is considered for CPTu spatial variability quantification. CPTu sounding records of normalized cone tip resistance, Qt_1, sleeve friction FR, as well as pore pressure B_q, are reported in Figure 5b. As reported by several studies (Peña et al. 2021, Collico et al. 2022, 2024), the first 15m depth are characterized by a strong alternation of clay-like, soil-mixture and sand-like soil units, making particularly challenging the detrending process. The CPTu variability is then quantified, assuming a squared exponential correlation model function for each parameter. As for the synthetic case study, a total of Q =3

latent function is assumed and the logarithm of Qt_1, FR and $1+B_q$ is taken to facilitate latent coefficients computation. To reduce the high computational cost of K_M matrix inversion, which is of the order $O(nT)^3$), the CPTu measurements are taken every 20 cm. Each parameter profile is then composed of 192 measurements for a total of 2883 training datapoints when considering the three parameters.

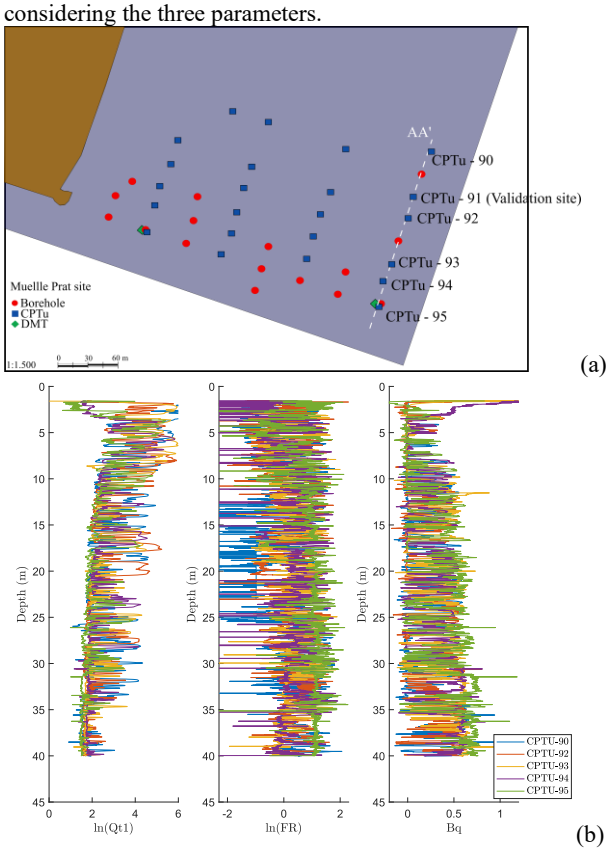

Figure 5. a) Overview of geotechnical campaigns at Muelle Prat site. b) CPTu profiles considered for MOGP training.

A visual representation of spatial variability for the three outputs in terms of predicted mean, according to Equation (8), is reported in Figure 6a, b, c. Summary of *hyperparameters* of trend and stochastics components are reported in Table 2. The profile of trend and stochastic components of each parameter at the validation location (i.e. CPTu-91) is reported in Figure 7.

Estimated *hyperparameters* in Table 2 describe the spatial characteristics and cross-output correlations of latent functions underlying the three CPTu parameters. For the stochastic component, the latent scale of fluctuation in the horizontal and vertical directions reveal distinct spatial behaviors across the three latent functions. For instance, latent function q_1 exhibits a very large δ_h, implying slow variation along horizontal distance, whereas q_2 and q_3 have shorter δ_h values indicating more localized variability. Vertically, all stochastic latent functions have relatively short δ_v suggesting more rapid variation with depth. For the trend component, the differences are more pronounced: latent function q_1 shows a very large δ_v compared to q_2 and q_3 indicating that this latent varies smoothly with depth. Conversely, δ_h is also large, suggesting broad-scale variation in both directions. By contrast, q_3 has a much smaller horizontal length scale implying more localized horizontal structure in the trend. Coregionalization matrices A_{stoch} and A_{trend} quantify the degree to which the latent functions contribute jointly to the different model outputs. Off-diagonal entries indicate substantial cross-output coupling for the corresponding latent.

