TU Delft

Delft University of Technology

Document Version
Final published version

Licence
CC BY

Citation (APA)
Durmaz, M., Guan, Z., & Hicks, M. A. (2026). CPT-based two-dimensional soil stratification using non-stationary mixture of Gaussian processes. *Computers and Geotechnics*, *200*, Article 108459. https://doi.org/10.1016/j.compgeo.2026.108459

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

Computers and Geotechnics 200 (2026) 108459

Contents lists available at ScienceDirect

Computers and Geotechnics

journal homepage: www.elsevier.com/locate/compgeo

Research Paper

CPT-based two-dimensional soil stratification using non-stationary mixture of Gaussian processes

Muhammet Durmaz, Zheng Guan, Michael A. Hicks *

Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands

ARTICLE INFO

Keywords:
Gaussian processes
Non-stationarity
Site characterization
Soil stratification
Spatial variability

ABSTRACT

In geotechnical practice, the cone penetration test (CPT) is widely used for soil stratification; however, the large spacing between soundings makes it difficult to construct accurate two-dimensional (2D) subsurface cross-sections. Conventional Gaussian Process (GP) models are effective for spatial interpolation under stationary assumptions, but real CPT data often exhibit strong non-stationarity due to geological layering, limiting the applicability of GP-based models in practice. To address this limitation, this paper develops a Mixture of Gaussian Processes (MGP) framework for explicitly modeling the distinct spatial features of different soil layers. The MGP integrates layer-specific GPs with a probabilistic layer identification mechanism, enabling joint representation of stratigraphic boundaries and within-layer variability. The framework is validated using synthetic ground profiles of varying stratigraphic complexity and applied to real CPT data from the Groningen region of the Netherlands. Results show that the MGP generally improves prediction accuracy and uncertainty characterization relative to the GP, particularly in inclined or more complex stratigraphic settings where layer-dependent variability and boundary effects are important.

1. Introduction

In geotechnical practice, the Cone Penetration Test (CPT) is the most widely used in-situ test for soil stratification, and is often interpreted through Robertson's soil behavior type (SBT) chart (Robertson, 2016; Lengkeek 2024). A reliable 2D stratigraphic cross-section plays a vital role in geotechnical design and analysis, particularly for linear infrastructures such as embankments, tunnels, and levees, where the length of the structure greatly exceeds its width. However, although CPTs provide almost continuous vertical measurements, the horizontal spacing between soundings is typically large (e.g., 20 m – 50 m), making it notoriously difficult to accurately characterize soil stratigraphy in two-dimensional (2D) cross-sections.

Delineation of 2D ground stratigraphy requires interpolating CPT measurements, such as soil behavior type index (I_c), along the horizontal direction (e.g., Hu and Wang, 2020; Varkey et al., 2023; Yao et al., 2025). Two distinct approaches have been adopted to exploit I_c for subsurface stratigraphic modeling: (1) continuous and (2) discrete (Phoon et al., 2022). In the continuous approach, the spatial variation of I_c values is first interpolated and then classified into SBT categories (e.g., Shuku and Phoon, 2021; Wang et al., 2020). In this approach, the entire ground is considered to be a random field conditioned on CPT locations. In contrast, the discrete approach (layer-based), directly models layer boundaries by considering the ground as comprising layers of different soil types with distinct stochastic characteristics (e.g., Shuku et al., 2020). Phoon et al. (2022) further argued that soil layer boundaries can be interpreted as points where the trend of the I_c profile changes. Following this concept, Shuku et al. (2020) applied a sparse Bayesian lasso method to detect layer boundaries using CPT data.

Various spatial interpolation methods have been applied in geotechnical characterization, including inverse distance weighting (IDW) (Rahman et al., 2021), Gaussian process regression (Yoshida et al., 2021), Bayesian compressive sampling (BCS) (Wang and Zhao, 2017), sparse Bayesian learning (SBL) (Ching and Phoon, 2017) and Geotechnical Lasso (Shuku and Phoon, 2023). Among these, GP regression has been extensively used in geotechnical site characterization because of its unique ability to model spatially correlated, uncertain, and sparse data (e.g., Lloret-Cabot et al., 2012; Vessia et al., 2020).

In GP regression, geotechnical data are typically described as the sum of (1) a trend component and (2) a spatially correlated residual, with the latter modeled as a stationary random field. However, evidence from the literature shows that different soil layers have distinct trend functions and spatial variability, resulting in highly non-stationary soil

* Corresponding author.
E-mail addresses: M.Durmaz@tudelft.nl (M. Durmaz), Z.GUAN@tudelft.nl (Z. Guan), M.A.Hicks@tudelft.nl (M.A. Hicks).

https://doi.org/10.1016/j.compgeo.2026.108459
Received 3 March 2026; Received in revised form 13 July 2026; Accepted 15 July 2026
Available online 27 July 2026
0266-352X/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Nomenclature

Abbreviations

CPT	Cone Penetration Test
EM	Expectation-Maximization
GMM	Gaussian Mixture Model
GP	Gaussian Process
IR	Identification Rate
MAE	Mean Absolute Error
MAP	Maximum a posteriori
MGP	Mixture of Gaussian Processes
MLL	Mean Log Loss
MTGP	Multi-Task Gaussian Process
RMSE	Root Mean Square Error
SBT	Soil Behavior Type

Symbols

$c(\cdot,\cdot)$	covariance function
$\mathbf{C}(\cdot,\cdot)$	covariance matrix
$\mathbf{C}_T$	task covariance matrix in MTGP
H	total depth of soil profile
I_c	soil behavior type index
$I_{c,m,n}$	observed I_c value at depth sample n of CPT m
$I_{c,i}^{obs}$	observed I_c value at test location i
$I_{c,k}^{*}$	predictive mean of the k^{th} GP at a prediction point
k	layer or GP component index
$l_{m,n}$	latent layer indicator for sample n of CPT m
m	CPT index
$m(.)$	GP trend function
$m_k(z)$	trend function of the k^{th} GP
M	total number of CPTs
$N_{k,m}$	number of observations assigned to layer k of CPT m
N_k	number of Monte Carlo realizations assigned to layer k
N_{test}	number of test observations
$p(\cdot)$	probability density function
$P(\cdot)$	probability
r	Monte Carlo realization index
R	total number of Monte Carlo realizations
$s_{k,m}$	standard deviation of the k^{th} GMM component of CPT m
$s_{k,m}^2$	variance of the k^{th} GMM component of CPT m
V^*	GP predictive variance at a new input
$V_{k,r}^*$	predictive variance of the k^{th} GP at test location r
w_0^k, w_1^k, w_2^k	quadratic trend coefficients of the k^{th} GP
x	horizontal coordinate
x_m	horizontal coordinate of CPT m
x_i, x_j	horizontal input locations used in the MTGP kernel
$\boldsymbol{y}$	vector of observed values in the GP formulation
y^*	GP predictive mean at a new input
z	depth coordinate
$z_{m,n}$	depth of sample n of CPT m
z_{min}^k	minimum depth of the observations assigned to layer k
z_{max}^k	maximum depth of the observations assigned to layer k
$\boldsymbol{\xi}$	input location vector
$\boldsymbol{\xi}_{m,n}$	input location of sample n of CPT m
$\boldsymbol{\xi}_*$	prediction location vector
$\boldsymbol{\theta}$	vector of GP hyperparameters
$\boldsymbol{\theta}^{GP,k}$	hyperparameter vector of the k^{th} GP
θ_v	vertical length scale
θ_h	horizontal length scale
$\mu_{k,m}$	mean depth of the k^{th} GMM component of CPT m
$\boldsymbol{\mu}_\pi$	mean vector in the MTGP prior for mixture weights
$\pi_{k,m}$	mixture weight or prior probability of layer k of CPT m
π_k^*	interpolated or sampled mixture weight of layer k at a prediction point
σ^2	variance of Gaussian process
σ_k^2	marginal variance associated with the k^{th} GP in the MAP step
τ	MTGP length scale

spatial variations (e.g., Cami et al., 2020; De Gast et al., 2021; Shi and Wang, 2022; Guan et al., 2024). Therefore, conventional GP models with a global trend function (e.g., linear or quadratic) and a single spatial variability term may produce inaccurate predictions and unrealistic uncertainty quantification.

To address this challenge, this study develops a Mixture of Gaussian Processes (MGP) framework that is capable of representing the non-stationary behavior inherent in multilayered soil deposits. The MGP handles non-stationarity by explicitly identifying distinct geological layers and modeling each layer as a separate random field with its own mean and spatial variability structure. This allows the model to capture both the abrupt transitions at layer boundaries and the within-layer continuity that cannot be represented by a single stationary GP. A finer stratigraphic resolution is then obtained by computing the predictive distribution of I_c values using the MGP and discretizing the resulting I_c field according to Robertson's SBT chart. Through this combination, the method integrates the advantages of both discrete (layer-based) and continuous (random-field-based) approaches, enabling a more accurate and interpretable delineation of stratigraphy. The framework is validated using synthetic and real CPT datasets to demonstrate its applicability and evaluate its performance for 2D geotechnical site characterization.

2. Methodology

Soil profiles are inherently stratified, and different layers commonly exhibit different property means, variances, and spatial correlation structures. As a result, the spatial variation of CPT-derived I_c values is typically non-stationary, especially near layer boundaries. A conventional Gaussian Process (GP), which assumes a single covariance structure over the entire domain, is therefore not well suited to represent both the within-layer continuity and the abrupt changes that occur across stratigraphic transitions. To address this limitation, this study adopts a Mixture of Gaussian Processes (MGP) framework.

In the present application, the basic idea is to divide the CPT profile into a small number of primary layers, model the spatial variability within each layer by a separate GP, and then combine these layer-specific predictions according to the probability that a point belongs to each layer. In this way, the model can simultaneously represent major stratigraphic boundaries and the spatial variability within individual layers.

Fig. 1 illustrates the overall structure of the proposed MGP framework for CPT-based site stratification. The framework consists of two main components: (1) a layer-assessment function, and (2) layer-specific Gaussian Processes (GPs). The layer-assessment function is used to identify the major layer boundaries at the CPT locations and to estimate, at locations between CPTs, the probability that a point belongs to each layer. In the proposed formulation, this function is represented by a one-dimensional Gaussian Mixture Model (GMM) at each horizontal location, with parameters that vary in the horizontal direction. The second component comprises the layer-specific GPs, which are used to represent the spatial variability of I_c within the major layers. In this way, the

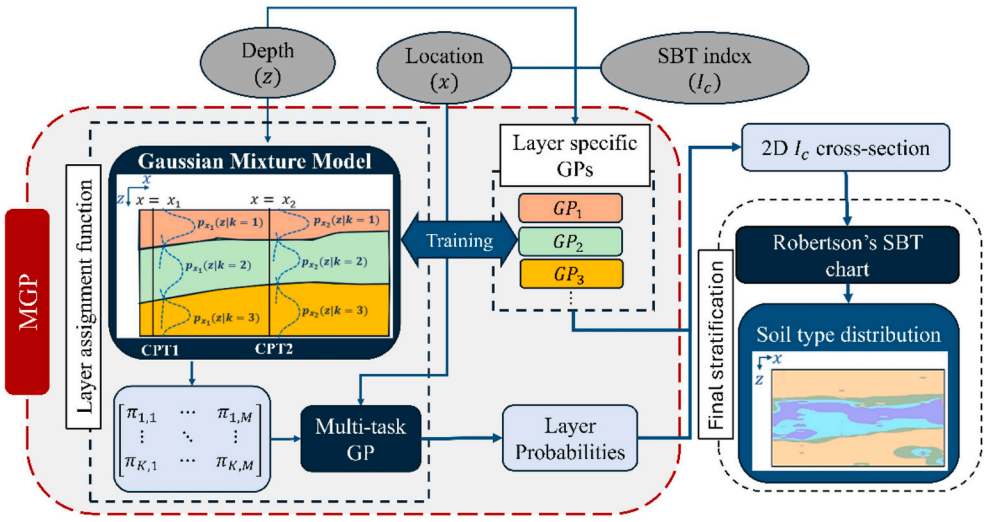


Fig. 1. Illustration of subsurface stratification using the Mixture of Gaussian Processes (MGP) model. Note: $\pi_{k,m}$ is the weight of the k^{th} layer at the m^{th} CPT.

framework separates the tasks of identifying the large-scale stratigraphic structure and modeling the within-layer spatial variation.

