Research Article   **Open Access**

Zb. Młynarek, J. Wierzbicki*, W. Wołyński

Use of functional cluster analysis of CPTU data for assessment of a subsoil rigidity

Abstract: This paper shows an example of the grouping of piezocone penetration test (CPTU) characteristics using functional data analysis, together with the results of clustering, in the form of a subsoil rigidity model. The subsoil rigidity model was constructed based on layer separation using the proposed method, as well as the *k*-means method. In the construction of the subsoil rigidity model, the constrained modulus M was applied. These moduli were determined from empirical relationships for overconsolidated and normally consolidated soils from Poland based on cone tip resistance.

Keywords: CPTU; cluster analysis; functional data; rigidity model.

1 Introduction

When designing foundations for wind farms, a decisive role is played by the regime of subsoil deformations. This is so because foundations for such structures are loaded by considerable torque. In order to ensure foundation stability, despite the relatively low stresses found in the subsoil, a key role is played by the magnitude and the lack of uniformity of the foundation settlement. Advanced computational methods such as the finite element method (FEM), which may be applied in the calculation of settlement and non-uniformity of the foundation slab settlement in a wind farm, are effective if a model of subsoil rigidity is constructed. The model is most frequently constructed based on constrained moduli of soils. This

model should show continuous and spatial changes in the constrained modulus of the subsoil under the entire foundation. The piezocone penetration test (CPTU), which is frequently supported by the flat dilatometer test (DMT), proves to be highly advantageous for the construction of such a model. The CPTU may also be used to determine the best location for the wind farm in a specific area. The best location is understood as such a zone in the area where the subsoil has the greatest rigidity (stiffness).

The construction of a model of subsoil rigidity, in which zones with homogeneous values of the constrained modulus are isolated using isolines, is based on the grouping of statistically non-different penetration characteristics recorded using the CPTU. An identical rule is binding for the separation of homogeneous areas in terms of rigidity of soil layers.

These two tasks may be solved using statistical methods of data clustering (e.g. Młynarek and Lunne [1] and Młynarek et al. [2]). Until now, data from CPTU were grouped as discrete values. This article presents a new concept for grouping of penetration characteristics as linear functions, which represent penetration characteristics. Such grouped data constitute the basis for the construction of rigidity models for the dimensioning of foundations for wind farms in the area covered by this study.

2 Geological characteristics of the tested area

Studies were conducted in the northern part of Poland in an area with a characteristic, young glacial geological structure (Figure 1). In an area of several hundred square kilometres, the subsoil to a depth of almost 200 m is composed of alternating beds of boulder clays from successive Pleistocene glaciations, separated by non-cohesive inter-glacial deposits.

Nine testing sites selected for analyses were arranged relatively uniformly over an area of approximately 1 km². Normally consolidated or slightly overconsolidated clays

*Corresponding author: J. Wierzbicki, Institute of Geology, Adam Mickiewicz University in Poznań, ul. Bogumiła Krygowskiego 12, 61-680 Poznań, Poland, E-mail: jwi@amu.edu.pl
Zb. Młynarek: Poznań University of Life Sciences, ul. Wojska Polskiego 28, 60-637 Poznań, Poland
W. Wołyński: Faculty of Mathematics and Computer Science, Adam Mickiewicz University in Poznań, ul. Umultowska 87, 61-614 Poznań, Poland

of the last stage of the youngest Wisła (Vistulian) glaciation with variable thicknesses are found within the depth of analyses, i.e. 16 m from the ground surface. These deposits are transformed into more consolidated boulder clays of the older stage of the Vistulian glaciation, occasionally separated from younger clays with sandy inter-beddings. In turn, deposits of the last glaciation are lying on strongly overconsolidated inter-glacial sands, constituting very good building subsoils but, unfortunately, also forming a barrier for static penetrations. Figure 2 presents the characteristic geotechnical and CPTU profiles for the analysed area.

Figure 1: Location of the test site in Poland.

3 Theoretical background

3.1 The concept for construction of a subsoil rigidity model

When calculating the volume and non-uniformity of the foundation slab settlement of a wind farm, in a vast majority of cases, secant constrained modulus is applied, which corresponds to the oedometric modulus. Secant moduli and their changes in subsoil with changes in geostatic stresses are determined based on CPTU penetration characteristics using separate formulas for soils with so-called drained conditions (coarse-grained soils) and for undrained (fine-grained soils, e.g. clays) and intermediate soils [3-5]. The first step in the construction of a subsoil rigidity model involves the isolation of statistically homogeneous layers in the context of drainage conditions. A convenient method to perform a preliminary qualification of homogeneous groups of soils is provided by the classification system proposed by Robertson [5].