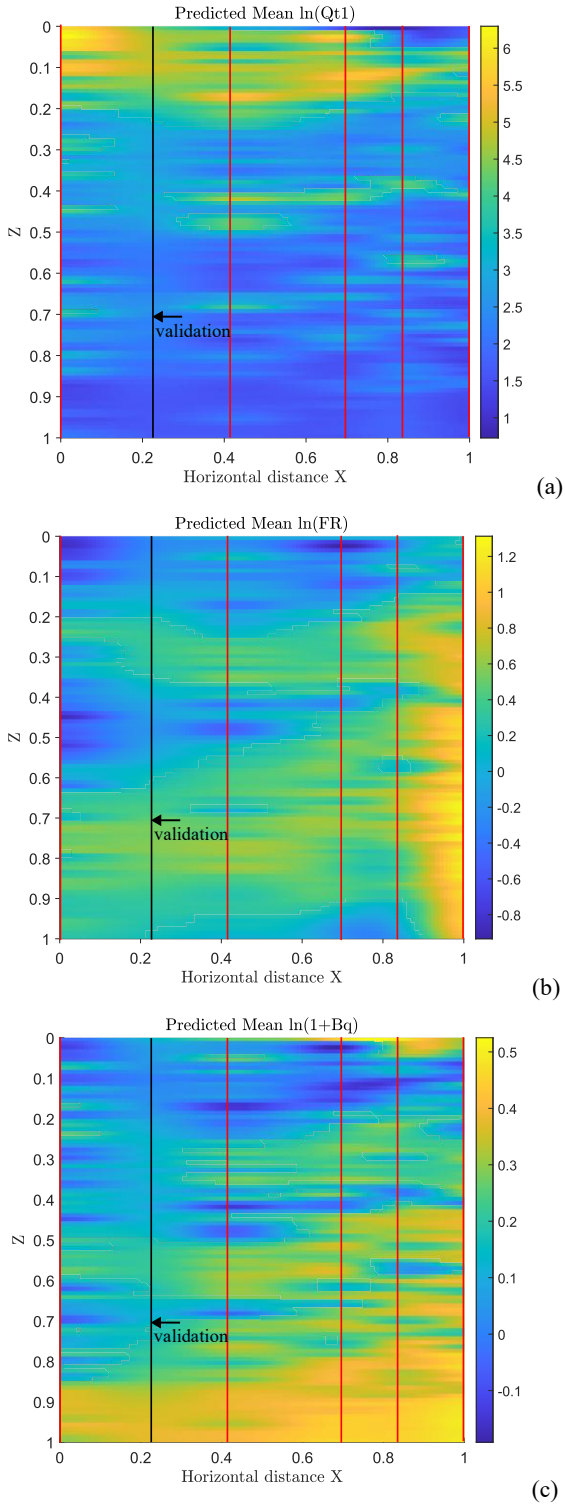

Figure 6. Spatial variability of normalized CPTu parameters: a) ln (Qt1). b) ln (FR). c) ln(1+Bq).

5 CONCLUSIONS

This work presents a multi-output Gaussian process with a superposition of random fields to jointly capture the trend and stochastic components of multiple CPTu parameters spatial variability. Preliminary results, obtained on a synthetic case study and then applied to a highly heterogeneous site, indicate that the proposed approach effectively models the spatial interdependence among CPTu parameters. However, when applied to complex lithological sequences—such as the one

examined—further refinements may be needed to improve trend identifiability. In addition, the analysis covers only a small portion of the full geotechnical campaign illustrated, for which non-negligible computational costs are already emerging. Stochastic variational learning could help manage such large datasets and enable the integration of additional in-situ tests. Most importantly, the approach should be extended to rigorously assess the statistical uncertainty in hyperparameter estimation, which has proven to be critical when only limited data are employed.

