



Development of an integrated game theory-optimization subground stratification model using cone penetration test (CPT) measurements

Mohammad Sadegh Farhadi¹ · Tim Länsivaara¹

Received: 31 July 2020 / Accepted: 12 December 2020 / Published online: 21 January 2021
© The Author(s) 2021

Abstract

The continuous cone penetration test (CPT) measurements provide an advantageous liable rapid tool for stratification and soil behavior classification that can be employed in the sustainable design of the infrastructures. However, the CPT measurements are often interpreted by geotechnical experts because of the involved complexities and uncertainties. In this study, a novel stratification and soil type behavior (SBT) classification model is developed to identify the transition and thicker layers by integrating the geotechnical knowledge with the three submodels of (a) locally estimated scatterplot smoothing (LOESS), (b) a game theory model known as Nash–Harsanyi (N–H) bargaining, and (c) grey wolf optimizer (GWO). The LOESS and integrated N–H bargaining-GWO models are, respectively, used to approximate the outliers in CPT measurements and identify the SBT and layer changes. Attractively, in the proposed model, the engineer has the opportunity to judge on the precision of the stratification profile regarding their own preferences in a project. Solving simple algebraic equations, high speed, identifying thick and the interlayer transition layers, and small required training data are the other advantages of the developed model. Finally, the applicability of the proposed model has been assessed in an example. The compared estimated and two other models' stratification profiles highlighted the potential of the proposed model to identify thin transition layers.

Keywords Subground stratification · Cone penetration test (CPT) · Game theory · Nash–Harsanyi bargaining model · Grey wolf optimizer (GWO)

List of symbols

Cone penetration test (CPT)

F_r	Normalized friction ratio
f_s	Sleeve friction
I_c	SBT index
q_t	Cone tip resistance
Q_m	Normalized cone resistance
SBT	Soil behavior type

Grey wolf optimizer (GWO)

\vec{A} and \vec{C}	Coefficient vectors
\vec{a}	Vector of numbers decreasing linearly from 2 to 0
\vec{D}	Distance criterion between wolves and prey
\vec{r}_1 and \vec{r}_2	Random vectors in [0, 1]

\vec{X}_p	Position vector of prey
\vec{X}	Position vector of a grey wolf
X_i	A grey wolf considered in computations
α, β and δ	Three best wolves approaching a prey, respectively
ω	Representing a wolf

Locally estimated scatterplot smoothing (LOESS) model parameters

a_0, a_1, a_2	Variables in linear and quadratic approximations
h	Neighborhood bandwidth for each data point
(x_i, y_i)	Coordinate of any data point
$\mu(x_i)$	Local regression estimate
$w(x_i)$	Smoothing weight of data point i in the neighborhood range of h
w	Weight function

Nash–Harsanyi bargaining model parameters

$D_N - H$	Nash–Harsanyi distance criterion
$D_{\alpha\text{-cut}}$	Layer change distance limit/precision regulating parameter
d_i	Disagreement point for player i

✉ Tim Länsivaara
tim.lansivaara@tuni.fi

Mohammad Sadegh Farhadi
mohammadsadegh.farhadi@tuni.fi

¹ Department of Civil Engineering, Faculty of Built Environment, Tampere University, Tampere, Finland

$d_{N-H(i,i-k)}$	Nash–Harsanyi distance between data points i and $i - k$
N–H	Nash–Harsanyi
n	Number of players
p_i	proportion of cooperation of each player
u_i	Objective function of player i
Ω	Overall profit
Others	
CPD	Change point detection
λ	Coefficient of regulating parameter

1 Introduction

A better knowledge of the subground state is beneficial in the sustainable design and development of the infrastructures' foundations. Cone penetration test (CPT) is a fast, repeatable, economical site investigation test in geotechnical engineering that provides almost continuous measurements in depth [37, 46]. Hence, its popularity is still growing among engineers. The continuous measurements of cone tip and sleeve frictional resistances in depth make it a promising method for subground stratification and soil behavior type (SBT) classification [8], although an appropriate interpretation may provide further engineering information [20, 37, 44, 48, 54, 64]. Meanwhile, the ranges of CPT measurements and the involved fluctuations for different SBTs are always influenced by several factors, such as the spatial variability of soils, cone size, and sensors' precision of measurements [2, 5, 28, 35, 37, 43]. They bring about both complexity and uncertainty in CPT-based soil stratification. Therefore, subground stratification has often been performed by the CPT experts. However, in the past 2 decades, numerous studies have tried to bridge the gap between mathematical and geotechnical knowledge for stratifying subsurface soils using CPT measurements.

To overcome the complexities and uncertainties of the measurements in stratification, vast research is performed on the determination of the CPT measurements ranges for different soils and SBT classes, as well as the development of the computational processing models to recognize the soil layers from measurements. Since the invention of CPT, SBT classification based on the measurements has been investigated numerous and some charts are proposed for this purpose [4, 6, 7, 19, 21, 22, 29, 43, 45, 47, 49, 51, 53, 58]. However, as a problem, the existence of uncertainties and probably outliers in CPT measurements is almost inevitable. The outlier identification in time series and the procedures to approximate them mathematically have been targeted in numerous studies since many years [1, 9, 10, 36, 57]. Some studies used the geotechnical principles and statistical methods simultaneously to detect and remove the outliers

[12, 20]. In the meantime, especially in the past 2 decades, engineers have sought computational methods to facilitate identifying subground layers and to substitute the required experts' knowledge. Zhang and Tumay [66] suggested the statistical and fuzzy sublayer identification approaches using the soil types almost similar to the ones in the unified soil classification system (USCS). The CPT tests conducted at the National Geotechnical Experimentation site (NGES) at Texas A&M University [56] were used to show their method's applicability. Hegazy and Mayne [25] presented the improvement of clustering methods over the previous statistical ones for the CPT-based soil classification. They showed that clustering could detect major changes within the stratigraphy, which might not be apparently visible. A probabilistic approach was developed by Jung et al. [30] to modify the soil identification charts based on the CPT data. Das and Basudhar [17] proposed self-organizing maps and fuzzy clustering techniques for identification of different layers. The estimated results were comparable with those of obtained from the cone classification chart, hierarchical and K-mean clustering techniques. Wang et al. [63] modelled the uncertainty in the CPT-based soil stratification and classification by means of the Bayesian approach and using the Robertson chart proposed in 1990 [45]. The proposed model was evaluated based on some real CPT data. Ching et al. [13] used the SBT index, I_c , in their proposed stratigraphic profiling approach. The layer boundaries were recognized at the relatively large change points in the I_c profile. They utilized the wavelet transform modulus maxima (WTMM) method, and 50 real CPT-based stratification profiles provided by experts to determine the layer change point criterion. Cao et al. [8] developed a Bayesian framework based on the SBT index, I_c , for the probabilistic soil stratification. The number and thickness of layers, and also their associated identification uncertainty were estimated. Wang et al. [62] proposed a semi-supervised clustering method built on a hidden Markov random field framework using boreholes and CPT sounding logs. Wang et al. [59] suggested an unsupervised Bayesian inferential framework integrated with the Robertson chart [45] to determine the strata and the corresponding SBTs. In brief, the literature indicates that the computational CPT-based subground stratification models were built on two general approaches. In the first approach, the consistency among the sequential CPT data points has been employed to find the thick layers, such as in the Bayesian theory and clustering models. On the other hand, in the second approach, the boundaries between the soil layers are determined based on the change points and especially sudden fluctuations detections in CPT results, such as in the wavelet model proposed by Ching et al. [13]. There are still some deficiencies with the existing methods, although great progress has been made so far in stratification models. Some proposed methods are time-consuming, and investigation is still being performed to

reduce their computation time. Almost all proposed models concentrated on seeking the strata featuring the high consistency of the corresponding measurements. But the transition layers between them are neglected or finding them may require longer computation time. In almost all methods, it has been tried to find the layer boundaries deterministically even though probabilistic methods were employed. However, due to some intrinsic uncertainties in CPT soundings which were not targeted focally in the proposed models, reporting the layers change boundaries and SBTs deterministically may be still a bit risky.

Hence, in this paper, integrating geotechnical knowledge with time series/signal smoothing, game theory, and optimization models, a totally different and novel model is proposed for identifying the strata depths and their SBTs. As the advantages of the proposed model, it runs rapidly, finds both the thick layers and the transition layers in between, and provides the possibility to tune the precision of stratification-SBT classification profile. The precision regulating parameter is implemented in the model to avoid deterministic presentation of the subground stratification and to provide the engineer the possibility to make their own judgement if required. In the proposed model, first, a local regression smoothing method is utilized to reduce the outliers' and uncertainties' impact within the CPT measurements. Then integrating the Nash–Harsanyi (N–H) model, as a game theory model, grey wolf optimizer (GWO), as an optimization model, and Robertson soil classification chart (1990), a model is proposed to identify the transition and then the thick layers in between, as well as their corresponding SBTs. Although the focus and novelty of the present study has been the developed model itself, after describing the model, the practicality of the model is verified based on training the model using only one CPT-based stratification profile provided by experts and comparing the results with two other published stratification models for other three testing CPT soundings.