The soil type was also verified by results of the grain size composition of samples, collected from laboratory analyses. The separation of soil layers in the first stage, according to the criterion of drainage conditions, was based on two parameters of the penetration characteristics, i.e. q_t – cone resistance, f_s – sleeve friction and geological data, which enabled the calculations of Q_{t1}, F_r and I_c coefficients, according to Robertson [5].

Discrete values for the three parameters specified herein were grouped using the cluster method [2]. The results of layer grouping and separation are discussed in Section 3. In the second stage of construction of the subsoil rigidity model, the values of the constrained modulus M were calculated using the following formulas.
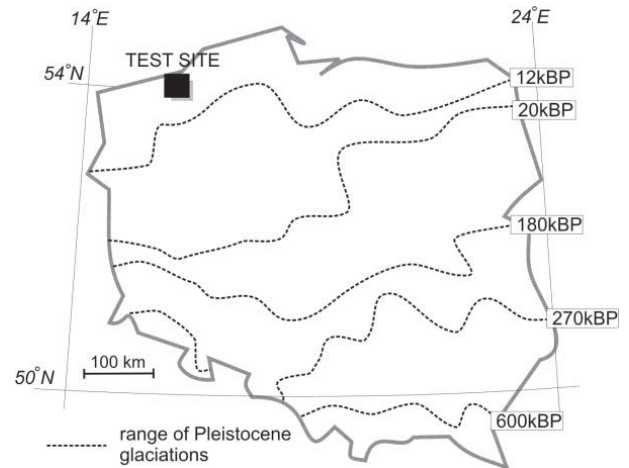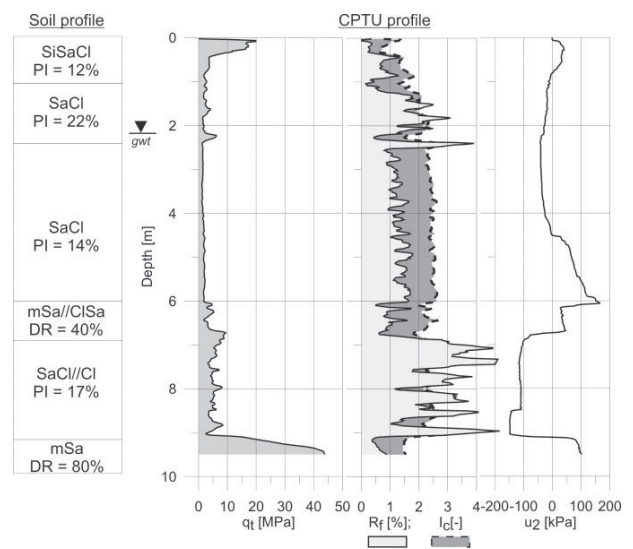
For soils with drained conditions [6],

Figure 2: Typical soil and CPTU profiles of the test site.

$$M = a \cdot q_c \qquad (1)$$

where a depends on the cone resistance.

For soils with undrained conditions [7],

$$M = 8.25(q_t - s_{v0}). \qquad (2)$$

A change in these moduli with a change in the vertical component of geostatic stress for each CPTU test was transformed to a curve, which was analogous to the characteristic $q_t = f(s_{vo}) \gg f(z)$, where z is the depth.

Curves for $M = f(z)$ were grouped using the functional data analysis method, as presented in Figure 3. The theoretical foundations for this method are given in the
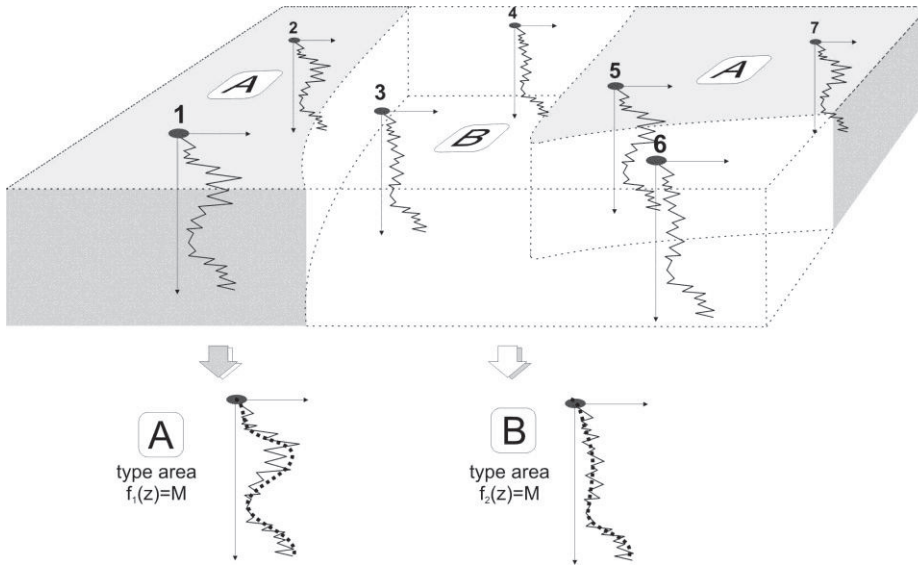
Figure 3: The division of a subsoil sample according to the concept of functional data analysis.

