Application of Data Mining Techniques in the Estimation of Mechanical Properties of Jet Grouting Laboratory Formulations over Time

Joaquim Tinoco and A. Gomes Correia and Paulo Cortez

Abstract Sometimes, the soil foundation is inadequate for constructions purpose (soft-soils). In these cases there is need to improve its mechanical and physical properties. For this purpose, there are several geotechnical techniques where Jet Grouting (JG) is highlighted. In many geotechnical structures, advance design incorporates the ultimate limit state (ULS) and the serviceability limit state (SLS) design criteria, for which uniaxial compressive strength and deformability properties of the improved soils are needed. In this paper, three Data Mining models, i.e. Artificial Neural Networks (ANN), Support Vector Machines (SVM) and Functional Networks (FN), were used to estimate the tangent elastic Young modulus at 50% of the maximum stress applied ($E_{tg50\%}$) of JG laboratory formulations over time. A sensitivity analysis procedure was also applied in order to understand the influence of each parameter in $E_{tg50\%}$ estimation. It is shown that the data driven model is able to learn the complex relationship between $E_{tg50\%}$ and its contributing factors. The obtained results, namely the relative importance of each parameter, were compared with the predictive models of elastic Young modulus at very small strain (E_0) as well as the uniaxial compressive strength (Q_u). The obtained results can help to understand the behavior of soil-cement mixtures over time and reduce the costs with laboratory formulations.

Joaquim Tinoco
Department of Civil Engineering/C-TAC, University of Minho, 4710, Guimarães, Portugal, e-mail: jabtinoco@civil.uminho.pt

A. Gomes Correia
Department of Civil Engineering/C-TAC, University of Minho, 4710, Guimarães, Portugal, e-mail: agc@civil.uminho.pt

Paulo Cortez
Department of Information systems/R&D Algoritmi Centre, University of Minho, 4710, Guimarães, Portugal, e-mail: pcortez@dsi.uminho.pt

1 Introduction

Given the growth of the human population and the finite resources of the planet Earth, we are forced to use soft-soils as a soil foundation. In these cases, there is need to improve its physical and mechanical properties. For this purpose, the Jet Grouting (JG) technology is widely used [6, 12], given it great versatility. This technology allows the improvements of the mechanical and physical properties of several types of soil, since grain to fine soils, and different shapes of treatment (columns, panels, etc.) can be obtained. In few words, this technology consists in injecting high speed grouting of water-cement mixture into the subsoil with or without other fluids (air or water). The fluids are injected through small nozzles placed at the end of a rod which is inserted until the intended depth. This rod is continually rotated and slowly removed up to the surface. At the end, a soil-cement mixture with better properties is obtained. Currently adopted JG methods can be classified according to the number of fluids injected into the subsoil: water-cement grout $\rightarrow$ Jet 1; air + grout $\rightarrow$ Jet 2 and water + air + grout $\rightarrow$ Jet 3.

This paper will focus the JG initial stage, where a set of laboratory formulations, which are function of the soil type to be treated and the design properties, are used to set the soil-cement mixture that will be used in the construction works. In particular, this study allows the definition of the grout water/cement ratio, the amount of cement for cubical meter of treated soil and the cement type, needed to satisfy the design and economical requirements. The remaining parameters that control the final characteristics of the JG elements (e.g. the speed of withdrawal and rotation of the rods) will be evaluated with the execution of test columns, given the difficulty to simulate such parameters in laboratory. However, despite of all advantages of JG technology, its design is essentially based on empirical methods that are often too conservative [5, 11]. As a result, the economy and the quality of the treatment can be affected. Thus, given the high potential of JG technology, is very important to develop more accurate and rational models to estimate the effects of the different parameters involved in JG process. This will allow reducing field tests, optimizing all the constructive process and obtaining a higher technical and economical efficiency.

In the other hand, powerful tools have emerged that allow extract useful information from large databases, i.e. Data Mining (DM) techniques. These techniques enable the exploration of complex relationships between several inputs and the target variable. Hence, given the high complexity inherent to JG process, due to the number of parameters involved and the heterogeneity of the soil, DM techniques are an interesting tool to explore JG data. Thus, in order to develop rational models and satisfy the current project requirements, three DM techniques were applied to estimate the mechanical properties of JG laboratory formulations. So, we started by developing models to predict the uniaxial compressive strength (Q_u) of JG laboratory formulations over time [15]. However, to deal with the serviceability state of the structure, deformability properties of the improved ground are also necessary. In this context, predictive models for deformability moduli, namely elastic Young modulus at very small strain (E_0), of JG laboratory formulations were also developed [16].

