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

doi: 10.32075/17ECSMGE-2019-0271

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

Optimising geotechnical correlations using Receiver Operating Characteristic (ROC) analysis

Optimisation des corrélations géotechniques à l'aide de l'analyse des caractéristiques de fonctionnement du récepteur (ROC)

L. Jones
SSE Renewables, Glasgow, United Kingdom

D. Rushton
East Point Geo, Norwich, United Kingdom

ABSTRACT: This paper aims to demonstrate how the Receiver Operating Characteristic (ROC) analysis can be applied for geotechnical diagnosis and how ROC results can be communicated in a geospatial context. A specific example is provided in which the ROC analysis is used to quantify the confidence in a calibration of geophysical amplitude response to areas of anthropogenic hard ground.

The ROC analysis is a statistical approach used to assess the efficacy of diagnostic evaluation procedures. To assess how well a diagnostic model performs, the true positive rate (TPR) from the diagnostic model is plotted against the false positive rate (FPR) to produce an ROC curve. The area under the curve (AUC) is used to assess model performance, with increasing AUC indicative of better model performance. To maximise AUC, it is necessary to determine the optimum operating point (OOP) on the ROC curve, which defines the threshold value which minimises the rate of misprediction for the model whilst optimising the rate of correct prediction.

This approach allows the risk of inaccurate prediction to be quantified. For example, it may be desirable to increase the threshold to reduce the chance of false positive predictions; the ROC approach allows this reduction in false positive predictions to be quantified and it also quantifies the reduction in true positive predictions associated with the same threshold increase.

RÉSUMÉ: Cet article montre comment l'analyse des caractéristiques de fonctionnement de récepteur (ROC) peut être appliquée au diagnostic géotechnique et comment les résultats ROC peuvent être communiqués dans un contexte géospatial. L'article inclue un exemple spécifique dans lequel l'analyse ROC est utilisée pour quantifier la confiance dans l'étalonnage de la réponse d'amplitude géophysique dans des zones de sol dur anthropique.

L'analyse ROC est une approche statistique utilisée pour évaluer l'efficacité des procédures d'évaluation de diagnostique. Pour évaluer l'efficacité du modèle de diagnostic, le taux de vrais positifs (TPR) du modèle de diagnostic est calculé en fonction du taux de faux positifs (FPR) pour produire une courbe ROC. L'aire sous la courbe (AUC) est utilisée pour évaluer la performance du modèle, l'augmentation de l'ASC indiquant une meilleure performance du modèle. Pour maximiser l'ASC, il est nécessaire de déterminer le point de fonctionnement optimal (POO) sur la courbe ROC, qui définit la valeur de seuil qui minimise le taux de mauvaise prédiction pour le modèle tout en optimisant le taux de prédiction correcte.

Cette approche permet de quantifier le risque de prédiction inexacte. Par exemple, il peut être souhaitable d'augmenter le seuil pour réduire le risque de prédictions faussement positives. L'approche ROC permet de quantifier cette réduction des prévisions faussement positives et quantifie également la réduction des prévisions positives réelles associées à la même augmentation de seuil.

Keywords: Receiver Operating Characteristic; Statistics

1 INTRODUCTION

Geomaterials, such as soil and rock, are inherently variable and it is common to encounter differences in their engineering properties on most sites. This intrinsic variability typically has an associated spatial component due to the effect of geological formative processes on geotechnical properties. Given this inherent variability of geomaterials, it is necessary to classify them to understand the material behaviours and the impact this has on geotechnical designs.

Classification of geomaterials is undertaken using both direct and indirect approaches. Direct approaches include methods in which sample characteristics are directly measured to classify the geomaterial. Common examples of these direct approaches are described by soil classification standards (ISO (2017a and 2017b); BSI (2015); ASTM (2017)) and rock classification standards (ISO (2017c); CIRIA C574 (Lord et al., 2002); ASTM (2008)). Indirect approaches include methods in which measured properties are correlated with soil classification properties or behaviour types. Common examples of indirect approaches include cone penetration test (CPT) correlations (e.g. Robertson, 2010), empirical correlations between classification tests and more advanced parameters, and geophysical correlations (such as inversion approaches, e.g. Vardy et al., 2018).