following section. The primary advantage of this method is the fact that in comparison to the clustering of discrete values of the constrained moduli, the curve accurately reflects the changes in rigidity in the isolated subsoil layer, these changes being connected not only with a change in the values of the stresses s_{v0} but also with the grain size composition, parameters of the state (such as liquidity index and density ratio) found in sediments of the subsoil and the overconsolidation ratio (OCR). The final effect of the construction of the model of rigidity was to isolate spatially homogeneous zones for values of the constrained modulus, which were separated by isolines with a constant value of the modulus. At this stage, the inverse distance weight (IDW) method was applied when constructing this model [2].

3.2 Assumptions for the algorithm using functional data analysis

In view of the indirect relationship of the values measured in the CPTU test (e.g. cone resistance) to the constrained modulus established in the subsoil, first of all, for the constructed algorithm, the value of modulus M as the function of depth has to be determined at each testing site (CPTU profile). This stage is based on a procedure for the identification of homogeneous geotechnical layers using cluster analysis, presented by Młynarek et al. [2]. Thus, Eqs. (1) and (2) are applied to determine the discrete values of moduli in individual layers based on cone resistance.

Such prepared data are then subjected to functional data analysis, in which it is assumed that each of the 'i' objects (assumed to be each CPTU) is described using a curve $M_i(z)$, $z\hat{I}$, $i = 1,2,...,N$ [8]. In the analysed case, z denotes depth, while M denotes the constrained deformation modulus.

In the determination of the scale of similarity between the analysed objects in the first step of functional data analysis, the following measure of distance between them is assumed:

$$d(M_i, M_j) = \sqrt{\int_Z (M_i(z) - M_j(z))^2 dz}. \qquad (3)$$

It is a distance in space $L_2(Z)$ (the space of square-integrable functions). In practice, data are observed at discrete points $z_1,...... z_t$, usually with some noise. In order to obtain the functional characteristics of data, smoothing or interpolation methods are used, considering that the true curve belongs to a finite dimensional space spanned by some basis functions.

Assume that each curve $M_i(z)$ ($i = 1,2,...,N$) has the following representation:

$$M_i(z) = \sum_{j=1}^{K} c_{ij}\phi_j(z) \qquad (4)$$

where $\{f_j\}$ is the basis function.

The most popular basis systems are as follows:
- the Fourier basis system (recommended for periodic functions),
- the splines basis system (recommended for non-periodic functions).

In this study, the splines basis system was used in the calculations. The coefficients c_{ij} are estimated by the least-squares method.

For each i, let $c_i = (c_{i1},...,c_{ik})'$, and

$$M_i = \begin{bmatrix} M_i(z_1) \\ M_i(z_2) \\ \vdots \\ M_i(z_L) \end{bmatrix}, \quad \Phi = \begin{bmatrix} \phi_1(z_1) & \cdots & \phi_K(z_1) \\ \phi_1(z_2) & \cdots & \phi_K(z_2) \\ \vdots & \cdots & \vdots \\ \phi_1(z_L) & \cdots & \phi_K(z_L) \end{bmatrix} \quad (5)$$

Then, we obtain the following expression:

$$\hat{c}_i = (\Phi'\Phi)^{-1}\Phi'M_i \quad (6)$$

The degree of smoothness of the function $M_i(z)$ depends on the values of K (small values of K cause more smoothing of the curves). The popular criteria for the optimum K selection are the following:
- the Bayesian information criterion (BIC) – applied in this analysis;
- the Akaike information criterion (AIC); and
- generalised cross-validation (GCV).

