Probabilistic soil strata delineation using DPT data and Bayesian changepoint detection

Stephen K. Suryasentana[1], Ph.D.
Myles Lawler[2], Ph.D.
Brian B. Sheil[3], Ph.D.
Barry M. Lehane[4], Ph.D.

Affiliations

[1] Lecturer, Department of Civil and Environmental Engineering, University of Strathclyde, 75 Montrose St, Glasgow G1 1XJ, UK.

[2] Independent Geotechnical Consultant, Ireland.

[3] RAEng Research Fellow, Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, UK.

[4] Winthrop Professor, Department of Civil, Environmental and Mining Engineering, University of Western Australia, 35 Stirling Hwy, Crawley WA 6009, Australia.

Full contact details of corresponding author

Stephen K. Suryasentana

stephen.suryasentana@strath.ac.uk

Main text word count: 2517

Figures: 5

Tables: 1

Mar 26, 2022

This is a peer-reviewed, accepted author manuscript of the following article: Suryasentana, S. K., Lawler, M., Sheil, B. B., & Lehane, B. M. (2023). Probabilistic soil strata delineation using DPT data and Bayesian changepoint detection. Journal of Geotechnical and Geoenvironmental Engineering, 149(4). https://doi.org/10.1061/jggefk.gteng-10843

Abstract

Soil strata delineation is a fundamental step for any geotechnical engineering design. The dynamic penetration test (DPT) is a fast, low cost in-situ test that is commonly used to locate boundaries between strata of differing density and driving resistance. However, DPT data are often noisy and typically require time-consuming, manual interpretation. This paper investigates a probabilistic method that enables delineation of dissimilar soil strata (where each stratum is deemed to belong to different soil groups based on their particle size distribution) by processing DPT data with Bayesian changepoint detection methods. The accuracy of the proposed method is evaluated using DPT data from a real-world case study, which highlights the potential of the proposed method. This study provides a methodology for faster DPT-based soil strata delineation, which paves the way for more cost-effective geotechnical designs.

Keywords

bayesian analysis, site investigation, soil classification

List of notation

N	DPT no. of blows
r_z	'Run length' random variable
$x_{1:n}$	Set of data $\{x_1, x_2, \ldots, x_{n-1}, x_n\}$
α, β	Parameters of inverse gamma distribution
p_{cp}	Changepoint probability threshold

Introduction

Soil strata delineation is a fundamental step for any geotechnical engineering design. Delineation divides the soil volume into separate layers of geological material deemed to belong to the same group. This process typically requires a time-consuming, manual interpretation of a combination of borehole data and associated in-situ and laboratory test results (Parry et al. 2014). It is highly desirable to develop a rapid approach that can delineate the soil strata automatically.

The cone penetration test (CPT) (Lunne et al. 1997) is an in-situ ground investigation method that is widely used for soil delineation by applying soil behaviour type (SBT) classification rules (e.g. Robertson 1990; Jefferies and Davies 1993; Schneider et al. 2008) to the measured CPT data. Other delineation approaches include fuzzy analysis (Zhang and Tumay 1999), clustering analysis (Hegazy and Mayne 2002; Depina et al. 2016), signal processing analysis (Ching et al. 2015) and statistical/Bayesian analysis (Wickremesinghe and Campanella 1991; Phoon et al. 2003; Wang et al. 2013, 2019, 2020; Li et al. 2016; Cao et al. 2019). Bayesian analysis has the advantages of being robust to noisy data and allowing quantification of uncertainty, although it tends to be computationally intensive.

The dynamic probing/penetration test (DPT) is a fast and low cost in-situ ground investigation method (BS 2005), which bears some similarities to both CPT and the standard penetration test (SPT). Like CPT, DPT uses a cylindrical steel cone penetrometer. However, DPT drives the cone into the ground using a hammer, and the measured result is the number of blows N for a given penetration (e.g. 100mm). The primary advantage of DPT over CPT is lower costs, faster speed of operation and applicability in terrains with poor accessibility. However, there are limited methods to interpret DPT results for soil strata delineation.

This paper aims to develop a method that enables fast soil strata delineation using DPT data. The proposed method uses Bayesian changepoint detection (BCPD) methods to detect abrupt changes in the soil data trends indicative of transitions between different soil strata. Unlike most Bayesian approaches, the proposed method is computationally efficient. Two BCPD methods are explored: (i) ‘online’, where each data point is processed as it becomes available and inferences are made without knowledge of future measurements (e.g. Fearnhead and Liu, 2007;

Adams and MacKay, 2007); and (ii) 'offline', where the entire DPT dataset is required before making inference (e.g. Barry and Hartigan, 1993; Stephens, 1994; Fearnhead, 2005, 2006).