Conventional practice is to utilise engineering judgement to classify geomaterial behaviour. Such an approach is inherently subjective and can result in notable discrepancies, which is why there has been increasing interest in utilising more objective and repeatable approaches to classify geomaterials.

Statistical methods offer a relatively more objective approach to classifying geomaterial behaviour and have seen increased application in recent years. DNV (2015) details recommended practice for the statistical representation of soil data and Baecher and Christian (2003) provide numerous examples of how statistical methods can be applied in a geotechnical context.

Moreover, with the increasing use of quantitative risk assessments to assess ground-related risks, statistical methods offer the distinct advantage of enabling risks to be mathematically quantified and value assigned accordingly. It should be noted that the use of statistical approaches does not replace the application of engineering judgement and should instead be viewed as merely another tool to assist engineering judgement.

One statistical approach that has seen relatively limited application in a geotechnical context, but which has significant potential, is the receiver operating characteristic (ROC) analysis. The ROC analysis is a statistical approach used to assess the efficacy of a diagnostic evaluation procedure. It has been widely used in areas of research such as electrical engineering, medicine, meteorology, machine learning and data-mining. In a geotechnical context, ROC analyses have been applied to assess the efficacy of predictive methods for pile driveability (Mens et al., 2012) and liquefaction (Maurer et al., 2015).

The following sections provide an overview of the ROC analysis method and detail an example ROC analysis which demonstrates how this approach can be used to assist with geospatial classification problems. In the example, the ROC analysis is used to calibrate and assess the efficacy of a predictive method for determining the presence of hard ground. The results are subsequently plotted within a geographical information system (GIS) to illustrate the locations where there is greatest confidence of encountering hard ground and which therefore pose the greatest risk to shallow foundation installation.

2 ROC ANALYSIS

For binary classification problems, such as whether a feature will or will not be encountered at a given location, there are two possible outcomes: a ‘positive’ result (i.e. the classifier is observed) and a ‘negative’ result (i.e. no classifier is observed). An associated diagnostic model will also provide either a ‘positive’ or ‘negative’ prediction. As such, the model can perform in one of four ways:

- A positive prediction that in reality is a positive result is a true positive (TP);
- A positive prediction that in reality is a negative result is a false positive (FP);
- A negative prediction that in reality is a positive result is a false negative (FN);
- A negative prediction that in reality is a negative result is a true negative (TN).

The above can be expressed using a confusion matrix (also known as a contingency table), as shown in Figure 1.

		Instance	
		Positive	Negative
Diagnostic Model Prediction	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Figure 1. Confusion matrix

When a diagnostic model is based on a continuous variable (e.g. a magnetic field, an electric charge, a seismic amplitude, etc.), continuous probability distributions of 'positive' and 'negative' results can be plotted as a function of that variable. In such cases, a threshold value is used in the diagnostic model to define the boundary of positive and negative predictions: values one side of a given threshold yield a ‘positive’ prediction whereas values on the other side of the threshold yield a ‘negative’ prediction. Depending on the distributions and threshold value used, different proportions of the above four outcomes will be achieved. Figure 2 illustrates this concept.

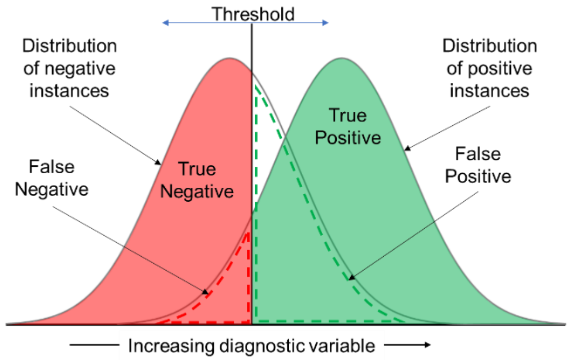

Figure 2. Illustration of a typical bimodal distribution from a continuously varying diagnostic model showing the effect of the threshold value on prediction success rate

When the distributions of ‘positives’ and ‘negatives’ overlap, setting the threshold too low will result in numerous false positives and setting the threshold too high will result in numerous false negatives. Both outcomes will have consequences (e.g. increased direct cost or increased risk), so it may be desirable to determine the optimum threshold value to balance the risks.