In the case of the N function $M_i(z)$, one common value of K is chosen as a modal value of the numbers $K_1, K_2,...,K_N$. After the smoothing processes, we have the following:

$$M_i(z) = \Phi(z)\hat{c}_i \quad (7)$$

where

$$\Phi(z) = [\phi_1(z),\ldots,\phi_K(z)] \quad (8)$$

Then, we obtain the following expression:

$$\begin{aligned} d(M_i, M_j) &= \sqrt{\int_Z (\Phi(z)\hat{c}_i - \Phi(z)\hat{c}_j)^2 dz} \\ &= \sqrt{(\hat{c}_i - \hat{c}_j)'\int_Z \Phi'(z)\Phi(z)dz(\hat{c}_i - \hat{c}_j)} \end{aligned} \quad (9)$$

Distances between the objects are calculated from Eq. (9). After they have been established, the final stage is initiated in the construction of the data analysis algorithm, again based on the conventional method of hierarchical cluster analysis [2]. As a result, we obtain the hierarchy of object similarity, which may be presented most comprehensibly in the form of a dendrogram.

4 Preparation of data for the functional data analysis

In accordance with assumptions presented in Section 3, in the first stage of the analysis, a preliminary division of each analysed CPTU profile into homogeneous layers was conducted in terms of soil behaviour. For this purpose, classification systems were used [5]. The analysis of similarity in terms of soil properties was performed based on two CPTU parameters q_t and R_f, as well as the soil behaviour index I_c [5]. Data collected from nine CPTUs were analysed in accordance with the assumptions of cluster analysis by the k-means method [2]. When assessing the effectiveness of separations, the applied procedure was conducted using both the statistical criteria according to Caliński-Harabasz [9], as in Eq. (10), and the criteria minimising the mean weighted coefficient of variation in the layer, proposed by Wierzbicki [10].

$$VRC(K) = \frac{tr(B(C_K))/(K-1)}{tr(W(C_K))/(n-K)} \quad (10)$$

where $tr(B(C_K))$ is the trace of the inter-cluster matrix, $tr(W(C_K))$ is the trace of the intra-cluster matrix; K is the number of clusters in a given step of analysis; and n is the number of data adopted in analysis.

In the case of the analysed data, the statistical criteria clearly showed the optimal number of isolated geotechnical layers to be nine (Figure 4a). The coefficient of variation criterion is not as clear cut but, in this case also, at the number of 8–9 layers, the first minimum is observed in the coefficient of variation in the layer (Figure 4b). In further analysis, the division of subsoil into nine geotechnical layers was adopted, and their location in the classification system according to Robertson [5] is presented in Figure 5. In the second stage of analysis, the values of the constrained modulus M were determined along the analysed CPTU profiles (Figure 6).

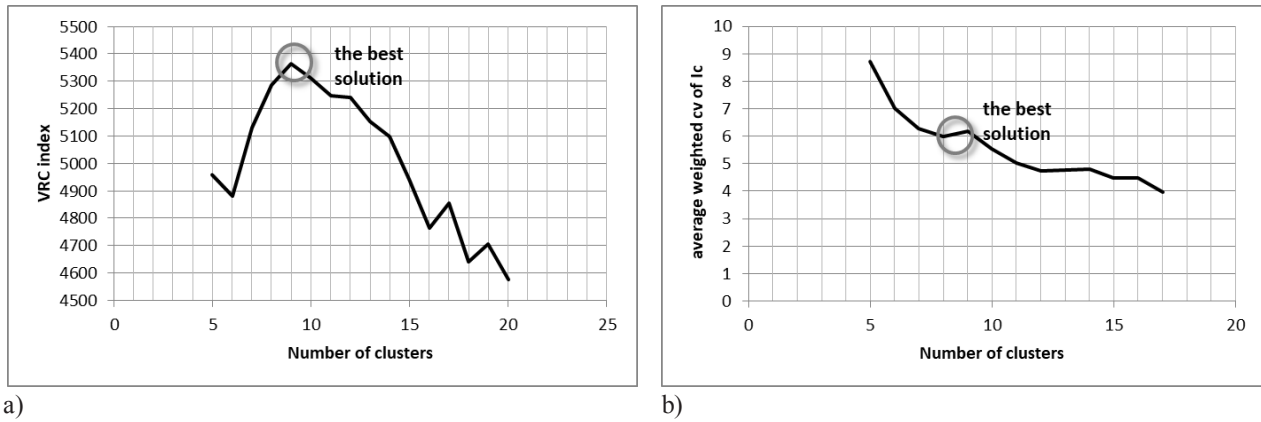a)   b)