2 The proposed model

The proposed model is encapsulated in Fig. 1. It is comprised of two modules. In the first module, the raw sleeve friction, f_s , and cone tip resistance, q_c , measurements are loaded, corrected by the penetrometer manufacturer's recommendations, modified into normalized friction ratio, F_r , and normalized cone tip resistance, Q_{tn} , using the equations proposed by Robertson [46], and smoothed by a local regression method named as locally estimated scatterplot smoothing, LOESS. The smoothing is proposed as a denoising method to reduce the sudden illogical fluctuations in measurement signals in depth. In the second module named as soil stratification module, soil layers are determined by

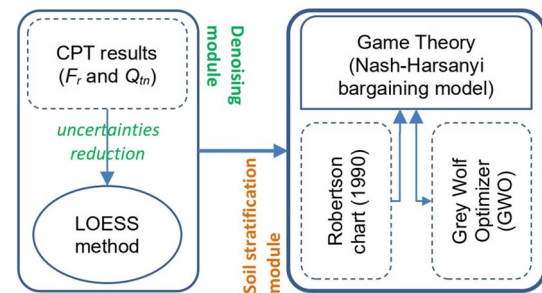


Fig. 1 Concise schematic illustration of the proposed model

integrating the N–H bargaining and GWO models with the Robertson SBT classification chart [45] as the geotechnical experience-based judgement.

The algorithm of the proposed model is explained in detail in Sect. 2.2; after the employed submodels are described mathematically in Sect. 2.1.

The proposed model was developed in MATLAB, 2019a (The Mathworks, Inc. MATLAB, Version 9.6 2019).

2.1 The utilized submodels

The four computational and geotechnical models or approaches used in the proposed model are described below:

2.1.1 LOESS smoothing method

Local regression estimation was independently introduced in several fields in the nineteenth and twentieth centuries [26, 52]. In 1979, Cleveland applied the robust locally weighted regression to smoothing scatterplots, and Cleveland and Devlin [14] extended the model to the multivariate case [34].

LOESS is a weighted regression method for smoothing a scatterplot, $(x_i, y_i) i = 1, 2, \dots, n$, in which the fitted value at x_k is the value of a polynomial fit to the data using weighted least squares. The underlying principle is that a smooth function can be well approximated by a low-degree polynomial in the neighborhood, $(-h, h)$, of any point x . For example, a local linear approximation is [14]:

$$\mu(x_i) \approx a_0 + a_1(x_i - x) \quad (1)$$

for $x_i - h < x_i < x_i + h$. A local quadratic approximation, which was used in this study, is

$$\mu(x_i) \approx a_0 + a_1(x_i - x) + \frac{a_2}{2}(x_i - x)^2. \quad (2)$$

The local approximation can be fitted by locally weighted least squares. The coefficient estimates a_0 , a_1 and a_2 were chosen to minimize:

$$\sum_{i=1}^n w(x_i)[Y_i - \mu(x_i)]^2 \tag{3}$$

Each local least squares problem defined $\hat{\mu}(x_i)$ at one point x ; if x was changed, the smoothing weights $w(x_i)$ would change, and so the estimates a_0 and a_1 .

A smooth weight function results in a smoother estimate [38]. The tricube weight function was used in the LOESS fitting procedure in this study [14–16]:

$$w(x_i) = \begin{cases} (1 - |\frac{x_i - x}{h}|^3)^3 & |u| < 1 \\ 0 & |u| > 1. \end{cases} \tag{4}$$

w is a weight function with the following properties: [14]:

1. $w(x) > 0$ for $|x| < 1$;
2. $w(-x) = w(x)$;
3. $w(x)$ is a nonincreasing function for $x \geq 0$;
4. $w(x) = 0$ for $|x| \geq 1$.

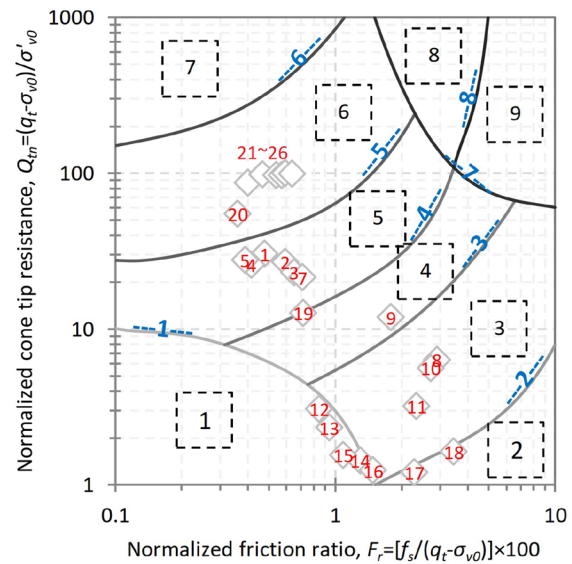
2.1.2 Robertson chart

Robertson [45] categorized soils, with respect to worldwide CPT results, based on their particles size and their behaviour. The outcome of their study has been the chart shown in Fig. 2. As can be seen, nine SBTs are defined in the chart. In addition, the data points labelled with numbers 1–26 show the distribution of a sequence of CPT measurements in a range of 50 cm in depth. The points 1–7 (excluding point 6 which has been an outlier located out of the chart range) and 20–26 are located in zones 5 and 6, respectively; and the points 8–19 are spread in zones 1–5.

The SBT boundaries proposed in the Robertson chart were identified based on the experiments, but not the analytical methods. Therefore, to use the chart in calculations, eight curves were fit to them. Jung et al. [30] assessed the exponential equations in a semi-logarithmic scale. In this study, fitting both polynomial and exponential equations to the boundary lines in both linear and log–log space were verified. Sum squared error (SSE), coefficient of determination (R^2), and root mean square error (RMSE) were used as the error criteria to select the best fitted curves. The best-fitted equations for each line (numbered on the Robertson chart in Fig. 2) and the corresponding error criteria are presented in Table 1. The small quantities of error indicate the well-fitted curves.

2.1.3 Nash–Harsanyi bargaining method

As a Game Theory model, the Nash–Harsanyi (N–H) bargaining model has been employed in different fields,



zone	Soil behavior type (SBT)
1	sensitive, fine grained
2	organic soils-peats
3	clays-clay to silty clay
4	silt mixtures-clayey silt to silty clay
5	sand mixtures-silty sand to sandy silt
6	sands-clean sand to silty sand
7	gravelly sand to sand
8	very stiff sand to clayey sand*
9	very stiff, fine grained*

*Heavily overconsolidated or cemented

Fig. 2 Soil behavior type (SBT) chart based on CPT data, proposed by Robertson [45]. The SBT boundary lines are numbered on the chart, and the SBT classification is stated underneath, in the table. The data points numbered as 1–26 have been measured at the Lukang test site, Taiwan, at the depth range of 3.65–4.9 m. The whole F_r and Q_{tn} measurements of the site are presented in Fig. 4

especially in the environmental and water management studies as a conflict-resolution model [18, 23, 24, 39]. Nash [42] proposed a bargaining model considering the cooperation among the players. This method maximizes the total profit of players through coalition and cooperation, considering each player an equal proportion of cooperation. In spite, in the asymmetric N–H model, the players share their different proportions of cooperation to obtain an agreement for the maximum overall profit. Indeed, both the individual and also the collective rationalities are considered in this method [24].

In the N–H model, n players with u_i objective functions, where i represents each player, and d_i is the disagreement points of players in the game. The overall profit, Ω , in the model can be written as

Table 1 The best-fitted curves to the Robertson chart [45] boundaries

Line no.	Equation	SSE	R ²	RMSE
1	$y = -2.474x^5 - 7.17x^4 - 7.066x^3 - 3.836x^2 - 1.721x + 0.4744$	0.0012	0.9995	0.00646
2	$y = 1.086x^4 - 0.8107x^3 - 0.0905x^2 + 0.8637x - 0.1521$	0.0004	0.9998	0.00352
3	$y = 0.4864x^3 + 0.2955x^2 + 0.7638x + 0.7346$	0.0023	0.9996	0.00768
4	$y = 1.204 \exp(0.5827x) + 0.0034 \exp(8.784x)$	0.0017	0.9997	0.00650
5	$y = 0.7409x^4 + 1.503x^3 + 1.286x^2 + 0.8894x + 1.806$	0.0022	0.9995	0.00655
6	$y = 0.5231x^3 + 1.518x^2 + 1.743x + 2.924$	0.0016	0.9995	0.00587
7	$y = -0.429x^4 - 1.195x^3 + 5.555x^2 - 6.068x + 3.917$	0.0045	0.9995	0.00904
8	$y = 258.4x^3 - 453.1x^2 + 268.6x - 51.57$	0.0069	0.9978	0.01475

x and y stand for $\log_{10}F_r$ and $\log_{10}Q_{in}$, respectively

$$\Omega = \max \prod_{i=1}^n (u_i - d_i)^{p_i} \tag{5}$$

subject to

$$u_i \geq d_i, \text{ where } i = 1, 2, \dots, n.$$

2.1.4 Grey wolf optimizer (GWO)

Among the numerous applied meta-heuristic optimization models [11, 55, 60, 61, 65], Mirjalili et al. [40] proposed a meta-heuristic search algorithm inspired by grey wolves (Canis lupus) attacking a prey. This optimization algorithm has been employed in different fields like soil mechanics, electric power system, image processing, and medicine [27, 31–33, 41, 67].

