

# Assessment of Bayesian Changepoint Detection Methods for Soil Layering Identification Using Cone Penetration Test Data

Stephen K. Suryasentana <sup>1,\*</sup>, Brian B. Sheil <sup>2</sup> and Myles Lawler <sup>3</sup>

<sup>1</sup> Department of Civil and Environmental Engineering, University of Strathclyde, Glasgow G1 1XQ, UK

<sup>2</sup> Laing O'Rourke Centre and Construction Engineering and Technology, Department of Engineering, University of Cambridge, Cambridge CB2 1TN, UK

<sup>3</sup> Independent Geotechnical Consultant, Dublin D02 F6N2, Ireland

\* Correspondence: stephen.suryasentana@strath.ac.uk

**Abstract:** This paper assesses the effectiveness of different unsupervised Bayesian changepoint detection (BCPD) methods for identifying soil layers, using data from cone penetration tests (CPT). It compares four types of BCPD methods: a previously utilised offline univariate method for detecting clay layers through undrained shear strength data, a newly developed online univariate method, and an offline and an online multivariate method designed to simultaneously analyse multiple data series from CPT. The performance of these BCPD methods was tested using real CPT data from a study area with layers of sandy and clayey soil, and the results were verified against ground-truth data from adjacent borehole investigations. The findings suggest that some BCPD methods are more suitable than others in providing a robust, quick, and automated approach for the unsupervised detection of soil layering, which is critical for geotechnical engineering design.

**Keywords:** Bayesian machine learning; ground modelling; site investigation; data driven

**Citation:** Suryasentana, S.K.; Sheil, B.B.; Lawler, M. Assessment of Bayesian Changepoint Detection Methods for Soil Layering Identification Using Cone Penetration Test Data. *Geotechnics* **2024**, *4*, 382–398. <https://doi.org/10.3390/geotechnics4020021>

Academic Editors: Md Rajibul Karim, Md. Mizanur Rahman and Khoi Nguyen

Received: 29 February 2024

Revised: 2 April 2024

Accepted: 3 April 2024

Published: 4 April 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The identification of soil layering is an essential task in geotechnical engineering, as it provides essential information for the design of various infrastructures such as foundations, tunnels, and roads. A soil layer is defined in this paper as a stratum of geological material that belongs to the same classification group. Identifying soil layering involves determining the number of layers and their thickness. Accurately identifying soil layering is crucial for designing foundations such as surface footings [1,2], piles [3–6], and suction caissons [7–10], as it would significantly affect engineering performance such as foundation stiffness [11]. This process is commonly achieved through a manual and often time-consuming interpretation of in situ site investigation measurements using techniques such as borehole sampling and cone penetration testing (CPT). Therefore, there is significant motivation in developing robust and automated interpretive tools for more reliable and objective interpretations of ground models.

The CPT [12] is an in situ testing method that is widely used in geotechnical engineering for soil characterisation, where a cone-shaped probe is pushed into the ground at a constant rate, and measurements of the cone resistance and sleeve friction are recorded continuously as the probe advances. These CPT measurements can be used directly in foundation design (e.g., [13–15]) and to identify soil layering by applying soil behaviour type (SBT) classification rules (e.g., [16–21]). Since the CPT results at each depth can be associated with a soil behaviour type, the boundaries of soil layers can be identified by changes in the soil behaviour type. This technique was improved by using kriging [22] to spatially interpolate the CPT results, enabling the identification of soil layering at unsampled locations.

Bayesian methods have gained significant attention in geotechnical engineering. They have been employed to assess the probability of slope failure [23–27] and to predict tunnel

deformation [28–32], excavation movements [33–37], pipe-jacking forces [38], and pile driveability [39]. The advantages of Bayesian methods in geotechnical engineering include the ability to integrate multiple sources of information, such as expert opinions, field data, and laboratory tests, to reduce uncertainties and improve decision-making. Bayesian methods have also become a popular means of identifying soil layers using CPT data owing to several benefits they offer, such as robustness to noisy measurements and the ability to handle uncertainty in a consistent mathematical framework. For instance, a Bayesian approach [40] was proposed that combines prior knowledge with CPT data to determine the most likely number and boundaries of statistically homogeneous soil layers. It was found that the results of this approach can be significantly affected by the prior knowledge used. As a result, it is recommended to prioritise high-quality, site-specific test data and use relatively uninformative priors when prior knowledge is not well justified. Another study [41] proposed a Bayesian model class selection approach that uses both CPT data and the SBT classification chart to determine the most probable number of soil layers. More recently, a Bayesian framework [42] was proposed for probabilistic soil stratification that uses the CPT-derived SBT index  $I_c$  to determine the most likely number and thicknesses of soil layers, as well as their associated identification uncertainty.

A distinct class of Bayesian methods that have been used to identify soil layer boundaries are Bayesian changepoint detection (BCPD) methods, which originate from the Bayesian machine learning community. Changepoints refer to sudden changes in the data that indicate transitions between different states. These changes divide a sequence of data into non-overlapping partitions, where it is assumed that the data within each partition are generated by the same statistical model. Therefore, the goal of BCPD is to detect points in a data series where the underlying data distribution has changed, indicating a changepoint. In the context of soil data, the changepoint would correspond to a significant change in soil behaviour and, thus, a soil layer boundary. A key advantage of BCPD methods is their computational efficiency, as it typically takes only a few seconds to predict the soil layer boundaries at a single location [43]. There are two types of BCPD methods: offline and online. Offline BCPD methods consider the entire dataset before making inferences about the changepoints, while online BCPD methods detect changepoints in real-time.

The BCPD method was first applied in geotechnical engineering as an offline BCPD method [44] that uses undrained shear strength  $s_u$  to identify layering structures in clayey ground. Recently, both online and offline BCPD methods [43] were proposed that use dynamic penetration test (DPT) data to identify soil layer boundaries between three soil classes (fine-grained soils, sand, and gravel). The BCPD method proposed in Ref. [44] provides the maximum a posteriori estimate of the soil layer boundaries, while the BCPD methods proposed in Ref. [43] provide the probability of a changepoint at each depth, which is then compared to a threshold to determine whether a changepoint has been detected. The main difference between the BCPD methods proposed in Refs. [43,44] is that the former is an ‘unsupervised’ method that requires no training process, while the latter is a ‘supervised’ method that requires a training process to calibrate the changepoint threshold. Unsupervised BCPD methods are generally more convenient to apply, as it takes significant time and effort to prepare the training data.

This paper addresses significant gaps in current research, particularly focusing on the challenges associated with Bayesian methods applied to CPT data for soil layer identification. The primary issue is that most existing Bayesian approaches are computationally demanding and time-consuming, which limits their adoption in favour of faster methods like the SBT classification method. Although the BCPD method offers a fast Bayesian approach, its application to CPT data remains unexplored. Consequently, it is uncertain if BCPD methods can effectively delineate soil layer boundaries using CPT data. Moreover, prior research on BCPD methods in geotechnical engineering has exclusively focused on univariate data, overlooking the potential use of multivariate data available from CPT measurements, such as tip resistance and sleeve friction. Thus, it is uncertain if utilizing

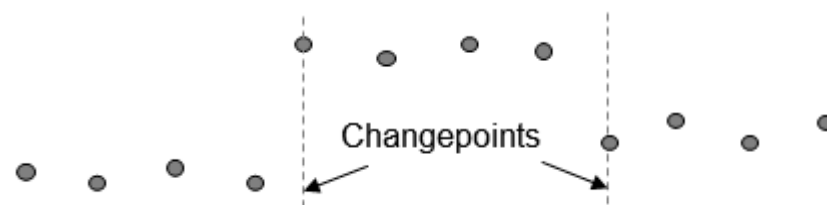
multiple data points could enhance changepoint detection accuracy. This paper aims to fill these research gaps.

The main aim of the current paper is to provide an assessment of the accuracy and computational efficiency of different unsupervised BCPD methods for soil layering identification using CPT data, in view of identifying the most effective BCPD method for this task. Four types of BCPD methods are assessed: a previously established offline univariate method [44] designed for identifying clay layers via undrained shear strength data, a new online univariate method introduced in this study, and multivariate versions of both methods that are capable of detecting changepoints using multiple data sources from CPT. The core task to be carried out by the BCPD methods is to divide the soil profile up into two categories: (i) predominantly fine-grained soils (e.g., clay and silt) and (ii) predominantly coarse-grained soils (e.g., sand and gravel). These categories are the two main groups of the USCS [45] soil classification system. These two soil categories have very different permeability, stiffness, and strength properties such that poor identification will have a negative impact on geotechnical design. The evaluation of the BCPD methods is carried out using real-world CPT data from a case study involving multi-layered sandy-clayey deposits. The BCPD predictions of soil layer boundaries are compared with the ground truth provided by the neighbouring borehole data.