Figure 4: Indication of optimal cluster number based on the Caliński-Harabasz criterion (a) and the mean weighted coefficient of variation for index I_c (b).
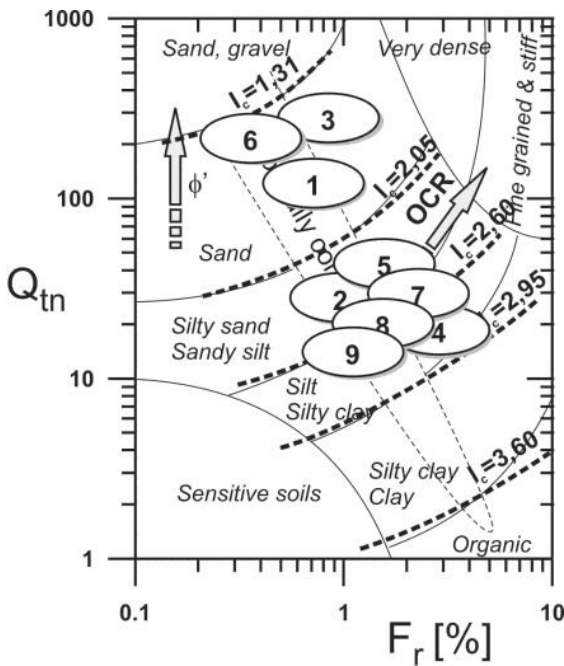

Figure 5: Location of the tested soils and the soil behaviour chart, according to Robertson [5].

5 Analysis of subsoil rigidity based on functional data analysis

The obtained dependencies of the modulus values on depth were adopted as a function of a single variable, and they were clustered in accordance with the assumptions presented in Section 3. In view of the assumed foundation depth for wind farms and the depth of the shallowest CPTU, the analysis was conducted in the depth range from 2 m to 8 m. Due to the varied thickness of the surface, which was a weak layer, apart from the investigation of the entire profile, analyses were additionally conducted separately in two depth ranges: 2–5 m and 5–8 m below the surface. In accordance with the principles of functional data analysis, the first step included smoothing of the function of the modulus depending on the depth, in each of the testing point providing the spline function. Calculations were performed for coefficient $K = 6$ (Figure 7). Recorded functions did not correspond in all cases to the values of the modulus, which were determined from the characteristics presented in Figure 6. Spline functions were also determined along the profile, thus describing the variation of the modulus for coefficient $K = 16$, providing an almost ideal representation of changes in the modulus values with depth (Figure 7).

In the second step of the functional data analysis, the smoothed spline functions were compared, using a hierarchical algorithm presented in Section 3. As a result, dendrograms were obtained, illustrating the clustering of individual objects depending on their similarity. The Ward method was applied in the clustering of objects. The shorter the distance, the more statistically significant are the similarities between the objects. Results of cluster analysis showed that the value of the adopted smoothing coefficient K had a limited effect on the composition of clusters and the order in which they appeared (Figure 8). In relation with the above, the results of the analysis were considered reliable, which were indicated by the statistical criterion, defining the value of coefficient $K = 6$. Based on the obtained dendrograms, an optimal division of the testing sites into homogeneous clusters was provided in accordance with the Caliński-Harabasz [9] criterion. This criterion was supported by information on the geological structure of the subsoil (Figure 8).
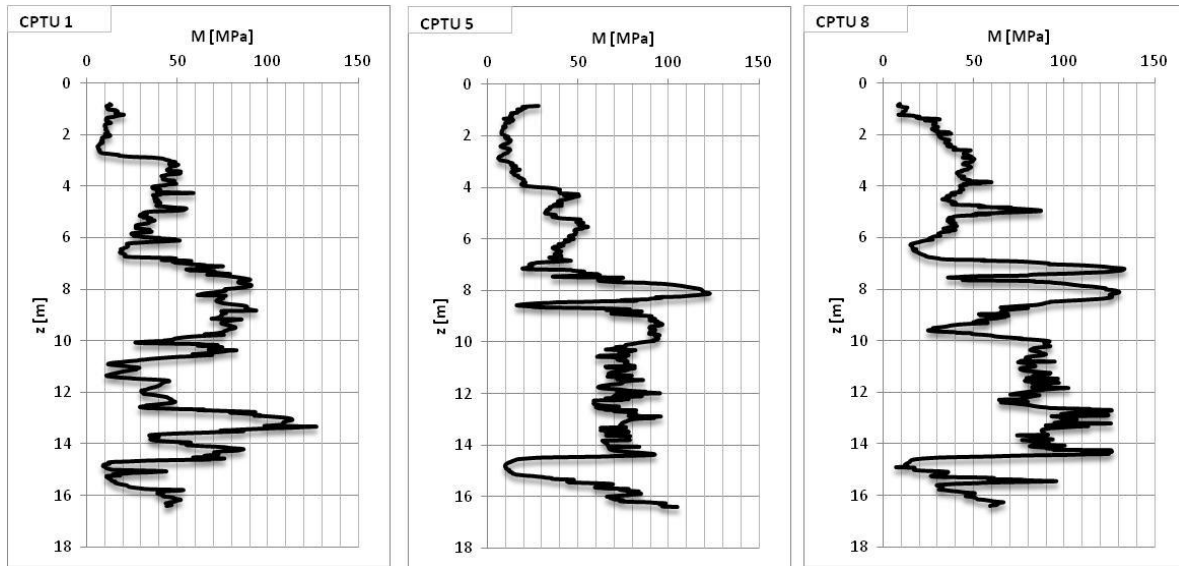
Figure 6: Changes in the constrained modulus M with depth along three arbitrarily selected CPTU profiles.
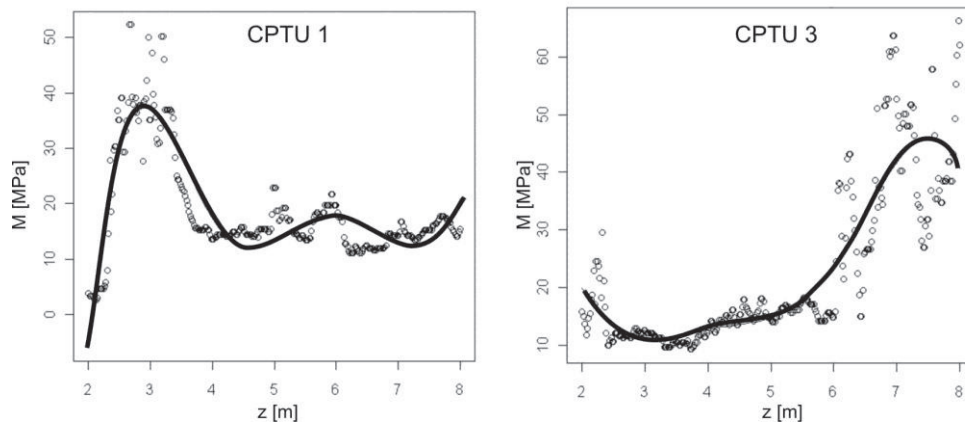