The first stage of the method is the joint calibration of the layer-assessment function and the layer-specific GPs at the CPT locations. This is carried out by fitting the GMMs and the layer-specific GPs simultaneously using the hard-cut Expectation-Maximization (EM) algorithm. Here, the GMMs are not used to model the distribution of I_c values directly; rather, they are used to represent the probabilistic vertical layer membership along each CPT profile through the depth coordinate. In this representation, the mean of each Gaussian component corresponds to the central depth of the associated layer, while its variance is related to the vertical extent of that layer. The purpose of this stage is to determine the optimal layer boundaries at the CPT locations and, at the same time, to estimate the GP parameters that are most consistent with those boundaries. Since the GMM parameters define the vertical layer distribution at each CPT, calibrating the GMM is equivalent to determining the most likely locations of the layer boundaries. The training stage therefore identifies both the primary stratigraphic structure and the corresponding layer-specific spatial variability.

After the training stage, the layer-specific GPs are fixed. The next step is to determine how the major layers continue between CPT locations. In the proposed framework, this is achieved by interpolating the GMM mixture weights in the horizontal direction. The details of this interpolation are presented in the following section. The outcome of this step is a probabilistic description of the major layering across the entire cross-section; i.e., for any prediction point, the probability of belonging to each layer can be estimated.

These layer probabilities are then combined with the calibrated layer-specific GPs to obtain the predictive distribution of I_c throughout the section. This yields a probabilistic I_c cross-section in which the predicted mean and uncertainty depend both on the spatial variability of the relevant layer-specific GP and on the uncertainty in layer membership. Consequently, the method can represent relatively smooth variation within layers while also producing increased uncertainty near stratigraphic transitions.

In the final stage, the predicted I_c cross-section is converted into a more detailed soil behavior distribution using the Robertson and Wride (1998) classification chart. This step refines the broad primary layering identified by the MGP into a finer soil behavior type (SBT) distribution, which is more suitable for geotechnical interpretation and subsequent engineering analysis.

The main components of the MGP are elaborated in the following sections.

2.1. Model structure

2.1.1. Brief review on GP

Gaussian Processes (GPs) can be viewed as an extension of probability distributions from random variables to functions (Rasmussen and Williams, 2006). Under this framework, the functions used to predict outputs are treated as samples from a multivariate Gaussian distribution defined by a covariance matrix **C**, such that

$$p(\boldsymbol{y}|\boldsymbol{\xi}) \sim N[\boldsymbol{m}(\boldsymbol{\xi}), \mathbf{C}(\boldsymbol{\xi}, \boldsymbol{\xi})] \tag{1}$$

Here, $\boldsymbol{y} = [y_1, y_2, y_3, \cdots y_n]$ represents the vector of n observed I_c values. The function $\boldsymbol{m}(\boldsymbol{\xi}) = [m(\boldsymbol{\xi}_1), m(\boldsymbol{\xi}_2), m(\boldsymbol{\xi}_3), \cdots m(\boldsymbol{\xi}_n)]$ denotes the mean of the GP, and $\mathbf{C}(\boldsymbol{\xi}, \boldsymbol{\xi}) = \left[c\left(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j\right)\right]_{n\times n}$ is the covariance matrix. Here, $\boldsymbol{\xi}$ is the input vector where each entry $\boldsymbol{\xi}_i = \{z_i, x_i\}$ represents a depth-horizontal coordinate pair (Fig. 1).

This study applies a single exponential autocorrelation function for defining the covariance matrix (Cami et al., 2020; Shuku and Phoon, 2023), where each element is given by

$$c(\xi_i, \xi_j) = \sigma^2 \exp\left(-\left(\frac{|z_i - z_j|}{\theta_v} + \frac{|x_i - x_j|}{\theta_h}\right)\right) \tag{2}$$

In this formulation, σ^2 and θ_v, θ_h are hyperparameters representing the variance and the length scales in the vertical and horizontal directions, respectively. To apply Gaussian processes in practice, it is essential to determine appropriate values for these hyperparameters. This is typically achieved through maximizing the marginal log likelihood in Eq. (3):

$$\log p(\boldsymbol{y}|\boldsymbol{\xi}, \boldsymbol{\theta}) = -\frac{1}{2}(\boldsymbol{y} - \boldsymbol{m})^{\mathrm{T}}\mathbf{C}^{-1}(\boldsymbol{y} - \boldsymbol{m}) - \frac{1}{2}\log|\mathbf{C}| - \frac{n}{2}\log(2\pi) \tag{3}$$

where $\boldsymbol{m} = \boldsymbol{m}(\boldsymbol{\xi})$, $\mathbf{C} = \mathbf{C}(\boldsymbol{\xi}, \boldsymbol{\xi}; \boldsymbol{\theta})$ and $\boldsymbol{\theta} = [\sigma, \theta_v, \theta_h]$ denotes the hyperparameters of the GP.

The covariance model in Eq. (2) is a separable, axis-aligned exponential covariance function. This form implies relatively rough, non-differentiable spatial fluctuations. It is adopted because exponential covariance models are commonly used in geotechnical spatial variability studies (Liu and Chen, 2010; Jiang and Huang, 2018; Varkey et al., 2023), where soil properties often show irregular short-scale variability. Smoother alternatives, such as squared-exponential or higher-order Matérn covariance functions, could also be used within the proposed MGP framework; however, comparing different covariance families is outside the scope of the present study.

After estimating the hyperparameters, the GP can be used to make predictions at a new input $\boldsymbol{\xi}_*$, with the predictive distribution also being Gaussian:

$$p(y^*|\boldsymbol{y},\boldsymbol{\xi},\boldsymbol{\xi}_*) \sim N[\mu^*, V^*] \tag{4}$$

Here, μ^* is the predicted mean and V^* is the associated predictive variance. These quantities are computed as follows (Rasmussen and Williams, 2006):

$$\mu^* = \mathbf{C}(\boldsymbol{\xi}_*,\boldsymbol{\xi})\mathbf{C}(\boldsymbol{\xi},\boldsymbol{\xi})^{-1}[\boldsymbol{y} - \boldsymbol{m}(\boldsymbol{\xi})] + m(\boldsymbol{\xi}_*) \tag{5}$$

$$V^* = \mathbf{C}(\boldsymbol{\xi}_*,\boldsymbol{\xi}_*) - \mathbf{C}(\boldsymbol{\xi}_*,\boldsymbol{\xi})\mathbf{C}(\boldsymbol{\xi},\boldsymbol{\xi})^{-1}\mathbf{C}(\boldsymbol{\xi},\boldsymbol{\xi}_*) \tag{6}$$

In this study, the GP model is formulated as a noiseless Gaussian process without an observation-noise (nugget) term. Consequently, the predictive variance reflects only interpolation uncertainty and does not account for measurement error or transformation uncertainty.

2.1.2. GMM for layer probability representation

In the MGP framework, the CPT measurements are assumed to be generated by K latent components, each corresponding to a major soil layer and represented by a layer-specific Gaussian Process (GP). To use this formulation, it is first necessary to determine which layer each observation is likely to belong to. This is the role of the gating function. In this study, the gating function proposed in the MGP formulations of Chen et al. (2014) and Zhou et al. (2019) is adopted and applied separately to each CPT profile.

Let $m = 1, \cdots, M$ denote the CPT index, $n = 1, \cdots, N_m$ denote the depth-sample index within CPT m, and $k = 1, \cdots, K$ denote the layer index. Each observation is represented by the pair $(\boldsymbol{\xi}_{m,n}, I_{c,m,n})$, where $\boldsymbol{\xi}_{m,n} = (z_{m,n}, x_m)$, $z_{m,n}$ is the depth coordinate, and x_m is the horizontal coordinate of CPT m. Let $l_{m,n}$ be a latent indicator variable such that $l_{m,n} = k$ means that the n^{th} sample of CPT m belongs to the k^{th} layer and is modeled by the k^{th} GP. In the present formulation, the indicator variables are assumed to follow a multinomial distribution within each CPT profile, i.e.,

$$P(l_{m,n} = k) = \pi_{k,m} \tag{7}$$

where $\pi_{k,m}$ is the prior probability that a depth point in CPT m belongs to layer k.

Conditional on the indicator variable, the depth coordinate is modeled by a Gaussian distribution:

$$p(z_{m,n}|l_{m,n} = k) \sim N\left(\mu_{k,m}, s^2_{k,m}\right) \tag{8}$$

where $\mu_{k,m}$ and $s^2_{k,m}$ are the mean and variance, respectively, of the k^{th} Gaussian component of CPT m. In this way, the depth-dependent layer membership at a given CPT location is represented by a one-dimensional Gaussian Mixture Model (GMM).

In this formulation, $\pi_{k,m}$ represents the prior probability that a randomly chosen depth point from CPT m belongs to layer k. Because CPT measurements are typically taken at nearly uniform depth intervals, $\pi_{k,m}$ can be interpreted as the proportion of the CPT profile occupied by that layer. Accordingly, the GMM provides a probabilistic representation of the vertical layer structure within each CPT, where each Gaussian component describes the location and spread of one layer with depth. Although the depth points belonging to a given layer at a CPT location may in general be regarded as approximately uniformly distributed within that layer interval, a Gaussian distribution is adopted here for the gating function. The reason is that the Gaussian form allows the layer membership probability to decrease smoothly from the center of the layer toward its boundaries during the iterative updating of layer assignments. In contrast, a uniform distribution would cause the membership probability to drop abruptly to zero at the layer boundary, leading to an unnecessarily sharp transition between adjacent layers.

Given the set of observations assigned to layer k, the corresponding I_c values are modeled by the k^{th} Gaussian process. Let $I_c^{(k)}$ denote the vector of I_c observations assigned to layer k, and let $\boldsymbol{\xi}^{(k)}$ denote the corresponding set of input locations. Then

$$p\left(\boldsymbol{I_c}^{(k)} \middle| \boldsymbol{\xi}^{(k)}\right) \sim N\left[\boldsymbol{m}_k, \mathbf{C}_k\left(\boldsymbol{\xi}^k, \boldsymbol{\xi}^k \middle| \boldsymbol{\theta}^{GP,k}\right)\right] \tag{9}$$

where $\boldsymbol{m}_k$ is the mean vector of the k^{th} GP evaluated at $\boldsymbol{\xi}^{(k)}$, $\mathbf{C}_k$ is the corresponding covariance matrix, and $\boldsymbol{\theta}^{GP,k}$ denotes the GP hyperparameters of layer k.

For each GP component, a quadratic trend function is adopted, since it is the highest-order trend function commonly used in the literature to represent homogeneous soil layers and has been shown to perform reliably even when the true trend is simpler, such as constant or linear (Ching et al., 2023). The trend function of layer k is written as

$$m_k(z_{m,n}) = w_0^k + w_1^k \bullet \left(z_{m,n} - z_{min}^k\right) + w_2^k \bullet \left(z_{m,n} - z_{min}^k\right)^2 \tag{10}$$

where w_0^k, w_1^k, and w_2^k are the trend coefficients of the k^{th} GP, and $z_{\min}$ is the minimum depth of the observations assigned to that component.

In this study, a separate 1D GMM is fitted to each CPT profile in order to represent the layer distribution at the observed CPT locations. After the layer assignments have been learned, the associated mixture weights are interpolated in the horizontal direction using a multi-task Gaussian Process (MTGP), which is described in Section 2.1.4.