The ROC analysis is a method to assess the efficacy of a diagnostic model for predicting instances of a classifier and can be used to calibrate diagnostic models so that the predictive performance is optimised. To assess how well a diagnostic model performs, the true positive rate (TPR) from the diagnostic model is plotted against the false positive rate (FPR) to produce an ROC curve. It should be noted that TPR and FPR are synonymous with the true positive and false positive probabilities, respectively. The following equations summarise how the outcomes presented in Figure 1 are calculated:

$$TPR = \frac{\text{TP}}{(TP+FN)} \tag{1}$$

$$FPR = \frac{\text{FP}}{(FP+TN)} \tag{2}$$

$$\text{TNR} = 1 - \text{FPR} \tag{3}$$

$$\text{FNR} = 1 - \text{TPR} \tag{4}$$

Figure 3 illustrates how an ROC curve is used to assess the efficacy of a diagnostic model. A diagonal line from coordinates (0,0) to (1,1) indicates random guessing as the rate of correct and incorrect prediction is equal. A point at (0,1) indicates a perfect diagnostic model with a threshold value which segregates all ‘positives’ and ‘negatives’. The area under the curve (AUC) is commonly used to assess the efficacy of a diagnostic model, with increasing AUC indicative of better model performance. To maximise AUC, it is necessary to determine the optimum operating point (OOP) on the ROC curve, which defines the threshold value which minimises misprediction (i.e. FPR + FNR).

3 CASE STUDY

3.1 Background

In this case study, areas of hard ground were encountered around recently installed infrastructure within an offshore site, but the spatial extent of the hard ground was unknown. The hard ground was detected at seabed or at very shallow depths below seabed following a reconnaissance survey campaign which included approximately 15 seabed CPTs and 5 boreholes. The presence of hard ground was not detected in earlier site investigation campaigns performed prior to installation of the existing infrastructure and it is believed that the hard ground was caused by drilling activities.