Figure 7: Graphs of the spline function in CPTU profiles 1 and 3, smoothing the course of dependencies of modulus M on depth, assumed for coefficient $K=6$.
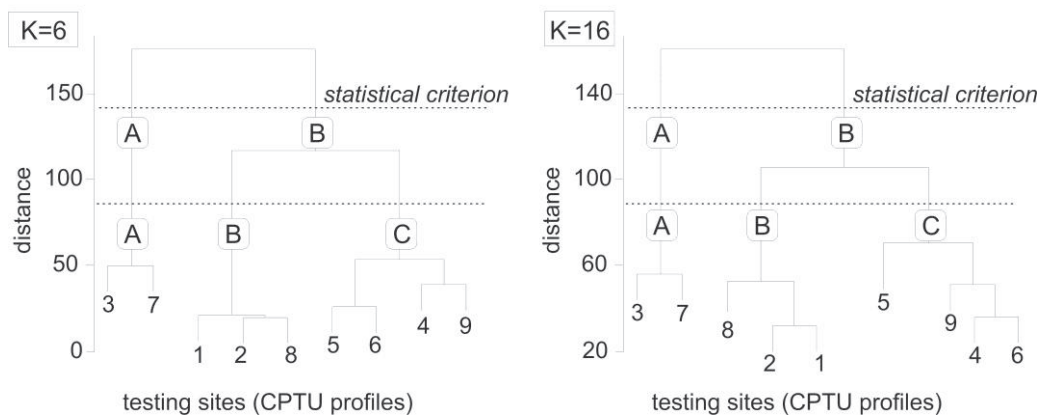

Figure 8: Dendrograms of the clustering hierarchy of curves $M= f(z)$ from the CPTU testing of sites within the range of depth of 2−8 m for the strongly smoothed function ($K=6$) and weakly smoothed function ($K=16$).
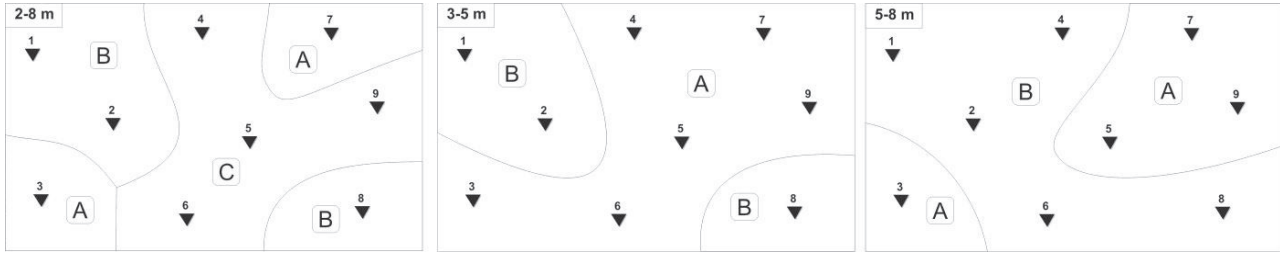
Figure 9: Division of testing area in view of the subsoil rigidity model for the depth range of 2−8 m, 2−5 m and 5−8 m.