2.1.3. Training Algorithm

The hard-cut EM algorithm (Chen et al., 2014; Zhou et al., 2019) is used to train the model. The EM algorithm is an iterative extension of the maximum likelihood (ML) method for problems involving hidden or unobserved variables. It alternates between two steps: the Expectation (E) step and the Maximization (M) step. In the E-step, the algorithm computes the expected value of the log-likelihood with respect to the conditional distribution of the latent variables, given the observed data and current parameter estimates. In the M-step, this expected log-likelihood is maximized to update the model parameters. The main steps of the hard-cut EM algorithm are as follows:

(1) Clusters are initialized using K-means clustering based on I_c and depth values, where the indicator variables are set based on the cluster assignments of the data points.

(2) **M-step:** The hyperparameters of the Gaussian Processes (GPs) for each layer are estimated using the maximum likelihood (ML) approach. To mitigate overfitting, the coefficients of the parametric mean function are assigned weakly informative normal priors, with their scales defined based on the layer-wise ranges of I_c and depth. This follows common Bayesian regularization practice for regression and Gaussian process mean functions, where prior scales are chosen to reflect the magnitude of the data and prevent overly flexible trends (Gelman et al., 2008; Theodoridis, 2020). Specifically, the intercept coefficient w_0^k is assigned a prior whose mean and standard deviation are equal to the sample mean and sample standard deviation of the I_c values currently assigned to layer k. For coefficients w_1^k and w_2^k, the prior means are set to zero, while their standard deviations are defined as $[I_{c,max}^{(k)} - I_{c,\,min}^{(k)}]/[z_{max}^{(k)} - z_{min}^{(k)}]$ for w_1^k and $[I_{c,max}^{(k)} - I_{c,\,min}^{(k)}]/[(z_{max}^{(k)})^2 - (z_{min}^{(k)})^2]$ for w_2^k. Maximum likelihood estimations of the GMM parameters, $\pi_{k,m}$, $\mu_{k,m}$, and $s^2_{k,m}$ are given by Chen et al. (2014) as follows:

$$\pi_{k,m} = \frac{N_{k,m}}{\sum_{k=1}^{K} N_{k,m}} \tag{11-a}$$

$$\mu_{k,m} = \frac{1}{N_{k,m}} \sum\nolimits_{n=1:l_{m,n}=k}^{N_{k,m}} z_{m,n} \tag{11-b}$$

$$s_{k,m}^2 = \frac{1}{N_{k,m}} \sum_{n=1:l_{m,n}=k}^{N_{k,m}} \left(z_{m,n} - \mu_{k,m}\right)^2 \tag{11-c}$$

Here, a constraint is imposed such that layers remain continuous in the vertical direction. This requires that, within a CPT log, all points belonging to the same layer appear in one contiguous depth interval, with no intervening points assigned to other layers. In addition, each CPT log contains non-repeating depth entries that are uniformly spaced between the minimum and maximum depths; in other words, the depth values are uniformly distributed along the sounding. Consequently, the mean and variance of the depth values within each layer can be approximated by those of a uniform distribution. Under these conditions, the GMM gate parameters can be expressed directly in terms of the mixture weights and the total depth of each CPT as

$$\mu_{k,m} = H\left(\sum_{j=1}^{k-1} \pi_{j,m} + \frac{1}{2}\pi_{k,m}\right) \tag{12-a}$$

$$s_{k,m}^2 = \frac{1}{12}\left(\pi_{k,m}H\right)^2 \tag{12-b}$$

where H denotes the total depth of each CPT profile. Equations (12-a) and (12-b) assume that the layer indices are ordered from the ground surface downward. Accordingly, once the clusters are initialized, the layer indices are assigned based on the average depth of each cluster. Hence, a lower k is given to the layers with lower average depth.

(3) **E-step**: The indicator variables are updated using the maximum a posteriori (MAP) criterion,

$$l_{m,n} = \underset{k}{\operatorname{argmax}}[\pi_{k,m} N\left(z_{m,n}|\mu_{k,m}, s_{k,m}^2\right) N\left(I_{c,m,n}|m_k\left(z_{m,n}\right), \sigma_k^2\right)] \tag{13}$$

where σ_k^2 denotes the marginal variance associated with the k^{th} GP under the current parameter estimates. Since this step updates only the latent layer assignments and does not update the model parameters, it constitutes the E-step of the hard-cut EM algorithm.

Steps (2) and (3) are repeated until the indicator variables no longer change between consecutive iterations. Note that the E-step and M-step are carried out simultaneously using the data from all CPTs. This is because the layer-specific GP parameters are shared across all CPTs and represent the common spatial variability of each layer, whereas the GMM parameters vary from one CPT to another, since they describe the location and vertical spread of the layers at each individual CPT.

The hard-cut EM algorithm assigns each observation deterministically to a single layer during the maximization step and therefore does not explicitly account for assignment uncertainty when estimating layer-specific GP hyperparameters. This approximation may introduce bias when layer-specific I_c distributions are not sufficiently separated, since GP parameters are then estimated using only the subset of observations assigned to each layer. In principle, this limitation can be alleviated by sampling-based approaches such as MCMC–EM (Wu and Ma, 2019), which propagate assignment uncertainty through the estimation procedure. However, such methods require computationally demanding MCMC simulations within each EM iteration together with convergence diagnostics, resulting in substantially increased computational cost. For this reason, the hard-cut EM formulation is adopted here to maintain computational efficiency. When layer-specific I_c distributions are reasonably separated, posterior membership probabilities are typically close to 0 or 1 for most of the points and the difference between hard and soft assignment becomes limited. Moreover, uncertainty in layer membership is explicitly reintroduced at the prediction stage through the probabilistic combination of layer-specific GP outputs using interpolated layer probabilities.

(4) **Prediction:** After training, the I_c value at a prediction point $\boldsymbol{\xi}_* = (z_*, \boldsymbol{x}_*)$ is represented as a weighted mixture of the predictions from the K layer-specific GPs as

$$p\left(I_c^*|\boldsymbol{\xi}_*\right) = \sum_{k=1}^{k=K} P(l_* = k|\boldsymbol{\xi}_*) N\left[I_c^*|\mu_k^*, V_k^*\right] \tag{14}$$

Here, $P(l_* = k|\boldsymbol{\xi}_*)$ is the posterior probability that the prediction point belongs to layer k, and μ_k^* and V_k^* are the predictive mean and variance, respectively, of the k^{th} GP at $\boldsymbol{\xi}_*$, computed using Eqs. (5) and (6).

The predictive distribution given in Eq. (14) is a Gaussian mixture. The corresponding predictive variance can be decomposed into two contributions: a within-layer component associated with spatial interpolation uncertainty inside each layer, and a between-layer component associated with uncertainty in layer membership. The second component becomes dominant near stratigraphic transitions where multiple layer probabilities are comparable. Therefore, predictive variance increases near layer boundaries. In contrast, stationary GP models do not explicitly represent this boundary uncertainty component.

2.1.4. Multi-task GP model

After the training phase, the GMM gate parameters at locations between CPTs are determined from the mixture weights associated with each layer. In the proposed framework, only the mixture weights are interpolated directly, while the corresponding GMM means and variances are subsequently obtained from Eqs. (12-a) and (12-b). To interpolate the mixture weights in the horizontal direction, a multi-output (multi-task) Gaussian Process (MTGP) model is used. The advantage of the MTGP is that it explicitly accounts for correlations between the mixture weights of different layers. The MTGP extends the standard Gaussian Process (GP) framework to jointly model multiple correlated outputs. The formulation adopted in this study follows Bonilla et al. (2007), which also forms the basis of the implementation in the GPyTorch library (Gardner et al., 2018).

For CPT m, let the vector of mixture weights be defined as $\left[\pi_{1,m}, \pi_{2,m} \cdots \pi_{K,m}\right]$, where x_m is the horizontal coordinate of CPT m. In the MTGP framework, the layer index $k = 1, \cdots, K$ is treated as the task index, and the joint distribution of the mixture weights over all CPT locations is modeled as a multivariate Gaussian distribution:

$$p(\boldsymbol{\pi}|\boldsymbol{x}) \sim N[\mu_\pi, \mathbf{C}_T \otimes \mathbf{C}(\boldsymbol{x}, \boldsymbol{x})] \tag{15}$$

where $\boldsymbol{\pi}$ is the stacked vector of all mixture weights, $\boldsymbol{\mu}_\pi$ is the corresponding mean vector, $\mathbf{C}_T$ is the task covariance matrix describing correlations between layers, $\mathbf{C}(\mathbf{x}, \mathbf{x})$ is the covariance matrix over the horizontal coordinates of the CPTs, and $\otimes$ denotes the Kronecker product. Compared with the standard GP formulation in Eq. (1), the additional task covariance term allows the MTGP to capture both spatial dependence and inter-layer dependence in a scalable manner. The underlying assumption is that the outputs, i.e., the mixture weights of different layers, are correlated and share the same horizontal spatial correlation structure.

In this study, the horizontal covariance function of the MTGP is taken as the squared exponential kernel,

$$c\left(x_i, x_j\right) = \exp\left(-\frac{\left|\mathrm{x_i} - \mathrm{x_j}\right|^2}{2\tau^2}\right) \tag{16}$$

where τ is the horizontal correlation length of the MTGP. The hyperparameters of the MTGP are trained by the maximum likelihood method. For the implementation of the MTGP and GP components in the MGP, the GPyTorch library (Gardner et al., 2018) has been employed. The Adaptive Moment Estimation (ADAM) optimizer is employed to maximize the likelihood functions, which is a stochastic gradient-based optimization algorithm that adaptively adjusts the learning rate for each parameter based on estimates of the first and second moments of the gradients.

After training the MTGP, the posterior probability of layer membership at a prediction point $\boldsymbol{\xi}_* = (z_*, \boldsymbol{x}_*)$ is estimated by Monte Carlo simulation. In each realization $r = 1, \cdots, R$, a sample of the mixture-

weight vector, $\pi_*^{(r)} = [\pi_1^{*(r)}, \pi_2^{*(r)}, \ldots \pi_K^{*(r)}]$, is drawn from the MTGP, and the remaining gate parameters are computed using Eqs. (12a) and (12b). Subsequently, layer labels are assigned according to the maximum a posteriori (MAP) criterion described in Eq. (17):

$$l_*^{(r)} = \underset{k}{\text{argmax}} \left[\pi_k^{*(r)} N\left(z_* \middle| \mu_k^{*(r)}, \left(s_k{}^{*(r)}\right)^2 \right) \right] \tag{17}$$

The probability of a layer's existence at a given point, $P(l_*= k|\boldsymbol{\xi}_*)$, is then approximated as the ratio N_k/R, where N_k is the number of Monte Carlo samples in which the point is classified into the k^{th} layer, and R is the total number of realizations, which is set to 500 in this research.

Note that the mixture weights must be non-negative and sum to one. However, the MTGP in Eq. (15) defines a Gaussian predictive distribution over the mixture weights, and Gaussian samples are not guaranteed to satisfy this simplex constraint. To address this issue, a clipping-and-renormalization procedure is adopted to map each sampled mixture-weight vector onto the admissible simplex before it is used in the MAP layer-assignment step. Specifically, negative sampled values are first truncated to zero, and the resulting weight vector is then normalized by its sum so that all entries are non-negative and sum to one. This procedure does not constitute a fully coherent probabilistic model on the simplex, but it has been used as a practical approximation for constructing admissible mixture proportions (Zhang et al., 2026). A more coherent probabilistic treatment, such as a logistic-normal formulation that maps unconstrained Gaussian variables onto the simplex using a softmax transformation, may be considered in future work (Aitchison and Shen, 1980).