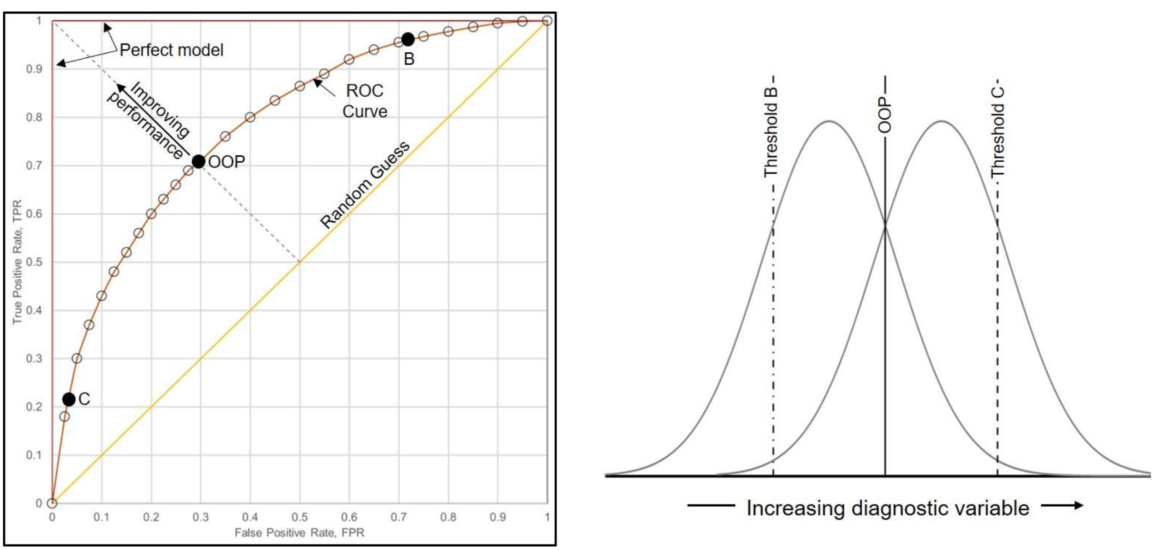

Figure 3: Illustration of how the ROC curve is used to assess the performance of the diagnostic test (left) and comparison to the distribution of the continuously varying diagnostic model (right). The position of three threshold values (OOP, B and C) are illustrated. The OOP represents the combination of the highest TPR and lowest FPR

Additional structures with shallow foundations were planned for installation at the site. The shallow foundations (mudmats with shallow skirts) were designed for the soft clay that is naturally occurring at the site. The hard ground presented a significant risk to the successful installation of the shallow founded structures, with potential for differential skirt penetration across the structure and resulting reduced performance, risk of overturning and possible misalignment of connections with other equipment.

The hard ground was clearly detected from CPT cone tip resistance (q_c) as a sharp increase in q_c, often resulting in CPT refusal, as shown in Figure 4. However, given the small area represented by a CPT test, an impractically large number of discrete CPT tests would be required to determine the extent of the hard ground with the level of detail required to microsite the additional structures away from hazardous areas. In areas where hard ground was detected from CPT q_c data, a high geophysical amplitude response was also observed in sub-bottom profiler data. The absolute amplitude that related to hard ground was less defined than the clear indications from CPT q_c. However, in contrast to discrete CPT data, it is possible to acquire spatially resolute site-wide coverage of (continuous) geophysical amplitude data.

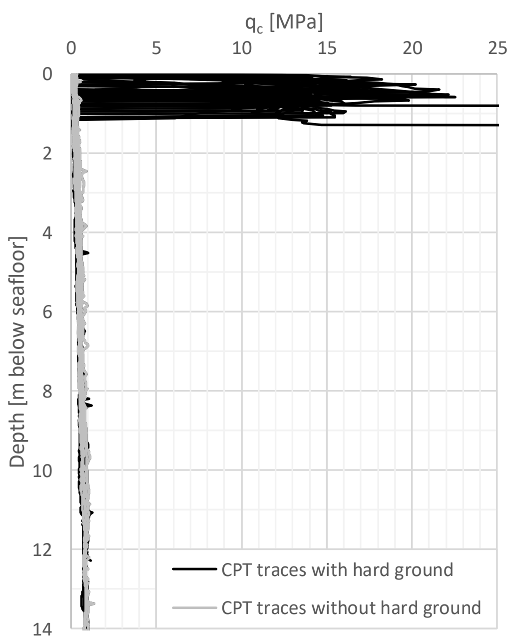

Figure 4: Selection of CPT traces where hard ground was and was not identified

Therefore, a chirp sub-bottom profiler mounted on a Remotely Operated Vehicle (ROV) was used to acquire data on a closely-spaced grid which was interpolated to give full sub-bottom profiler coverage across the area of interest. An ROV-based seabed CPT was used to acquire approximately 100 CPTs in a grid-based acquisition pattern aimed to target areas where hard ground was expected but also aimed to target areas for best calibration of the amplitude data. Figure 5 presents a map of the seismic amplitude data and CPT locations.

