Proceedings of the 20th International Conference on Soil Mechanics and Geotechnical Engineering– Rahman and Jaksa (Eds)

Statistical analyses of variability of sensitive soils and effects on roads geotechnical design

Analyses statistiques de la variabilité des sols sensibles et des effets sur la conception géotechnique des routes

Emilia Milutinovici
Prospect Technical Studies, Romania, emiliamilutinovici@gmail.com

Daniel Mihailescu
Doctoral School of Geology, Faculty of Geology and Geophysics, University of Bucharest, Romania

ABSTRACT: Expansive soils have large occurrences in Romania and therefore they increase the complexity of geological and geotechnical conditions for designing roads and motorways. Their ubiquity, can incur large variabilities between geologically similar deposits and unexpected similarities between geologically different ones. A detailed statistical method was used in order to obtain the proper data for geotechnical parameters. Three selection criteria have been used: geological, statistical and technical criteria. More than 3000 complex drilling sheets and laboratory results from almost 100 geotechnical studies were processed during the work. Based on the testing procedure, it establishes the variability across the data series and the degree of uncertainty.=

RÉSUMÉ: Les sols expansifs sont très présents en Roumanie et par conséquent augmentent la complexité des conditions géologiques et géotechniques pour la conception des routes et des autoroutes. Leur omniprésence peut entraîner de grandes variabilités entre des dépôts géologiquement similaires et des similarités inattendues entre des dépôts géologiquement différents. Une méthode statistique détaillée a été utilisée afin d'obtenir les données appropriées pour les paramètres géotechniques. Trois critères de sélection ont été utilisés : des critères géologiques, statistiques et techniques. Plus de 3000 fiches de forage complexes et les résultats de laboratoire de près de 100 études géotechniques ont été traités au cours des travaux. Sur la base de la procédure de test, on établit la variabilité des séries de données et le degré d'incertitude

KEYWORDS: sensitive soils, characteristic values, statistics, geotechnical design

1 INTRODUCTION.

Sensitive soils are cohesive soils, very active when they are in contact with water, suffering a significant variation of volume when the water content varies. They are known as high swelling and contraction soils (HSCS), expansive or contractive soils and they are part of the difficult soil's category. Although they have a large occurrence in Romania, was never described a set of characteristic values for geotechnical parameters. The scope of this paper is to obtain a set of characteristic values for principal geotechnical parameters, using statistics evaluation for a large number of input data. In the process of selecting the input data 3 principal criteria have been used. Natural criteria, including geology and geomorphology was the first one, the second was the technical criteria, meaning type of drilling, sampling and laboratory procedure and the third was human criteria (human factor), meaning drilling team, lab test team. For selecting the input data, we use information from tens of geotechnical studies all managed by the authors. The objectives of geotechnical studies were especially linear construction like roads, motorways, airport runways or water supplies pipes.

2 WORK METHOD

2.1 *Data set selection criteria*

Choosing the data set is very important for the accuracy of the results. In order to obtain a useful result, it is important that the data can be correlated, to be of great importance for the pursued purpose and that the result of the research can be reproduced on a larger or a smaller scale in a subsequent experiment. Three types of criteria were taken into consideration, such as: geological criteria (stratigraphic, depositional), statistical criteria (amount of data, variability) and technical criteria (based on the harvesting and processing techniques).

2.1.1 *Geological criteria*

The data selection has been based on the following geological criteria:
- Geological age: cuaternar
- Depositional environment: sedimentary
- rock types: soft and cohesive rocks
- dificult rock types: HSCS high swelling and contraction soils
- depth measured from the surface: 6.00 m

2.1.2 *Statistical criteria*

Taking into account that the main objective of this research is to determine the values, the following statistical input data have been settled upon
- statistical population: soft and cohesive rocks, with HSCS character from Romania
- variables: the value of geotechnical, physical and mechanical parameters.
- statistical population parameters: calculation values and derived characteristic values
- selection type: simple selection

2.1.3 *Technical criteria*

Other than the objective criteria for choosing a data set that will be analyzed, there are also subjective criteria, which take into account the human influence, such as the equipment that was used, methods for sampling and testing and methods of interpretation (Milutinovici, E., 2016). That is why I considered that it is necessary to take into account the following:
- the area of investigation: for each data set all the samples have been harvested from the same area.