The proposed method divides the soil profile up into three dissimilar soil categories: (i) predominantly fine-grained soils (e.g. clay, silt), (ii) predominantly sand, and (iii) predominantly gravel. These soil categories have very different permeability, stiffness and strength properties such that poor identification will have a negative impact on optimal geotechnical design. The proposed method bears some similarities to that of Zhang and Tumay (1999), who applied fuzzy analysis to CPT data to identify three soil categories, although the methodology and nature of the data are different. The performance of the proposed BCPD methods are evaluated using a real-world case study.

Methodology

Changepoints are abrupt changes in data, which typically represent transitions between states, as shown in Fig. 1. Given a sequence of data, these changepoints split the data into a set of non-overlapping partitions, where it is assumed that the data within a partition are generated by the same model. While many changepoint detection methods are available (Reeves et al. 2007; Aminikhanghahi and Cook 2017; Truong et al. 2020), this paper focuses on Bayesian changepoint detection (BCPD) methods.

Online Bayesian changepoint detection

The first method investigated in this paper is an online BCPD method (Adams and Mackay 2007), denoted 'BCPD-ON'. In the following exposition, the notation $x_{1:n}$ refers to the set of data $\{x_1, x_2, \ldots, x_{n-1}, x_n\}$. BCPD-ON estimates the probability of a changepoint at a given depth based only on data processed up to that depth. It does so by computing the probability distribution of a random variable called the 'run length' r_z, which represents the length of the current data partition. Each new data point either (a) comes from the same distribution, in which case the parameter estimates of the current distribution is updated using Bayes' theorem and r_z increases by one, or (b) it belongs to a new distribution which means a changepoint occurs and the new distribution will reset back to the prior distribution and r_z resets to zero. When the most probable value of r_z is zero, it is likely that there is a changepoint at depth z, the probability of which is equivalent to the posterior probability of $r_z = 0$:

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

The posterior distribution of the run length i.e. $p(r_z | x_{1:z})$ can be calculated as:

$$p(r_z | x_{1:z}) = \frac{p(r_z, x_{1:z})}{p(x_{1:z})} \tag{2}$$

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:

$$\begin{aligned} p(r_z, x_{1:z}) &= \sum_{r_{z-1}} p(r_z, x_z, | r_{z-1}, x_{1:z-1}) p(r_{z-1}, x_{1:z-1}) \\ &= \sum_{r_{z-1}} p(r_z | r_{z-1})\, p(x_z | r_{z-1}, \boldsymbol{x}_z^r)\, p(r_{z-1}, x_{1:z-1}) \end{aligned} \tag{3}$$

where $\boldsymbol{x}_z^r$ is the set of data associated with the run length r_z. $p(r_{z-1}, x_{1:z-1})$ is a recursive term, which represents the previous iteration of Eq. 3 at depth $z - 1$. $p(r_z | r_{z-1})$ is the conditional distribution of the run length. Finally, $p(x_z | r_{z-1}, \boldsymbol{x}_z^r)$ is the posterior predictive distribution and it can be calculated analytically by assuming that the data point x_z comes from some probability distribution (e.g. Gaussian) and by adopting conjugate priors. More details about these calculations can be found in Adams and Mackay (2007).

Offline Bayesian changepoint detection

The second method investigated in this paper is an offline BCPD method (Fearnhead 2005, 2006) denoted ‘BCPD-OFF’, which was previously employed by Houlsby and Houlsby (2013) for clay layer delineation using undrained shear strength data. BCPD-OFF is based on a recursive algorithm that computes the posterior probability distribution exactly over the location of changepoints. This is significantly more efficient than previous Markov Chain Monte Carlo (MCMC) approaches for computing the posterior (e.g. Punskaya et al. 2002).

In this case, the data within each partition are modelled by some probability distribution, with distribution parameters independent of those determined for other partitions. Let c_j represent the jth changepoint. The posterior distribution of c_j is $p\left(c_j \mid x_{1:n}\right)$. The probability of a changepoint occurring at depth z can be calculated as:

$$p(\text{changepoint at } z \mid x_{1:n}) = \sum_{j=1}^{z} p\left(c_j = z \mid x_{1:n}\right) \tag{4}$$

where all possible scenarios of 1 to z changepoints thus far are considered. This approach differs from that of Houlsby and Houlsby (2013), which first identifies the *maximum a posteriori* (MAP) number of changepoints and then the conditional MAP locations of the changepoints. This modification makes the outputs of BCPD-OFF and BCPD-ON identical, thereby allowing direct comparisons.