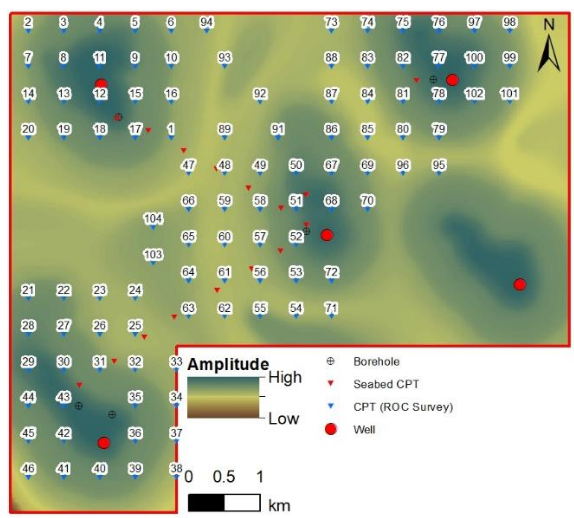

Figure 5: Map of seismic amplitude data and geotechnical investigation locations

Comparison of the geophysical amplitude data with CPT q_c data showed that the geophysical amplitude could be used to predict the presence of hard ground. However, due to the scatter in the dataset, it was unclear what threshold value of geophysical amplitude would best distinguish between the presence or absence of hard ground.

An effective way of optimising the geotechnical correlation is to use ROC analysis.

3.2 ROC Method

An ROC analysis was performed to assess the efficacy of different geophysical amplitude thresholds for predicting the presence of hard ground and to determine the optimum amplitude threshold. The competing diagnostic tests were the different amplitude thresholds and the index test results were the q_c values exceeding 5 MPa in the top 2 m below seafloor. Accordingly, true and false positives were scenarios where hard ground is predicted by geophysical amplitude, but were and were not observed from q_c response, respectively.

At each CPT location, geophysical amplitude was taken from the amplitude data gridded at 1m horizontal resolution. This gridding process reflects the resolution of the geophysical data, which returns an aggregated response from a small elliptical footprint (rather than a single point). The gridding process also helps to account for potential positioning errors or local anomalies. This amplitude data was subsequently tabulated alongside the CPT test data with a binary classification indicating whether or not hard ground was considered present based on the definition mentioned above.

An ROC analysis was performed using this dataset by first specifying threshold geophysical amplitudes at intervals of 50. If hard ground was interpreted to be present based on CPT results and the amplitude was greater than the threshold value, values of 1 were assigned; where these criteria were not met, values of 0 were assigned. Values of TN, FN, TP and FP were subsequently calculated and operating characteristics were calculated using Equations (1) to (4). The rate of misprediction (FPR + FNR) was also calculated.

The TPR and FPR were subsequently plotted together to produce an ROC curve. The OOP was the geophysical amplitude with the lowest rate of misprediction.

3.3 Results

Figure 7 presents the ROC analysis results. The OOP for this dataset was an amplitude of 850. At this OOP, the TPR was 84% and the FPR was 15%. The rate of misprediction was thus 31%.

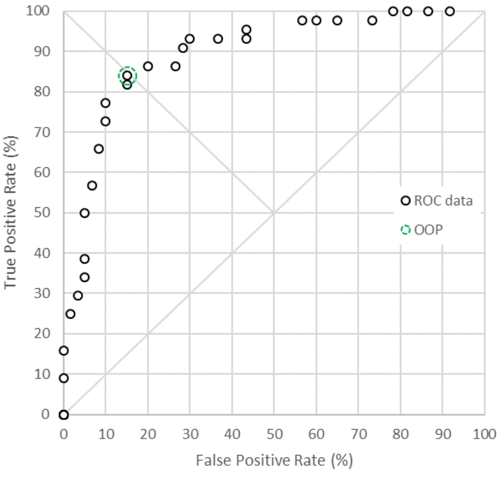

Figure 7: Results from ROC analysis

To understand how the selection of different amplitude thresholds (i.e. different TPR and FPR) influences the spatial extent of the problem, several contours indicating different amplitudes were plotted. Figure 8 presents a map with contours for the OOP and amplitudes of 300 and 1350, which have misprediction rates of 75%.