- harvesting method: The harvesting method has been unitary. The same drilling equipment has been used, by the same team of operators.
- laboratory testing method: all the laboratory tests have been performed in the same laboratory by the same team of technicians.
- corrections applied before laboratory processing: there have been no corrections applied and no statistically interpreted data has been used. All the values of the geotechnical parameters are the values put forward by the laboratory in which the tests have been performed.

2.2 *Data selection*

For the data selection process, around 70 geotechnical studies have been analyzed in the preliminary stage (all the geotechnical studies have been coordinated by the author) for objectives in areas where high swelling and contraction soils are found, the selected studies have been for linear constructions (roads, airport runways, utility networks). During the selection process, more than 3000 drilling files and lab analysis files have been analyzed.

Following the studies and drilling sheets analysis, 25 data sets have been compiled, containing all types of terrain that follow the aforementioned selection criteria. The 25 data sets are evenly distributed across the territory of Romania, in the areas where the difficult terrain types that form the object of this research (HSCS).

In this research paper the obtained results for 3 geotechnical parameters: consistency index, free swelling, oedometric modulus are described.

2.3 *Processing data through statistical analysis*

The 25 data sets have been analyzed from a statistical point of view, and so was the data set compiled from all the values of the selected geotechnical parameters, named *master set*.

2.3.1 *Evaluation of homogeneity, experimental histogram*

The evaluation of homogeneity for the selected data has been made with the help of the experimental histogram. The establishment of frequency classes and the tracing of the histogram have been established for the „*master set*". The number of frequency intervals has been determined by using Sturges's formula. (Scrădeanu M, 2015). The determined factors were: the value of each frequency interval, and also the number of frequency intervals.

Table 1. Frecquency classes for free swelling (UL) and oedometric modulus (M2-3)

UL (%)			M2-3 (kPa)		
Interval	Frequency	Cumulative %	Interval	Frequency	Cumulative %
66.67	1	0.07	1111.0	3	0.55
104.65	922	67.87	3361.0	5	1.46
142.63	357	94.12	5611.0	39	8.58
180.61	63	98.75	7861.0	75	22.26
218.59	7	99.26	10110.9	146	48.91
256.57	3	99.49	12360.9	126	71.90
294.55	0	99.49	14610.9	112	92.34
332.53	2	99.63	16860.9	36	98.91
More	5		19110.9	3	99.45
			21360.9	2	99.82
			More	1	100.00

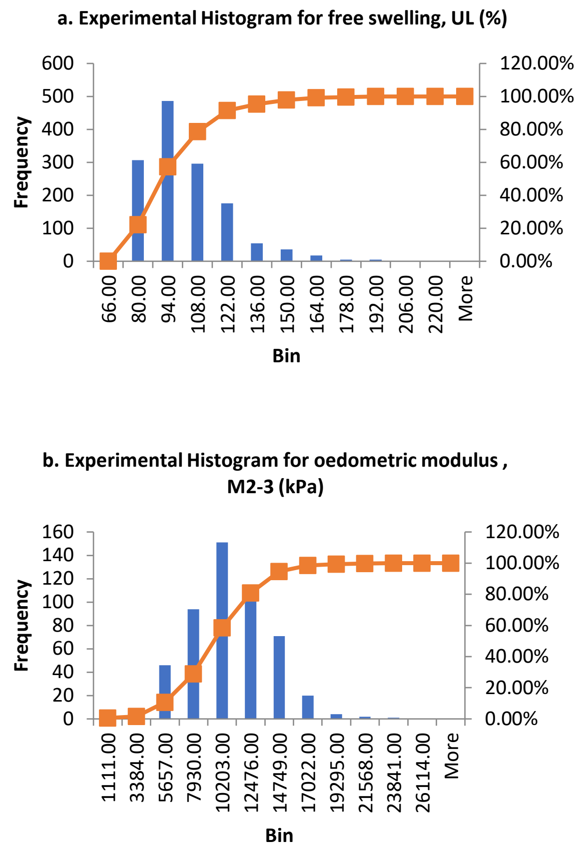

Figure 1. Experimental histogram and cumulative frequency diagram for a- free swelling (UL) and b-oedometric modulus (M2-3)

The histograms exemplified in Figure 1 are both unimodal (single peak histograms), which shows the homogenous character of the selection. In the case of a homogenous selection, the analysis of univariate global variability is followed with the entire data set of values. (Scrădeanu & Popa, 2001).