As a result, models of subsoil rigidity were obtained, resulting from the division of subsoil into zones characterised by a similarity of function $f(z)=M$ for the three adopted depth ranges, i.e. for the entire subsoil zone (2–8 m), as well as the two depth ranges of 2–5 m and 5–8 m (Figure 9).

Figures 8 and 9 show that a simultaneous analysis of the entire profile requires adoption of three isolations (A, B and C), while the a priori adoption of two depth ranges simplifies the solution to two separations (A and B). However, it needs to be stressed that these isolations are separate for individual depth ranges. In each of the isolations, the included testing points are characterised by a statistically significant similarity of values of modulus M and the trend of its changes with depth. Thus, solutions obtained at this stage of analysis make it possible to categorise the investigated area in terms of not only the absolute values of the modulus M but also the distribution of subsoil rigidity with depth. In order to obtain discrete values of modulus M in a given point of the analysed space, its value needs to be calculated in accordance with the function specified for a given testing point. Another, more general solution may be to determine the mean course of function $f(z)=M$ in a given area and adoption of this distribution as a characteristic functional form of changes in modulus M in this region. These functions may replace the mean values of parameter M, typically used in designing the foundation as values characteristic of a given geotechnical layer. In this case, instead of a mean value determined in a given depth range, which is assigned to the mean value of the vertical component of geostatic stress s_{v0} in the layer, the designer has a notation of the function of this parameter depending on location in space. The obtained solutions make it possible to compare models of subsoil rigidity constructed on the basis of deterministic specific and averaged characteristic data for the isolated soil layers and data recorded in the function form. For this purpose, Figure 10 presents three models of subsoil rigidity, prepared in a selected section, based on the IDW interpolation technique [2].

Assuming the reference to be provided by the subsoil rigidity model based on direct values of modulus M determined from Eqs. (4) and (5) (individualised analysis of values of M at every 2 cm of the profile) (Figure 10a), we may clearly see differences in the structure of this model in comparison to that of the model based on mean values for the isolated soil layers (Figure 10b). The solution based on mean values significantly overestimates subsoil rigidity in a considerable part of the analysed area and fails to identify softer subsoil zones. In turn, the reference solution is of limited practical use due to the high number of data, which need to be included in the analysis of subsoil deformability (analysis at every 2 cm depth). Such drawbacks are not observed for solutions based on means of the function $M = f(z)$, determined for individual isolations separated by functional data analysis. In this case, we obtained both M values noted in the form of means of the functions universal for the isolated subsoil areas (Figure 11) and the distribution of rigidity corresponding to a considerable extent of the reference image (Figure 10 c and d).

Mean values obtained for the functions of changes in constrained moduli with depth for isolated areas divide the entire testing area into zones with varied rigidity levels. Such a division has a highly important practical advantage, as it facilitates the selection of the best location for the planned wind farms.

6 Conclusions

The proposed new clustering method for CPTU characteristics is a highly effective method to isolate homogeneous subsoil zones in the context of rigidity or strength. This paper presented the application of the functional data analysis method, combined with cluster analysis, in the construction of the subsoil rigidity model in a two-dimensional system of coordinates. This method also makes it possible to isolate, in the analysed area, spatial, three-dimensional areas, which statistically are
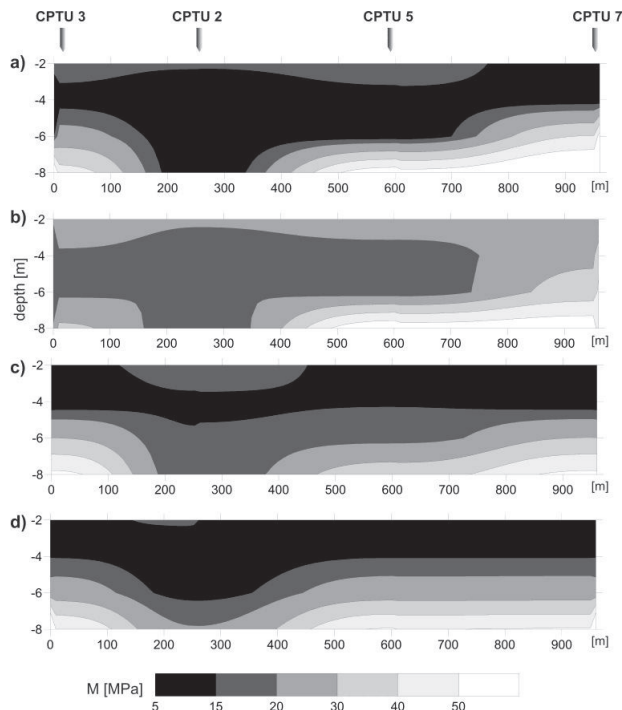
Figure 11: Graphs of mean of the function *f(z)=M* for isolations A, B and C assumed in the depth range of 2–8 m.

