Exploitation of the Existing Soil Data for Establishment of Geotechnical Maps

Hamhami. Mouloud, Trouzine. Habib, Ghembaza. Moulay Smaine, Asroun. Aissa

Abstract**—** The geotechnical maps are currently working tools, more and more essential, for building professionals from diverse backgrounds (engineers, planners, urban ...).

Our goal is to study the feasibility of such cards, based on geotechnical data, taken from soil studies of existing projects.

We present in this communication a exploratory statistical analysis tools to study the spatial variability of the forms of distribution, correlations between geotechnical parameters. Some examples of exploratory statistical treatment have been discussed in this manuscript.

Keywords**—**geotechnical database, geotechnical cards, statistical tools.

I. INTRODUCTION

ON all the structures with sinister, disorders related to foundations represent 7.8% of the total number of claims reported, which include those related to structure, framing, roofing ... disorders related to foundations represent a cost approximately 12.3% of total disorder. Disorders related to structures, minimal compared to other sources of disorder, accounted for 5.1% of the claims, or 6.6% of the total cost. This last point is to be taken with caution, because major disasters including the cost of repair is high (> € 150 000), are not considered [1].

Over 2000 listed sinister [2]. related to soil shows that the foundations disorders come from:

- at 82% to a misunderstanding of the soil, which are distributed as follows: 25% are the result of a misunderstanding of the behavior of heterogeneous foundations (shallow and deep foundations coupling), equally with a misunderstanding of the role of the embankments and the rest from a misunderstanding of modifications on clay soils by water variations,

- And at 18% of the presence of unstable soil in depth, the runtime errors, ...

Hamhami. Mouloud. is with Laboratoire de Génie Civil & Environnement, Sidi-Bel Abbes university, Faculté de Technologie, BP 89 DZ-22000 Sidi Bel Abbes - Algeria phone: 00(213)555250038 ; e-mail:mouloudhm@yahoo.fr.

Trouzine. Habib. is with Laboratoire de Génie Civil & Environnement, Sidi-Bel Abbes university,Faculté de Technologie - Algeria ; (e-mail: h_trouzine@yahoo.fr).

Ghembaza. M. Smaine. is with Laboratoire de Génie Civil & Environnement, Sidi-Bel Abbes university, Faculté de Technologie, - Algeria; (e-mail: Ghembaza_moulay@yahoo.fr).

Asroun. Aissa. is with Laboratoire de Génie Civil & Environnement, Faculté de Technologie, - Algeria ; (e-mail: a_asroun@yahoo.fr)..

The repairs related to foundations are more expensive because they require a recovery in infrastructure[1].

The vast majority of disorders in foundations relates shallow foundations, footings or slab (93.3%), which represents 91.9% of the total cost of damage and disorders. This disorder results from an inadequate embedment of foundations, a bad choice of foundation type, which result in half of the cases of uncertainty, a partial or total ignorance of the subgrade. That information gap is the result of a lack of resources and / or negligence.

If we still master of the materials used to erect a building, we must support structure on natural element on which we can do nothing, the ground. In the past, we always managed to choose locations that obviously should not pose any problem. But nowadays, it is not possible to proceed in this manner because it is important to build very quickly with high productivity, building on diverse sites.

To build a structure, the engineer needs to know in the most perfect way as possible and in every detail, the environment in which it will work. He must make a site reconnaissance. This notion implies first of all an idea of physical description of the material itself with all the subtleties that includes: appearance, color, texture, structure .... This purely physical description must be complemented by determining the spatial distribution of different categories of land: layer thickness, slope, stratification and the presence of water..

In practice the engineer responsible for the recognition always manages to be on the side of safety by taking the lower values. This state of mind, which could be more depending on experience and on knowledge of the engineer, has the disadvantage of being very wasteful and not based on any scientific rigor.

II. PROBLEM DEFINITION

We observed that in the programming phase of development projects initiated by different structures in the areas of equipment, civil engineering, hydraulics, field constraints are rarely considered. These constraints are compounded when they are discovered after site installation. This situation results to delays in construction time, additional costs not covered in departure budgets and sometimes even threaten the realization of projects. This is not the exclusive responsibility of owners of projects, because they are unable to incur the expenses of soil studies in this preliminary study.