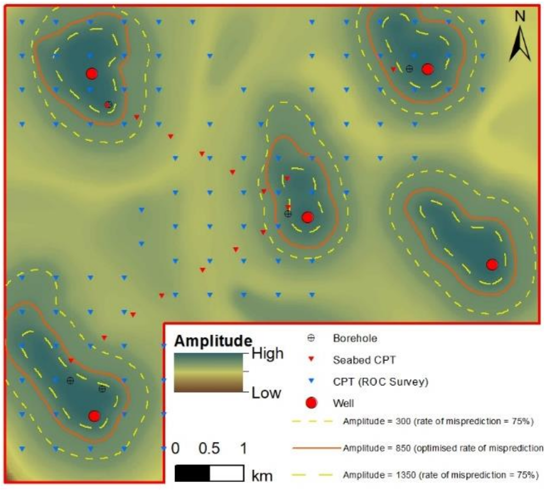

Figure 8: Spatial context for ROC analysis results

3.4 Discussion

The results convey how the selection of a different amplitude threshold affects the balance of risk and direct cost. Moreover, increasing the amplitude threshold would reduce direct costs as the spatial extent of the hazardous zone reduces, thereby decreasing the area requiring remedial work. However, this would result in increased risk of encountering hard ground unexpectedly. Conversely, decreasing the amplitude threshold would increase the direct cost (increased hazardous area) but reduce risk. Communicating this in a spatial context facilitates understanding of the spatial variation of risk.

As shown in Figure 8, the analysis results can be used to define zones with a common risk associated with them. Depending on preferences for direct cost or risk, the zone can be extended or reduced and the residual risk quantified.

The results also allow a prediction of hard ground to be made in areas of the site where discrete CPT traces were not performed. This helps to reduce direct costs since fewer CPT traces are required. However, as for the selection of an appropriate amplitude threshold, direct costs are often balanced against risks; in this case, that the model would be more reliable if more CPT data were acquired.

To confirm whether sufficient CPT traces were obtained to develop a reliable model, the ROC analyses were performed with a varying number of CPT results considered. CPT traces for these smaller ROC analyses were randomly selected and the analyses repeated numerous times (N = 10, 100 and 1000) to obtain an average result for a given number of CPT traces. This repetition was necessary due to the randomised selection of CPT/amplitude pairs and to assess whether results were converging.

Figure 9 presents the predicted OOP with number of CPT traces considered. Figure 10 presents the normalised rate of misprediction (the rate of misprediction in percent divided by the number of tests) with the number of CPT traces considered.

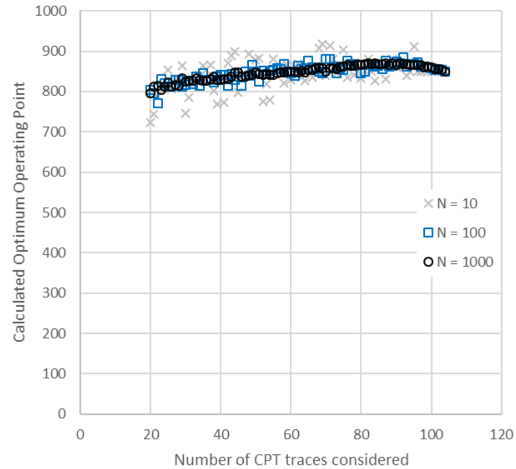

Figure 9: OOP prediction with number of CPT results

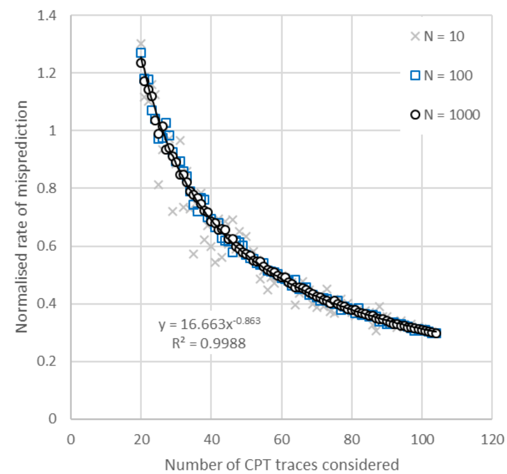

Figure 10: Normalised rate of misprediction with number of CPT results

As shown in Figure 9, the OOP results converge to a value close to 850 after around 80 CPTs are considered in the ROC analysis. Given that the ROC analyses considered amplitude thresholds at intervals of 50, this suggests that a sufficient number of CPTs were performed because the results had essentially converged.