$p\left(c_j \mid x_{1:n}\right)$ in Eq. 4 is obtained by marginalising out the previous changepoints:

$$p\left(c_j \mid x_{1:n}\right) = \int p\left(c_j, \dots, c_1 \mid x_{1:n}\right) dc_{j-1} \dots dc_1 \tag{5}$$

As the probability of a changepoint is assumed to be dependent only on the previous changepoint, the integrand in Eq. 5 can be calculated as:

$$p\left(c_j, \dots, c_1 \mid x_{1:n}\right) = p\left(c_j \mid c_{j-1}, x_{1:n}\right) p\left(c_{j-1} \mid c_{j-2}, x_{1:n}\right) \dots p\left(c_2 \mid c_1, x_{1:n}\right) p\left(c_1 \mid x_{1:n}\right) \tag{6}$$

Each of the terms on the right hand side of Eq. 6 can be calculated exactly and efficiently using the recursive algorithm described in Fearnhead (2005, 2006).

Case study

The proposed BCPD methods are evaluated using a case-study involving multi-layered alluvial deposits, consisting of sands, silts, clays, and gravels. This case study is based on the Deutsche Bahn AG (German Rail) ‘DB46/2’ project, which is an expansion line from Emmerich to Oberhausen in Germany. A complex three-dimensional (3D) ground model for this project has been documented in Prinz (2019). This paper considers 26 DPT tests from the case study: 20 (approximately 77% of the dataset) are randomly selected for calibration of the priors and hyperparameters for BCPD-OFF and BCPD-ON; the remaining 6 DPT locations (labelled ‘T1’ to ‘T6’) are used for testing to evaluate the performance of the calibrated methods. A plan map of the DPT calibration and test locations is shown in Fig. 2.

Expert predictions are also made for each DPT location, where the soil strata are identified among the three soil categories defined in the introduction. These expert predictions were extracted from the 3D ground model that was developed separately for the case study (Prinz 2019). This ground model was based on careful, manual interpretation of both the DPT data and the borehole data in an integrated manner, ensuring no conflicts between the interpretation of the soil layering boundaries based on both types of data (e.g. the soil stratification interpreted from the DPT data should be consistent with that observed from a neighbouring borehole). Fig. 3 shows a typical DPT profile from one of the DPT locations and its corresponding expert prediction of the soil strata. The proposed BCPD methods will be applied to DPT data only.

Calibration

For both BCPD-OFF and BCPD-ON, the data in each partition are assumed to be normally distributed with unknown mean μ and variance σ^2. Therefore, the DPT data were preprocessed using a Freeman-Tukey transformation (Freeman and Tukey 1950): $N_{\mathrm{transformed}} = \sqrt{N} + \sqrt{N+1}$ where N represents the raw DPT blowcount data. This transformation is typically used to make discrete count data better approximate a normal distribution (Mosteller and Youtz 2006; Lin and Xu 2020). To test for normality of the transformed data, the Shapiro-Wilk test (Shapiro and Wilk 1965) was applied to the transformed data in each soil layer at DPT locations where neighbouring borehole data is available to determine the approximate locations of the soil layer boundaries. The p-values obtained are greater than 0.05 and thus the null hypothesis that the transformed data is normally distributed is not rejected. Following Houlsby and Houlsby (2013), the variance σ^2 is assumed to follow an inverse gamma distribution and the distribution parameters α = 1.8 and β = 0.38 are obtained by curve-fitting the cumulative distribution of the variance for the DPT calibration dataset, as shown in Fig. 4.

Outputs of interest for both BCPD-ON and BCPD-OFF are the probabilities of a changepoint occurrence at each depth (i.e. using Eq. 1 and Eq. 4, respectively). When the changepoint probability exceeds a predefined threshold p_{cp}, the soil is considered to have changed category at this depth. The optimal value of p_{cp} is dependent on the method adopted (BCPD-ON or BCPD-OFF) and is calibrated as a hyperparameter. For each method, a grid search is implemented within the set of trial $p_{\mathrm{cp}} = \{0.1, 0.15, 0.2, 0.25, 0.3, \ldots, 0.8, 0.85, 0.9\}$ to identify the value of p_{cp} that achieve the best match with the expert predictions for the soil stratification at each DPT calibration location. To quantify the match with expert predictions, the accuracy measure, F1 score, is adopted,

$$\text{F1 score} = 2(\text{Precision} * \text{Sensitivity})/(\text{Precision} + \text{Sensitivity}) \tag{7}$$

where Precision = True Positive/(True Positive + False Positive) and Sensitivity = True Positive/(True Positive + False Negative). True Positive (TP) is the number of times an expert prediction for soil layer boundary has been correctly identified, while False Positive (FP) is the number of times an expert prediction for soil layer boundary has been incorrectly identified. False Negative (FN) is the number of times an expert prediction for soil layer boundary has not been identified. A higher F1 value indicates a better match with the expert predictions. As the predicted boundaries based on the DPT data are not expected to exactly match the expert predictions, this paper considers a soil layer boundary to be correctly identified if the DPT-predicted boundary is within a distance of 1m from the expert prediction for a boundary. The grid search exercise gives the optimal values of p_{cp} = 0.45 and 0.4 for BCPD-OFF and BCPD-ON respectively.