2.2. Performance metrics

To assess the predictive capability of the proposed MGP model, appropriate performance metrics are essential. In this study, two widely used deterministic accuracy measures, Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), are employed. These metrics are well-established in the literature and straightforward to interpret. In addition, for geotechnical applications, the accuracy of the uncertainty quantification is particularly important. Shuku and Phoon (2023) suggested using two additional metrics for Gaussian Process (GP) models: the coverage percentage (i.e., the proportion of data points falling within the 95% confidence interval) and the width of the confidence interval. A model with a higher coverage percentage and narrower confidence interval is generally preferred. However, when one model achieves a higher coverage but also has a wider confidence interval than another, these metrics may yield conflicting conclusions. To address this limitation, we adopt the Mean Log Loss (MLL) metric proposed by Rasmussen and Williams (2006). MLL provides a unified quantitative measure that implicitly balances both narrow confidence intervals and high coverage ratios. It is computed as the mean negative logarithm of the predictive density evaluated at the observed test values. In the proposed MGP framework, the predictive distribution is given by the Gaussian mixture defined in Eq. (14). Therefore, the predictive density at each test location is obtained as the weighted sum of the layer-specific Gaussian predictive densities using the corresponding layer probabilities $P(l_i= k|\boldsymbol{\xi}_*^i)$. The MLL is then computed by averaging the logarithm of this predictive density over all test observations, and can be expressed as:

$$MLL = -\frac{1}{N_{test}} \sum_{i=1}^{N_{test}} log p\left(I_{c,i}^{obs} \middle| \boldsymbol{\xi}_*^i \right) \tag{18}$$

$$p\left(I_{c,i}^{obs} \middle| \boldsymbol{\xi}_*^i \right) = \sum_{k=1}^{K} P(l_i= k|\boldsymbol{\xi}_*^i) \frac{1}{\sqrt{2\pi V_{k,i}^*}} \exp\left(-\frac{\left(I_c^{obs} - \mu_{k,i}^* \right)^2}{2V_{k,i}^*} \right) \tag{19}$$

Here, μ_k^* and V_k^* denote the predictive mean and variance, respectively, of the k^{th} GP component at the i^{th} test location $\boldsymbol{\xi}_*^i$.

Another performance metric used in this study is the Identification Rate (IR). It is computed by comparing the predicted and actual soil behavior type (SBT) distributions after classifying the ground according to Robertson and Wride (1998). The IR is defined as the ratio of the number of correctly predicted soil types to the total number of data points at the test locations.

3. Synthetic data example

A synthetic data experiment was conducted to evaluate the performance of the proposed MGP model under controlled conditions where the true subsurface layering and spatial statistics are known. Three distinct geological layering geometries were considered, each consisting of three layers in the synthetic profile. The first geometry represents the simplest case, with perfectly horizontal layer boundaries. In the second geometry, the thickness of the middle layer remains constant, but the layer boundaries are inclined. In the third geometry, the middle layer tapers out and disappears. For all scenarios, the layer boundaries were modeled as linear functions of the form $a + bx$, where the parameters a and b for each layer are given in Table 1.

For each layer, synthetic I_c values were generated based on the statistical properties summarized in Table 2. Correlation lengths for the layers were selected following values reported in the literature (Phoon et al., 2022; Shuku and Phoon, 2023), while the means and the variances were chosen to minimize substantial overlap between the I_c distributions of different layers, as summarized in Table 2. A single exponential correlation function, as defined in Eq. (2), was employed to model the spatial correlation. To create a more challenging test case, the synthetic data were generated with a constant mean (i.e., no within-layer trend), whereas the individual GPs in the MGP model employed a quadratic mean function. This setup serves as a test of the model's resistance to overfitting, allowing evaluation of its performance when the underlying trend order is less than quadratic.

The virtual ground models were generated with a spatial resolution of 0.10 m in the vertical direction and 1.0 m in the horizontal direction, and resulted in grids that extend 10 m in depth and 20 m horizontally. Each virtual ground, therefore, contains $100 \times 20 = 2000$ data per scenario. The resulting I_c fields for the three virtual grounds are shown in Fig. 2. To emulate site investigation data from CPTs, the synthetic I_c values were sampled at four horizontal locations: i.e., at $x = 2.5$ m, 7.5 m, 12.5 m, and 17.5 m, as shown by the dashed lines in Fig. 2. This sampling produced 400 data points for model training, while the remaining points were used to evaluate the interpolation performance of the MGP model.

The estimated trend coefficients from the three-layer MGP model, trained on CPT data for all three scenarios, are presented in Table 3. The results indicate that the coefficients of the linear and quadratic terms are either very small in magnitude or have opposite signs, effectively yielding a constant mean in each layer, consistent with the underlying statistical structure of the synthetic data. This confirms the MGP model's ability to avoid overfitting when the true mean structure is simpler than the assumed quadratic form. Fig. 3 shows the initial layering determined by K-means clustering, which is used to initialize the MGP model. The K-means results already capture the main stratigraphic structure at the CPT locations and provide a reasonable initial partition of the data into three layers. Fig. 4 illustrates the estimated trends together with the sampled CPT observations. In the trend plots, different colors represent the layers to which the data points are assigned. Comparison of Figs. 3 and 4 shows that the MGP model refines the initial K-means-based

Table 1
Equations of the layer boundaries in Fig. 2. In the table, z represents depth (m).

	Scenario 1	Scenario 2	Scenario 3
1st layer boundary	$z = 2$	$z = 2 + 0.1x$	$z = 2 + 0.1x$
2nd layer boundary	$z = 6$	$z = 6 + 0.1x$	$z = 6 - 0.1x$

Table 2
Input parameters for the synthetic soil profile.

	mean (μ)	stdv. (σ)	θ_v (m)	θ_h (m)
Layer 1	1.3	0.1	0.4	10
Layer 2	3.2	0.2	1.2	15
Layer 3	2.1	0.15	0.8	12

layering and provides highly accurate identification of layer boundaries at the CPT locations, closely matching the true stratigraphic boundaries in the synthetic profiles.

The interpolation performance of the MGP compared to that of a single GP is summarized in Table 4. In most cases, the MGP achieved lower interpolation errors than the GP, with the only exception being a slightly higher MLL value for Scenario 1. This difference arises because RMSE and MAE evaluate only the accuracy of the predictive mean, whereas MLL evaluates the full predictive distribution and is therefore sensitive to both mean prediction and variance calibration. According to Eq. (5), the predictive mean is a weighted combination of the observed data, with weights determined by the covariance structure. Since the horizontal scale of fluctuation is larger than the vertical one, observations at similar depth have a stronger influence on the mean prediction. As a result, correct layer assignment has a greater effect on RMSE and MAE, particularly when the layer boundaries are inclined and the horizontal heterogeneity increases. In contrast, for fixed hyperparameters, the predictive variance depends only on the calibrated covariance structure. When the ground is divided into layers, fewer data are available to estimate each layer-specific covariance model, which increases statistical uncertainty in the covariance calibration. This interpretation is consistent with the general identifiability issue of covariance parameters: when the available data provide limited information about the correlation structure, especially in the horizontal direction, maximum-likelihood estimates may not fully represent the uncertainty in the scale of fluctuation. In Scenario 1, where the layering is horizontally simple, the benefit of separating observations by layer is limited, and this reduction in calibration efficiency can lead to a slightly higher MLL for the MGP despite improved RMSE and MAE. For Scenarios 2 and 3, however, the performance gap between the two models increases as the layer geometry becomes inclined, indicating that accurate boundary assessment becomes more important when stratigraphy deviates from horizontal layering. The evolution of the mixture weights in the horizontal direction is shown in Fig. 5. These mixture weights are obtained by MTGP interpolation and are used to generate the layer probabilities between CPT locations. The results show that the weights remain nearly constant in Scenario 1, whereas they vary progressively in Scenarios 2 and 3 in accordance with the inclination of the layer

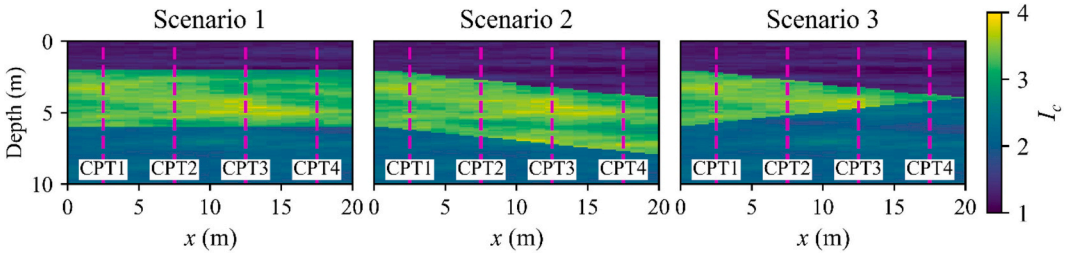


Fig. 2. Spatial distribution of I_c for virtual ground profiles (vertical dashed lines show the sampling locations for training data).

Table 3
Trend coefficients that were obtained after fitting a 3-layer MGP.

	Scenario 1			Scenario 2			Scenario 3		
	w_0	w_1	w_2	w_0	w_1	w_2	w_0	w_1	w_2
Layer 1	1.28	0.014	-0.016	1.28	0.005	0.001	1.28	0.005	0.002
Layer 2	3.20	0.040	-0.014	3.23	0.011	0.003	3.27	0.03	-0.004
Layer 3	2.10	0.042	-0.002	2.17	0.014	-3e-4	2.12	0.014	3e-4

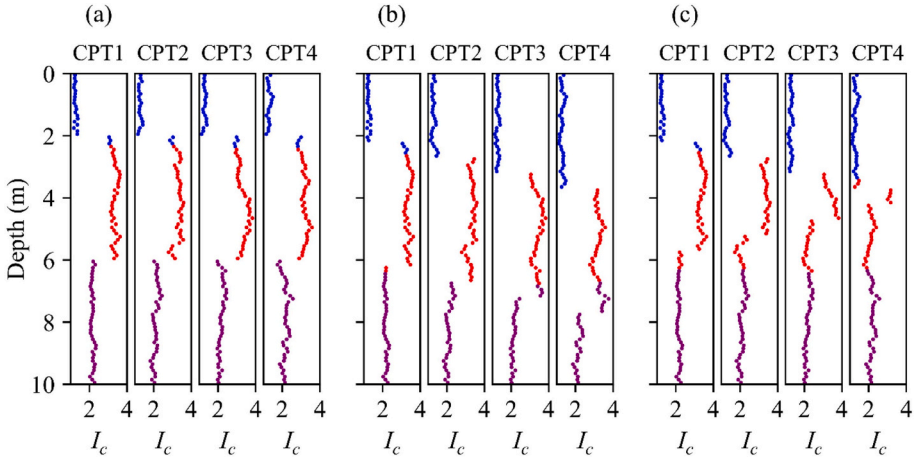


Fig. 3. Initial layer assignment after K-means clustering: (a) Scenario 1, (b) Scenario 2, (c) Scenario 3. Different marker colors indicate different layers.

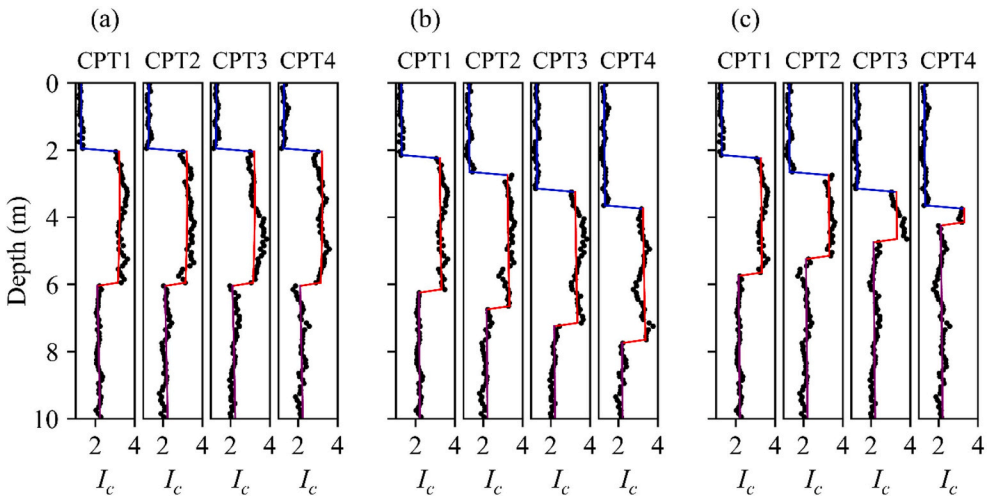


Fig. 4. The data used for training and the global trend (solid lines) that were fitted by the MGP model with 3 components: (a) Scenario 1, (b) Scenario 2, (c) Scenario 3.