The only solution to this situation is to provide planners a maximum of data on selected sites through maps summarizing the main features of the construction zone. But for several years hundreds of studies of soil were conducted through SIDI-BEL-ABBES city. These studies, which represent tens of millions of dinars, are stored in very precarious conditions after project completion. Following a flood in SIDI-BEL-ABBES city, dozens of reports have been destroyed in the archives of laboratories. Preliminary examination of the data allowed us to see that there are two ways to approach the work:

The first approach is to directly exploit this data by putting the results of geotechnical studies directly on maps trying to show the characteristics of each region [3]. This approach has several disadvantages including:

a-The density of information was variable from one zone to another, and concentrated in major cities. Large areas were largely unexplored making the results not representative.

b-The nature and quality of information contained in the reports require a pretreatment to present reliable results.

c-This procedure can not be used as a tool for future study.

The second alternative, is more attractive is to create a general tool that can be used to support any similar work.

In this order of ideas we have adopted an approach that includes the following steps:

1. Identification of existing geotechnical investigations taking into account the nature and the density of available information. The scattering of data between several structures, the absence of archiving organizations, the difficulty of the spatial positioning of studies (some studies are definitely unusable because of the absence of any indication on the situation of projects) .

2. Exploration of the statistical tools that can be used in treatment.

3. Implementing management tools of the database. The database is of significant interest to people involved (developers, architects, planners, researchers ....). Moreover, these data thus stored and presented in an orderly manner lend themselves more easily to a scientific treatment. Nevertheless, we must find an answer to the following questions:

- What information to collect and how to store them in the manner best suited to the use and probabilistic modeling? Because we want to build a geotechnical database for different types of information that can be stored in the space. Test data carried out on site or in the laboratory include: test conditions, the test curves [4].

- How to deal with spatial information? The Information on soils can be both numerous and inadequate, and with a very diverse nature. Furthermore it should be characterized as variables in space. Their modeling in the random functions can present difficult problems for sampling and interpretation of measurements. Moreover a large number of factors influence the measurements, increasing their dispersion. We are thus obliged to call the set of techniques in probabilistic and statistical treatment [5].

How to take this into account in the calculation of structures? Consider these random functions of space in the calculations of structures is very important if you want a realistic uncertainty induced by various design parameters on the behavior of structures. The designer will want to know what parameters are randomly heavy introducing greater uncertainty in the final results. so that it can act on them [6].

Finally, a very important aspect that interests us merits discussion namely the experimental conditions for data collection that can have a significant influence on the measured parameters. In this domain, two trends are evident: the first of these trends and the most scientific, lies at the pure research is to select sites and devices by placing them in ideal experimental conditions. Thus we can make a limited number of measures unless it has abundant credit. another trend, more pragmatic uses routine tests executed on sites and compares them, even to satisfy a strong dispersion in the results. This procedure closer conditions for future use. This allows to collect a larger amount of information and then proceed by statistical analysis. This is the aim of our work.lease submit your manuscript electronically for review as e-mail attachments. When you submit your initial full paper version, prepare it in two-column format, including figures and tables.

III. PRESENTATION OF STATISTICAL ANALYSIS TOOLS

Statistics has several objects: descriptive, exploratory, decision or modeling. depending on whether one seeks to represent data structures, confirm or clarify a theoretical model or predict. We are interested in the topic of modeling and more particularly to linear methods and those are reduced to the linear case therefore involved parametric methods in which linear combinations of variables.

These variables can be quantitative or qualitative. This criterion will determining the type of method or model to implement. We can use simple techniques to summarize the characteristics (central tendency, dispersion) of a variable density estimate (box plot, histogram, nonparametric estimation) and to detect outliers () or to show the need for a transformation (log, root ...). They also assess the relationships between pairs of variables of the same type quantitative (correlation coefficient, scatter) or qualitative (Cramer, Tchuprow) or different types (correlation ratio, parallel boxplots) [7]-[8]. Rudimentary or trivial side of these tools should not lead to neglect in favor of an immediate implementation of much more sophisticated methods therefore much more sensitive to outliers or asymmetrical distributions. They are an essential prerequisite to a first approach of data. Other methods to reduce the size to summarize a table (nxp) large and reveal its characteristics: the principal component analysis (PCA) for quantitative variables, analysis of single or multiple connections (MCA) for qualitative variables. The factorial discriminant analysis (FDA) for judging the quality of discrimination of a set of quantitative variables to explain a typology described by a variable.

IV. STATISTICAL TREATMENT OF GEOTECHNICAL DATA.

A. Study of Statistical Distributions of Soils Properties

Table I shows the statistical results of some parameters measured in clay samples taken randomly from a few studies concerning the agglomeration of Sidi Bel Abbes

These results do not have a direct interest but are indicators for future investigations. The reading of results shows that the intrinsic parameters or nature are very few variables. The Parameters with mechanical meaning, are highly variable.