The analyzed data sets have been classified in the category of moderate and severe asymmetries, for the most part, the asymmetries have been positive, however in some cases negative asymmetries have also been identified, although no case of extreme asymmetry has been identified.

Also in this stage, the variability diagrams have been traced with the purpose of: determining the distribution type, identifying extreme values and determining whether or not to eliminate them. (Scrădeanu M., 2015). Using the *Trendline* function in Microsoft Excel, the lines/curves of approximation have been traced. For every approximation the R^2 value has been determined. Usually, the R values obtained are quite large values, close to the maximum value (1).

2.3.2 *Correspondence test*

The correspondence test has the purpose of verifying the applicability of a hypothesis, in the context of research. Keeping in mind the fact that one of the main purposes of this research, it is important to determine the way in which the analyzed value data are being distributed in comparison to the average value, especially that in geological practice, the average value is the most common way of determining the characteristic value.

It starts by determining two hypotheses and using the chi-square test, the correspondence between the two hypotheses is verified.

The first hypothesis is a null hypothesis (Ho – null hypothesis) and it is a commonly accepted hypothesis, whereas the second hypothesis represents the result of the research and it usually rejects the null hypothesis. (Glen, 2020).

Table 2. Comparison between experiment and theoretical frequency fo free swelling (UL)

Frequency class no	Frequency class	Experimental frequency ni	Theoretical frequency, npi
1	66.67	1.00	342.60
2	104.65	922.00	620.55
3	142.63	357.00	594.54
4	180.61	63.00	135.06
5	218.59	7.00	0.02
6	256.57	3.00	0.00
7	294.55	0.00	2.04
8	332.53	2.00	0.00
	More	5.00	0.00

The larger X^2 will reject the hypothesis for which X^2 is smaller. The verification has been made for all value datasets of the geotechnical parameters of the master set. Some of the value data sets have confirmed the null hypothesis, the other sets contradicting it (Table 3). Most value data sets contradict the null hypothesis

Table 3. Hypothesis concordance

Geotechnical parameter	X^2_t	X^2_{exp}	conclusion
Free swelling, UL (%)	16.9	>1000	Ho rejected/ H1 accepted
Oedometric modulus, M2-3 (kPa)	18.31	17.09	Ho rejected/ H1 accepted

2.3.3 *Analysis of suspicious values, the Chauvenet criteria*

After going through the first stages, we have verified the extreme values and their belonging in the variability resulting from the dispersion of the normal distribution model for the frequency of values that are being processed.

Due to the fact that not all distributions of the analized value sets are normal, before analizing the extreme values, all values are normalized. The place that extreme values have in the accepted values interval has been looked into only for normalized values. When normalizing the values, one must take into account intensity of the asimetry of the experimental histogram.

After normalizing the values, the variability diagram is traced again, this time for the normalized values, with the helo of wich the extreme values are isolated.

2.4 *Determining the final numeric characteristics*

Even though there are few noticeable differences between the numerical characteristics, determined during the analysis stages (Table 4) the statistical side of the research is still necessary, no matter how laborious it is, because it brings an extra degree of certainty regarding the applied methods and the usage of the characteristic values in geotechnical design.

3 RESULTS

3.1 *Characteristic values*

The notion of a characteristic value is a guideline from European standards, Eurocode series (ASRO, 2004). As a result, for any geotechnical study determining the characteristic values of the geotechnical parameters becomes mandatory for any terrain that is in any way influenced by construction.

The higher and lower characteristic values have been determined for the „master set" and for the 25 value sets of the geotechnical parameters, for each parameter. (Figure **2**).

Even though in geotechnical practice, the tendency is to tacking into account for design calculus the lower characteristic value, knowing the upper characteristic value is also very important. There are situations where to set the characteristic value of the geotechnical parameter as the lower characteristic value is a far too conservative approach, which can lead to large-scale geotechnical works and therefore very high costs. In this situation, knowing the upper characteristic value, and being in a safe confidence interval, the average value of the geotechnical parameter can be chosen as the characteristic value of the geotechnical parameter.