The novel contributions of this paper are as follows: (i) it is the first study to evaluate the effectiveness of different BCPD methods for identifying soil layers using CPT data; (ii) it pioneers the application of BCPD methods to multivariate geotechnical data, exploring whether this approach offers advantages over the traditional use of univariate data in previous research; and (iii) it develops two new unsupervised online BCPD methods suitable for univariate and multivariate data, termed BCPD-ON and BCPD-ON-MV, respectively.

## 2. Methodology

Changepoint detection has been used in various fields, including medical condition monitoring, climate change detection, speech analysis, and image analysis [46–50]. Figure 1 illustrates the concept of changepoints for an arbitrary data series.



**Figure 1.** Illustration of a series of data points (shown as grey markers). There are three distinct partitions in the data series, separated by changepoints. The changepoints are detected at locations where abrupt changes are observed.

Numerous studies have focused on developing new methods for detecting changepoints, as summarised in [51–58]. These methods can be broadly classified as either offline or online. Offline methods (e.g., [59,60]) use the entire dataset to detect changepoints, whereas online methods (e.g., [61,62]) process each data point in real-time to detect a changepoint as soon as it occurs. This paper assesses the effectiveness of offline and online unsupervised BCPD methods for identifying soil layering. Univariate and multivariate data series from CPT are used to evaluate both methods. An overview of each BCPD method is provided below, where it is assumed that the dataset being interrogated contains one-dimensional ‘depth-series’ data indexed by depth. To clarify the notation used in this paper,  $x_{1:n}$  refers to the set of data  $\{x_1, x_2, \dots, x_{n-1}, x_n\}$  and  $p(A|B)$  refers to the probability of event  $A$  occurring given that event  $B$  has occurred.

### 2.1. Offline BCPD

This paper investigates the offline BCPD method [59,60], which was previously used in geotechnical engineering [43] to delineate clay layers based on undrained shear strength data. Offline Bayesian changepoint detection algorithms aim to pinpoint moments when the statistical characteristics of data significantly shift. Unlike online methods that process data points in real-time as they are received, offline approaches analyse the complete dataset in a single operation after all data have been gathered. This analysis involves comparing every possible set of changepoint locations to identify the most probable arrangement based on the data. Tackling the entire dataset simultaneously, particularly with large datasets, demands substantial computational effort. The offline BCPD method studied in the current paper employs a recursive algorithm to efficiently calculate the posterior probability distribution for changepoint locations, a notable improvement over earlier Markov Chain Monte Carlo methods [63]. This recursive approach simplifies the calculation process by breaking down the overall problem into smaller, more manageable sub-problems, allowing for a more feasible and efficient determination of changepoint probabilities. Detailed explanations of this algorithm can be found in the literature [43,59,60], and the following section briefly describes the key equations behind the algorithm. The data within each partition are modelled using a probability distribution, with parameters that are independent of those determined for other partitions. Let  $c_j$  represent the  $j$ th changepoint. The posterior distribution of  $c_j$  is  $p(c_j | x_{1:n})$ . The probability of a changepoint occurring at depth  $z$  can be calculated as follows:

$$p(\text{changepoint at } z | x_{1:n}) = \sum_{j=1}^z p(c_j = z | x_{1:n}) \quad (1)$$

where the summation is due to the possibility of there being 1 to  $z$  changepoints thus far at depth  $z$ .  $p(c_j | x_{1:n})$  in Equation (1) is obtained by marginalising out the previous changepoints:

$$p(c_j | x_{1:n}) = \int p(c_j, \dots, c_1 | x_{1:n}) dc_{j-1} \dots dc_1 \quad (2)$$

As the probability of a changepoint is assumed to be dependent only on the previous changepoint, the integrand in Equation (2) can be calculated as follows:

$$p(c_j, \dots, c_1 | x_{1:n}) = p(c_j | c_{j-1}, x_{1:n}) p(c_{j-1} | c_{j-2}, x_{1:n}) \dots p(c_2 | c_1, x_{1:n}) p(c_1 | x_{1:n}) \quad (3)$$

Each of the terms on the right-hand side of Equation (3) can be calculated exactly using the recursive algorithm described in Ref. [59], which is briefly outlined here. Let  $L(i, j) = p(x_{i:j} | \text{partition from } i \text{ to } j)$  be the likelihood of the data in the partition from depth  $i$  to  $j$ , and  $g(n) = 1/n$  be the prior distribution of a changepoint. In the context of CPT data,  $L(i, j)$  can be interpreted as the likelihood of the CPT data at depth  $i$  to  $j$  belonging to the same soil layer. Let  $Q(z) = p(x_{z:n} | \text{changepoint at } z - 1)$  be the marginal probability of the data from depth  $z$  to the end, given that there is a changepoint at depth  $z - 1$ :

$$\begin{aligned} Q(z) &= \sum_{j=z}^{n-1} p(x_{z:n} | \text{next changepoint at } j) + p(x_{z:n} | \text{no more changepoint}) \\ &= \sum_{j=z}^{n-1} L(z, j) Q(j + 1) g(n) + L(z, n) g(n) \end{aligned} \quad (4)$$

The presence of  $Q(j + 1)$  indicates that Equation (4) is a recursive algorithm. To calculate  $Q(z)$  for all depths, Equation (4) should be applied recursively, starting from depth  $n - 1$  and working backwards. Thereafter, each of the terms on the right-hand side of Equation (3) is calculated as follows:

$$\begin{aligned}
 p(c_j|c_{j-1}, x_{1:n}) &= \frac{p(c_j, x_{c_{j-1}+1:n}|c_{j-1})}{p(x_{c_{j-1}+1:n}|c_{j-1})} \\
 &= \frac{p(x_{c_{j-1}+1:c_j}|c_j, c_{j-1})p(x_{c_j+1:n}|c_j, c_{j-1})p(c_j|c_{j-1})}{p(x_{c_{j-1}+1:n}|c_{j-1})} \\
 &= \frac{L(c_{j-1} + 1, c_j)Q(c_j + 1)g(n)}{Q(c_{j-1} + 1)}
 \end{aligned} \tag{5}$$

### Maximum a Posteriori Estimations of Changepoints

In Ref. [43], a changepoint is defined to have occurred if the probability calculated by Equation (1) exceeds some user-defined probability threshold. This threshold is to be determined by a training process that relates the changepoint probability to some ground-truth data such as expert judgment of a soil layer boundary. To avoid this training process, the current study will calculate the maximum a posteriori locations of the changepoints using the Viterbi algorithm [59], as follows. Let  $Q_m(z) = p(x_{z:n}|\text{changepoint at } z - 1, \text{ maximum a posteriori changepoint for } z:n)$  be the probability of the data from depth  $z$  to the end, given that there is a changepoint at depth  $z - 1$  and there is a maximum a posteriori estimate of the next changepoint between depth  $z$  to the end.  $Q_m(z)$  can be calculated as follows:

$$\begin{aligned}
 Q_m(z) &= \max_{j \in z:n} \{p(x_{z:n}|\text{next changepoint at } j)\} \\
 &= \max_{j \in z:n} \{L(z, j)Q_m(j + 1)g(n)\}
 \end{aligned} \tag{6}$$

The presence of  $Q_m(j + 1)$  indicates that Equation (6) is a recursive algorithm. Similar to Equation (4), Equation (6) should be applied recursively, starting from depth  $n - 1$  and working backwards. Let  $j_m(z) = \operatorname{argmax}_{j \in z:n} \{p(x_{z:n}|\text{next changepoint at } j)\}$  be the maximum a posteriori estimate of the next changepoint from depth  $z$  to the end. After  $Q_m(z)$  has been calculated for all depths, the maximum a posteriori estimates of the changepoints can be determined recursively as follows. First, let  $c_1 = j_m(1)$ . Then, let  $c_2 = j_m(c_1 + 1), c_3 = j_m(c_2 + 1) \dots, c_j = j_m(c_{j-1} + 1)$  until the end of the data is reached.