The social hierarchy of wolves is modelled mathematically considering alpha, α , beta, β , delta, δ , and omega, ω , wolves. The three best solutions are considered as α , β and δ wolves, respectively, and the other possible solutions are considered as ω wolves. In a hunting (optimization) process, ω wolves follow the three best wolves.

In brief, different steps of hunting are encoded mathematically as below [40]:

- Encircling prey:

$$\vec{D} = |\vec{C} \cdot \vec{X}_p(t) - \vec{X}(t)|, \tag{6}$$

$$\vec{X}(t + 1) = \vec{X}_p(t) - \vec{A} \cdot \vec{D}, \tag{7}$$

$$\vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{a}, \tag{8}$$

$$\vec{C} = 2\vec{r}_2, \tag{9}$$

where \vec{A} and \vec{C} are coefficient vectors, \vec{X}_p is the position vector of the prey, \vec{X} indicates the position vector of a grey wolf, \vec{r}_1 and \vec{r}_2 are random vectors in [0, 1], and \vec{a}

vectors' components are linearly decreased from 2 to 0 over the course of iterations.

- Hunting:

The α , β , and δ wolves have better estimation of the potential solution and the ω wolves follow them. This is mathematically implemented as

$$\begin{aligned} \vec{D}_\alpha &= |\vec{C}_1 \cdot \vec{X}_\alpha - \vec{X}| \\ \vec{D}_\beta &= |\vec{C}_2 \cdot \vec{X}_\beta - \vec{X}| \\ \vec{D}_\delta &= |\vec{C}_3 \cdot \vec{X}_\delta - \vec{X}| \end{aligned} \tag{10}$$

$$\begin{aligned} \vec{X}_1 &= \vec{X}_\alpha - \vec{A}_1 \cdot (\vec{D}_\alpha) \\ \vec{X}_2 &= \vec{X}_\beta - \vec{A}_2 \cdot (\vec{D}_\beta) \\ \vec{X}_3 &= \vec{X}_\delta - \vec{A}_3 \cdot (\vec{D}_\delta) \end{aligned} \tag{11}$$

$$\vec{X}(t + 1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3}. \tag{12}$$

- Attacking prey:

Approaching a prey is modelled by decreasing the value of \vec{a} over the course of iterations.

- Search for prey (exploration):

To avoid the local optimum finding, \vec{A} with random values greater than 1 and smaller than -1 oblige the search agent to diverge from the prey and search for the solution globally. To find an applicable knowledge of the whole process of GWO, the reader may refer to a pseudo-code presented in [40].

2.2 The algorithm of the proposed model

The stepwise procedure of the proposed model is presented in Fig. 3. As can be observed, CPT acquisition process, the derived CPT data denoising, and stratification identification modules are illustrated step-by-step schematically and the details are presented below.

Observing the sequence of F_r and Q_{in} series on the Robertson chart (like the points shown in Fig. 2) revealed that

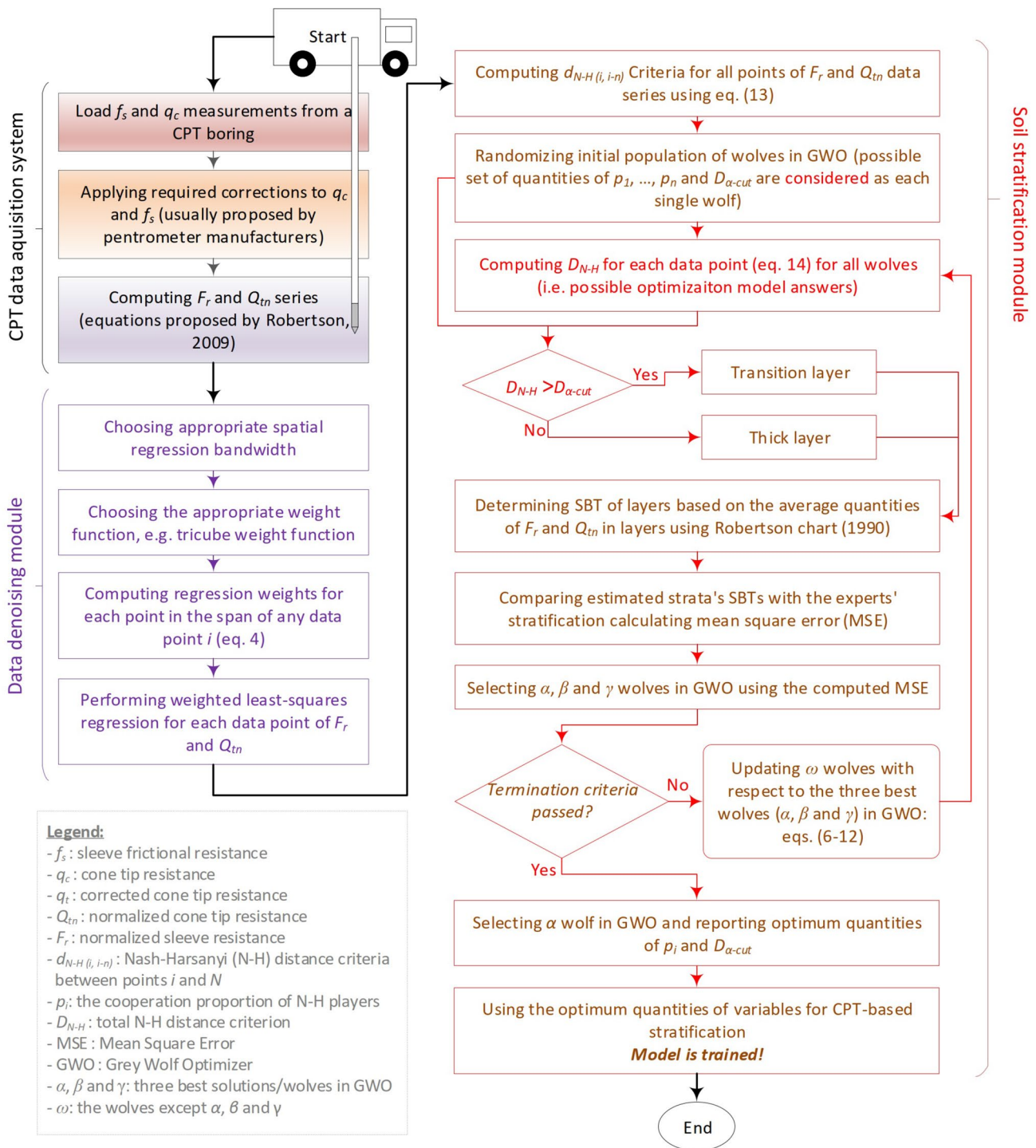


Fig. 3 Schematic stepwise illustration (flowchart) of the developed model

the main problem with the stratification and determination of SBT directly from the Robertson chart was reading different SBTs for two adjacent measurements located close to each other on the chart but on different sides of a boundary line. For example, in Fig. 2, it can be observed for points 14 and 16. But in a thick layer, the succeeding points on the

Robertson chart locate close to each other as observed in Fig. 2 for points 20–26. That is why some previous studies used clustering methods for finding thick layers. However, if only the Robertson chart is used for stratification and SBT behavior determination, so many layers with different SBTs will be determined. To clarify, the data derived from Lukang

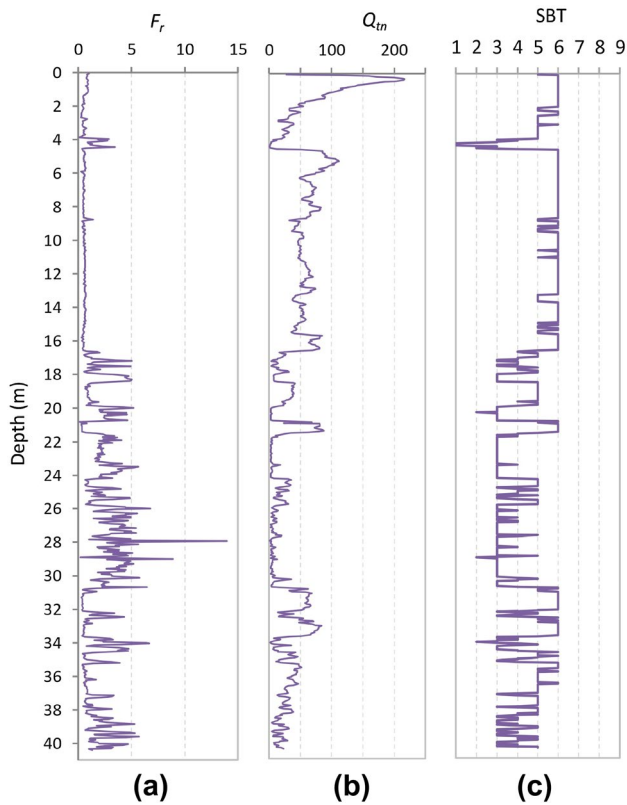


Fig. 4 F_r and Q_{tn} data sets of Lukang case, Taiwan, and the SBT determined for each recorded point in depth based on the Robertson chart [45]

case, Taiwan, is presented as F_r and Q_{tn} in Fig. 4a and b, and the corresponding stratification is presented in Fig. 4c. As can be observed, there are a lot of fluctuations in the SBT graph and too many layers with different SBTs have been identified. Such stratification causes problems in a geotechnical numerical analysis. Two main causes for these high fluctuations were identified as

- (a) there were a lot of fluctuations with different ranges in both F_r and Q_{tn} , and
- (b) two proximate points on Robertson chart may locate on different sides of a SBT boundary line.