Table 4
Comparison of the error metrics obtained by GP and 3-layer MGP.

	Scenario 1		Scenario 2		Scenario 3	
	GP	MGP	GP	MGP	GP	MGP
RMSE	0.092	**0.087**	0.143	**0.098**	0.148	**0.122**
MAE	0.072	**0.068**	0.092	**0.070**	0.096	**0.084**
MLL	**-0.974**	-0.862	-0.525	**-0.750**	-0.498	**-0.825**
IR	**0.86**	**0.86**	0.80	**0.86**	0.81	**0.86**

boundaries. The uncertainty associated with the interpolated mixture weights is very small and visually unnoticeable, which is consistent with the fact that the location of the true layer boundaries in the synthetic data vary linearly with horizontal distance. This difference is also evident when comparing the I_c predictions from the GP (Figs. 6a–c) and the MGP (Figs. 6d–f). While the GP produces smooth transitions near layer boundaries, the MGP captures sharper changes, better representing the higher variability typically present in these regions of the virtual ground.

Similar trends are observed in terms of the identification rate (IR). The IR values of the GP and MGP are nearly identical for Scenario 1 but show greater differences for the other two scenarios. Figs. 7a-c illustrate the true soil type distribution based on the synthetic data, whereas Figs. 7d-f and 7g-i show the predicted soil type distributions obtained from the GP and MGP models, respectively. Misclassifications in the GP predictions occur more frequently near the transition zones between major layers, whereas the MGP provides a more accurate delineation of soil types.

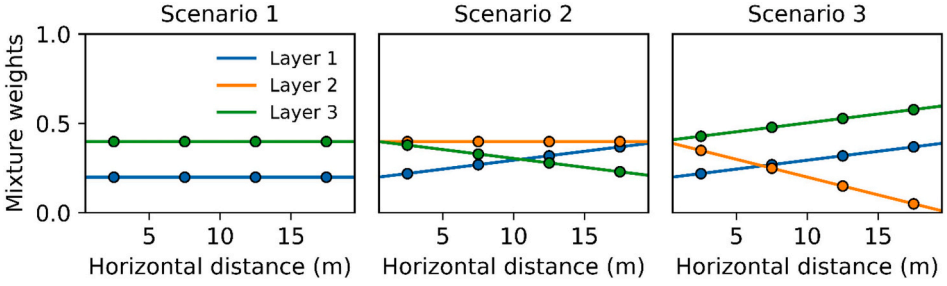


Fig. 5. Change of layer GMM mixture weights with horizontal distance after MTGP interpolation. In Scenario 1, the mixture weights of layer 2 and layer 3 overlap.

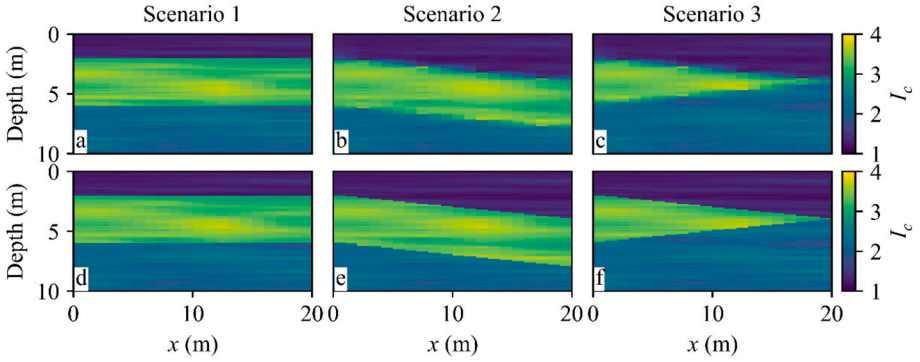


Fig. 6. Comparison of interpolation results using three-layer MGP and Gaussian process: a, b, c, and d, e, f show interpolation of I_c using GP and MGP, respectively.

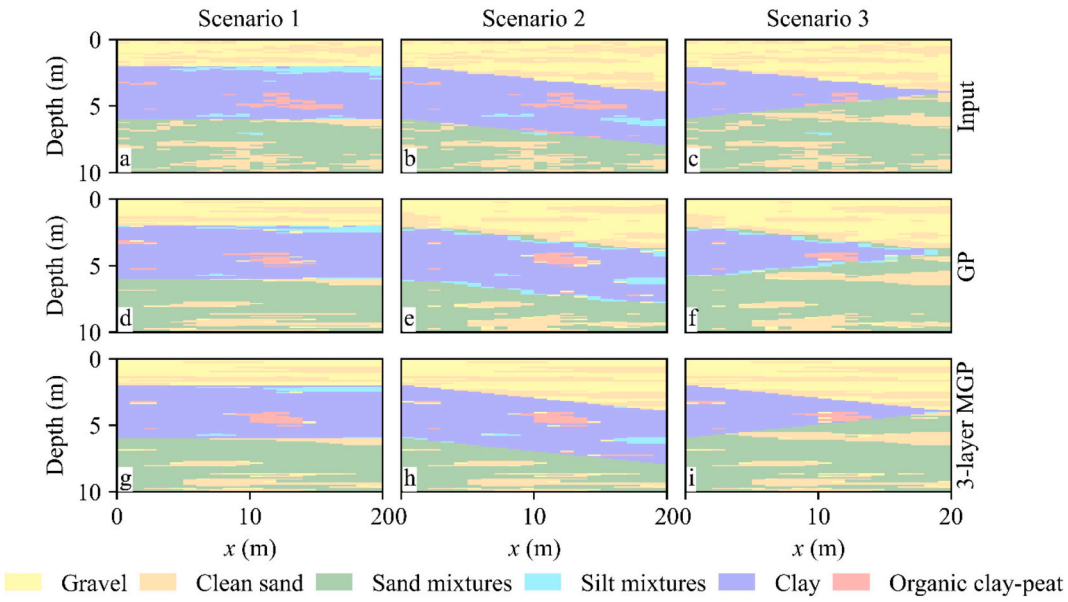


Fig. 7. Comparison of the soil behaviour type distributions of the input data, GP model, and three-layer MGP model.

4. Implementation with real CPT data

The proposed MGP model was applied to real CPT data in order to evaluate its performance under real site conditions. A total of ten CPT I_c profiles from the Groningen region in the Netherlands were used in this study. The relative spatial locations of the CPTs are shown in Fig. 8, providing an overview of the investigation layout. Fig. 9 presents the corresponding I_c profiles together with the interpreted soil types based on Robertson and Wride (1998) to give an insight into the stratigraphic variability at the site.

In the synthetic data experiments, the true number of components was predefined, allowing the MGP model to be trained with this known value. However, for real ground conditions, the correct number of components is not known in advance and must be determined from the data. Therefore, K-fold cross-validation is applied to determine the number of layers in the MGP model. To implement this approach, CPT3

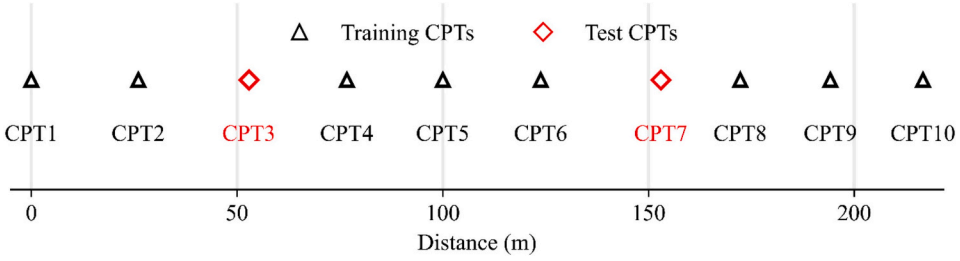


Fig. 8. Layout of CPTs.

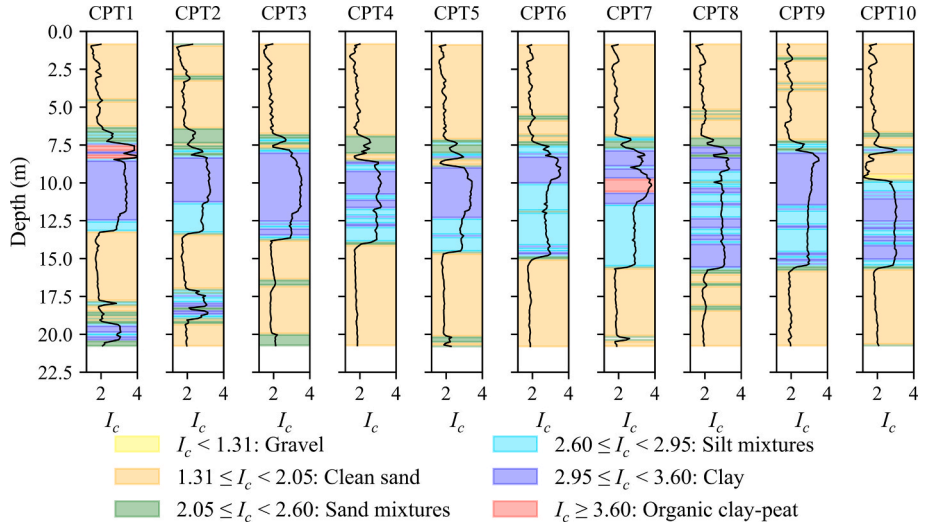


Fig. 9. I_c distribution and classification of the CPT data that were used for implementation.

Table 5
Mean error results obtained by each fold in cross-validation.

# of layers	RMSE	MAE	MLL
1	0.274	0.192	0.344
2	0.283	0.197	0.300
3	**0.263**	**0.178**	**0.199**
4	0.285	0.195	0.674

and CPT7 were removed from the dataset and reserved as test data; these profiles were also excluded from the cross-validation process. Among the remaining CPTs, one profile between CPT1 and CPT10 was removed at each fold to serve as a validation set. The model was then trained with different numbers of components, and the RMSE was calculated by comparing the mean estimates of the MGP model with the observed I_c values at the validation locations. The three-component model yielded the lowest mean RMSE, MAE and MLL as shown in Table 5, and was therefore selected as the most appropriate representation of the data. The four-layer model gives slightly worse RMSE and MAE than the three-layer model, while its MLL shows a significant increase. This indicates poorer calibration of the predictive variance, most likely because the additional component is not supported by the available CPT profiles.

It is observed that all CPTs exhibit a consistent stratigraphic pattern in which a clayey layer lies between two sandy layers, except for CPT1 and CPT2. This intermediate clayey layer generally occurs between depths of approximately 7 m and 15 m, although its exact position varies from one CPT to another. This recurring structure supports the use of three main layers in the MGP model. A second, deeper clayey layer is observed between about 17.5 m and 20 m in CPT1 and CPT2; however, this feature is absent in the other CPTs. Because it is only captured at two isolated locations, there is insufficient information to establish a continuous stratigraphic boundary or meaningful spatial statistics for this deeper clayey layer, and therefore, it is not identified as a separate component in the MGP.