Table 4. Numerical characteristics for free swelling before and after ealimination of extreme values determined by Chauvenet cryteria

Numerical characteristics	Inoput data	Date prelucrate
Average, %	102.0	100.5
Median, %	96.6	96.6
Mode, %	88.3	88.3
Min, %	11.3	70.0
Max, %	500.0	200.0
Interval, %	488.7	130.0
No of values	1361	1345.0
Standard deviation	30.0	21.6

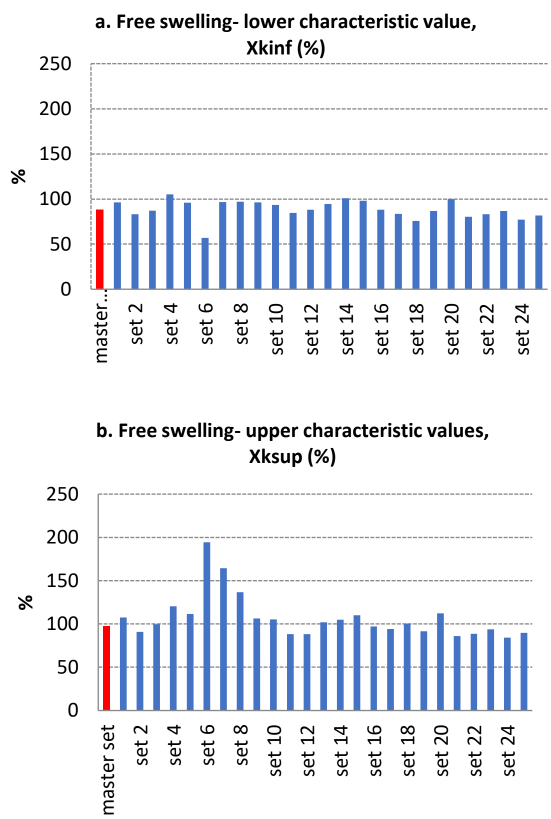

Figure 2. Characteristics values for free swelling a. inferior value; b. superior value

Table 5. Result for characteristic values

No	Geotechnical parameter	Variation coefficient	Variation interval	Characteristics values
1	Free swelling, UL (%)	0.21	57.28-143.67	93.78
2	Oedometric modulus, M2-3, (kPa)	0.30	4058.06-16627.38	9380.85

4 CONCLUSIONS

High swelling and contraction soils are widely spread around Romania. The large presence of these deposits has led to a good delimitation of their space, as well as a categorizing taking into account the swelling potential. On the other hand, their existence on large surfaces, directly affects the quality and cost of constructions, especially the linear type (highways, roads, airport runways, utility networks), for which the foundations are on large surfaces and which imply a lot of maintenance work. Even though they are very common as foundation terrain for a large amount of construction types, in standard technical literature there is no data set of characteristic values for the geotechnical parameters that describe these types of rocks, as it exists for other difficult soil types such as for soils sensitive to moisture (loessoid soils) For the data selection, in the preliminary stage around 70 geotechnical studies have been analyzed (all the studies have been conducted by the author) for the objectives situated in the areas in which high swelling and contraction soils, the selected studies have been for linear constructions (highways, roads, airport runways, utility networks). Across the data selection more than 3000 complex drill sheets have been analyzed, also containing laboratory analysis. The data selection that has been analyzed in this research was based on geological, statistical and technical criteria. Following the analysis of geotechnical studies and drilling sheets, 25 data sets have been selected, containing the types of terrain that fit into the aforementioned selection criteria. The 25 data sets are uniformly distributed across Romania, in the areas in which the types of difficult terrain, that are the object of this research, are found (HSCS).

One of the main purposes of this research is defining a set of characteristic values for the main geotechnical parameters of high swelling and contraction soils (HSCS). In order to reach this objective, the processing has been made for a data set in which all 25 individually selected data sets have been taken over. This data set was named ***"master set".*** The *master set* consists of values for 24 geotechnical parameters and it is the set that was used for the main statistical processing, normalizing, concordance testing and correlation testing.

The statistical analysis was carried out in several stages: stage I: data string variant 1 (DI) determination of numerical characteristics for the original data; stage II: data string variant 2 (cC), applied only to the master set, with correction of the data string by removing the values on the Chauvenet criterion, for which they were performed: homogeneity assessment, experimental histogram and variability plot; testing the normality of the frequency distribution with the X^2 test; analysis of suspect values and their elimination using the Chauvenet criterion; determination of numerical characteristics (mean values and confidence interval).