To solve the former problem, the LOESS method was applied, and the N–H model integrated with GWO was proposed to solve the latter problem.

The LOESS method was used to reduce the impacts of outliers and uncertainties in the measurements—or in other words, to denoise the measurements signals. The local bandwidth of local regression may depend on the uncertainties originating from the CPT equipment, soil type, operator’s experience, etc.—which has been out of the scope of the present study and needs more detailed study. Therefore, as can

be observed in Fig. 3, the bandwidth of the LOESS method is not proposed to be optimized with the GWO model. The LOESS method is applied to both F_r and Q_{tn} , individually.

The proximity of the sequential data points on the Robertson chart has been considered as the basic for thick layer identification. In other words, the sequentially distant point(s) were considered as the transition layer(s) between the thick ones. To implement this concept computationally, a distance criterion between the sequential data points and also a distance limit (denoted hereafter as $D_{\alpha-cut}$) ought to be determined for recognizing the layer change. Provided that only the distance between two adjoining measurements is considered, the outliers and some uncertain measurements may bring about problems with stratification. Hence, as an example, for points 1–26 shown in Fig. 2 (which were measured at 3.65–4.9 m depth in Lukang test site), the distances between each point, i , and the $i - 1$, $i - 2$, $i - 4$ and $i - 10$ points are computed and shown in Fig. 5a–d as $d_{N-H(i,i-1)}$, $d_{N-H(i,i-2)}$, $d_{N-H(i,i-4)}$ and $d_{N-H(i,i-10)}$. The distances between the points i and k are computed as

$$d_{N-H(i,i-k)} = ((\log_{10} F_{r(i)} - \log_{10} F_{r(i-k)})^2 + ((\log_{10} Q_{tn(i)} - \log_{10} Q_{tn(i-k)})^2)^{0.5}, \tag{13}$$

where $F_{r(i)}$, $F_{r(i-k)}$, $Q_{tn(i)}$ and $Q_{tn(i-k)}$ represent the normalized friction ratio of points i and $i - k$, respectively.

In Fig. 5a, if the layer change distance limit, $D_{\alpha-cut}$, is assumed equal to 0.6, two transition layers can be recognized from the depth ranges 3.85–3.95 and also 4.50–4.55 m, where $d_{N-H(i,i-1)} > D_{\alpha-cut}$. Figure 5c and d shows that the depth of these two transition layers changes a bit if $d_{N-H(i,i-2)}$ and $d_{N-H(i,i-4)}$ will be used. In addition, Fig. 5d shows that the transition layers have moved about 50 cm deeper and also shape of $d_{N-H(i,i-10)}$ is changed totally in contrast to Fig. 5a–c. Since there is a conflict among the four distance criteria to find the layer change depths, the N–H bargaining model is used as a conflict-resolution model to determine the proportion of cooperation (p_i in Eq. 5) of the four considered distances as game players. In an initial example, following Eq. (5), a parameter named as the N–H total distance, $D_N - H$, can be computed as $D_N - H = \prod_{k=1}^n (d_{N-H(i,i-k)})^{p_k}$, where, n is the number of distance criteria, or in other words, the players number in the N–H model.

Assuming the power quantities, i.e. proportions of cooperation of 0.8, 0.6, 0.4 and 0.2, for the distance criteria, $[d_{N-H(i,i-1)}]^{0.8} \times [d_{N-H(i,i-2)}]^{0.6} \times [d_{N-H(i,i-4)}]^{0.4} \times [d_{N-H(i,i-10)}]^{0.2}$ is computed for the measurement points 1–26 and illustrated in Fig. 5e. Changing the proportion of cooperation, i.e. power p_i in Eq. (5), and $D_{\alpha-cut}$, different transition layers and the thick layers in between can be recognized.

The compromise among the players (Ω in Eq. 5), i.e. the maximum total profit of the n players, which means the best stratification, can be reached if the $n + 1$ variables of

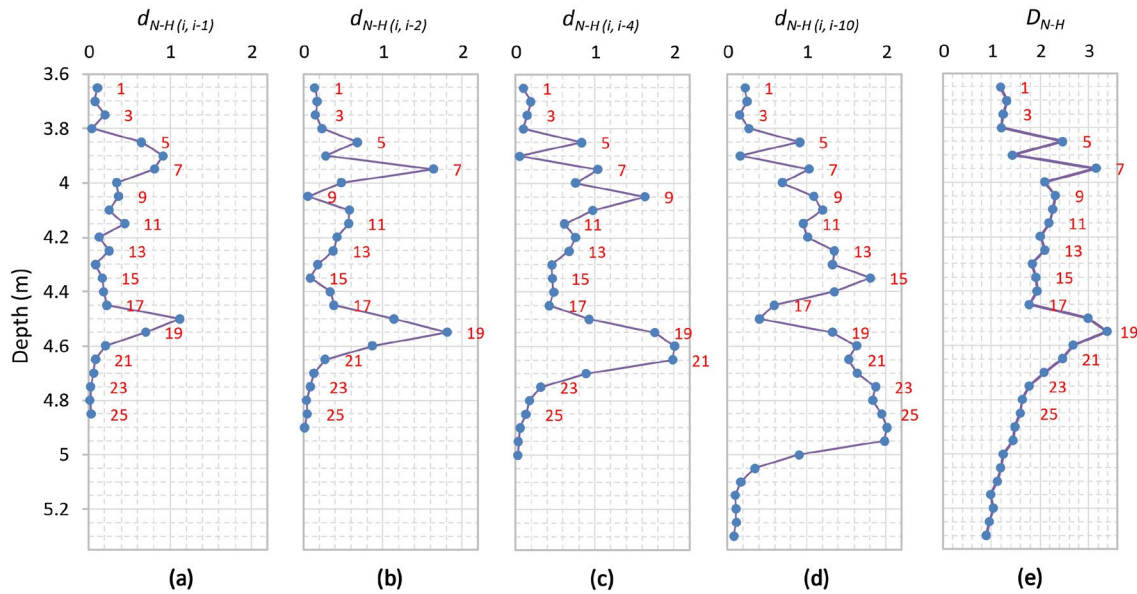


Fig. 5 The four players considered for the initial clarification example in the N–H bargaining model as four distance criteria (a–d, represented by $d_{N-H(i,i-1)}$, $d_{N-H(i,i-2)}$, $d_{N-H(i,i-4)}$ and $d_{N-H(i,i-10)}$, respec-

tively), and their total profit in the N–H bargaining, represented by D_{N-H} in e. A section of measurements from the depth range 3.6–5.3 m from Lukang site, Taiwan, is presented here

the model (p_1, p_2, p_3, p_4 in Eq. 5, and also $D_{\alpha-cut}$) can be computed based on some observed reliable stratification results approved by experts/soil sampling. In other words, the N–H model should be trained using some optimization model. Due to its high potential for solving various optimization problems, GWO is used as the optimization model. Indeed, the bargaining process in the N–H bargaining model is executed through the GWO model. For finding the optimum quantities of the five variables, random quantities are considered as the initial population of wolves and in different iterations, as explained in the previous section. Then they are modified to find the optimum quantities of variables by minimizing the mean square error (MSE) calculated by comparing the model SBT estimations and observed reliable stratification SBTs. In computations, SBTs are used as the numbers proposed by Robertson [45].

There are some notes worth mentioning:

- (i) After finding the transition and thick layers, the SBT of the layers are considered as the average of the SBTs of points of each layer on the Robertson chart.
- (ii) The averaging bandwidth and weight function of the LOESS method can be also considered as variables in the optimization process.
- (iii) Although only four distance criteria are considered in this study, a sensitivity analysis through optimization can be performed to find the best distance criteria.
- (iv) The focus and also the innovation of the present study is the proposed model itself as a basic tool for strati-

fication. It can be modified to include the impacts of more geotechnical uncertainty origins, such as the impact of the interlayer weak/stiff soils on CPT measurements.

3 Applying the developed model to case studies

As an example to show the applicability of the developed model, only the CPT data from Lukang case study was utilized for training the model. Then the trained model was applied to three other test data from NGES, Yuanlin, and WuFeng case studies. For the Lukang test site, the stratification profile approved by the CPT experts was available. The profile was in accordance with the WTMM stratification approach proposed by Ching et al. [13]. Another benefit of using the Lukang CPT stratification for training the model was that it contained both thick and thin layers and also low and highly fluctuated F_r and Q_{ln} data sections.

It should be mentioned that the problem and data provided in this paper were the question of International Society for Soil Mechanics and Geotechnical Engineering (ISSMGE), TC304, student contest held in conjunction with the annual European Safety and Reliability conference (ESREL), Hannover, Germany, in 2019.

The role of the submodels and modules in the developed model are explained in the following sections.

3.1 Data denoising module

The smoothing LOESS method with the local regression bandwidth of 1% of the whole measured signal was applied to both F_r and Q_{tn} data. This bandwidth was selected visually in this example such that the measured signals of F_r and Q_{tn} would not change a lot but just some noises were decreased. Albeit, a preliminary sensitivity analysis on the impact of the bandwidth on the stratification results was performed also. For clarification, a section of the smoothed F_r and Q_{tn} in the depth range of 20–25 m is illustrated in Fig. 6. As can be seen, the whole data and the trends have not been changed considerably after smoothing. But the outliers considered as the sudden sharp flocculating points have been approximated closer to the neighbouring points. It can be seen specifically in Fig. 6a at the depth range of approximately 21.5–22 m. An example of removing sudden sharp fluctuations can be observed in Fig. 6b at the depth of 20.9 m approximately. In conclusion, the denoising module has smoothed the input signals and removed the probable outliers, but the measurement signals are not changed drastically.