Fig. 10 presents the results of the MGP model after training with all available CPT data except for the data reserved for the test set. The observed I_c values and the computed quadratic trends are shown, with each layer represented in a distinct color for both the data and the fitted trend. The computed trend parameters and hyperparameters of the GPs for each layer are summarized in Table 6. The results show that the coefficients of the linear and quadratic terms are very small for the first and third layers, indicating that the trends in these layers are negligible. In contrast, the middle layer exhibits a slight quadratic trend, consistent with its more complex soil composition. For comparison, the parameters obtained from a traditional Gaussian Process approach, where a single GP is used to represent the entire profile, are listed in the last row of Table 6. A key observation is that both the horizontal and vertical scales of fluctuation are smaller in the MGP model than in the single GP model. This difference arises because, when a single GP is applied to the entire ground, the heterogeneity introduced by multiple layers inflates the estimated scale of fluctuation, whereas the MGP framework accounts for stratification explicitly and thus yields more realistic estimates. On the other hand, the horizontal scales of fluctuation reported in Table 6 should be interpreted with caution. In contrast to the vertical scale, which is informed by many measurements along each CPT profile, the horizontal scale is inferred from a limited number of CPT locations. As a result, the marginal likelihood may be relatively flat with respect to the horizontal scale of fluctuation, and the maximum-likelihood estimates should be regarded as indicative correlation parameters rather than precisely identified site constants. Since the present study uses maximum-likelihood estimates, uncertainty in the covariance hyper-parameters is not propagated into the predictive variance.

When the layering is inspected closely, it is seen that the first layer predominantly consists of clean sands and sand mixtures, the second layer is composed mainly of silt mixtures and clays, and the third layer again primarily comprises clean sands and sand mixtures. Fig. 10 indicates that the fitted boundaries at the CPT locations are geologically more consistent for the transition between the second and third layers than for the transition between the first and second layers. The latter remains less distinct because sandy and clayey soils are interbedded in several CPTs. For instance, at CPT5, a clean sand interval occurs within an otherwise clayey–silty section, complicating the identification of a clear boundary between the first and second layers. It should be noted that uncertainty in the layer determination at the CPT locations is not represented in the present model. Once the layer thicknesses at the CPT locations are determined by the hard-cut EM procedure, they are treated as fixed in the subsequent interpolation stage.

The most probable distribution of the layers across the site is shown in Fig. 11a, while the corresponding uncertainty in the most probable layering is presented in Fig. 11b. The layers described in Fig. 11 consist of mixed soil types, meaning that parts of a single layer can exhibit

Table 6
Trend and covariance parameters of each layer of MGP and Gaussian process.

	ω_0	ω_1	ω_2	σ^2	$\theta_v(m)$	$\theta_h(m)$
Layer 1	1.89	-0.011	4.60e-4	0.052	1	69.6
Layer 2	2.53	0.181	-1.88e-2	0.169	1.5	26.6
Layer 3	2.06	-2.45e-2	2.20e-3	0.108	1.8	33.5
GP	2.18	2.28e-2	-1.46e-3	0.426	5.4	128.6

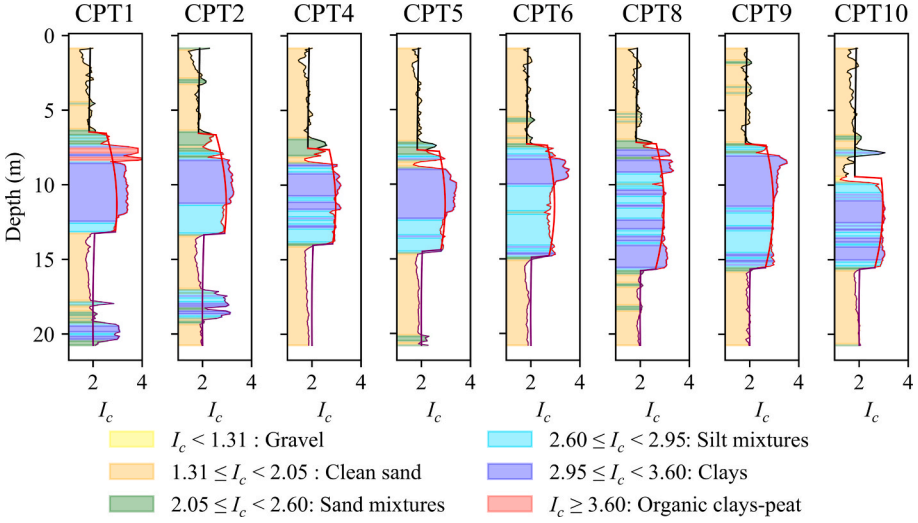


Fig. 10. Layering and mean fitting results on training CPT data. Layers are indicated by different colors of the trend lines.

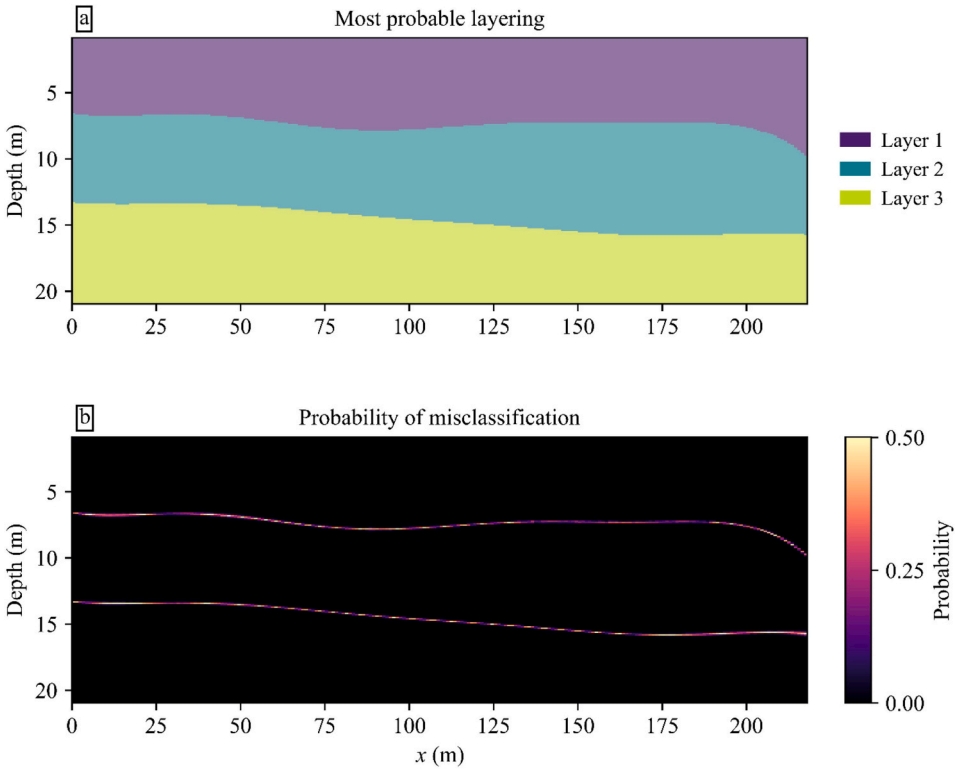


Fig. 11. The most probable layering (a) and uncertainty in layer boundaries (b) for the 3-layer MGP.

different behaviors. This is particularly evident in the third layer, where clayey soils are embedded within sandy deposits (see Fig. 10). For this reason, directly inputting such broad-scale layering into geotechnical analyses may not be appropriate. Nevertheless, working with larger-scale layering is still valuable, as it enables more accurate estimation of spatial variability and improves interpolation accuracy. To achieve consistency in numerical analyses such as finite element modeling, however, a finer-scale representation of the layering is needed, and this can be obtained by using the I_c predictions from the MGP model. Fig. 12a presents the mean I_c distribution across the site, while Fig. 12b shows the variance of the MGP estimation. Based on these spatial distributions, the most probable soil behavior type classification is derived and illustrated in Fig. 12c, where features such as the clay inclusions within the sandy third layer (observed in Fig. 10) are explicitly captured and can therefore be incorporated into geotechnical analyses. The corresponding uncertainty in the most probable soil type distribution is shown in Fig. 12d. To compute these uncertainties, 500 realizations of the I_c field were generated by sampling from the predictive distribution of the MGP model and classifying each realization according to the Robertson and Wride (1998) soil behavior chart. From these samples, the probability of each soil type at every location was calculated using the same approach described in Section 2.1.

For comparison, equivalent results obtained with a single GP model are shown in Fig. 13. The first notable difference between the two approaches appears in the variance distribution of their I_c estimations. In the GP model (Fig. 13b), the variance depends primarily on the distance from the CPTs. In contrast, the variance in the MGP model (Fig. 13b) reflects not only the distance to the CPTs but also the point's layer assignment and its proximity to layer boundaries.

As shown in Fig. 12b, the variance of the MGP predictions increases near layer boundaries, which is consistent with geological reality, where stratigraphic transitions are inherently uncertain. Beyond boundary effects, a key feature of the MGP results is the clear contrast in uncertainty levels between different layers. In particular, the uncertainty of I_c in the top sandy layer is noticeably smaller than that in the middle clayey–silty layer. This behavior is geotechnically reasonable and reflects the inherent variability observed in the CPT data: the top layer exhibits relatively uniform I_c values with low scatter across CPTs, whereas the middle layer shows substantially higher variability due to its mixed soil composition and frequent alternation between silts and clays. Because the MGP assigns a separate GP to each layer, it is able to learn and propagate these layer-specific variance characteristics into the prediction and uncertainty estimates.

Although the most probable soil type distributions obtained from the GP and MGP appear broadly similar (Figs. 12c and 13c), differences emerge at the transition between the second and third layers. In this zone, the stationary GP produces a relatively thick sand-mixture transition (highlighted by red dashed ellipses in Fig. 13c), whereas the MGP yields a thinner transition zone that more closely matches the CPT observations. This difference arises because the MGP framework can accommodate steep changes in the data by allowing both the mean and variance to vary across layers, rather than enforcing a single stationary structure. The contrast becomes even more pronounced in the uncertainty maps of the soil type distribution (Figs. 12d and 13d). In the MGP results, the uncertainty is distinctly higher in the middle clayey–silty layer and lower in the sand-dominated layers, directly mirroring the layer-dependent CPT variability. In contrast, the single GP model produces relatively uniform uncertainty estimates that depend primarily on the distance from CPTs and fail to distinguish variability between soil layers. Consequently, the MGP provides a more realistic and geotechnically interpretable representation of both stratigraphic variability and the associated uncertainties.