Results

Fig. 5 shows the soil strata predictions determined using BCPD-OFF and BCPD-ON for the 6 DPT test locations. The BCPD changepoint probability predictions are shown in the figure as grey lines and a soil strata boundary is identified when these predictions exceed p_{cp}.
From this figure, it can be observed that both BCPD-OFF and BCPD-ON perform well for most locations, where the predicted soil strata boundaries are similar to the expert predictions. The exception to this is Location T3, where the expert prediction for the soil strata is very complex, and both BCPD methods only detect some of the soil strata boundaries. Nevertheless, the overall performance is encouraging as the BCPD predictions agree well with the expert predictions, despite using information only from the local DPT data. Some of the soil strata

boundary detections are noteworthy (e.g. see Fig. 5d), as they are not obvious from manual inspection of the noisy DPT data alone.

Comparing the two BCPD methods, it is evident that BCPD-OFF is the more sensitive of the two, as it can detect more soil strata boundaries (e.g. at locations T3 and T4), despite having a higher p_{cp} than BCPD-ON. However, this increased sensitivity comes with the drawback of producing more false positives (see Figs. 5b, c). To quantify the accuracy of both methods, their F1 scores are calculated based on Eq. 7, as detailed in Table 1. BCPD-ON has a slightly higher F1 score than BCPD-OFF, indicating that BCPD-ON has a slightly better balance of precision and sensitivity. In terms of computational efficiency, BCPD-ON has the advantage of being much faster than BCPD-OFF (on average, BCPD-ON takes approximately 0.03 seconds to process each DPT location, while BCPD-OFF takes approximately 5 seconds).

A key highlight is that both BCPD-OFF and BCPD-ON could detect soil strata boundaries quickly and automatically without manual intervention. This makes them helpful to industry practitioners for extracting additional insights from the DPT data to complement their current workflow for identifying soil strata. A useful application of the approach could be, for example, to assist the design of large-scale foundation projects such as solar farms. Engineers could be faced with up to 1000 DPT locations in one project, and this approach provides a consistent, automated and rapid way to interpret the soil stratigraphy.

When applying these BCPD methods to a new site, a calibration process should be carried out to obtain site-specific values for both the priors and the p_{cp} hyperparameter; this should provide improved soil layer boundary detection results. Site-specific calibration should not be an issue as DPT tests are typically carried out in conjunction with borehole tests. However, if calibration data is not available at the new site, the calibrated parameters in this paper may be used for preliminary analysis, using the BCPD methods to highlight potential locations of soil layer boundaries through the 'spikes' in the changepoint probability. However, caution is advised as a non-site specific calibration of p_{cp} and the priors will affect the precision of the soil layer boundary detections. To investigate the sensitivity of the calibration to the number of DPT tests, the calibration results (i.e. the calibrated values for $\alpha, \beta, p_{\mathrm{cp}}$) were determined using random selections of 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 DPT tests. The analysis indicates that when 5 or more

DPT tests are used for calibration, the calibrated p_{cp} values are the same and the calibrated values of α, β change by less than 4% from the values used in the current study. However, caution should be advised against taking this as a general rule as these results may be specific to the dataset used in the current study. Furthermore, in this study, each DPT test location is near a calibration location. The effect of the distance between the calibration and test locations on the predictive accuracy of the BCPD methods has not been evaluated in this study. Further research is required with a comprehensive study, involving a larger database of DPT data from a wider range of sites, to provide more definitive answers to the above questions and to obtain values of the priors and hyperparameter more suited to general use across different sites.

Conclusion

This paper proposes a fast, automatic Bayesian approach for soil strata delineation using DPT data. The proposed approach is based on the concept of offline and online Bayesian changepoint detection, which allows both retrospective and real-time soil strata delineation. Its reliability and utility have been evaluated using DPT data from a real-world case study. The proposed approach is very fast to run and provides additional insights from the DPT data for a more robust soil strata identification solution.

Acknowledgements

The authors would like to acknowledge Deutsche Bahn AG and Aloys Kisse, Dr.-Ing. for the use of the test data for research purposes. Oriol Ciurana, OSI is gratefully acknowledged in the development of the 3D ground model referred to herein. The third author is funded by the Royal Academy of Engineering under the Research Fellowship scheme.

Competing interests statement

Competing interests: The authors declare there are no competing interests.

Funding statement

Funding: The authors declare no specific funding for this work.

Data availability statement

Some or all data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

References

Adams, R. P., and MacKay, D. J. (2007). Bayesian online changepoint detection. arXiv preprint arXiv:0710.3742.

Aminikhanghahi, S., and Cook, D. J. (2017). A survey of methods for time series change point detection. Knowledge and information systems, 51(2), 339-367.