3.2 Soil stratification module

Similar to the initial example in Sect. 2, the same four distance criteria were considered as the N–H players. Bargaining among them was modeled with the GWO based

on the training data from Lukang site and a compromise was reached. The optimum cooperation proportion of each player, p_i , and $D_{\alpha-cut}$ were 0.7865, 0.5046, 0.3255, 0.1691 and 0.0875, respectively. The lowest mean square error (MSE) and highest correlation coefficient (R^2) for the optimum results were 0.3403 and 0.9149, respectively. Figure 7 compares the optimum estimated stratification with the training stratification. Interestingly, the developed model has identified more (thin) layers based on the implemented distance criteria, although it was trained based on the compared stratification profile. However, the model’s sensitivity regarding the global optimum proportion of cooperation of players, i.e. globally optimized p_i in Eq. (5), in the N–H submodel has not been in the scope of this study and may be investigated in future.

3.3 LOESS regression bandwidth’s impact

The impact of either applying or not applying the LOESS method with the spatial regression bandwidth of 1% on the optimum stratification profile is compared in Fig. 8. As can be observed, without applying the LOESS method, more thin layers were identified in the stratification profile, especially where more fluctuations exist in the input data (shown in Fig. 2) and the SBT data points on the Robertson chart are located close to the boundary lines. Hence, application of the LOESS submodel affects the final stratification results.

It is believed that the LOESS spatial regression bandwidth may depend on some factors like the precision of penetrometer, cone penetration rate, SBT of each soil layer, etc., which may be studied in future.

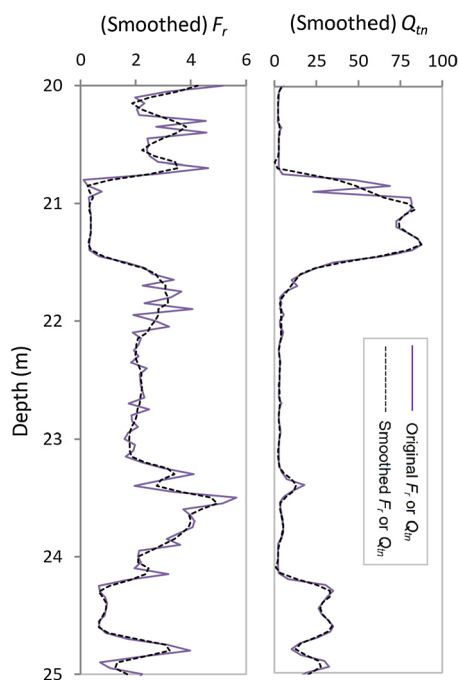
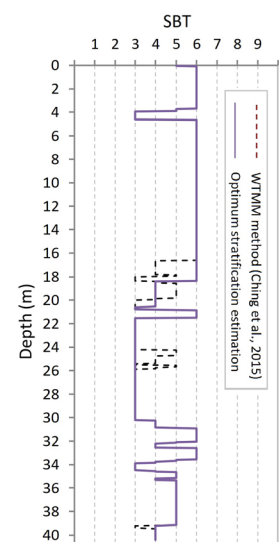


Fig. 6 Comparison of original and smoothed F_r and Q_{tn} in depth (magnified view for the depth range between 20 and 25 m) for Lukang case, Taiwan

Fig. 7 Comparison of stratification profiles by the proposed model and experts, i.e. WTMM method [13], for Lukang case, Taiwan



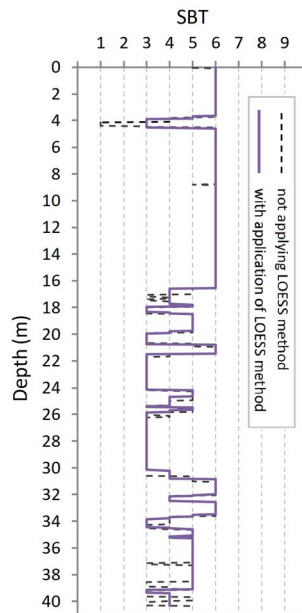


Fig. 8 Comparison of stratification results for both cases of applying and not applying the LOESS smoothing method to the Lukang case CPT measurements

3.4 $D_{\alpha\text{-cut}}$ as a regulating parameter

There are many parameters that may affect the required precision of subground stratification. Assume that a building is going to be constructed on soft soil layers and the geotechnical designer needs to evaluate the settlement. The building may be either a private villa to be used only on weekends or the main hospital of a large city. The engineer might consider the hospital with a higher importance factor in contrast to the private villa, because settlement may cause malfunctioning of the hospital instruments and affect a considerably higher number of peoples' lives. Therefore, it will be logical to have a more precise soil stratification profile for the hospital to perform the geotechnical analysis. In addition, some other factors such as the background/credit of the consultant or contractor company and nonhomogeneity of soil can influence the required precision of detected subground layers and their SBTs. In this regard, the developed model includes a precision regulating parameter, $D_{\alpha\text{-cut}}$, that can be used for altering the CPT-based subground stratification based on the geotechnical engineer designer's mindset.

The normalized $D_N - H$ -depth graph for Lukang case study is presented in Fig. 9. The layer change N-H distance limit, $D_{\alpha\text{-cut}}$, line is also illustrated in the figure. Figure 9 shows that lower $D_{\alpha\text{-cut}}$ usually identifies more transition layers and consequently, more layers in the stratification profile of a CPT log and vice versa.

To discuss about different quantities of $D_{\alpha\text{-cut}}$, a multiplier, λ , of the optimum quantity of $D_{\alpha\text{-cut}}$ is defined as

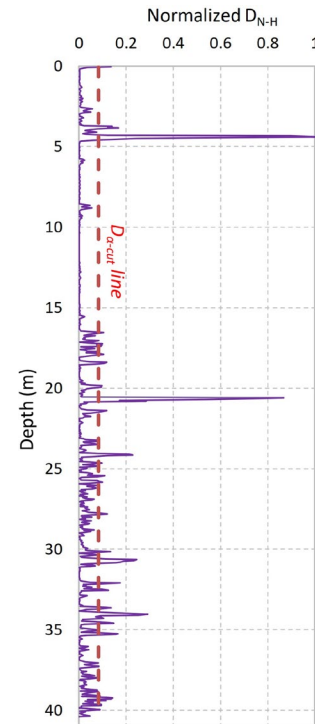
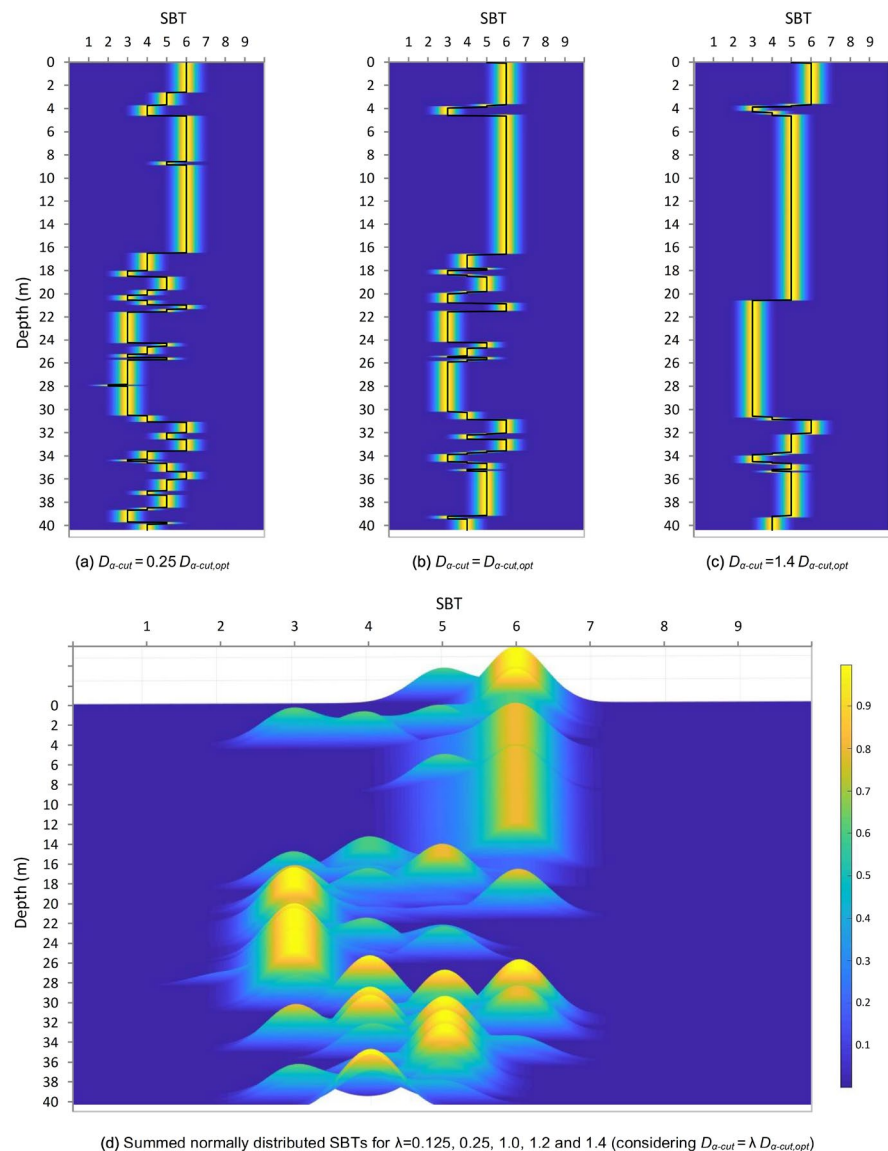