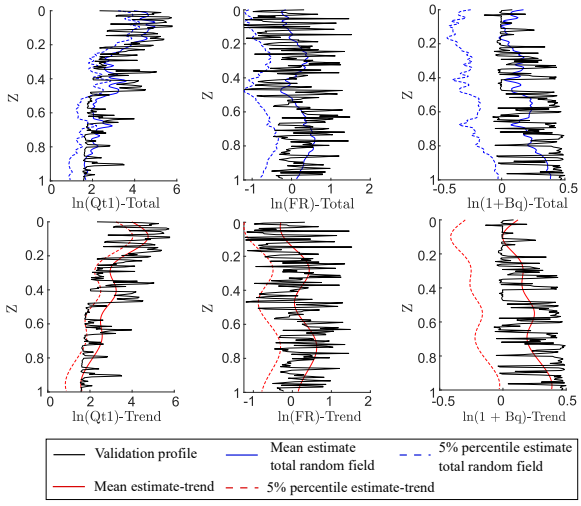

Figure 7. Estimated trend and total random field for the three-parameter at validation location.

Table 2. Estimated *hyperparameters* of each latent function q.

Hyperparameters	Estimated
$\delta_{h\_q1}\ \delta_{h\_q2}\ \delta_{h\_q3}$	70.7, 22.5, 26
$\delta_{v\_q1}\ \delta_{v\_q2}\ \delta_{v\_q3}$	0.96, 0.68, 0.38
A_{stoch}	$\begin{bmatrix} 0.38 & -0.17 & -0.33 \\ -0.17 & 0.13 & 0.09 \\ -0.33 & 0.09 & 0.36 \end{bmatrix}$
$\delta_{h\ tr\_q1}\ \delta_{h\ tr\ q2}\ \delta_{h\_tr\ q3}$	192, 186, 42
$\delta_{v\_tr\ q1}\ \delta_{v\_tr\ q2}\ \delta_{v\_tr\ q3}$	14.4, 4.48, 5.51
A_{tr}	$\begin{bmatrix} 0.83 & -0.24 & -0.44 \\ -0.24 & 0.53 & 0.33 \\ -0.44 & 0.33 & 0.51 \end{bmatrix}$
$\varepsilon_{q1}\ \varepsilon_{q2}\ \varepsilon_{q3}$	0.41, 0.74, 0.59

6 REFERENCES

Betz, W., Papaioannou, I., & Straub, D. (2016). Transitional Markov chain Monte Carlo: observations and improvements. *Journal of Engineering Mechanics*, *142*(5), 04016016.

Cami, B., Javankhoshdel, S., Phoon, K. K., & Ching, J. (2020). Scale of fluctuation for spatially varying soils: estimation methods and values. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, *6*(4), 03120002.

Ching, J. and Phoon, K. K. 2017. Characterizing uncertain site-specific trend function by Sparse Bayesian Learning, Journal of Engineering Mechanics, ASCE, 143(7), 04017028.

Ching, J., Phoon, K.K., Beck, J.L., and Huang, Y. 2017a. Identifiability of geotechnical site-specific trend functions. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, 845 Part A: Civil Engineering, 3(4): 04017021

Ching, J., Phoon, K.K., and Sung, S.P. 2017b. Worst case scale of fluctuation in basal heave analysis involving spatially variable clays. Structural Safety, 68: 28-42.

Ching, J., and Phoon, K.K. 2018. Impact of Autocorrelation Function Model on the Probability of Failure. Journal of Engineering Mechanics, 145(1): 04018123.

Ching, J., Phoon, K.K., Stuedlein, A.W., and Jaksa, M. 2019. Identification of sample path smoothness in soil spatial variability. Structural Safety, 81: 101870.

Ching, J., Huang, W. H., and Phoon, K. K. 2020. Three-dimensional probabilistic site characterization by sparse Bayesian learning. ASCE Journal of Engineering Mechanics, under 858 review.