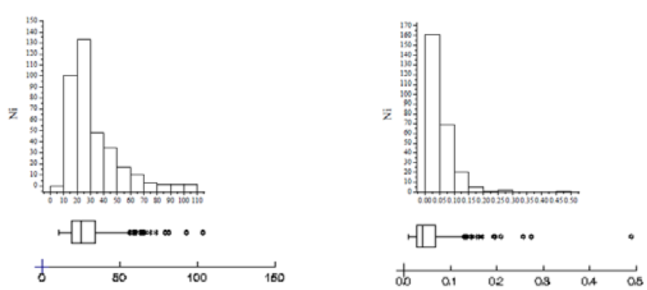
Many researchers [9]-[11]. have fitted various probability distributions for soil properties. Popescu et al. [12]. concluded that (1) most soil properties exhibit skewed, non-Gaussian distributions, and (2) each soil property can follow different probability distributions for various materials and sites. Therefore, in addition to mean and variance, it is also necessary to have more information about probability distributions of soil properties

TABLE I
DISTRIBUTION OF SOME SOIL PROPERTIES THROUGH SIDI-BEL-ABBES CITY

	w (%)	w_L (%)	w_P (%)	Ip (%)	γ_t (T/m^3)	γ_d (T/m^3)	Sr (%)	γ_s (T/m^3)	c_{uu} (kg/cm^2)	φ_{uu} (degrès)	Pc (kg/cm^2)	Ct	Cg
mean	22.97	64.30	28.84	36.06	1.98	1.63	87.28	2.56	0.80	26.93	2.55	0.19	0.06
Standard error	0.51	1.61	0.79	0.97	0.01	0.01	0.56	0.01	0.04	0.85	0.07	0.01	0.00
median	20.29	53.90	24.70	30.90	1.99	1.67	89.00	2.56	0.70	26.00	2.30	0.14	0.04
mode	17.90	93.00	20.00	18.90	2.05	1.73	100.00	2.52	0.80	30.00	2.20	0.09	0.03
Standard deviation	9.77	30.37	14.68	18.05	0.12	0.21	10.70	0.06	0.53	11.40	1.17	0.24	0.05
variance	95.38	922.60	215.54	326.16	0.02	0.04	114.52	0.00	0.28	130.00	1.38	0.06	0.00
Variation coefficient	0.43	0.47	0.51	0.50	0.06	0.13	0.12	0.02	0.66	0.42	0.46	0.93	0.89
Kurstosis	3.17	0.79	3.44	0.93	9.47	0.80	8.12	3.75	7.38	-0.59	4.93	57.06	25.28
Asymmetry coefficient	1.59	1.10	1.65	1.04	-1.66	-0.88	-1.92	-0.66	1.75	0.03	1.52	6.70	3.87
range	54.34	173.63	93.47	99.30	1.12	1.08	100.00	0.35	4.08	49.00	9.34	2.69	0.48
minimum	8.56	19.20	10.18	5.70	1.07	0.94	10.00	2.35	0.00	3.00	0.26	0.01	0.01
maximum	62.90	192.83	103.65	105.00	2.19	2.02	110.00	2.70	4.08	52.00	9.60	2.70	0.49
sum	8498	22762	10064	12585.2	507.62	601.43	31682.0	102.27	144.60	4819.7	649.5	48.93	14.29
number	370	354	349	349	257	370	363	40	181	179	255	259	259

w : Water content, w_L : liquid Limit, w_P : Plastic Limit, *Ip:* plasticity index, γ_t : Total unit weight, γ_d : Dry unit weight, *Sr* : Degree of saturation, γ_s : unit weight of solid, c_{uu} : soil cohesion, φ_{uu} : internal friction angle, *Pc* : preconsolidation pressure, *Ct* : settlement coefficient, *Cg* : swelling coefficient

Histograms and box plots of these values are used to evaluate the distribution of the various parameters, Fig.1. It is customary to compare the distribution of the parameters with the distribution of theoretical laws. applying transformations on the variables (square root, logarithm, ...) is often applied to improve the symmetry of the distribution, or to moderate the effect of outliers or specific values.



(a) Plastic Limit (b) coefficient of swelling

Fig. 1 Examples of histograms and box plots of soil parameter

B. Correlation and Regression Analysis

The parameters studied in the previous application have they interactions? Table II shows the correlation coefficients among the parameters taken pairwise. Atterberg limits (liquid Limit w_L, Plastic Limit w_P, plasticity index *Ip*), have strong correlations between parameters. However, the relationship between mechanical parameters (soil cohesion c_{uu} , internal friction angle φ_{uu} and preconsolidation pressure *Pc)* and other parameters are very low. It will be necessary to sort values to generate coherent sets for each soil.