Fig. 9 Normalized $D_N - H$ as the total profit of players in Nash-Harsanyi bargaining model and the illustration of α -cut distance, $D_{\alpha\text{-cut}}$ for Lukang case, Taiwan

$\lambda \times D_{\alpha\text{-cut, opt}}$. Figure 10a–c shows the identified stratification profile for Lukang case for the three quantities of $\lambda = 0.25, 1.0$ and 1.4 . The SBT stratification contours are illustrated in these figures. Gaussian distribution was used to draw contours for each SBT number. It can be observed that the SBT graph for $\lambda = 0.25$ and 1.0 are almost similar. However, higher number of layers are identified for the depth range between 34 and 40 m for $\lambda = 0.25$. Comparing $\lambda = 0.25$ and 1.4 graphs reveals that in addition to recognizing lower number of layers, the $\lambda = 1.4$ graph shows some contrasts with the other two graphs in Fig. 10a and b: the SBT for the approximate depth range of 5–17 m has been 5 while it has been 6 for the same depth according to the two other graphs. To overcome this problem, the stratification results for different λ quantities can be summed.

Simply, in this example, the five stratification profiles for $\lambda = 0.125, 0.25, 1.0, 1.2$ and 1.4 were summed. Just to have a better view of the derived profiles, a Gaussian distribution was assumed for each SBT number at each profile. The SBT number at any depth was considered as the mean of the distribution function. The three-dimensional (3D) probability-based colored presentations of stratification profiles can help the engineer to select the appropriate numbers and depths of layers and their corresponding SBTs as well. Figure 10d illustrates the 3D summation of the stratification profiles for the five quantities of λ . A glance over this figure may induce

Fig. 10 Regulating parameter impact on the precision of sub-ground stratification estimated for Lukang case, Taiwan



a general insight for an engineer that there are generally three layers of soil within the approximate depth ranges of 0–16, 16–30.5 and 30.5–40 m with some transition layers/lenses in between. However, if the graph is scanned more precisely, more layers may be considered in the subground numerical analyses.

3.5 Comparing the proposed and other CPT-based stratification models for different case studies

In this example, the developed model has been compared with two different approaches for determining subground stratigraphy based on CPT data. The first approach identifies layers with respect to the consistency among the data in each layer, for instance, the Bayesian model proposed by Wang et al. [63]. The second approach identifies layers searching for their boundaries. Such methods may be considered as

change point detection (CPD) methods [3]. An example for this method can be the WTMM method proposed by Ching et al. [13].

Figure 11a–d shows the comparison of the developed model with the WTMM [13] and Bayesian [63] models for four data sets of Lukang, NGES, Yuanlin, and WuFeng cases. Generally, the identified layers boundaries and SBTs have been almost the same for the three models. However, the developed model identified more thin layers than the other methods. Aminikhanghahi and Cook [3] performed a survey on methods for time series change point detection and finally stated that “Evaluating the significance of the detected change point is an important open issue for unsupervised methods. Currently, most existing methods compare detecting changes scores with a threshold value to determine whether change occurs or not. Selecting the optimal threshold value is difficult. The values may be

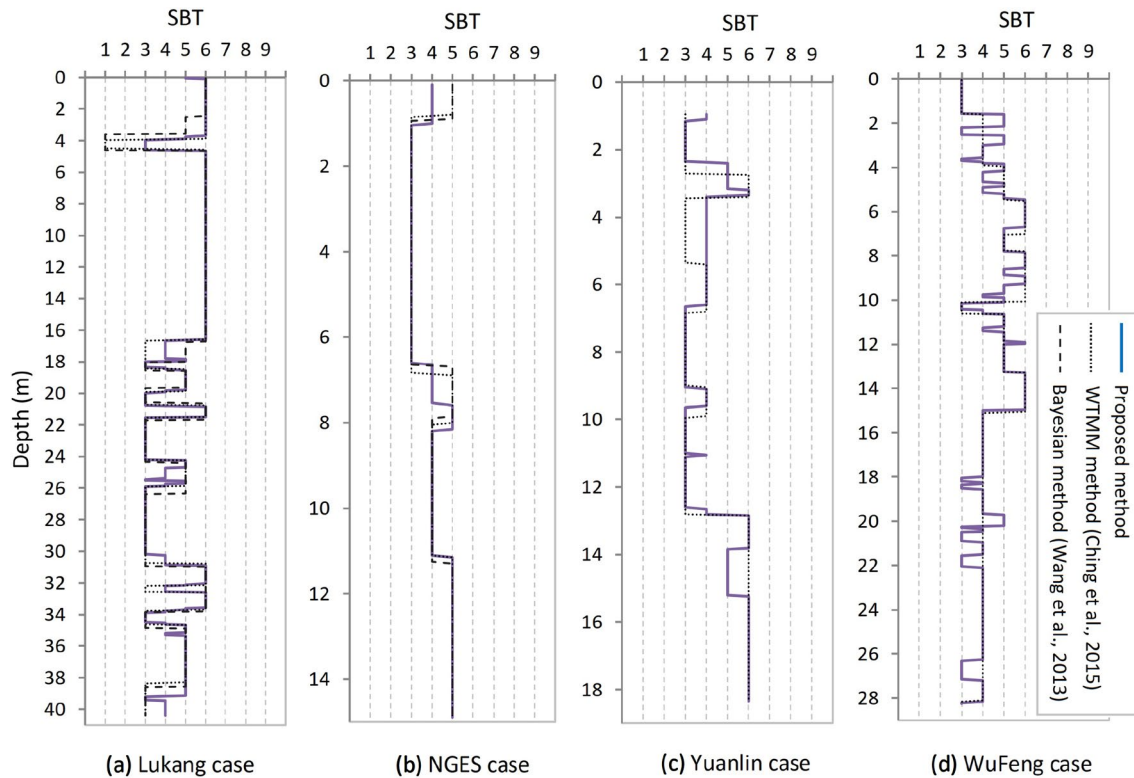


Fig. 11 Comparison of the proposed model with two other CPT-based stratification models for different CPT results

application dependent and they may change over time. Developing statistical method to find significant change point based on previous values may offer greater autonomy and reliability”. This was the problem that Ching et al. [13] dealt with through a probabilistic study on 50 expert-based stratifications deciding on the threshold to introduce some data variations as subground layer changes. Therefore, in this study, the variable parameters, i.e. p_i and $D_{\alpha\text{-cut}}$, were optimized based on their study results. At the meantime, the underlying N–H distance criteria in the developed model brought the capability to find the transition layers among the thick layers. This can be regarded as an advantage of the developed model compared to the two other models.

It is interesting that the trained model in this example, which was trained only based on one data set, estimated the stratification profiles generally similar to the Bayesian and WTMM methods for NGES, Yuanlin, and WuFeng sites, as the testing data sets. It shows that the developed model works as a semi-supervised model. Another advantage of the developed model was the very fast computation time, specially compared to the Bayesian method.

The three- and two-dimensional contour plots of the stratification profiles for the three testing cases, are illustrated in Fig. 12a–c. For NGES case, Fig. 12b showed that WTMM and Bayesian models suggested 5 layers while the proposed model estimated 6 layers. However, a glance over Fig. 12b

may inspire an engineer to define 4 or even 3 layers in numerical analysis for NGES site. On the other hand, looking deeper into these figures, more layers can be identified.

Figure 12d shows a 3D presentation of the Yuanlin case stratification estimate. It is illustrated here as an example just to show how an engineer can observe the results. Rotating this figure and seeing different layers with different probabilities of SBT helps the engineer to decide easier on the details of the identified layers and their corresponding SBTs. For future studies, the impact of variations of some parameter, like I_c , defined by Robertson and Wride [50], may provide the engineer a more precise probabilistic 3D view of SBTs.

As final words, it should be mentioned that the sensitivity analysis on some factors, like the local regression bandwidth of LOESS method, different combinations of $d_N - H$ in the N–H model, and the regulating parameter $D_{\alpha\text{-cut}}$ requires further investigation.