The overall interpolation accuracy was further evaluated using the test data (CPT3 and CPT7). The performance metrics of the MGP model are summarized in Table 7, together with those obtained from the GP model and the non-stationary GP model proposed by Yoshida et al. (2021) for comparison. In the approach of Yoshida et al. (2021), CPT data are modeled as the superposition of two Gaussian processes: a smooth zero-mean GP representing the global trend and a non-smooth zero-mean GP representing the residual component. Following their

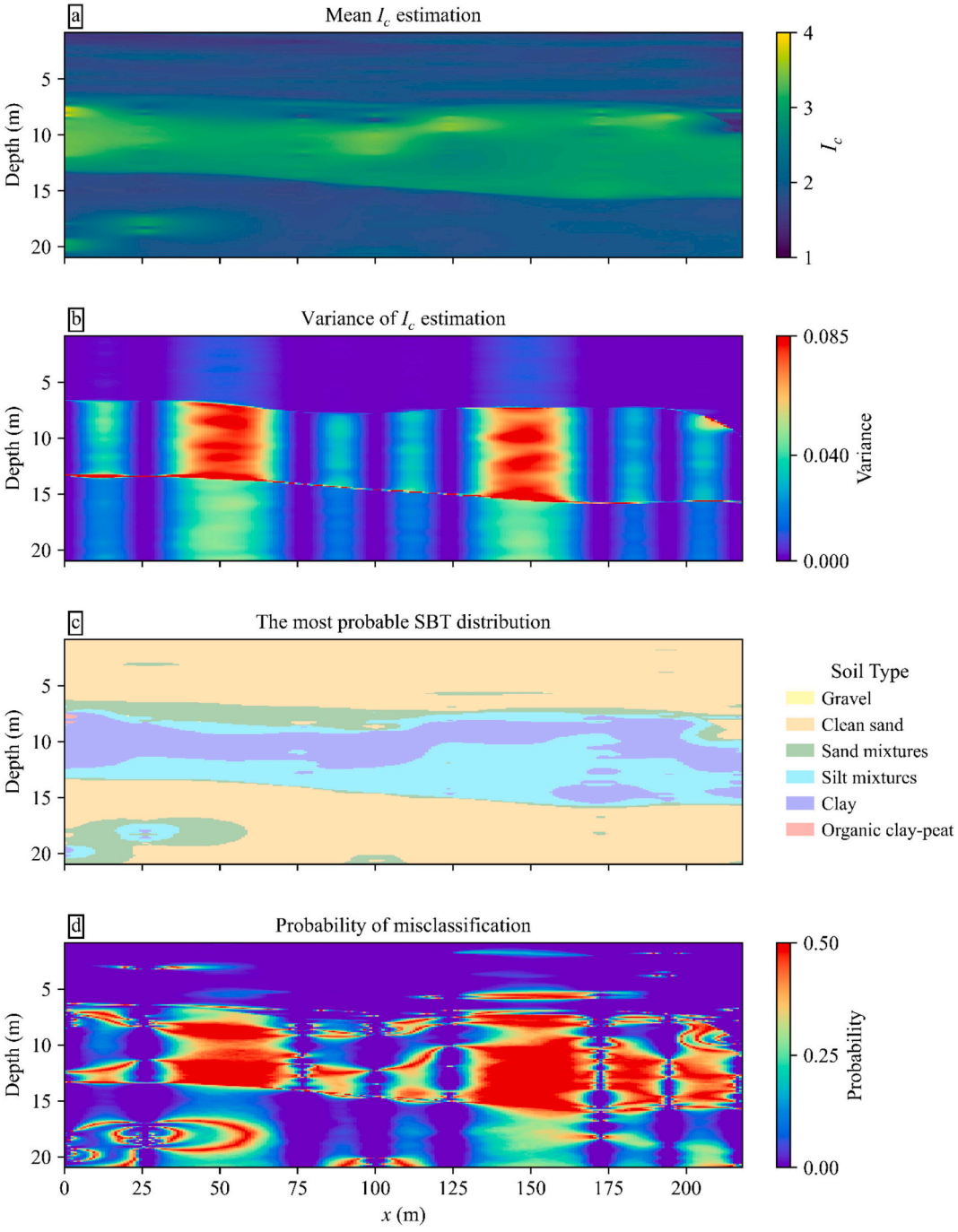


Fig. 12. Prediction results by 3-layer MGP for the entire ground profile.

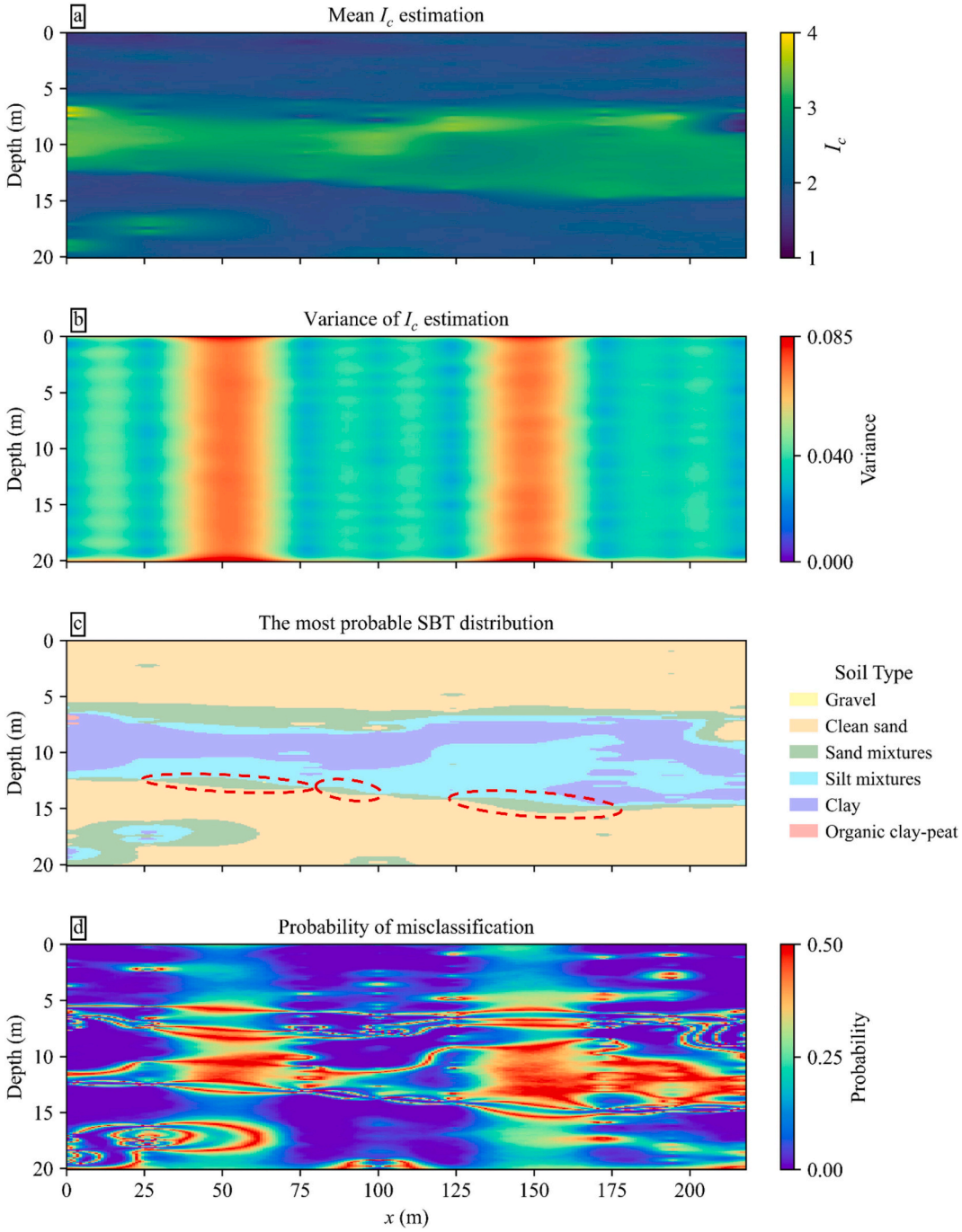


Fig. 13. Prediction results by GP for the entire ground profile.

Table 7
Error metrics calculated on test data.

	RMSE	MAE	MLL	IR
Three-layer MGP	**0.27**	**0.186**	**-0.127**	**0.74**
GP	0.286	0.207	0.171	0.72
Non-stationary GP	**0.27**	0.206	0.141	0.68

recommendation, a squared exponential covariance function was used for the trend component and a single exponential covariance function for the residual component. The results of this comparison are shown in Fig. 14 and Table 7. On the test data, the MGP consistently outperformed the stationary GP, showing a slightly lower RMSE and MAE, a higher IR, and a substantially lower MLL. Compared with the non-stationary GP, the three-layer MGP yielded the same RMSE but a lower MAE and MLL, and higher IR values. Fig. 14 presents the predictions from all models together with the observed data at the test CPTs. The mean prediction of the non-stationary GP appears to change smoothly near layer boundaries, even more smoothly than that of the stationary GP, which may be due to the smooth global trend assumption. This behaviour explains why the non-stationary GP yields a lower IR, despite performing better than the stationary GP in terms of the other error metrics. In addition, the uncertainty bounds produced by the non-stationary GP remain nearly uniform with depth, similar to those of the stationary GP, since both the trend and residual components are modeled as global processes. Although the confidence intervals produced by the non-stationary GP match the observations better than those of the stationary GP and result

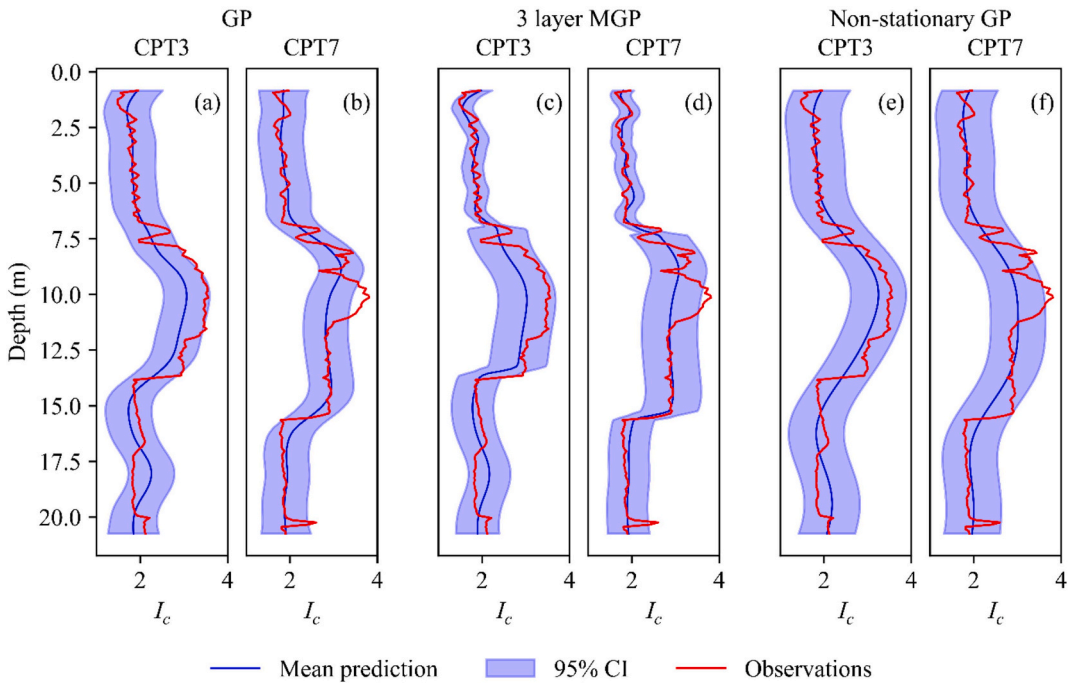


Fig. 14. Comparison of the predictions of GP (a, b), 3-layer MGP (c, d), and non-stationary GP (e,f) on test CPTs.

in lower MLL values, their width remains nearly uniform with depth. In contrast, the uncertainty bounds produced by the MGP vary according to the assigned layer. This allows the MGP to represent changes in uncertainty associated with stratigraphic transitions more effectively. This conclusion is further supported by the significantly lower MLL values achieved by the MGP model (Table 7). Since MLL measures the consistency between the predictive distribution and the observed test data, its reduction indicates that the MGP framework provides a more reliable quantification of uncertainty.

5. Conclusions

This study has presented a Mixture of Gaussian Processes (MGP) framework for two-dimensional site characterization using CPT data. By integrating Gaussian Processes with probabilistic layer estimation, the model addresses the key limitation of conventional GP approaches; their inability to represent non-stationarity arising from stratified soils. A central distinction between the two methods lies in their uncertainty quantification. While a single GP produces uncertainty estimates that depend almost exclusively on the distance from CPTs, the MGP incorporates both the proximity to stratigraphic boundaries and the layer-dependent spatial variability, yielding a more realistic and geotechnically meaningful representation of subsurface uncertainty.