Barry, D., and Hartigan, J. A. (1993). A Bayesian analysis for change point problems. Journal of the American Statistical Association, 88(421), 309-319.

BS (2005). Geotechnical Investigation and Testing–Field Testing–Part 2; Dynamic Probing. BS EN ISO 22476-2, BSI, London, UK.

Cao, Z. J., Zheng, S., Li, D. Q., and Phoon, K. K. (2019). Bayesian identification of soil stratigraphy based on soil behaviour type index. Canadian Geotechnical Journal, 56(4), 570-586.

Ching, J., Wang, J.-S., Juang, C.H., and Ku, C.-S. (2015). Cone penetration test (CPT)-based stratigraphic profiling using the wavelet transform modulus maxima method. Canadian geotechnical journal, 52(12): 1993-2007.

Depina, I., Le, T.M.H., Eiksund, G., and Strøm, P. (2016). Cone penetration data classification with Bayesian Mixture Analysis. Georisk: Assessment and management of risk for engineered systems and geohazards, 10(1): 27-41.

Fearnhead, P. (2005). Exact Bayesian curve fitting and signal segmentation. IEEE Transactions on Signal Processing, 53(6), 2160-2166.3, 747–758.

Fearnhead, P. (2006). Exact and efficient Bayesian inference for multiple changepoint problems. Statistics and computing, 16(2), 203-213.

Fearnhead, P., and Liu, Z. (2007). On-line inference for multiple changepoint problems. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 69(4), 589-605.

Freeman, M. F., and Tukey, J. W. (1950). Transformations related to the angular and the square root. The Annals of Mathematical Statistics, 607-611.

Hegazy, Y.A., and Mayne, P.W. (2002). Objective site characterization using clustering of piezocone data. Journal of Geotechnical and Geoenvironmental Engineering, 128(12): 986-996.

Houlsby, N. M. T., and Houlsby, G. T. (2013). Statistical fitting of undrained strength data. Géotechnique, 63(14), 1253-1263.

Jefferies, M. G. and Davies, M. P. (1993). Use of CPTU to estimate equivalent SPT N60. Geotech. Test. J. 16, No. 4, 458–468.

Li, J., Cassidy, M. J., Huang, J., Zhang, L., and Kelly, R. (2016). Probabilistic identification of soil stratification. Géotechnique, 66(1), 16-26.

Lin, L., and Xu, C. (2020). Arcsine-based transformations for meta-analysis of proportions: Pros, cons, and alternatives. Health Science Reports, 3(3), e178.

Lunne, T., Robertson, P. K. and Powell, J. J. M. (1997). Cone penetration testing in geotechnical practice. London, UK: Blackie Academic and Professional.

Mosteller, F., and Youtz, C. (2006). Tables of the Freeman-Tukey transformations for the binomial and Poisson distributions. In Selected Papers of Frederick Mosteller (pp. 337-347). Springer, New York, NY.

Parry, S., Baynes, F.J., Culshaw, M.G., Eggers, M., Keaton, J.F., Lentfer, K., Novotny, J. and Paul, D. (2014). Engineering geological models - an introduction: IAEG commission 25. Bull. Engng. Geol. Environ., 73, 689–706.

Phoon, K.-K., Quek, S.-T., and An, P. (2003). Identification of statistically homogeneous soil layers using modified Bartlett statistics. Journal of Geotechnical and Geoenvironmental Engineering, 129(7): 649-659.

Prinz, I., (2019). Digitale Baugrundmodelle: BIM in der Geotechnik Erfahrungen und Ableitungen aus dem Projekt Ausbaustrecke Emmerich – Oberhausen (ABS 46/2), einem BIM-Piloten der Deutschen Bahn. Geotechnik 22.

Punskaya, E., Andrieu, C., Doucet, A., and Fitzgerald, W. J. (2002). Bayesian curve fitting using MCMC with applications to signal segmentation. IEEE Transactions on signal processing, 50(3), 747-758.

Reeves, J., Chen, J., Wang, X. L., Lund, R., and Lu, Q. Q. (2007). A review and comparison of changepoint detection techniques for climate data. Journal of applied meteorology and climatology, 46(6), 900-915.