Figure 10 shows that the normalised rate of misprediction decreases with an increase in the number of CPTs considered and is represented well using a power law function. This indicates diminishing returns with inclusion of new CPTs in the model. As the normalised rate of misprediction for the full ROC analysis (i.e. the analysis including all CPTs) is on a relatively flat part of this curve, this also suggests that a sufficient number of CPTs were performed.

While the case study presented here demonstrates one application of the ROC method, it appears a suitable approach for a variety of geoscience applications. The ROC method is particularly suitable for site-specific calibrations of correlations and predictive models which use threshold values to predict an outcome.

The ROC method can also be extended to multiple classification problems, such as correlating another parameter (e.g. thickness of hard ground). Such examples would aim to optimise the volume under the surface (VUS) instead of AUC as there are multiple correlations.

4 CONCLUSIONS

ROC analyses can be performed to optimise the predictive performance of geotechnical correlations. Results from ROC analyses enable the model performance to be quantified by calculating the true positive and false positive probabilities and the rate of misprediction.

Results from ROC analyses can also be used for probabilistic hazard zoning and assessing the spatial variation of risk. This allows the spatial extent of direct cost and risk to be quantified.

A method for assessing whether sufficient data has been obtained for the predictive model is also presented. The normalised rate of misprediction is found to be represented well by a power law.

The ROC method is an efficient approach which has potential for further uses in a geotechnical context. It is also a suitable method for multiple classification problems.

5 REFERENCES

ASTM, 2017. *D2487-17: Standard Practice for Classification of Soils for Engineering Purposes (Unified Soil Classification System)*. ASTM International, West Conshohocken, PA, USA.

ASTM, 2008. *D5878-08: Standard Guides for Using Rock-Mass Classification Systems for Engineering Purposes.* ASTM International, West Conshohocken, PA, USA.

Baecher, G.B. and Christian, J.T., 2003. *Reliability and statistics in geotechnical engineering*, John Wiley & Sons Ltd., Chichester, UK.

BSI, 2015. BS 5930:2015 *Code of Practice for Site Investigations*. BSI, London, UK.

DNV, 2015. *DNV-RP-C207 Statistical representation of soil data*. Oslo: Det Norske Veritas AS (DNV).

ISO, 2017a. *ISO 14688-1:2017 Geotechnical Investigation and Testing – Identification and Classification of Soil – Part 1: Identification and Description*. ISO, Geneva, Switzerland.

ISO, 2017b. *ISO 14688-2:2017 Geotechnical Investigation and Testing – Identification and Classification of Soil – Part 2: Principles for a classification*. ISO, Geneva, Switzerland.

ISO, 2017c. *ISO 14689:2017 Geotechnical investigation and testing — Identification, description and classification of rock*. ISO, Geneva, Switzerland.

Lord, J.A., Clayton, C.R.I. and Mortimore, R.N., 2002. *CIRIA C574 Engineering in chalk*. London: CIRIA.

Maurer, B.W., Green, R.A., Cubrinovski, M., and Bradley, B.A., 2015. Assessment of CPT-based methods for liquefaction evaluation in a liquefaction potential index framework. *Géotechnique*, 65 (5), pp. 328-336.

Mens, A.M.J., Korff, M., and Van Tol, A.F., 2012. Validating and improving models for vibratory installation of steel sheet piles with field observations. *Geotechnical and Geological Engineering*, **30** (5), pp. 1085-1095.

Robertson, P.K., 2010. Soil behaviour type from the CPT: an update. *2nd International Symposium on Cone Penetration Testing, CPT'10*, Huntington Beach, CA, USA

Vardy, M., Clare, M., Vanneste, M., Forsberg, C. F., and Dix, J., 1998. Seismic Inversion for Site Characterisation: When, Where and Why Should We Use It? *Proceedings of the Offshore Technology Conference*, OTC-28730-MS, Houston, Texas.