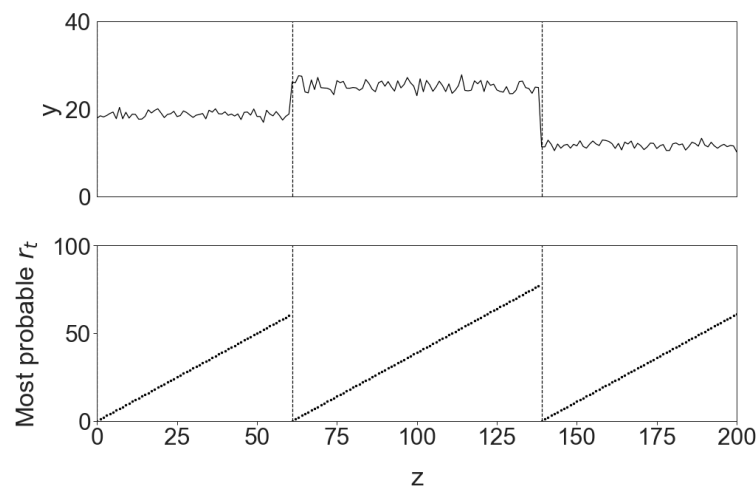
The authors of Refs. [59,60] considered only univariate data series, while another study [64] expanded on that work by considering multivariate data series. The offline BCPD methods for univariate and multivariate data are referred to in this paper as ‘BCPD-OFF’ and ‘BCPD-OFF-MV’, respectively. The main difference in the implementation of these two methods is the modelling of the likelihood and conjugate prior of the data, which will be discussed later in this paper.

### 2.2. Online BCPD

The second BCPD method investigated in this paper is based on the online BCPD method proposed in Ref. [61], which estimates the probability of a changepoint at a given depth based on data processed up to that depth. This method computes the probability distribution of a random variable called the ‘run length’  $r_z$ , which represents the amount of data between the current depth  $z$  and the last changepoint. The reason behind the need for the run length variable is that it encapsulates essential information about the recent history of the data stream in a single statistic. Online changepoint detection processes data sequentially as they arrive. It aims to detect changepoints as soon as possible after they occur, without the benefit of future observations. In contrast, offline methods can consider the data in their entirety, applying algorithms that optimise over all possible changepoints simultaneously, thus not requiring the concept of run length. The goal of online changepoint detection methods is to update beliefs about the presence of a changepoint with each new data point. By maintaining a probability distribution over the run length variable, the

algorithm can make immediate decisions about the likelihood that a changepoint has just occurred. This allows for real-time detection, and is more efficient than offline changepoint detection methods where all data must be analysed together.

The changepoints divide the sequence of data into non-overlapping partitions, where the length of each partition is  $r_z$ . Each new data point either belongs to the same distribution (and  $r_z$  increases by one), or it belongs to a new distribution (which means a changepoint occurs and  $r_z$  resets to zero). Therefore, a spike in the probability of  $r_z = 0$  suggests the likely presence of a changepoint at depth  $z$ . If  $r_z$  increases by one, the new data point will update the parameter estimates of the current distribution using Bayes' theorem. Otherwise, the new distribution resets back to the prior distribution. Figure 2 illustrates how the (known) changepoints for an arbitrary data sequence coincide with the locations where the most probable value of  $r_z$  is 0.



**Figure 2.** Illustration of the development of the most probable  $r_z$  for a sequence of data (shown in bottom subfigure) and how the changepoints coincide with the locations where the most probable  $r_z$  is 0. Here,  $y$  is the measured quantity and  $z$  is depth (for depth series data). The dashed lines represent the known locations of the changepoints.

The probability of a changepoint at depth  $z$  is equivalent to the posterior probability of  $r_z = 0$  at depth  $z$ , as follows:

$$p(\text{changepoint at } z | x_{1:z}) = p(r_z = 0 | x_{1:z}) \tag{7}$$

The posterior distribution of the run length  $p(r_z | x_{1:z})$  in Equation (7) is calculated using the recursive algorithm described in Ref. [61], which is briefly explained here.  $p(r_z | x_{1:z})$  in Equation (7) can be calculated as follows:

$$p(r_z | x_{1:z}) = \frac{p(r_z, x_{1:z})}{p(x_{1:z})} \tag{8}$$

where  $p(x_{1:z}) = \sum_{r_z} p(r_z, x_{1:z})$ . The joint distribution  $p(r_z, x_{1:z})$  can be calculated using the following recursive relationship:

$$\begin{aligned} p(r_z, x_{1:z}) &= \sum_{r_{z-1}} p(r_z, x_z, |r_{z-1}, x_{1:z-1}) p(r_{z-1}, x_{1:z-1}) \\ &= \sum_{r_{z-1}} p(r_z | x_z, r_{z-1}, x_{1:z-1}) p(x_z | r_{z-1}, x_{1:z-1}) p(r_{z-1}, x_{1:z-1}) \\ &= \sum_{r_{z-1}} p(r_z | r_{z-1}) p(x_z | r_{z-1}, x_{(z-1-r_z):z-1}) p(r_{z-1}, x_{1:z-1}) \end{aligned} \tag{9}$$

The following describes how each of the three terms in the last line of Equation (9) is calculated. First,  $p(r_{z-1}, x_{1:z-1})$  is a recursive term, which represents the previous iteration

of Equation (9) at depth  $z - 1$ . Second,  $p(r_z|r_{z-1})$  is the conditional prior of the run length:

$$p(r_z|r_{z-1}) = \begin{cases} 1/\kappa & \text{if } r_z = 0 \\ 1 - 1/\kappa & \text{if } r_z = r_{z-1} + 1 \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

where  $\kappa$  is a parameter that controls the sensitivity of the changepoint occurrence. Larger values of  $\kappa$  mean that stronger evidence is required to support a higher changepoint probability. To allow for a fair comparison with the offline BCPD method,  $\kappa$  is set as the number of data points in order to match the prior probability of changepoints  $g(n)$  for the offline BCPD method. Finally,  $p(x_z|r_{z-1}, x_{(z-1-r_z):z-1})$  is the posterior predictive distribution of the data point  $x_z$  based on the likelihood of the data partition. This is calculated analytically due to the use of conjugate priors, as described later in this paper.

### Maximum a Posteriori Estimations of Changepoints

The online BCPD method in Ref. [43] defines a changepoint to have occurred if the probability calculated by Equation (7) exceeds some calibrated user-defined probability threshold. To avoid the training process required for this threshold, the maximum a posteriori estimates of the changepoints are used instead. As the algorithm for obtaining these estimates is not provided in Ref. [61], the current study proposes the following new algorithm to identify them.

Let  $M_z$  be the event that the maximum a posteriori estimate of changepoints has occurred prior to depth  $z$ . Let  $C_{MAP}(z) = p(r_z, M_z, x_{1:z})$  be the joint probability of the current run length and the maximum a posteriori estimate of the changepoints prior to depth  $z$ . It can be calculated as follows:

$$C_{MAP}(z) = \max_{j \in 1:z-1} \{p(r_z = z - j | x_{1:z})p(r_j = 0 | x_{1:j})C_{MAP}(j - 1)\} \quad (11)$$

where  $C_{MAP}(0) = 1$ . Equation (11) should be applied recursively, starting from the first data point and working forward until  $C_{MAP}(z)$  is calculated for all depths. At any depth  $z$ , the maximum a posteriori estimate of the set of changepoints prior to depth  $z$  can be determined recursively using the values of  $j$  obtained in each recursive calculation. This process can be explained by estimating the maximum a posteriori locations of the changepoints for the exemplar data series in Figure 1. Suppose  $z = 12$  (i.e., the maximum a posteriori estimations is made considering all the data points in Figure 1), the maximum a posteriori estimate of the set of changepoints is  $\{5,9\}$  and it is obtained by applying Equation (11) repeatedly until the following recursive path produces the maximum value for  $C_{MAP}(12)$ :

$$C_{MAP}(12) = p(r_{12} = 3 | x_{1:12})p(r_9 = 0 | x_{1:9})C_{MAP}(8) \quad (12)$$

where

$$C_{MAP}(8) = p(r_8 = 3 | x_{1:8})p(r_5 = 0 | x_{1:5})C_{MAP}(4) \quad (13)$$

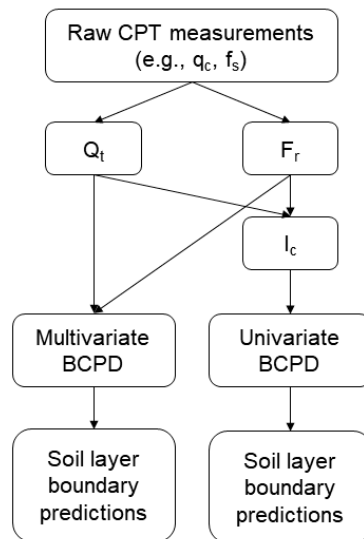
$$C_{MAP}(4) = p(r_4 = 3 | x_{1:4})p(r_1 = 0 | x_{1:1})C_{MAP}(0) \quad (14)$$

The online BCPD methods for univariate and multivariate data are referred to in this paper as ‘BCPD-ON’ and ‘BCDP-ON-MV’, respectively. The main difference in the implementation of these two methods is the modelling of the likelihood and conjugate prior of the data, which will be discussed in the next section.