Results from the synthetic experiments and real CPT data show that the MGP generally improves interpolation accuracy and probabilistic performance relative to the standard GP, although the improvement is not uniform across all metrics and scenarios. The most important improvement is reflected in the lower mean log-loss (MLL) values obtained in the inclined and more complex synthetic scenarios and in the real CPT application, indicating that the predictive distributions generated by the MGP provide better probabilistic calibration in non-stationary stratigraphic settings. In Scenario 1, where the layer boundaries are horizontal and the stratigraphic geometry is simple, the benefit of the MGP is less pronounced, and the MLL is slightly worse than that of the GP. This indicates that the advantage of the MGP becomes most relevant when layer boundaries are inclined or geometrically complex, where layer-dependent variability and boundary uncertainty play a larger role. Note that although a quadratic trend was assumed within each layer, the synthetic tests verified that the model does not overfit when the true underlying trend is simpler, underscoring the robustness of the proposed framework.

The proposed MGP framework is most suitable for stratigraphic settings in which the major layer ordering is laterally preserved and the layer boundaries vary smoothly across the section. Under these conditions, the layer-specific I_c fields and the mixture weights can be interpolated effectively by GP and MTGP models, with their covariance structure and horizontal correlation lengths learned from the data by marginal likelihood maximization. However, the method is less suitable for deposits with abrupt stratigraphic unconformities, strong erosional truncation, or layer reordering, because the smooth interpolation of mixture weights cannot explicitly represent such discontinuities. In such cases, denser CPT spacing or the incorporation of additional geological or geophysical information may be required to characterize more complex subsurface conditions. In addition, when the CPT profiles mainly exhibit a common monotonic trend and the remaining non-stationarity is weak, a single-component GP may be sufficient after detrending, reducing the advantage of the proposed mixture formulation.

Finally, it should be noted that CPT measurements have finite measurement support because the cone senses a finite volume of soil rather than an ideal point response (Ahmadi and Robertson, 2005; Boulanger and Dejong, 2018). As a result, thin layers or sharp boundaries may be vertically averaged in the measured I_c profile. This support effect can contribute to apparent boundary smoothing and may also influence the estimated autocorrelation parameters. Consequently, part of the boundary uncertainty represented by the MGP may reflect the finite support of the CPT measurement, rather than true geological uncertainty alone. A full change-of-support treatment is left for future work.

CRediT authorship contribution statement

Muhammet Durmaz: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Zheng Guan:** Writing – review & editing, Supervision. **Michael A. Hicks:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research was carried out within the research program RESET (Reliable Embankments for the Safe Expansion of rail Traffic), funded by ProRail.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the authors used ChatGPT (OpenAI) to assist with grammar checking, language editing, and the preparation of initial plotting scripts for visualizing the original data. The tool was not used to generate, modify, or interpret scientific data, nor to create AI-generated figures or images. All figures were produced by the authors from the original data, and the authors reviewed and edited the output as needed and take full responsibility for the content of the published article.

Data availability

Data will be made available on request.

References

Ahmadi, M.M., Robertson, P.K., 2005. Thin-layer effects on the CPT qc measurement. Canadian Geotechnical Journal 42 (5), 1302–1317. https://doi.org/10.1139/t05-036.

Aitchison, J., Shen, S.M., 1980. Logistic-normal distributions: Some properties and uses. Biometrika 67 (2), 261–272. https://doi.org/10.1093/biomet/67.2.261.

Bonilla, E.V., Chai, K.M.A., Williams, C.K.I., 2007. Multi-task Gaussian process prediction. In: Advances Neural Information Processing System 20.

Boulanger, R.W., DeJong, J.T., 2018. Inverse filtering procedure to correct cone penetration data for thin-layer and transition effects. Cone penetration testing. CRC Press 2018, 25–44.

Cami, B., Javankhoshdel, S., Phoon, K.K., Ching, J., 2020. Scale of fluctuation for spatially varying soils: Estimation methods and values. ASCE-ASME J. Risk Uncertain. Eng. Syst., A. Civ. Eng. 6 (4). https://doi.org/10.1061/ajrua6.0001083.

Chen, Z., Ma, J., Zhou, Y., 2014. A precise hard-cut EM algorithm for mixtures of Gaussian processes. In: International conference on intelligent computing, pp. 68–75.

Ching, J., Phoon, K.K., 2017. Characterizing uncertain site-specific trend function by sparse Bayesian learning. J. Eng. Mech. 143 (7). https://doi.org/10.1061/(ASCE)EM.1943-7889.0001240.

Ching, J., Uzielli, M., Phoon, K.K., Xu, X., 2023. Characterization of autocovariance parameters of detrended cone tip resistance from a global CPT database. J. Geotech. Geoenviron. Eng. 149 (10). https://doi.org/10.1061/jggefk.gteng-11214.

De Gast, T., Vardon, P.J., Hicks, M.A., 2021. Assessment of soil spatial variability for linear infrastructure using cone penetration tests. Géotechnique 71 (11), 999–1013. https://doi.org/10.1680/jgeot.19.SiP.002.

Gardner, J.R., Pleiss, G., Bindel, D., Weinberger, K.Q., Wilson, A.G., 2018. GPyTorch: Blackbox matrix-matrix Gaussian process inference with GPU acceleration. Adv. Neural Inf. Process. Syst. 31. https://gpytorch.ai.

Gelman, A., Jakulin, A., Pittau, M.G., Su, Y.S., 2008. A weakly informative default prior distribution for logistic and other regression models. Ann. Appl. Stat. 2 (4), 1360–1383. https://doi.org/10.1214/08-AOAS191.

Guan, Z., Wang, Y., Phoon, K.K., 2024. Dictionary learning of spatial variability at a specific site using data from other sites. J. Geotech. Geoenviron. Eng. 150 (9), 04024072. https://doi.org/10.1061/JGGEFK.GTENG-12408.

Hu, Y., Wang, Y., 2020. Probabilistic soil classification and stratification in a vertical cross-section from limited cone penetration tests using random field and Monte Carlo simulation. Comput. Geotech. 124. https://doi.org/10.1016/j.compgeo.2020.103634.

Jiang, S.H., Huang, J., 2018. Modeling of non-stationary random field of undrained shear strength of soil for slope reliability analysis. Soils Foundations 58 (1), 185–198.

Lengkeek, A., 2024. CPT based classification with focus on organic soils. In: 7th International Conference on Geotechnical and Geophysical Site Characterisation, pp. 18–21.

Liu, C.N., Chen, C.H., 2010. Estimating spatial correlation structures based on CPT data. Georisk 4 (2), 99–108. https://doi.org/10.1080/17499511003630504.

Lloret-Cabot, M., Hicks, M.A., van den Eijnden, A.P., 2012. Investigation of the reduction in uncertainty due to soil variability when conditioning a random field using kriging. Géotech. Lett. 2 (3), 123–127. https://doi.org/10.1680/geolett.12.00022.

Phoon, K.K., Ching, J., Shuku, T., 2022a. Challenges in data-driven site characterization. Georisk 16 (1), 114–126. https://doi.org/10.1080/17499518.2021.1896005.

Phoon, K.K., Shuku, T., Ching, J., Yoshida, I., 2022b. Benchmark examples for data-driven site characterisation. Georisk 16 (4), 599–621. https://doi.org/10.1080/17499518.2022.2025541.

Rahman, M.H., Abu-Farsakh, M.Y., Jafari, N., 2021. Generation and evaluation of synthetic cone penetration test (CPT) data using various spatial interpolation techniques. Can. Geotech. J. 58 (2), 224–237. https://doi.org/10.1139/cgj-2019-0745.

Rasmussen, C.E., Williams, C.K.I., 2006. Gaussian Processes for Machine Learning. MIT Press.

Robertson, P.K., Wride, C.E., 1998. Evaluating cyclic liquefaction potential using the cone penetration test. Can. Geotech. J. 35 (3), 442–459. https://doi.org/10.1139/t98-017.

Robertson, P.K., 2016. Cone penetration test (CPT)-based soil behaviour type (SBT) classification system—an update. Can. Geotech. J. 53 (12), 1910–1927. https://doi.org/10.1139/cgj-2016-0044.

Shuku, T., Phoon, K.K., 2021. Three-dimensional subsurface modeling using geotechnical lasso. Comput. Geotech. 133. https://doi.org/10.1016/j.compgeo.2021.104068.

Shuku, T., Phoon, K.K., 2023. Comparison of data-driven site characterization methods through benchmarking. ASCE-ASME J. Risk Uncertain. Eng. Syst., A. Civ. Eng. 9 (2). https://doi.org/10.1061/ajrua6.rueng-977.

Shuku, T., Phoon, K.K., Yoshida, I., 2020. Trend estimation and layer boundary detection in depth-dependent soil data using sparse Bayesian lasso. Comput. Geotech. 128. https://doi.org/10.1016/j.compgeo.2020.103845.

Shi, C., Wang, Y., 2022. Assessment of reclamation-induced consolidation settlement considering stratigraphic uncertainty and spatial variability of soil properties. Canadian Geotechnical Journal 59 (7), 1215–1230. https://doi.org/10.1139/cgj-2021-0349.

Theodoridis, S., 2020. Learning in parametric modeling: Basic concepts and directions. In: Machine Learning: A Bayesian and Optimization Perspective, 2nd ed. Academic Press, Oxford, pp. 67–120.

Varkey, D., Hicks, M.A., van den Eijnden, A.P., 2023. Predicting subsurface classification in 2D from cone penetration test data. Transp. Geotech. 43. https://doi.org/10.1016/j.trgeo.2023.101128.

Vessia, G., Di Curzio, D., Castrignanò, A., 2020. Modeling 3D soil lithotypes variability through geostatistical data fusion of CPT parameters. Sci. Total Environ. 698. https://doi.org/10.1016/j.scitotenv.2019.134340.

Wang, Y., Zhao, T., 2017. Statistical interpretation of soil property profiles from sparse data using Bayesian compressive sampling. Géotechnique 67 (6), 523–536. https://doi.org/10.1680/jgeot.16.P.143.

Wang, Y., Hu, Y., Zhao, T., 2020. Cone penetration test (CPT)-based subsurface soil classification and zonation in two-dimensional vertical cross section using Bayesian compressive sampling. Can. Geotech. J. 57 (7), 947–958. https://doi.org/10.1139/cgj-2019-0131.

Wu, D., Ma, J., 2019. An effective EM algorithm for mixtures of Gaussian processes via the MCMC sampling and approximation. Neurocomputing 331, 366–374. https://doi.org/10.1016/j.neucom.2018.11.046.

Yao, J.C., Wang, Y., Guan, Z., Senetakis, K., 2025. Cone penetration test (CPT)-based soil classification and stratification with consideration of data cross-correlation and noises. Acta Geotech. https://doi.org/10.1007/s11440-025-02732-6.

Yoshida, I., Tomizawa, Y., Otake, Y., 2021. Estimation of trend and random components of conditional random field using Gaussian process regression. Comput. Geotech. 136. https://doi.org/10.1016/j.compgeo.2021.104179.

Zhou, Y., Zhao, X., Lin, K.P., Wang, C.H., Li, L., 2019. A Gaussian process mixture model-based hard-cut iterative learning algorithm for air quality prediction. Appl. Soft Comput. 85. https://doi.org/10.1016/j.asoc.2019.105789.

Zhang, W., Brown, E.L., Usmani, A., Earland, N., Kang, M., Olelewe, C., Viswanathan, A., Chauhan, P.S., Steen, C.B., Jeon, H.S., Avagyan, S., Alahi, I., Semenkovich, N.P., Schwab, J.C., Sachs, C.M., Qaium, F., Harris, P.K., Cai, Q., Gentles, A.J., Knight, J., Graham, R.P., Bacchiocchi, A., Lucas, P.C., Fields, R.C., Sznol, M., Halaban, R., Chen, D.Y., Chaudhuri, A.A., Newman, A.M., 2026. Non-invasive profiling of the tumour microenvironment with spatial ecotypes. Nature 654, 1076–1087. https://doi.org/10.1038/s41586-026-10452-4.