Figure 10: Models of subsoil rigidity composed on the basis of constrained modulus *M* along a section selected in the investigated area, determined by IDW based on (a) values of moduli determined from Eqs. (4) and (5), (b) mean values of moduli for isolated soil layers, (c) the mean of the function *f(z)=M* for isolations in the depth range of 2–8 m and (d) the mean of the function *f(z)=M* for isolations in depth ranges of 2–5 m and 5–8 m.

References

[1]   Młynarek, Z., Lunne, T. (1987). Statistical estimation of homogeneity of North Sea overconsolidated clay. In: *Proceedings of International Conference on Statistical and Application Probability*, Vancouver, Canada.

[2]   Młynarek, Z., Wierzbicki, J., Wołyński, W. (2007). Efficiency of selected statistical criteria in determination of geotechnical parameters from CPTU. *Studia geotechnica et Mechania, 29*(1-2), 137-149.

[3]   Lunne, T., Robertson, P.K., Powell, J.J.M. (1997). Cone penetration testing in geotechnical practice. Reprint by *E & FN Spon*, London.

[4]   Młynarek, Z. (2007). Site investigation and mapping in urban area. In: *Geotechnical Engineering in Urban Environments*. Vol. 5. Edited by V. Cuellar et al.. Millpress, Rotterdam, pp. 175-202.

[5]   Robertson, P.K. (2009). Interpretation of cone penetration tests – a unified approach. *Canadian Geotechnical Journal, 46*(11), 1337-1355.

[6]   Lunne, T., Christophersen, H.P. (1983). Interpretation of cone penetrometer data for offshore sands. In: *Proceedings of the Offshore Technology Conference*, Richardson, Texas. Paper No. 4464.

[7]   Kulhawy, F.H., Mayne, P.W. (1990). Manual on estimating soil properties for foundation design. *Electric Power Research Institute*, EPRI, Palo Alto, CA, August 1990.

[8]   Ramsey, J.O., Silverman, B.W. (2005). *Functional Data Analysis*. 2nd edn. Springer, New York.

[9]   Caliński, T., Harabasz, J. (1974). A dendrite method for cluster analysis. *Communication in Statistics, 3*, 1-27.

[10] Wierzbicki, J. (2007). Determination of homogenous geotechnical layers in strongly laminated soil by means of CPTU and cluster analysis. In: *Geotechnical Engineering in Urban Environments*, Vol. 5. Edited by V. Cuellar, et al. Millpress, Rotterdam, pp. 575-579.

most similar in terms of constrained deformation moduli (Figure 9). Analysis of rigidity models for subsoil profiles, which were constructed using functional data analysis based on curves describing the function *M=f(z)* showed several highly important advantages of this method in comparison with the structure of the model based on conventionally isolated layers using discrete, mean values of modulus *M*. The primary advantage is that the model constructed by functional data analysis provides a comprehensive area of changes in modulus *M* in the isolated layer, while simultaneously clearly identifying strengthening and weakening zones. The range of these zones is specified by isolines, separating soil zones in the subsoil differing in rigidity. In the case of construction of a model based on mean values of modulus *M*, the weakening or strengtheni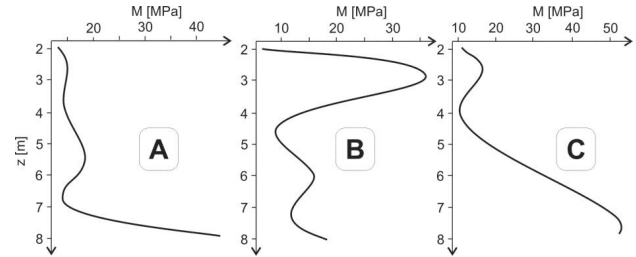ng zones are strictly dampened or they decline (Figure 10).