Robertson, P. K. (1990). Soil classification using the cone penetration test. Can. Geotech. J. 27, No. 1, 151–158.
Schneider, J. A., Randolph, M. F., Mayne, P. W. and Ramsey, N. R. (2008). Analysis of factors influencing soil classification using normalized piezocone tip resistance and pore pressure parameters. J. Geotech. Geoenvir. Engng 134, No. 11, 1569–1586.
Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). Biometrika, 52(3/4), 591-611.
Stephens, D. A. (1994). Bayesian retrospective multiple-changepoint identification. Journal of the Royal Statistical Society: Series C (Applied Statistics), 43(1), 159-178.
Truong, C., Oudre, L., and Vayatis, N. (2020). Selective review of offline change point detection methods. Signal Processing, 167, 107299.
Wang, Y., Huang, K., and Cao, Z. (2013). Probabilistic identification of underground soil stratification using cone penetration tests. Canadian Geotechnical Journal, 50(7), 766–776.
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. Canadian Geotechnical Journal, 57(7), 947-958.
Wang, H., Wang, X., Wellmann, J. F., and Liang, R. Y. (2019). A Bayesian unsupervised learning approach for identifying soil stratification using cone penetration data. Canadian Geotechnical Journal, 56(8), 1184-1205.
Wickremesinghe, D., and Campanella, R. (1991). Statistical methods for soil layer boundary location using the cone penetration test. Proc. ICASP6, Mexico City, 2: 636-643.
Zhang, Z., and Tumay, M.T. (1999). Statistical to fuzzy approach toward CPT soil classification. Journal of Geotechnical and Geoenvironmental Engineering, 125(3): 179-186.

Table 1 Accuracy calculations for the BCPD-OFF and BCPD-ON soil layer boundary predictions

	TP	FP	FN	Precision	Sensitivity	F1 score
BCPD-OFF	12	3	2	0.80	0.857	0.827
BCPD-ON	10	0	4	1.00	0.714	0.833

Figures

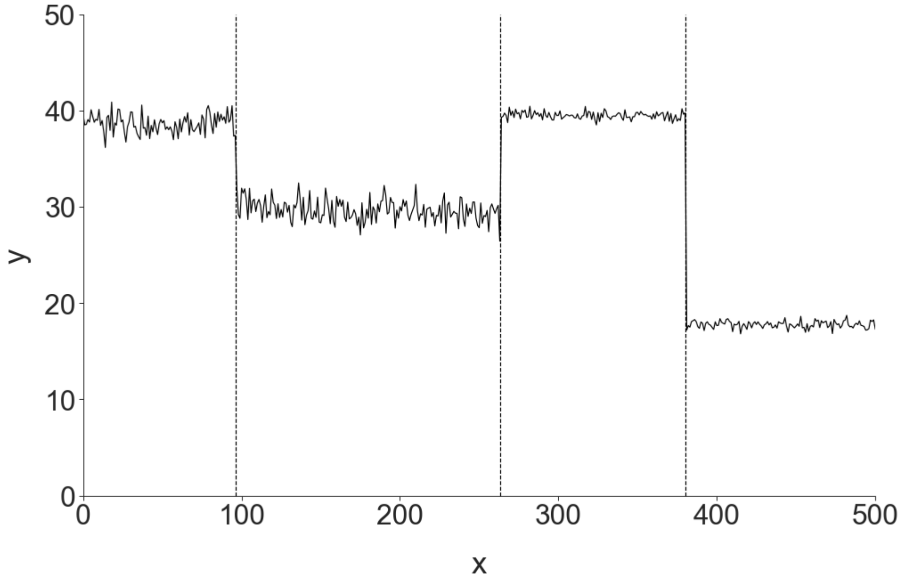

Fig. 1 Illustration of a sequence of data with abrupt changes, where y is the measured quantity and x is the index. The dashed lines represent the locations of the changepoints.