### 2.3. CPT Data Case Study

The CPT data used in this paper were acquired from a ground investigation carried out in Brandenburg, Germany. The ground conditions consist of layers of coarse-grained soils (predominantly sands, including gravelly sand) and fine-grained soils (silts and

clay). The water table at the site is approximately 6.5 m to 7.5 m below ground level. The ground investigation included conducting CPTs and borehole sampling at discrete locations, and there is at least one borehole positioned within 3 m of each CPT location. This means that each CPT dataset had a corresponding ‘ground-truth’ profile to evaluate the soil layering predictions. In total, five pairs of CPTs and borehole data are evaluated in this paper. The five CPT locations are termed CPT01, ..., CPT05 in this paper. A workflow summarising the application of the BCPD methods to the raw CPT measurements is shown in Figure 3.



**Figure 3.** Workflow summarising the steps from the raw CPT measurements to the soil layer boundary predictions by the univariate and multivariate BCPD methods.

The raw CPT measurements (e.g., cone tip resistance  $q_c$  and sleeve friction  $f_s$ ) are typically normalised for soil classification purposes. For example, the Robertson (2009) SBT classification method [18] uses the following normalised versions of the raw CPT measurements:

$$Q_t = \frac{q_t - \sigma_{v0}}{\sigma'_{v0}} \tag{15}$$

$$F_r = \frac{f_s}{q_t - \sigma_{v0}} \tag{16}$$

$$B_q = \frac{u_2 - u_0}{q_t - \sigma_{v0}} \tag{17}$$

where  $Q_t$  is the normalised cone resistance,  $F_r$  is the normalised friction ratio,  $\sigma'_{v0}$  and  $\sigma_{v0}$  are the in situ vertical effective and total stress, respectively, and  $B_q$  is the pore pressure ratio. The parameter  $q_t = q_c + (1 - a)u_2$  is the total cone end resistance corrected for the presence of the pore pressure filter, where  $a$  is cone area ratio.  $\sigma'_{v0}$  and  $\sigma_{v0}$  are the in situ vertical effective and total stress, respectively. The authors in Ref. [20] propose a soil classification index  $I_c$  that combines  $Q_t$  and  $F_r$  to approximate the SBT boundaries as follows:

$$I_c = \sqrt{(3.47 - \log_{10}(Q_t))^2 + (1.22 + \log_{10}(F_r))^2} \tag{18}$$

Table 1 gives the  $I_c$  ranges that correspond to the different SBT zones. For the purpose of this paper, fine-grained soils are defined as soils that belong to the Robertson (2009) SBT zones 2, 3, and 4, while coarse-grained soil are defined as the Robertson (2009) SBT zones



5, 6, and 7. The boundary between fine- and coarse-grained soils corresponds to an  $I_c$  value of approximately 2.6.

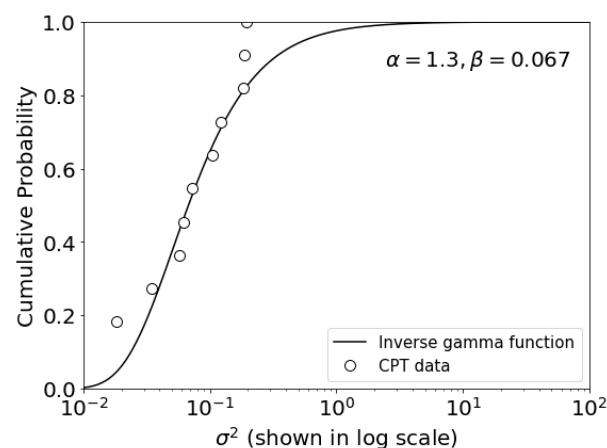
**Table 1.** Soil behaviour type zones and their corresponding  $I_c$  ranges.

SBT Zone	$I_c$ Range	Soil Mixture Description
9	-	Stiff fine grained
8	-	Stiff sand to clayey sand
7	<1.31	Gravelly sand to dense sand
6	1.31–2.05	Clean sand to silty sand
5	2.05–2.6	Silty sand to sandy silt
4	2.6–2.95	Clayey silt to silty clay
3	2.95–3.6	Silty clay to clay
2	>3.6	Organic soils
1	-	Sensitive soils

$I_c$  is an empirical index that incorporates information from both  $Q_t$  and  $F_r$ . However, it is uncertain if soil layering identification would be more effective by using  $Q_t$  and  $F_r$  directly. Therefore, this paper will assess the effectiveness of using  $I_c$  as a univariate data input and  $Q_t$  and  $F_r$  as multivariate data inputs into the BCPD methods.

#### 2.4. Priors for Univariate and Multivariate BCPD Methods

The likelihood of univariate and multivariate data is modelled using a Gaussian distribution and multivariate Gaussian distribution, respectively. Conjugate prior distributions are adopted for the data, which allow for efficient analytical calculations of the posterior distributions [65–68]. The same likelihood and conjugate prior distributions are adopted for both the online and offline BCPD methods. Following Ref. [44], a normal-inverse gamma distribution  $NIG(\mu_h, \lambda, \alpha, \beta)$  is adopted as the conjugate prior for univariate data. As the BCPD methods are proposed as unsupervised methods, the prior hyperparameters values for the normal-inverse gamma distribution are not calibrated using the CPT data considered in this paper and are simply set to the values determined in Ref. [43]:  $\alpha = 1, \beta = 0.1, \lambda = 1, \mu_h = 0$ . Perturbation of these hyperparameters values indicates that the final results of the current study are generally not sensitive to their values. Retrospective calibration analysis using the CPT  $I_c$  data provides the optimal values of  $\alpha = 1.3, \beta = 0.067$ , as shown in Figure 4, but the use of these optimal values did not change the univariate BCPD predictions of the soil layer boundaries in this paper.



**Figure 4.** Comparison of the best-fit inverse gamma cumulative distribution with the actual cumulative distribution of the variance in the CPT  $I_c$  data within each soil layer (known from the borehole data).

Following Ref. [64], a normal-inverse Wishart distribution  $NIW(\boldsymbol{\mu}_0, \lambda_0, \mathbf{V}_0, N_0)$  is adopted as the conjugate prior for multivariate data, and weakly informative prior hyperparameter values of  $N_0 = d$ ,  $\mathbf{V}_0 = \hat{\sigma}^2 \mathbf{I}$ ,  $\lambda_0 = 1$ ,  $\boldsymbol{\mu}_0 = \mathbf{0}$  are adopted, where  $d$  is the dimensions of the multivariate data (i.e., two for the current study),  $\hat{\sigma}^2$  is the mean of the empirical variance pooled across all the data, and  $\mathbf{I}$  is the identity matrix. Employing weakly informative priors gently steers the early stages of the analysis without dictating the outcome. This approach allows the data to play a significant role in shaping the final results.