4 Conclusions

The present study has concentrated on the development of a model for subground stratification and soil behaviour type (SBT) classification based on the cone penetration test (CPT) measurements. The proposed model consists of two denoising and soil stratification modules. In the first

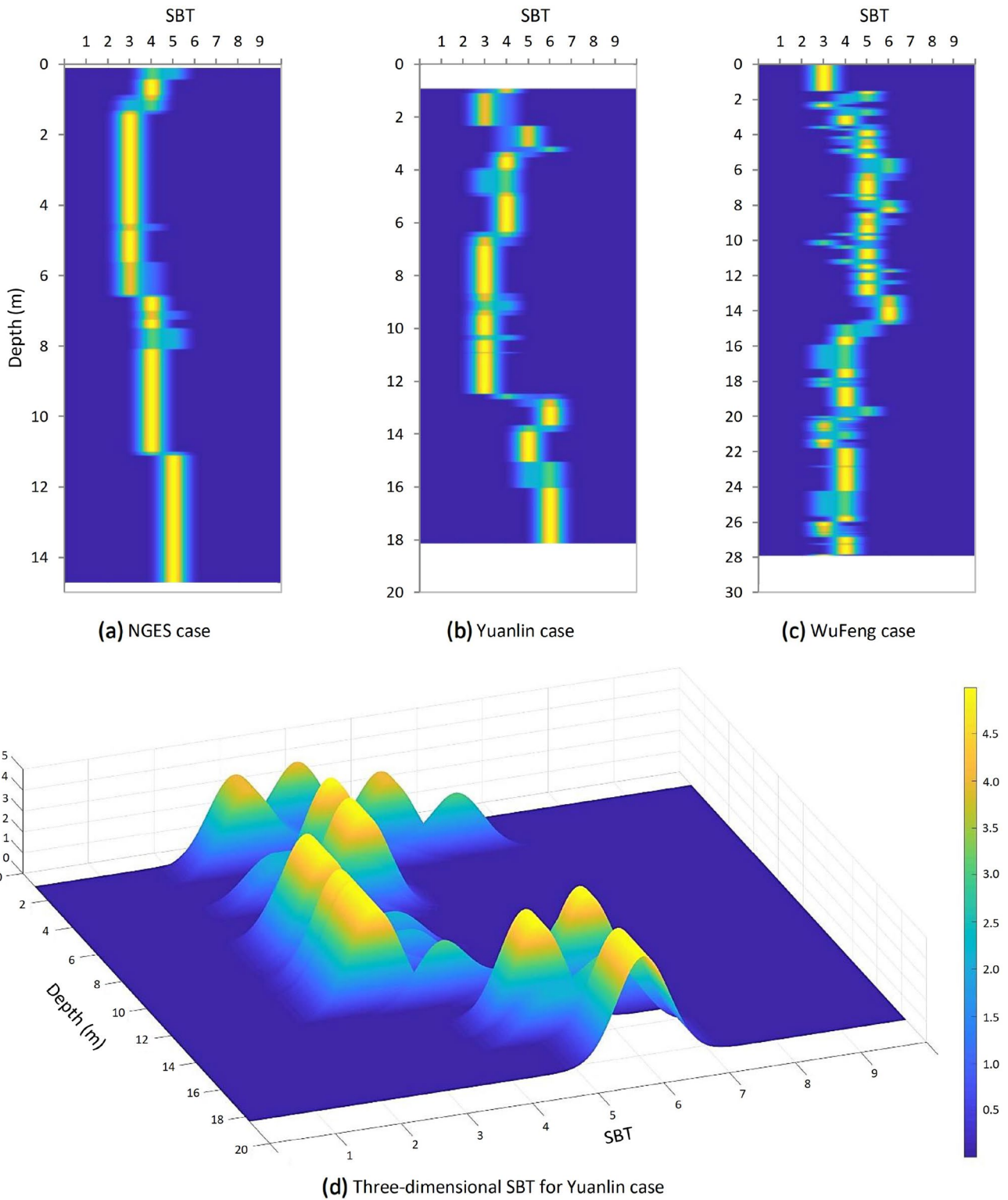


Fig. 12 Two- and three-dimensional plots of summed normally distributed SBTs for $\lambda = 0.125, 0.25, 1.0, 1.2$ and 1.4 (considering $D_{\alpha\text{-cut}} = \lambda \times D_{\alpha\text{-cut,opt}}$) for NGES, Yuanlin, and WuFeng cases

module, The CPT measurements are loaded and denoised by the locally estimated scatterplot smoothing (LOESS) method to approximate the probable outliers. As the main novelty of this study, in the second module, the SBT classification chart proposed by Robertson [45] is integrated with the Nash–Harsanyi (N–H) bargaining model to discover the subground stratification and the corresponding SBTs of the strata. The bargaining among the sequential distance criteria between the measured data points—considered as players in the N–H bargaining model—is simulated by the grey wolf optimizer (GWO) model to calculate the optimum proportion of cooperation of the N–H players. Generally, the transition layers are identified through the large variations in normalized friction ratio, F_r , and normalized cone tip resistance, Q_{tn} , locations on the Robertson chart. Inversely, the thick layers are identified through small variations of F_r and Q_{tn} on the chart. In an example, the practicality of the proposed model was verified. Only one CPT expert-based training stratification profile was utilized and for other three CPT measurements, the stratification profiles were estimated. They compared well with two other previously published models' estimations. The main advantages of the proposed model include the following:

1. It is a rapid stratification model, especially compared to the probabilistic stratification methods like the Bayesian inference models. Using a normal laptop, the computation time may take less than a minute.
2. Simple algebraic equations are solved in the proposed model. Therefore, the mathematics behind the model is not sophisticated.
3. In addition to the thick soil layers, the transition layers can also be identified using the proposed model.
4. It is not required that the engineer prescribes the number of layers before running the model. However, by adjusting the precision regulating parameter, $D_{\alpha\text{-cut}}$, they can control the number of strata and the corresponding SBTs regarding their own preferences.
5. Much data are not required for training the developed model.
6. Two- and three-dimensional presentations of subground stratification profiles provided from combining the stratification profiles with different $D_{\alpha\text{-cut}}$ support the engineering judgement about the details of subground layers.

After all, it should be noted that the proposed model requires further development although its application provided comparable stratification profiles with the other published methods.

Acknowledgements Professors K.K. Phoon, J. Ching, W. Puła, and G. Vessia are appreciated for holding ISSMGE TC304, student contest, held in ESREL conference, Hannover, 2019, and also for providing

CPT data. Besides, provision of the WTMM method's source code by J. Ching has been really appreciable.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Abraham B, Chuang A (1989) Outlier detection and time series modeling. *Technometrics* 31(2):241–248
2. Alshibli KA, Okeil AM, Alramahi B, Zhang Z (2009) Statistical assessment of repeatability of CPT measurements. In: International Foundation Congress and Equipment Expo 2009, Orlando, Florida, United States, pp 87–94
3. Aminikhanghahi S, Cook DJ (2017) A survey of methods for time series change point detection. *Knowl Inf Syst* 51(2):339–367
4. Begemann H (1965) The friction jacket cone as an aid in determining the soil profile. In: Proceedings of the 6th international conference on soil mechanics and foundation engineering (SMFE), Montreal, Canada, vol 2, pp 17–20
5. Bong T, Stuedlein AW (2018) Effect of cone penetration conditioning on random field model parameters and impact of spatial variability on liquefaction-induced differential settlements. *J Geotech Geoenviron Eng* 144(5):04018018
6. Campanella R, Robertson P, Gillespie D, Grieg J (1985) Recent developments in in-situ testing of soils. In: Proceedings of the 11th international conference on soil mechanics and foundation engineering (SMFE), San Francisco, United States, pp 849–854
7. Campanella RG, RG C, PK R et al (1982) Pore pressures during cone penetration testing. In: Proceedings of 2nd European symposium on penetration testing, ESOPT-2, Amsterdam, Netherlands, vol 2, pp 507–512
8. Cao ZJ, Zheng S, Li DQ, Phoon KK (2019) Bayesian identification of soil stratigraphy based on soil behaviour type index. *Can Geotech J* 56(4):570–586
9. Chang I, Tiao GC, Chen C (1988) Estimation of time series parameters in the presence of outliers. *Technometrics* 30(2):193–204
10. Chen H, Shen J (2018) Denoising of point cloud data for computer-aided design, engineering, and manufacturing. *Eng Comput* 34(3):523–541
11. Chen H, Zhang Q, Luo J, Xu Y, Zhang X (2020) An enhanced bacterial foraging optimization and its application for training kernel extreme learning machine. *Appl Soft Comput* 86:105884
12. Ching J, Phoon KK, Li KH, Weng MC (2019) Multivariate probability distribution for some intact rock properties. *Can Geotech J* 56(8):1080–1097