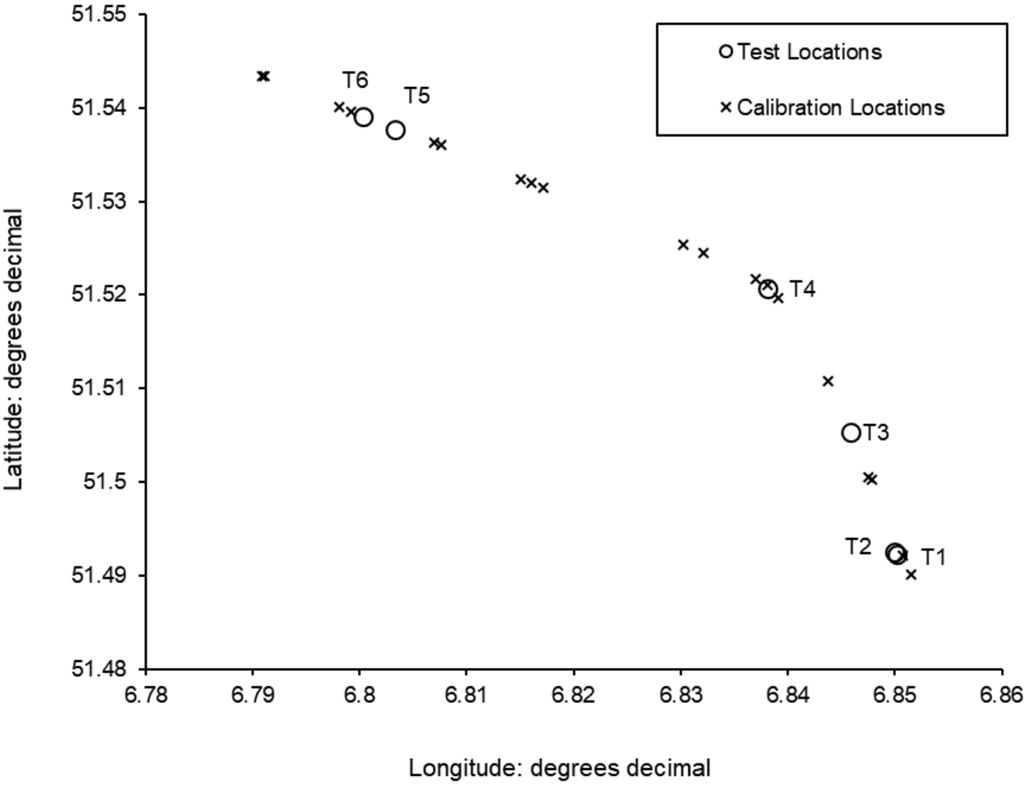

Fig. 2 Locations of DPT dataset used for calibration and testing of the BCPD methods.

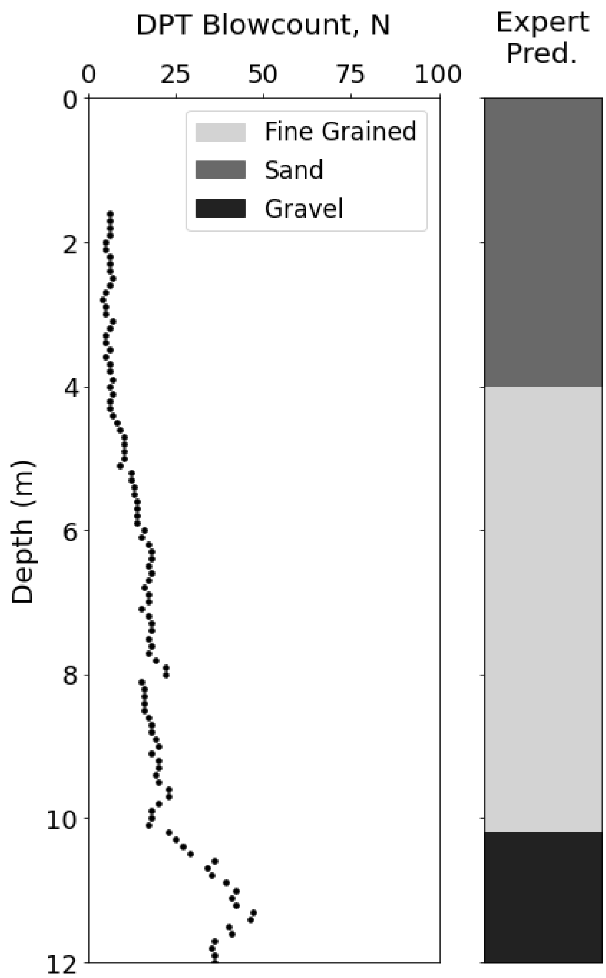

Fig. 3 Exemplar DPT profile showing the development of the DPT blowcount, N, with depth. The expert prediction for the soil strata at this location is also shown, where the soil categories are shown in the legend.