### 2.5. Performance Metrics

The performance of the proposed BCPD methods used in this study is evaluated in terms of accuracy and computational efficiency. To quantify the accuracy of the methods, several accuracy metrics are used as follows:

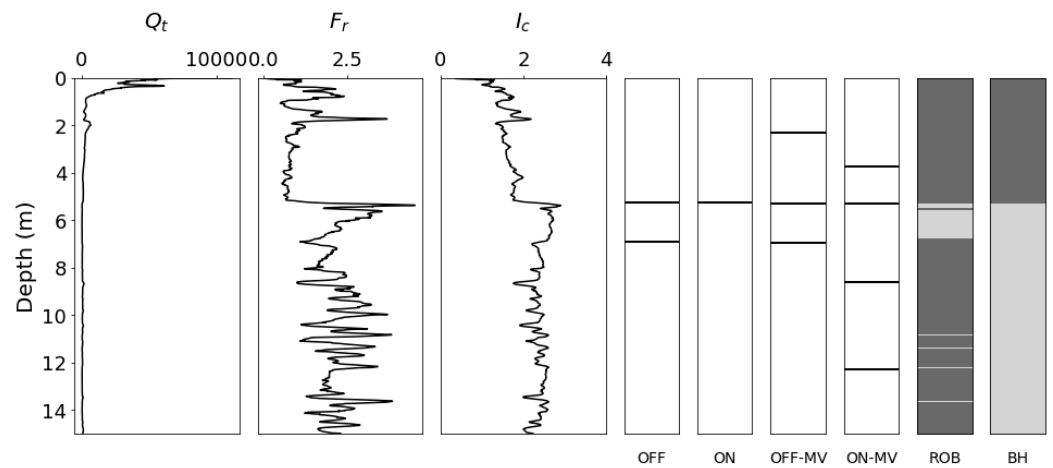
- True Positive (TP)—the number of times the method has correctly identified a soil layer boundary;
- False Positive (FP)—the number of times the method has incorrectly identified a soil layer boundary;
- False Negative (FN)—the number of times the method has failed to identify a true soil layer boundary;
- Precision =  $TP / (TP + FP)$ ;
- Sensitivity =  $TP / (TP + FN)$ ;
- F1 score =  $2(\text{Precision} \times \text{Sensitivity}) / (\text{Precision} + \text{Sensitivity})$ .

Precision, sensitivity, and F1 score are common composite metrics comprising TP, FP, and FN; a higher value indicates better performance. Precision is a measure of how accurate a model is when it detects something, sensitivity is a measure of how good a model is at detecting something, while the F1 score is an overall measure of how good a model is at both detecting something and the accuracy of those detections. It is important to note that due to the potential deviation of up to 3 m between the borehole location and the CPT location, as well as the unaccounted cone sensing and development distances [69], the predicted boundaries based on the CPT data are not expected to exactly align with the boundaries indicated by the borehole data. Consequently, this study considers a soil layer boundary to be correctly identified if the CPT-predicted boundary falls within a distance of 1m from the borehole-indicated soil layer boundary. To evaluate the computational efficiency of the methods, the average computational times required to process each CPT location by the proposed BCPD methods are calculated and compared.

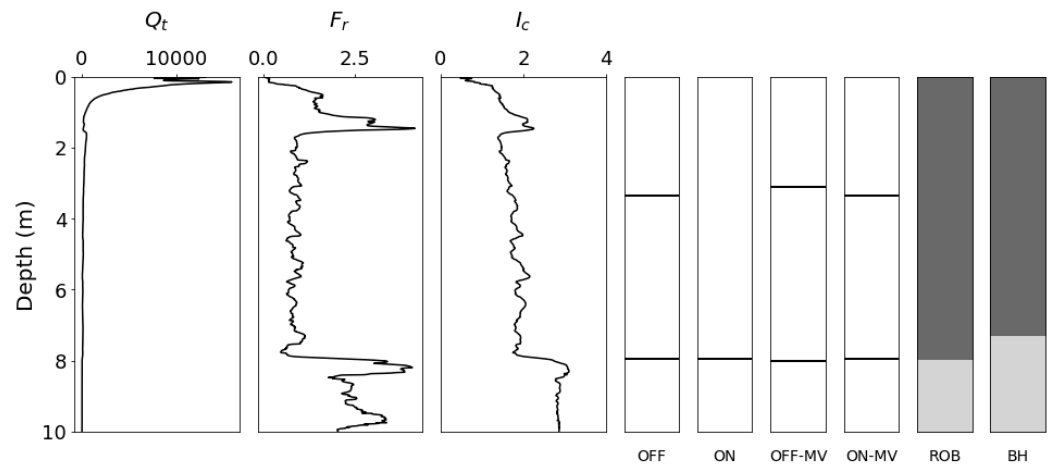
## 3. Results

This section presents the soil layering predictions for the CPT locations using the BCPD methods. The BCPD predictions are compared with soil layering predictions obtained from the  $I_c$  form of the Robertson (2009) SBT classification method, where an  $I_c$  value of 2.6 is considered the threshold between fine- and coarse-grained soils.

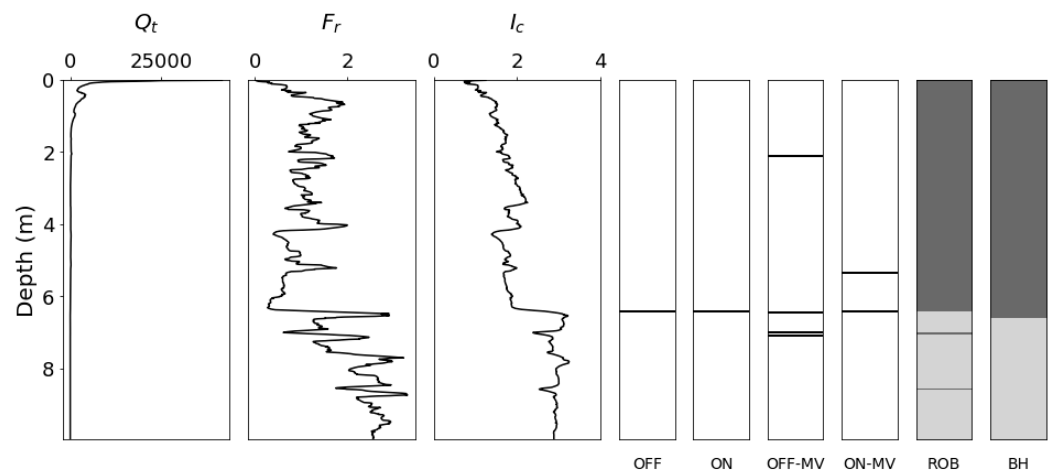
Figures 5–9 compare the BCPD predictions, the Robertson (2009) predictions, and the expert interpretation of the corresponding borehole data for CPT locations 1 to 5. This expert interpretation approach is consistent with the validation approach described in Ref. [44]. It represents the existing manual and time-consuming process required to determine information on soil layering, and is considered as the industry standard. Overall, these figures demonstrate that BCPD-ON provides the most accurate soil layering predictions, correctly identifying almost all soil layer boundaries as determined from the borehole data. It is worth noting that the BCPD methods are unsupervised, meaning they are not trained using any data to discern the locations of the soil layer boundaries. Nevertheless, the detected change points in BCPD approximately coincide with the soil layer boundaries from the borehole data.



**Figure 5.** Comparison of soil layer boundary predictions by the different BCPD methods (shown as horizontal black lines) with the corresponding Robertson (2009) predictions (labelled as ‘ROB’) and ground truth provided by neighbouring borehole data (labelled as ‘BH’), at location CPT01. Fine- and coarse-grained soils are shown as light and dark grey colours, respectively.



**Figure 6.** Comparison of soil layer boundary predictions with the borehole data at location CPT02.



**Figure 7.** Comparison of soil layer boundary predictions with the borehole data at location CPT03.

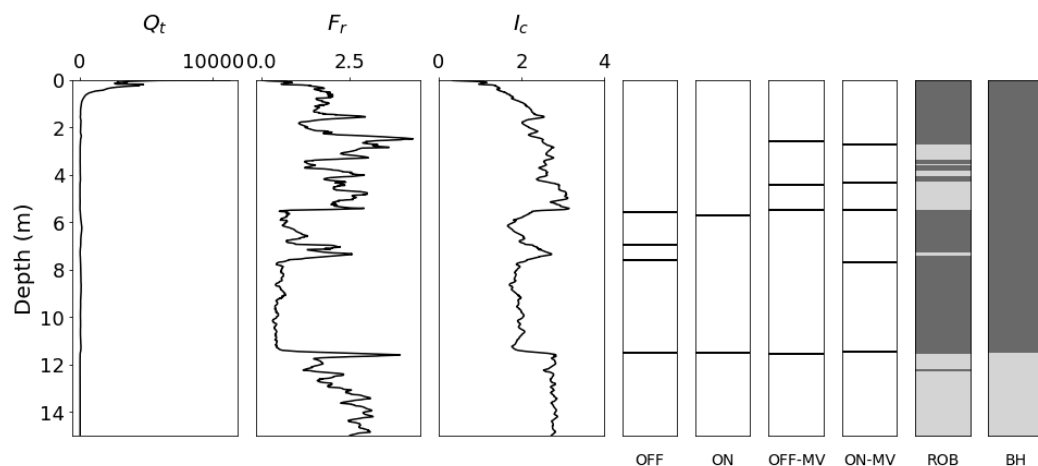


Figure 8. Comparison of soil layer boundary predictions with the borehole data at location CPT04.