Ching, J., Yoshida, I., & Phoon, K. K. (2023). Comparison of trend models for geotechnical spatial variability: Sparse Bayesian learning vs. Gaussian process regression. *Gondwana Research*, *123*, 174-183.

Collico, S., Arroyo, M., DeVincenzi, M., Rodriguez, A., & Deu, A. (2022). Probabilistic delineation of soil layers using Soil Behavior Type Index. In *Cone Penetration Testing 2022* (pp. 332-338). CRC Press.

Collico, S., & Arroyo, M. (2023). Bayesian Mixture Analysis of a global database to improve unit weight prediction from CPTu. *Engineering Geology*, *327*, 107353.

Collico, S., Arroyo, M., & Devincenzi, M. (2024). A simple approach to probabilistic CPTu-based geotechnical stratigraphic profiling. *Computers and Geotechnics*, *165*, 105905.

Collico, S., Spagnoli, G., & Monforte, L. (2025). Automating site characterization from pile field data. *Computers and Geotechnics*, *187*, 107498.

Deu, A., Marti, X., Peña, S., Tarragò, D., Gens, A., & Devincenzi, M. (2021). DMT, CPTU and laboratory tests comparison for soil classification and strength parameters of deltaic soft soils in Barcelona Port. In *Proceedings of the 6th International Conference on Site Characteriza tion ISC-6 Budapest*.

DeGroot, D. J., & Baecher, G. B. (1993). Estimating autocovariance of in-situ soil properties. *Journal of Geotechnical Engineering*, *119*(1), 147-166.

Lloret-Cabot, M.F.G.A., Fenton, G.A., and 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.

Liu, H., Cai, J., & Ong, Y. S. (2018). Remarks on multi-output Gaussian process regression. *Knowledge-Based Systems*, *144*, 102-121.

Liu, H., Wu, K., Ong, Y. S., Bian, C., Jiang, X., & Wang, X. (2023). Learning multitask Gaussian process over heterogeneous input domains. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *53*(10), 6232-6244.

Peña, S. Deu, A., & Devincenzi, M. (2021) Practical approach for soil characterization with multivariate analysis. In *Proceedings of the 6th International Conference on Site Characteriza tion ISC-6 Budapest*.

Pena F. & Asanza E. (2021). TÉCNICAS DE APRENDIZAJE AUTOMÁTICO PARA CARACTERIZACIÓN Y PERFILADO DEL TERRENO. APLICACIÓN PRÁCTICA AL CPTU. *Geotecnia*, (152), 585-615.

Robertson, P. K. (2016). Cone penetration test (CPT)-based soil behaviour type (SBT) classification system—an update. *Canadian Geotechnical Journal*, *53*(12), 1910-1927.

Tian, M., Li, D.Q., Cao, Z.J., Phoon, K.K., and Wang, Y. 2016. Bayesian identification of random field models using indirect test data. Engineering Geology, 210, 197-211.

Uzielli, M., Vannucchi, G., and Phoon, K.K. 2005. Random field characterisation of stress normalised cone penetration testing parameters. Géotechnique 55 (1): 3–20.

Vanmarcke, E.H., 1983. Random Fields: Analysis and Synthesis, MIT Press, Cambridge, USA.

Wang, Y., and Zhao, T. 2016. Interpretation of soil property profile from limited measurement 1373 data: a compressive sampling perspective. Canadian Geotechnical Journal, 53(9): 1547-1559.

Wang, Y. and Zhao, T. 2017. Statistical interpretation of soil property profiles from sparse data using Bayesian Compressive Sampling. Geotechnique, 67(6), 523-536.

Yoshida, I., Tomizawa, Y., & Otake, Y. (2021). Estimation of trend and random components of conditional random field using Gaussian process regression. *Computers and Geotechnics*, *136*, 104179.

Zinas, O., Papaioannou, I., Schneider, R., & Cuéllar, P. (2025). Multivariate Gaussian Process Regression for 3D site characterization from CPT and categorical borehole data. *Engineering Geology*, *352*, 108052.