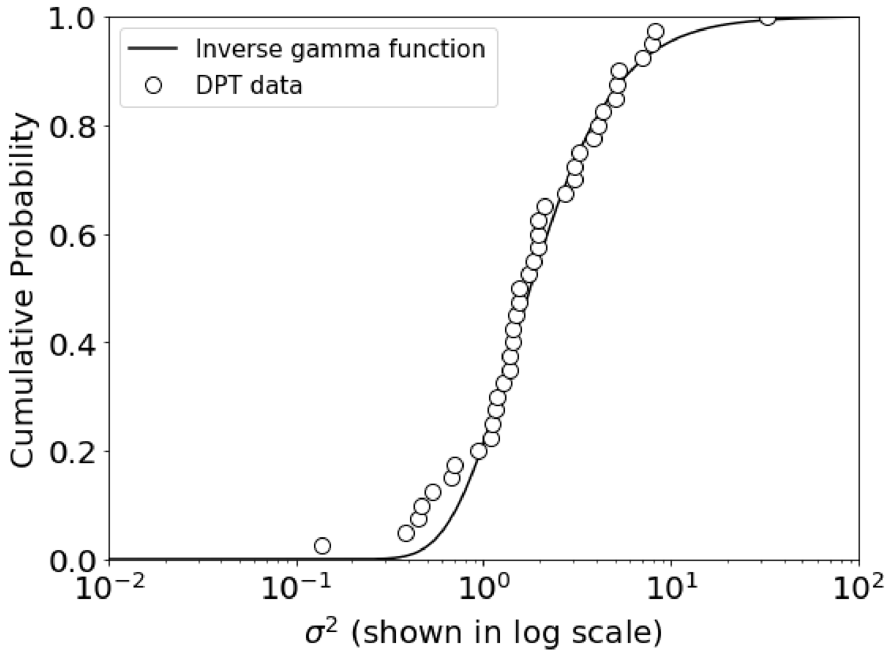

Fig. 4 Cumulative distribution of the variance of the transformed N data within each soil strata identified in the DPT calibration dataset, compared with the inverse gamma cumulative distribution with $\alpha = 1.8, \beta = 0.38$.

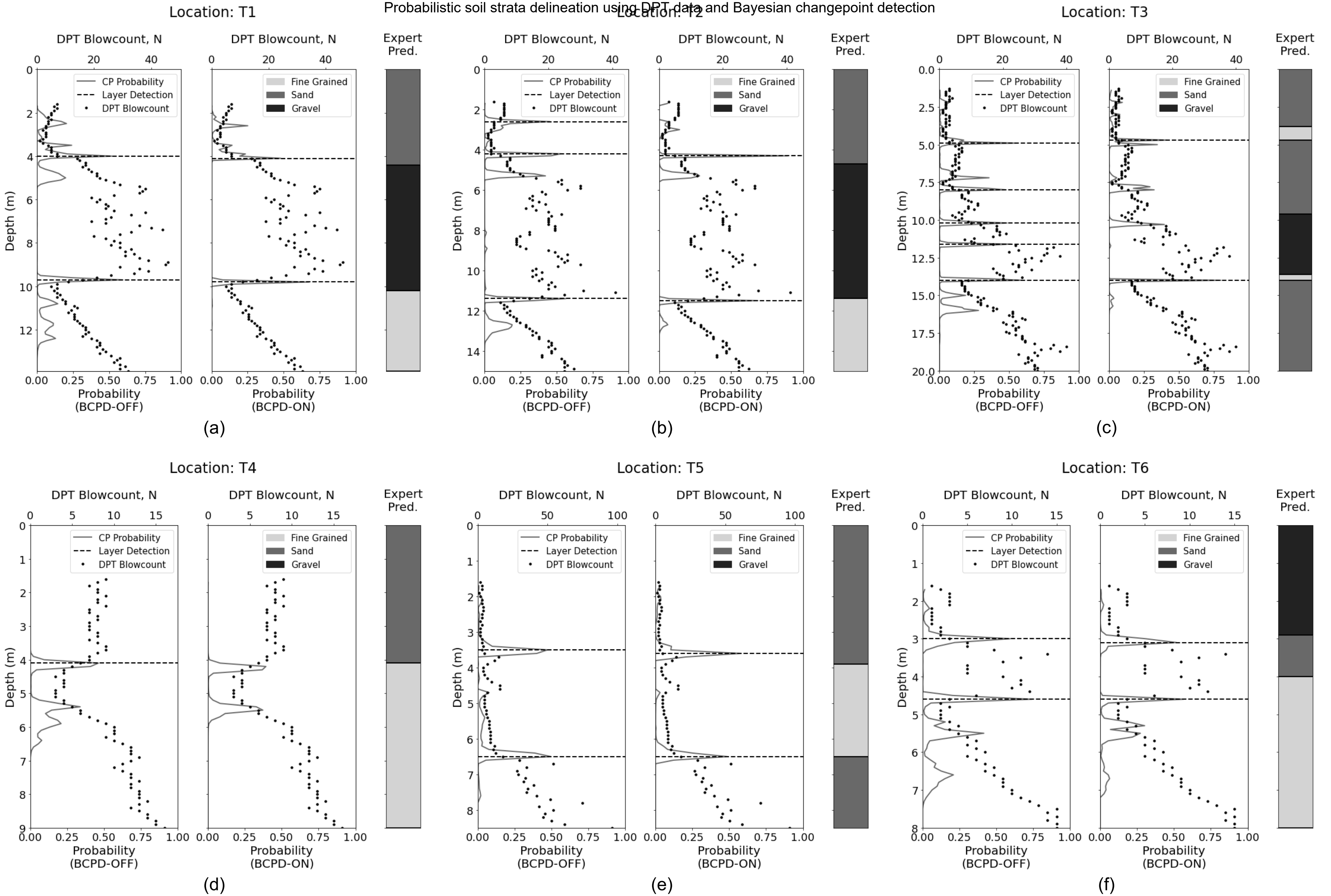

Fig. 5 Comparison of soil strata boundaries (shown as horizontal black lines) predicted by BCPD-OFF and BCPD-ON, with the expert predictions, at locations T1 to T6.