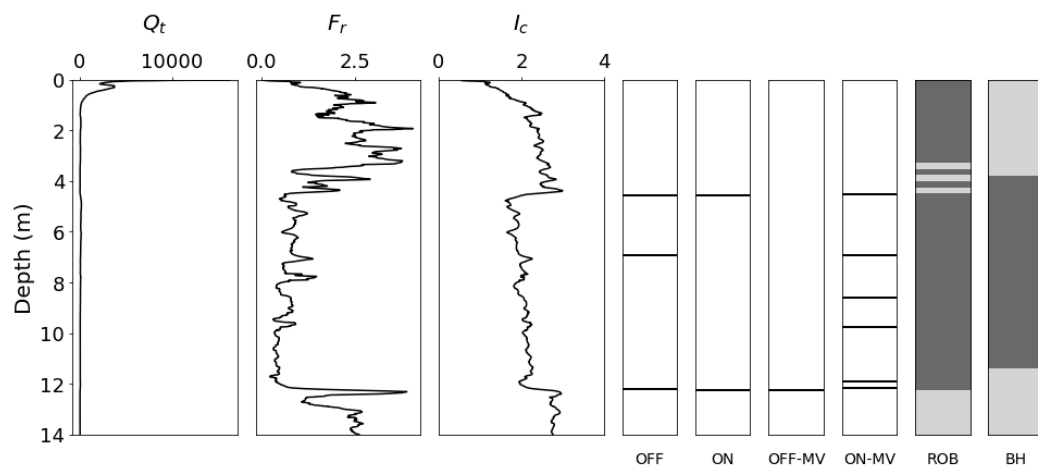


Figure 9. Comparison of soil layer boundary predictions with the borehole data at location CPT05.

BCPD-OFF also predicts the same soil layer boundaries as BCPD-ON but with more false soil layer boundary predictions. As for the multivariate counterparts of BCPD-ON and BCPD-OFF, it can be observed that BCPD-ON-MV and BCPD-OFF-MV generate even more false soil layer boundary predictions compared to their univariate counterparts.

Additionally, Figures 5–9 reveal that the Robertson (2009) method is prone to predicting numerous thin layers. This tendency is noticeable at depths where the computed  $I_c$  values are close to the assumed boundary between fine- and coarse-grained soils such that a small amount of noise can trigger the prediction of multiple thin layers.

*Comparison of Performance Metrics*

The accuracy metrics for the different BCPD methods are shown in Table 2.

Table 2. Accuracy metrics for the soil layer boundary predictions by the different methods.

Method	True Positive	False Positive	False Negative	Precision	Sensitivity	F1 Score
BCPD-OFF	6	5	0	0.545	1	0.706
BCPD-ON	6	1	0	0.857	1	0.923
BCPD-OFF-MV	5	8	1	0.385	0.833	0.526
BCPD-ON-MV	6	12	0	0.333	1	0.5
Robertson (2009)	6	18	0	0.25	1	0.4

Among these methods, the BCPD-ON method consistently outperforms the others, achieving the highest scores across all metrics. It notably achieves the highest F1 score of 0.923, indicating a strong balance between sensitivity and precision. Comparatively, all the BCPD methods evaluated in this study exhibit higher F1 scores than the Robertson (2009) method, which has the lowest precision due to its tendency to make numerous predictions of thin-layer soils. However, it is important to acknowledge that, depending on the geotechnical application, this model's behaviour may be desirable in certain geotechnical applications. For example, in slope stability assessments, it is crucial to identify very thin, free-draining layers that might not be easily detected using borehole data alone. These layers play a significant role in promoting excess pore pressure dissipation.

Table 3 presents the average computational time required to process CPT data for each location using the different BCPD methods. Using a computer with an Intel i7 2.8 GHz processor (eight central processing units) and 8 GB of RAM, the BCPD-ON method demonstrates the highest efficiency, processing each CPT location in approximately 0.044 s. This is about 60 times faster than the next fastest method, BCPD-ON-MV. Generally, the multivariate BCPD methods are slower than the univariate BCPD methods.

**Table 3.** Time taken for the soil layer boundary predictions for each location in the dataset.

Method	Time Taken Per CPT Location (s)
BCPD-OFF	2.46
BCPD-ON	0.044
BCPD-OFF-MV	3.32
BCPD-ON-MV	2.38

#### 4. Discussion

The study found that univariate BCPD methods are generally more effective in identifying soil layering from CPT data compared to their multivariate counterparts. The composite index  $I_c$  was found to be more suitable than the joint behaviours of  $Q_t$  and  $F_r$  for predicting soil layer boundaries using BCPD. This is likely because the formulation of the  $I_c$  index was derived through calibration against external soil classification databases and, as such, incorporates valuable prior knowledge for soil layering identification. In contrast, the multivariate BCPD methods that directly use the  $Q_t$  and  $F_r$  data do not benefit from this prior knowledge, which may explain their comparative disadvantage. Among the BCPD methods, BCPD-ON was found to be the most accurate and computationally efficient, making it suitable for real-time predictions of soil layering during live testing.

Compared to the Robertson (2009) method, the proposed BCPD methods are more robust to noisy data and less prone to thin layer predictions since they do not rely on pre-established rules for distinguishing between different soil layers. These BCPD methods are fast and require minimal manual interpretation of the CPT data. Moreover, they do not require any training data, which eliminates the need to manually interpret the CPT data before applying the BCPD methods. This aligns with the trend towards more data-driven and automated approaches for geotechnical analyses [70–73]. However, for some geotechnical applications, it may be critical to identify all possible soil layers. In such cases, using the supervised BCPD methods proposed in Ref. [43] is appropriate as it provides the probability of soil layer boundaries at every depth. This allows the user to define a lower changepoint probability threshold to minimise the chance of missing true soil layer boundaries (potentially at the expense of more false soil layer boundary predictions).

While this study provides valuable insights into the performance of different unsupervised BCPD methods for soil layering identification using CPT data, there are some limitations that should be considered. One limitation is that the BCPD methods can only identify the boundaries of soil layers and not the specific soil type within each layer, although this can potentially be remedied by using the Robertson (2009) SBT rules to identify the dominant soil type within each layer. Moreover, the autocorrelation of the data in each

partition is not modelled for computational efficiency reasons. Modelling the autocorrelation of the data in each partition using a Gaussian process may enhance the accuracy of the predictions, although the increase in computational complexity would be a trade-off. Finally, the stochastic properties of heterogeneous geomaterials [74,75] mean that their strength and other properties can vary significantly within a material layer. Thus, the effectiveness of the BCPD method in distinguishing between such geomaterials is uncertain. These limitations highlight several areas for future research, including the development and evaluation of new BCPD methods that address these specific challenges.

## 5. Conclusions

This paper presents an assessment of different unsupervised BCPD methods for distinguishing between fine- and coarse-grained soil layers using CPT data. It introduces a novel unsupervised online BCPD method and benchmarks it against an existing unsupervised offline BCPD method, using both univariate and multivariate CPT data.

The key findings are as follows:

1. Univariate BCPD methods (using  $I_c$  data) are generally more accurate and computationally efficient than their multivariate counterparts (using  $Q_t$  and  $F_r$  data) in identifying soil layer boundaries using CPT data.
2. The newly developed univariate online BCPD method demonstrates the highest accuracy and computational efficiency.
3. This research underscores the advantage of unsupervised BCPD methods, which forego the need for training data and manual analysis, contributing to the advancement of fast, automated Bayesian geotechnical analysis techniques.