The numerical characteristics determined were arithmetic mean, median, modulus, standard deviation, minimum value, maximum value, range between minimum and maximum value.

For variant 2 of the data string (cC), for the master set, univariate global variability analysis was performed. The homogeneity of the data was assessed graphically by plotting the experimental histogram, the variability plot and the distribution plot. Analytically, the normality of the frequency distribution was tested using the X2. It was observed that the larger the number of data, the lower the probability of obtaining a normal distribution of values.

The analysis of the suspect values and their membership in the variability resulting from the dispersion of the normal frequency distribution model of the values being worked with was performed. Due to the fact that not all distributions of the analyzed value sets are normal, before analyzing the outliers, normalization of the values was performed. The membership of suspect (extreme) values to the accepted range was investigated only for normalised values. The normalisation of the values is done according to the intensity of the asymmetry of the experimental histogram. Following the transformations for distribution normalization, the variability plot of the normalized values was redrawn, with which the extreme values were isolated. After isolating the extreme values, they were checked to see if they were within the tolerance range. If the extreme (suspect) value analysed fell within the range of normal fluctuations of the values, then they were kept in the range of values for which the analysis is continued, even if their presence in the range of values appeared suspect from the distribution graph. If the values did not fall within the allowed range of fluctuation, then they were removed from the range of values for which the analysis continues. The analyses performed showed little variation between the numerical characteristics determined for the two types of strings used. The higher the number of values in a string analysed, the more the variations between the numerical characteristics determined for the three strings of values were blurred. For example, in the case of the numerical characteristic mean value, for geotechnical parameters for which there is a number of analysed values in the order of thousands, the difference between the average values determined for the three string variants (DI, fO and cC) is between 0.12% and 3.87% (relative to the maximum value of the three calculated means), where the minimum difference is obtained for dry density and the maximum difference is obtained for natural moisture. For geotechnical parameters for wich the no of values are tens or hundreds order, the difference between average values determined for all three hypothesis is between 0.3 and 8.0 %,, wher the minimum difference is obtained for friction angle determined by consolidated – drained direct shearing test and the maximum difference is obtained for friction angle from consolidated – undrained direct shearing test.

The coefficient of variation was determined for each geotechnical parameter. For some of the geotechnical parameters, coefficients of variation higher than the theoretical coefficients of variation recommended for similar clay soils but not classified as difficult soils were obtained. The largest difference between the determined coefficient of variation and the recommended coefficient of variation for similar clay rocks but not falling into the category of difficult soils was observed for the geotechnical parameters for which the fewest values were processed. In general, the difference does not exceed 25% of the value of the theoretical coefficient of variation, but can reach differences > 60% (as in the case of swelling pressure or internal friction angle).

After eliminating the suspect values, the characteristic values of the geotechnical parameters analyzed were determined. Both upper and lower characteristic values were determined. In geological practice geotechnical calculation is done taking into account the worst case situation, therefore the lower characteristic values are chosen. However, knowing the upper characteristic and design values determined by the research helps to increase the confidence in the values considered, in some cases the choice of lower values is too severe and involves additional design costs.

5 ACKNOWLEDGEMENTS

I want to pay my respect to my teacher prof. dr. dipl. eng. Cristian Marunteanu for his exceptional support during my PhD process.

6 REFERENCES

ASRO. (2004). SR EN 1990-Eurocod: Bazele proiectării structurilor. *ASRO Standard Român*, pp. 1-76.

ASRO. (2007). *SR EN ISO 22475-1. Cercetări și încercări geotehnice. Prelevări si măsurari piezometrice. Partea 1: Principii de execuție.* București: Asociația de Standardizare din România.

Glen, S. (2020). *Null Hypothesis Definition and examples. How to state*. Retrieved April 2020, from Statistics How to: https://www.statisticshowto.com/probability-and-statistics/null-hypothesis/

Milutinovici, E. (2016, November 25). *Basic correlations between geotechnical parameters*. Retrieved January 09, 2020, from https://appliedgeophysics.ro: https://appliedgeophysics.ro/wp-content/uploads/2016/11/03-GEOSCIENCE2016_Milutinovici.pdf

Scrădeanu, D., & Popa, R. (2001). *Geostatistică aplicată.* București: Editura Universității din București.

Scrădeanu, M. (2015). *Modelarea statistică a geostructurilor.* București: Editura Printech.