TABLE II
COEFFICIENT OF SIMPLE CORRELATIONS TAKEN PAIRWISE

	w	w_L	w_P	Ip	c_{uu}	φ_{uu}	Pc
w	1.00						
w_L	0.86	1.00					
w_P	0.80	0.90	1.00				
Ip	0.78	0.94	0.70	1.00			
c_{uu}	0.03	0.16	0.19	0.10	1.00		
φ_{uu}	-0.36	-0.25	-0.18	-0.28	-0.14	1.00	
Pc	0.33	0.43	0.43	0.36	0.01	-0.08	1.00

Using regression involves testing hypotheses that : the linear model is correct, the errors are normally distributed with zero mean and the errors are mutually independent, the errors are constant variances. The following example is the case of a simple linear regression between the plasticity index IP and the water content w. The calculations are given in the table 3 Correlation is significant at the 5% level. It is evident that there is a strong relationship between IP and w. Column T is used to test the hypothesis that the theoretical correlation coefficient is zero. If the hypotheses of regression are met, this report follows a Student distribution with n-2 degrees of freedom. The value of P gives the probability that T is greater than the absolute value calculated by t. Assumptions appear to be satisfied, and we can conclude that there is a linear relationship between IP and w.

The Fig.2 représent an analysis of variance, analysis the total variation in distinguishing a variation due to the regression and variation due to residues. The sum of squared residuals is obtained by the method of least squares. The coefficient R = 0,612 means that 61.20% of the total variation in IP, is explained by the explanatory variable w. Statistics DURBIN-FISHER, and first-order autocorrelation are used to test the independence of errors. Calculations are performed with the program *SYSTAT V2.*

Pearson correlation matrix / *. Probability matrix*

	IP	W			IP	W
IP	1.000			IP	0.000	
W	0.782	1.000		W	0.000	0.000
sample size		312				

Dep Var: IP N: 312 Multiple R: 0.782 Squared Multiple R: 0.612
Adjusted Squared Multiple R: 0.611 Standard Error Of Estimate: 11.422

Variable	*Coefficient*	*Std Error*	*Std Coef*	*Tolerance*	*T*	*P*
Constant	2.420	1.662	0.000	---	1.46	0.146
w	1.460	0.066	0.782	1.000	22.11	0.000

Analysis of Variance

Source	*Sum-of-Squares*	*DF*	*Mean-Square*	*F-Ratio*	*P*
Regression	63780.671	1	63780.671	488.880	0.000
Residual	40443.450	310	130.463		

DURBIN-WATSON D STATISTIC 1.500
FIRST ORDER AUTOCORRELATION .249

Fig. 2 Regression of plasticity index (I_P) by the water content (w)

Correlation is significant at the 5% level. It is evident that there is a strong relationship between IP and w. Column T is used to test the hypothesis that the theoretical correlation coefficient is zero. If the hypotheses of regression are met, this report follows a Student distribution with n-2 degrees of freedom. The value of P gives the probability that T is greater than the absolute value calculated by t. Assumptions appear to be satisfied, and we can conclude that there is a linear relationship between IP and w. The table of analysis of variance, analysis the total variation in distinguishing a variation due to the regression and variation due to residues. The sum of squared residuals is obtained by the method of least squares. The coefficient R = 0,612 means that 61.20% of the total variation in IP, is explained by the explanatory variable w. Statistics DURBIN-FISHER, and first-order autocorrelation are used to test the independence of errors.

If the linear regression is not satisfactory, we can search for a quadratic equation to links the two variables. Figure 3 shows the different regressions, linear and quadratic, between the plasticity index IP and other parameters, with the regression equations and correlation coefficient for each case.

The graphs in Fig. 3, indicate that in cases where linear regression is valid, using the quadratic regression (or higher order) is not useful because there is no significant improvement in results. In the case of a very high dispersion of data, the use of a regression with the original data does not give satisfactory results. The use of transformations of variables (square root, logarithm, ...) may reduce the effect of this dispersion which could result in better results.

V.CONCLUSION

The Geotechnical maps becomes a basis of work, indispensable for any action of planning, forecasting, research...