**Author Contributions:** Conceptualization, S.K.S. and B.B.S.; methodology, S.K.S. and B.B.S.; software, S.K.S.; validation, S.K.S., B.B.S., and M.L.; formal analysis, S.K.S.; resources, S.K.S., B.B.S., and M.L.; data curation, S.K.S., B.B.S., and M.L.; writing—original draft preparation, S.K.S.; writing—review and editing, B.B.S. and M.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** Some data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

**Acknowledgments:** The second author is funded by the Royal Academy of Engineering under the Research Fellowship scheme.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Suryasentana, S.K.; Mayne, P.W. Simplified method for the lateral, rotational, and torsional static stiffness of circular footings on a nonhomogeneous elastic half-space based on a work-equivalent framework. *J. Geotech. Geoenvironmental Eng.* **2022**, *148*, 04021182.
2. Mayne, P. W., & Poulos, H. G. Approximate displacement influence factors for elastic shallow foundations. *Journal of geotechnical and geoenvironmental engineering*, **1999**, *125*, 453–460.
3. Gupta, B. K., & Basu, D. Offshore wind turbine monopile foundations: Design perspectives. *Ocean Engineering*. **2020**, *213*, 107514.
4. Doherty, P., & Gavin, K. Laterally loaded monopile design for offshore wind farms. *Proceedings of the institution of civil engineers*, **2012**, *165*, 7–17.
5. Byrne, B.W.; Houlsby, G.T.; Burd, H.J.; Gavin, K.G.; Igoe, D.J.; Jardine, R.J.; Martin, C.M.; McAdam, R.A.; Potts, D.M.; Taborda, D.M.; et al. PISA design model for monopiles for offshore wind turbines: Application to a stiff glacial clay till. *Géotechnique* **2020**, *70*, 1030–1047.
6. Byrne, B.W.; Aghakouchak, A.; Buckley, R.M.; Burd, H.J.; Gengenbach, J.; Houlsby, G.T.; McAdam, R.A.; Martin, C.M.; Schranz, F.; Sheil, B.B.; et al. PICASO: Cyclic lateral loading of offshore wind turbine monopiles. In Proceedings of the 4th International Symposium on Frontiers in Offshore Geotechnics (ISFOG 2021), Austin, TX, USA, 8–11 November 2021.
7. Houlsby, G. T., Ibsen, L. B., & Byrne, B. W. Suction caissons for wind turbines. *Frontiers in Offshore Geotechnics: ISFOG*, Perth, WA, Australia, **2005**, 75–93.

8. Byrne, B., Houlsby, G., Martin, C., & Fish, P. Suction caisson foundations for offshore wind turbines. *Wind Engineering*. **2002**, *26*, 145–155.
9. Suryasentana, S.K.; Byrne, B.W.; Burd, H.J.; Shonberg, A. Simplified model for the stiffness of suction caisson foundations under 6DoF loading. In Proceedings of the SUT OSIG 8th International Conference 2017, London, UK, 12–14 September 2017, 554–561.
10. Suryasentana, S.K.; Byrne, B.W.; Burd, H.J.; Shonberg, A. An elastoplastic 1D Winkler model for suction caisson foundations under combined loading. In *Numerical Methods in Geotechnical Engineering IX*; CRC Press: Boca Raton, FL, USA, 2018, 973–980.
11. Suryasentana, S.K.; Burd, H.J.; Byrne, B.W.; Shonberg, A. Modulus weighting method for stiffness estimations of suction caissons in layered soils. *Géotechnique Lett.* **2023**, *13*, 97–104.
12. Lunne, T.; Robertson, P.K.; Powell, J.J.M. *Cone Penetration Testing in Geotechnical Practice*; Blackie Academic and Professional: London, UK, 1997.
13. Jamiolkowski, M.; Ghionna, V.N.; Lancellotta, R.; Pasqualini, E. New correlations of penetration tests for design practice: Proc 1st International Symposium on Penetration Testing, ISOPT-1, Orlando, 20–24 March 1988V1, P263–296. Rotterdam: A A Balkema, 1988. *Int. J. Rock Mech. Min. Sci. Geomech. Abstr.* **1990**, *27*, A91. [https://doi.org/10.1016/0148-9062\(90\)95078-F](https://doi.org/10.1016/0148-9062(90)95078-F).
14. Sakleshpur, V.A.; Prezzi, M.; Salgado, R.; Zaheer, M. CPT-Based Geotechnical Design Manual, Volume 2: CPT-Based Design of Foundations—Methods (Joint Transportation Research Program Publication No. FHWA/IN/JTRP2021/23); Purdue University: West Lafayette, IN, USA, 2021. <https://doi.org/10.5703/1288284317347>.
15. Suryasentana, S.K.; Lehane, B.M. Verification of numerically derived CPT based py curves for piles in sand. In Proceedings of 3rd International Symposium on Cone Penetration Testing, Las Vegas, NV, USA, 13–14 May 2014, 3–29.
16. Robertson, P. K. Soil classification using the cone penetration test. *Canadian geotechnical journal*. **1990**, *27*, 151–158.
17. Robertson, P. K., & Wride, C. E. Evaluating cyclic liquefaction potential using the cone penetration test. *Canadian geotechnical journal*. **1998**, *35*, 442–459.
18. Robertson, P.K. Interpretation of cone penetration tests—A unified approach. *Can. Geotech. J.* **2009**, *46*, 1337–1355.
19. Robertson, P. K. (2010). Soil behaviour type from the CPT: an update. In 2nd International symposium on cone penetration testing (Vol. 2, No. 56, p. 8). Huntington Beach: Cone Penetration Testing Organizing Committee.
20. Jefferies, M.G.; Davies, M.P. Use of CPTU to estimate equivalent SPT N60. *Geotech. Test. J.* **1993**, *16*, 458–468.
21. Schneider, J.A.; Randolph, M.F.; Mayne, P.W.; Ramsey, N.R. Analysis of factors influencing soil classification using normalized piezocone tip resistance pore pressure parameters. *J. Geotech. Geoenviron. Engng.* **2008**, *134*, 1569–1586.
22. Li, J.; Cassidy, M.J.; Huang, J.; Zhang, L.; Kelly, R. Probabilistic identification of soil stratification. *Géotechnique* **2016**, *66*, 16–26.
23. Fattahi, H.; Zandy Ilghani, N. Slope stability analysis using Bayesian Markov chain Monte Carlo method. *Geotech. Geol. Eng.* **2020**, *38*, 2609–2618.
24. Contreras, L.F.; Brown, E.T. Slope reliability and back analysis of failure with geotechnical parameters estimated using Bayesian inference. *J. Rock Mech. Geotech. Eng.* **2019**, *11*, 628–643.
25. Huang, M.L.; Sun, D.A.; Wang, C.H.; Keleta, Y. Reliability analysis of unsaturated soil slope stability using spatial random field-based Bayesian method. *Landslides* **2021**, *18*, 1177–1189.
26. Chivatá Cárdenas, I. On the use of Bayesian networks as a meta-modelling approach to analyse uncertainties in slope stability analysis. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*. **2019**, *13*, 53–65.
27. Jiang, S. H., Papaioannou, I., & Straub, D. Bayesian updating of slope reliability in spatially variable soils with in-situ measurements. *Engineering Geology*. **2018**, *239*, 310–320.
28. Feng, X.; Jimenez, R. Predicting tunnel squeezing with incomplete data using Bayesian networks. *Eng. Geol.* **2015**, *195*, 214–224.
29. Feng, X., Jimenez, R., Zeng, P., & Senent, S. Prediction of time-dependent tunnel convergences using a Bayesian updating approach. *Tunnelling and Underground Space Technology*. **2019**, *94*, 103118.
30. Janda, T.; Šejnoha, M.; Šejnoha, J. Applying Bayesian approach to predict deformations during tunnel construction. *Int. J. Numer. Anal. Methods Geomech.* **2018**, *42*, 1765–1784.
31. Zheng, H., Mooney, M., & Gutierrez, M. Updating model parameters and predictions in SEM tunnelling using a surrogate-based Bayesian approach. *Géotechnique*. **2023**, 1–13.
32. Fang, G., Nilot, E. A., Li, Y. E., Tan, Y. Z., & Cheng, A. Quantifying tunneling risks ahead of TBM using Bayesian inference on continuous seismic data. *Tunnelling and Underground Space Technology*. **2024**, *147*, 105702.
33. Jin, Y.; Biscontin, G.; Gardoni, P. Adaptive prediction of wall movement during excavation using Bayesian inference. *Comput. Geotech.* **2021**, *137*, 104249.
34. Sun, Y.; Huang, J.; Jin, W.; Sloan, S.W.; Jiang, Q. Bayesian updating for progressive excavation of high rock slopes using multi-type monitoring data. *Eng. Geol.* **2019**, *252*, 1–13.
35. He, L., Liu, Y., Bi, S., Wang, L., Broggi, M., & Beer, M. Estimation of failure probability in braced excavation using Bayesian networks with integrated model updating. *Underground Space*. **2020**, *5*, 315–323.
36. Hsein Juang, C., Luo, Z., Atamturktur, S., & Huang, H. Bayesian updating of soil parameters for braced excavations using field observations. *Journal of Geotechnical and Geoenvironmental Engineering*. **2013**, *139*, 395–406.
37. Lo, M. K., & Leung, Y. F. Bayesian updating of subsurface spatial variability for improved prediction of braced excavation response. *Canadian Geotechnical Journal*. **2019**, *56*, 1169–1183.
38. Sheil, B.B.; Suryasentana, S.K.; Templeman, J.O.; Phillips, B.M.; Cheng, W.C.; Zhang, L. Prediction of pipe-jacking forces using a Bayesian updating approach. *J. Geotech. Geoenviron. Eng.* **2022**, *148*, 04021173.