13. Ching J, Wang JS, Juang CH, Ku CS (2015) Cone penetration test (CPT)-based stratigraphic profiling using the wavelet transform modulus maxima method. *Can Geotech J* 52(12):1993–2007
14. Cleveland WS (1979) Robust locally weighted regression and smoothing scatterplots. *J Am Stat Assoc* 74(368):829–836
15. Cleveland WS, Devlin SJ (1988) Locally weighted regression: an approach to regression analysis by local fitting. *J Am Stat Assoc* 83(403):596–610
16. Cleveland WS, Loader C (1996) Smoothing by local regression: principles and methods. In: Härdle W., Schimek M.G. (eds) *Statistical theory and computational aspects of smoothing. Contributions to statistics*. Physica-Verlag HD. https://doi.org/10.1007/978-3-642-48425-4_2
17. Das SK, Basudhar PK (2009) Utilization of self-organizing map and fuzzy clustering for site characterization using piezocone data. *Comput Geotech* 36(1–2):241–248
18. Dinar A, Hogarth M (2015) Game theory and water resources critical review of its contributions, progress and remaining challenges. *Found Trends Microecon* 11(1–2):1–139
19. Douglas B (1981) Soil classification using electric cone penetrometer. American Society of Civil Engineers, ASCE, In: *Proceedings of the conference on cone penetration testing and experience*, St. Louis, United States, pp 209–227
20. D'Ignazio M, Phoon KK, Tan SA, Länsivaara TT (2016) Correlations for undrained shear strength of finnish soft clays. *Can Geotech J* 53(10):1628–1645
21. Eslami A, Alimirzaei M, Aflaki E, Molaabasi H (2017) Deltaic soil behavior classification using CPTU records—proposed approach and applied to fifty-four case histories. *Mar Georesour Geotechnol* 35(1):62–79
22. Eslami A, Fellenius BH (2004) Cpt and Cptu data for soil profile interpretation: review of methods and a proposed new approach. *Iran J Sci Technol Trans B Eng* 28(1):69–86
23. Farhadi S, Nikoo MR, Rakhshandehroo GR, Akhbari M, Alizadeh MR (2016) An agent-based-nash modeling framework for sustainable groundwater management: a case study. *Agric Water Manag* 177:348–358
24. Fu J, Zhong PA, Zhu F, Chen J, Wu Yn, Xu B (2018) Water resources allocation in transboundary river based on asymmetric nash–harsanyi leader–follower game model. *Water* 10(3):270
25. Hegazy YA, Mayne PW (2002) Objective site characterization using clustering of piezocone data. *J Geotech Geoenviron Eng* 128(12):986–996
26. Henderson R (1916) Note on graduation by adjusted average. *Trans Actuar Soc Am* 17:43–48
27. Himanshu N, Kumar V, Burman A, Maity D, Gordan B (2020) Grey wolf optimization approach for searching critical failure surface in soil slopes. *Eng Comput*. <https://doi.org/10.1007/s00366-019-00927-6>
28. Hird C, Springman SM (2006) Comparative performance of 5 cm² and 10 cm² piezocones in a lacustrine clay. *Geotechnique* 56(6):427–438
29. Jefferies M, Davies M (1991) Soil classification by the cone penetration test: discussion. *Can Geotech J* 28(1):173–176
30. Jung B, Gardoni P, Biscontin G (2007) Probabilistic soil classification based on cone penetration tests. In: *Proceedings of the 10th International Conference on applications of statistics and probability in civil engineering, ICASP10, Tokyo, Japan*, pp 451–452
31. Kamboj VK, Bath S, Dhillon J (2016) Solution of non-convex economic load dispatch problem using grey wolf optimizer. *Neural Comput Appl* 27(5):1301–1316
32. Kaveh M, Chayjan RA, Taghinezhad E, Gilandeh YA, Younesi A, Sharabiani VR (2019) Modeling of thermodynamic properties of carrot product using alo, gwo, and woa algorithms under multi-stage semi-industrial continuous belt dryer. *Eng Comput* 35(3):1045–1058
33. Khairuzzaman AKM, Chaudhury S (2017) Multilevel thresholding using grey wolf optimizer for image segmentation. *Expert Syst Appl* 86:64–76
34. Kimmel RK, Booth DE, Booth SE (2010) The analysis of outlying data points by robust locally weighted scatter plot smooth: a model for the identification of problem banks. *Int J Oper Res* 7(1):1–15
35. Liao T, Mayne P (2007) Stratigraphic delineation by three-dimensional clustering of piezocone data. *Georisk* 1(2):102–119
36. Ljung GM (1993) On outlier detection in time series. *J R Stat Soc Ser B (Methodological)* 55(2):559–567
37. Lunne T, Powell JJ, Robertson PK (1997) *Cone penetration testing in geotechnical practice*. CRC Press, Boca Raton
38. Macaulay FR et al (1931) *The smoothing of time series*. NBER Books, Delhi
39. Madani K (2011) Hydropower licensing and climate change: insights from cooperative game theory. *Adv Water Resour* 34(2):174–183
40. Mirjalili S, Mirjalili SM, Lewis A (2014) Grey wolf optimizer. *Adv Eng Softw* 69:46–61
41. Moayedi H, Nguyen H, Foong LK (2019) Nonlinear evolutionary swarm intelligence of grasshopper optimization algorithm and gray wolf optimization for weight adjustment of neural network. *Eng Comput*. <https://doi.org/10.1007/s00366-019-00882-2>
42. Nash J (1953) Two-person cooperative games. *Econom J* 21(1):128–140
43. Robertson P (2016) Cone penetration test (CPT)-based soil behaviour type (SBT) classification system—an update. *Can Geotech J* 53(12):1910–1927
44. Robertson P, Campanella R (1983) Interpretation of cone penetration tests. Part II: Clay. *Can Geotech J* 20(4):734–745
45. Robertson PK (1990) Soil classification using the cone penetration test. *Can Geotech J* 27(1):151–158
46. Robertson PK (2009) Interpretation of cone penetration tests—a unified approach. *Can Geotech J* 46(11):1337–1355
47. Robertson PK (2010) Soil behaviour type from the CPT: an update. In: *Proceedings of 2nd International symposium on cone penetration testing, CPT'10, Huntington Beach, California, USA*, pp 2–56
48. Robertson PK, Campanella R (1983) Interpretation of cone penetration tests. Part I: sand. *Can Geotech J* 20(4):718–733
49. Robertson PK, Campanella R, Gillespie D, Greig J (1986) Use of piezometer cone data. In: *Proceedings of the ASCE special conference In Situ '86: Use of in situ tests in geotechnical engineering*, ASCE, New York, United States, pp 1263–1280
50. Robertson PK, Wride C (1998) Evaluating cyclic liquefaction potential using the cone penetration test. *Can Geotech J* 35(3):442–459
51. Sanglerat G, Nhim T, Sejourne M, Andina R (1974) Direct soil classification by static penetrometer with special friction sleeve. In: *Proceedings of the first European symposium on penetration testing, ESOPT-1, Stockholm, Sweden*, pp 337–344
52. Schiaparelli GV (1867) Sul modo di ricavare la vera espressione delle leggi della natura dalle curve empiriche. *Il Nuovo Cimento* (1855–1868) 26(1):171–190
53. Schmertmann JH (1978) *Guidelines for cone penetration test: performance and design*. Tech. rep., United States. Federal Highway Administration
54. Selänpää J, Di Buò B, Haikola M, Länsivaara T, D'Ignazio M (2018) Evaluation of existing CPTU-based correlations for the undrained shear strength of soft Finnish clays. In: *Cone penetration testing IV: Proceedings of the 4th International Symposium on Cone Penetration Testing (CPT 2018), Delft, Netherlands, CRC Press*: pp 185–191

55. Shen L, Chen H, Yu Z, Kang W, Zhang B, Li H, Yang B, Liu D (2016) Evolving support vector machines using fruit fly optimization for medical data classification. *Knowl Based Syst* 96:61–75
56. Titi HH, Tumay MT (2000) Cone penetration tests at the nges-texas A&M University: clay site. In: *Geotechnical special publication GSP 93, National geotechnical experimentation sites*, ASCE, pp 186–205
57. Tsay RS (1986) Time series model specification in the presence of outliers. *J Am Stat Assoc* 81(393):132–141
58. Vos J (1982) The practical use of CPT in soil profiling. In: *Proceedings of 2nd European symposium on penetration testing, ESOPT-2, Amsterdam, Netherlands, vol 2*, pp 933–939
59. Wang H, Wang X, Wellmann JF, Liang RY (2019) A Bayesian unsupervised learning approach for identifying soil stratification using cone penetration data. *Can Geotech J* 56(8):1184–1205
60. Wang M, Chen H (2020) Chaotic multi-swarm whale optimizer boosted support vector machine for medical diagnosis. *Appl Soft Comput* 88:105946
61. Wang M, Chen H, Yang B, Zhao X, Hu L, Cai Z, Huang H, Tong C (2017) Toward an optimal kernel extreme learning machine using a chaotic moth-flame optimization strategy with applications in medical diagnoses. *Neurocomputing* 267:69–84
62. Wang X, Wang H, Liang RY, Liu Y (2019) A semi-supervised clustering-based approach for stratification identification using borehole and cone penetration test data. *Eng Geol* 248:102–116
63. Wang Y, Huang K, Cao Z (2013) Probabilistic identification of underground soil stratification using cone penetration tests. *Can Geotech J* 50(7):766–776
64. Woollard M, Storteboom O, Coto Loria M, Vargas Herrera L (2015) Additional parameters measured in a single CPT using click-on modules. In: *Proceedings of geotechnical engineering for infrastructure and development: XVI European conference on soil mechanics and geotechnical engineering, ECSMGE, Edinburgh, Scotland*, pp 3783–3788
65. Xu Y, Chen H, Luo J, Zhang Q, Jiao S, Zhang X (2019) Enhanced moth-flame optimizer with mutation strategy for global optimization. *Inf Sci* 492:181–203
66. Zhang Z, Tumay MT (1999) Statistical to fuzzy approach toward CPT soil classification. *J Geotech Geoenviron Eng* 125(3):179–186
67. Zhao X, Zhang X, Cai Z, Tian X, Wang X, Huang Y, Chen H, Hu L (2019) Chaos enhanced grey wolf optimization wrapped elm for diagnosis of paraquat-poisoned patients. *Comput Biol Chem* 78:481–490

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.