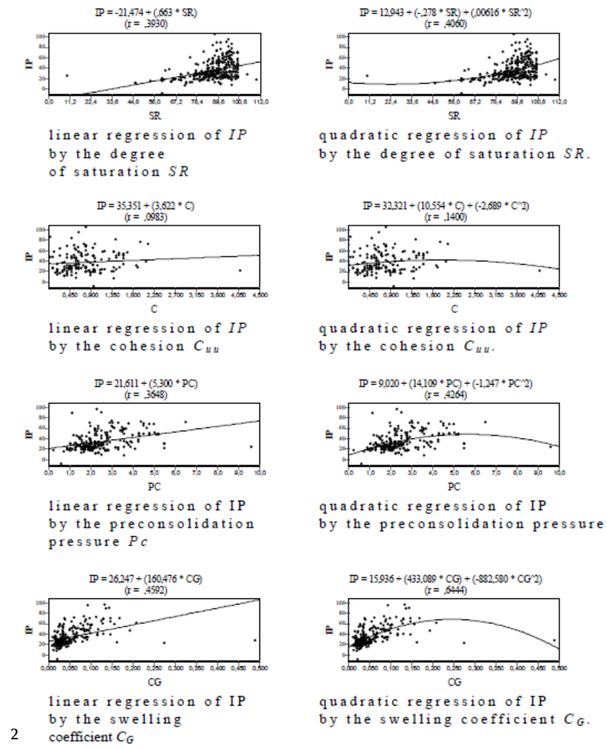


Fig. 3 Linear and quadratic regression of IP by several parameters.

Our work was to establish the preliminary geotechnical map of the town of Sidi Bel Abbes, was oriented towards research of a methodology and development tools without which any attempt to achieve a reliable geotechnical map, remain laborious , if not utopian. We have listed the difficulties of gathering geotechnical data. The quality of these data remain below the the requirements of a rigorous scientific work. However, we have chosen, even if there was no other alternative, to use these data and try to get maximum profit. We have presented a review of the most used statistical and probabilistic analysis tools . These techniques, developed for other science disciplines (economics, medicine, biology, ...) are the only tools that can be implemented to take advantage of the existing data. We discussed the issue of storage and the management of databases. We tried to demonstrate, through examples, the logic of the application of statistical reasoning on a data, which a priori does not permi.

Research in this area remains curiously, very few. It will be necessary put in place multidisciplinary research teams (geotechnical, statistics, computer science) that can solve this problem.

REFERENCES

[1] Gueyffier M.-F., Cloarec G.- *Observatoire de la qualité de la construction, Agence Qualité Construction*, 2003, 101p.

[2] Logeais L. - La Pathologie des fondations, *Edition Le Monit*eur, 1982, 158p.

[3] HAMHAMI M. « Etablissement de la carte géotechnique de la ville de Sidi Bel Abbès », *Mémoire de Magister, Université des Sciences et technologies d'Oran.* 2003, pp55-70

[4] FAVRE J-L. «Valorisation de l'information géotechnique », *Actes du colloque bilan et perspectives du GRECO GEOMATERIAUX*, 1989, pp. 239-243.

[5] BACCONNET C. (1989) « Géostatistique et géotechnique: Optimisation de compagnes », *Actes du colloque bilan et perspectives du GRECO GEOMATERIAUX*, pp 247-250.

[6] FRY.J.J. «Contrôle de la Qualité: Intérêt de la Géostatistique», *Actes du colloque bilan et perspectives du GRECO GEOMATERIAUX*, 1989, pp 244-246.

[7] BACCINI A. et BESSE P. « Statistique Descriptive Multidimensionnelle », Laboratoire De Statistique et Probabilités UMRS C5583.

[8] BERGERET F.; BESSE P. «Simulated annealing, weighted simulated annealing and genetic algorithm at work», *Computational Statistics and Data Analysis*, vol. 12, 1997, pp 447-465.

[9] Lumb, P. "The variability of natural soils." *Canadian Geotechnical Journal*, vol.3, pp. 74–97. 1966.

[10] Schultze, E. "Frequency distributions and correlations of soil properties." *In 1st International Conf. Appl. Stat. Prob. Soil Struct. Engr*., 1971, pp. 371-387.

[11] Harr, M.E. "Mechanics of particulate media: a probabilistic approach." *McGraw-Hill, New York*, 1977, 543 pages.

[12] Popescu, R., Prevost, J.H., and Deodatis, G.. "Spatial variability of soil properties: two case studies." *Geotechnical Earthquake Engineering and SoilDynamics Geotechnical Special Publication ASCE*, No. 75, 1998a, pp. 568-579.