39. Buckley, R.; Chen, Y.M.; Sheil, B.; Suryasentana, S.; Xu, D.; Doherty, J.; Randolph, M. Bayesian optimization for CPT-based prediction of impact pile drivability. *J. Geotech. Geoenviron. Eng.* **2023**, *149*, 04023100.
40. Cao, Z.; Wang, Y. Bayesian approach for probabilistic site characterization using cone penetration tests. *J. Geotech. Geoenviron. Engng ASCE* **2013**, *139*, 267–276.
41. Wang, Y.; Huang, K.; Cao, Z. Probabilistic identification of underground soil stratification using cone penetration tests. *Can. Geotech. J.* **2013**, *50*, 766–776.
42. Cao, Z.J.; Zheng, S.; Li, D.Q.; Phoon, K.K. Bayesian identification of soil stratigraphy based on soil behaviour type index. *Can. Geotech. J.* **2019**, *56*, 570–586.
43. Suryasentana, S.K.; Lawler, M.; Sheil, B.B.; Lehane, B.M. Probabilistic soil strata delineation using DPT data and Bayesian changepoint detection. *J. Geotech. Geoenviron. Eng.* **2023**, *149*, 06023001.
44. Houlsby, N.M.T.; Houlsby, G.T. Statistical fitting of undrained strength data. *Géotechnique* **2013**, *63*, 1253–1263.
45. ASTM D2487-17; Standard practice for classification of soils for engineering purposes (Unified Soil Classification System). ASTM International: West Conshohocken, PA, USA, 2011. Available online: [www.astm.org](http://www.astm.org) (accessed on 15 March 2024). <https://doi.org/10.1520/D2487-11>.
46. Reeves, J.; Chen, J.; Wang, X.L.; Lund, R.; Lu, Q.Q. A review and comparison of changepoint detection techniques for climate data. *J. Appl. Meteorol. Climatol.* **2007**, *46*, 900–915.
47. Beaulieu, C., Chen, J., & Sarmiento, J. L. Change-point analysis as a tool to detect abrupt climate variations. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences.* **2012**, *370*, 1228–1249.
48. Gallagher, C., Lund, R., & Robbins, M. Changepoint detection in climate time series with long-term trends. *Journal of Climate.* **2013**, *26*, 4994–5006.
49. Bates, B. C., Chandler, R. E., & Bowman, A. W. Trend estimation and change point detection in individual climatic series using flexible regression methods. *Journal of Geophysical Research: Atmospheres.* **2012**, *117*(D16).
50. Lund, R. B., Beaulieu, C., Killick, R., Lu, Q., & Shi, X. Good practices and common pitfalls in climate time series changepoint techniques: A review. *Journal of Climate.* **2023**, *36*, 8041–8057.
51. Polunchenko, A. S., & Tartakovsky, A. G. State-of-the-art in sequential change-point detection. *Methodology and computing in applied probability.* **2012**, *14*, 649–684.
52. Niu, Y. S., Hao, N., & Zhang, H. Multiple change-point detection: a selective overview. *Statistical Science.* **2016**, 611–623.
53. Van den Burg, G. J., & Williams, C. K. An evaluation of change point detection algorithms. arXiv preprint arXiv:2003.06222.
54. Aminikhanghahi, S.; Cook, D.J. A survey of methods for time series change point detection. *Knowl. Inf. Syst.* **2017**, *51*, 339–367.
55. Truong, C.; Oudre, L.; Vayatis, N. Selective review of offline change point detection methods. *Signal Process.* **2020**, *167*, 107299.
56. Barry, D.; Hartigan, J.A. A Bayesian analysis for change point problems. *J. Am. Stat. Assoc.* **1993**, *88*, 309–319.
57. Stephens, D.A. Bayesian retrospective multiple-changepoint identification. *J. R. Stat. Soc. Ser. C Appl. Stat.* **1994**, *43*, 159–178.
58. Chib, S. Estimation and comparison of multiple change-point models. *J. Econom.* **1998**, *86*, 221–241.
59. Fearnhead, P. Exact Bayesian curve fitting and signal segmentation. *IEEE Trans. Signal Process.* **2005**, *53*, 2160–2166.
60. Fearnhead, P. Exact and efficient Bayesian inference for multiple changepoint problems. *Stat. Comput.* **2006**, *16*, 203–213.
61. Adams, R.P.; MacKay, D.J. Bayesian online changepoint detection. arXiv **2007**, arXiv:0710.3742.
62. Fearnhead, P., & Liu, Z. On-line inference for multiple changepoint problems. *Journal of the Royal Statistical Society Series B: Statistical Methodology.* **2007**, *69*, 589–605.
63. Punsakaya, E.; Andrieu, C.; Doucet, A.; Fitzgerald, W.J. Bayesian curve fitting using MCMC with applications to signal segmentation. *IEEE Trans. Signal Process.* **2002**, *50*, 747–758.
64. Xuan, X.; Murphy, K. Modeling changing dependency structure in multivariate time series. In Proceedings of the 24th International Conference on Machine Learning, Corvallis, OR, USA, 20–24 June 2007; 1055–1062.
65. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2006.
66. Murphy, K. P. *Machine learning: a probabilistic perspective.* MIT press. 2012.
67. Bolstad, W. M., Curran, J. M. Introduction to Bayesian statistics. *John Wiley & Sons.* 2016.
68. Ellison, A. M. An introduction to Bayesian inference for ecological research and environmental decision-making. *Ecological applications.* **1996**, *6*, 1036–1046.
69. Boulanger, R.W.; DeJong, J.T. Inverse filtering procedure to correct cone penetration data for thin-layer and transition effects. In Cone Penetration Testing 2018: Proceedings of the 4th International Symposium on Cone Penetration Testing (CPT'18), Delft, The Netherlands, 21–22 June 2018; CRC Press: Boca Raton, FL, USA, 2018; 25.
70. Doherty, J. P., & Lehane, B. M. An automated approach for optimizing monopile foundations for offshore wind turbines for serviceability and ultimate limit states design. *Journal of Offshore Mechanics and Arctic Engineering.* **2018**, *140*, 051901.
71. Stuyts, B.; Suryasentana, S. Applications of data science in offshore geotechnical engineering: State of practice and future perspectives. In Proceedings of the 9th SUT Offshore Site Investigation and Geotechnics Conference 2023, London, UK, 12–14 September 2023.
72. Kadlíček, T., Janda, T., Šejnoha, M., Mašín, D., Najser, J., & Beneš, Š. Automated calibration of advanced soil constitutive models. Part I: hypoplastic sand. *Acta Geotechnica.* **2022**, 1–18.
73. Suryasentana, S.K.; Burd, H.J.; Byrne, B.W.; Shonberg, A. Automated procedure to derive convex failure envelope formulations for circular surface foundations under six degrees of freedom loading. *Comput. Geotech.* **2021**, *137*, 104174.



- 
74. Li, K.Q.; Chen, Q.M.; Chen, G. Scale dependency of anisotropic thermal conductivity of heterogeneous geomaterials. *Bull. Eng. Geol. Environ.* **2024**, *83*, 73.
  75. Liu, Y.; Lee, F.H.; Quek, S.T.; Chen, E.J.; Yi, J.T. Effect of spatial variation of strength and modulus on the lateral compression response of cement-admixed clay slab. *Géotechnique* **2015**, *65*, 851–865.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.