Yet, the best parameter to analyze the deformability properties of soil-cement mixtures is the tangent elastic young modulus at 50% of the maximum applied stress ($E_{tg50\%}$). Thus, DM techniques were also applied to estimate this property over time.

In this work, the performance of the predictive models of E_0 and Q_u are summarized and the predictive capacity of $E_{tg50\%}$ by application of DM techniques, namely Artificial Neural Network (ANN), Support Vector Machine (SVM) and Functional Networks (FN) are exposed. Moreover, the key parameters in Q_u, E_0 and $E_{tg50\%}$ estimation were identified, compared and discussed.

2 Materials and Jet Grouting Laboratory Data

Three datasets were used to train and test the predictive models of each studied property. All data were prepared at University of Minho, under a huge laboratory experimental program, to analyze the influence of several parameters in Q_u, E_0 and $E_{tg50\%}$ of JG material. These mechanical properties were obtained in unconfined compression tests with on sample strain instrumentation [3]. In a non-linear stress-strain relationship different moduli can be defined. For this work, tangent elastic young modulus at 50% of the maximum applied stress was adopted since, is a key geotechnical parameter that better defines the deformability properties of soil-cement mixtures. Table 1 shows the number of records of each dataset used during the training phase of each predictive model as well as the number of different formulations in each dataset.

Table 1 Number of records and different formulations of each dataset used during the training of each predictive model

	Q_u	E_0	$E_{tg50\%}$
Number of records	175	188	49
Number of formulations	35	9	8

Based on expert knowledge about soil-cement mixtures [13] and after some experiments, the following input parameters were selected : Water/Cement ratio - W/C; Age of the mixture - t; Coefficient related with the cement type - s; Relation between the mixture porosity and the volumetric content of cement - $\mathrm{n}/(\mathrm{C_{iv}})^{\mathrm{d}}$; Cement content of the mixture - %C; Percentage of sand - %Sand; Percentage of silt - %Silt; Percentage of clay - %Clay and Percentage of organic matter - %OM.

The basic statistics of the numerical parameters used in Q_u and E_0 datasets are described in [15, 16] respectively. Table2 shows the basic statistics for $E_{tg50\%}$. The geotechnical soil properties were evaluated using laboratory tests. While all of the soils were classified as fine grained soils they have different percentages of sand, silt, clay and organic matter. A detailed classification of soils can be found in [15].

All formulations were prepared with cement CEM I 42.5R, CEM II 42.5R and CEM IV/A 32.5R

Table 2 Synopsis of the numerical input parameters in $E_{tg50\%}$

Soil	Parameter	Minimum	Maximum	Mean	Standard Deviation
	W/C	0.69	1.11	0.98	0.12
	t(days)	28.00	84.00	64.57	19.13
	$n/(C_{iv})^d$	38.73	73.81	61.46	6.55
clay	%C	24.19	64.86	44.39	11.79
	%Sand	0.00	39.00	14.10	13.68
	%Silt	33.00	57.00	50.00	8.27
	%Clay	22.50	45.00	35.71	7.45
	%OM	0.40	8.30	3.71	2.43

3 Data Mining Models and Evaluation Measures

3.1 Data Mining Models

Three Data Mining models where trained to estimate Q_u [15], E_0 [16] and $E_{tg50\%}$, of JG laboratory formulations over time.

ANN mimic some basic aspects of brain functions [9], which processes information by means of interaction among several neurons. We adopted the most popular model, the multilayer perceptron that contains only feedforward connections, with one hidden layer with H processing units. The general model of the ANN is:

$$\hat{y} = W_{o,0} + \sum_{j=I+1}^{o-1} f\left(\sum_{i=1}^{I} X_i \cdot W_{j,i} + W_{j,0}\right) \cdot W_{o,i} \tag{1}$$

where $W_{j,i}$ represents the weight of the connection from neuron j to unit i, f is a logistic function $1/(1+e^{(-x)})$, and I is the number of input neurons. To chose the best value of H, we used a grid search within $\{2, 4, ..., 10\}$ [8].

SVM was initially proposed for classification tasks [7]. After the introduction of the ε-insensitive loss function, it was possible to apply SVM to regression tasks [14]. SVM has theoretical advantages over ANN, such as the absence of local miminima in the learning phase. The main idea of the SVM is to transform the input data into a high-dimensional feature space by using a nonlinear mapping. For this purpose, the popular Gaussian kernel was adopted:

$$k(x,x') = e^{\left(-y \cdot \|x-x'\|^2\right)}, y > 0 \tag{2}$$

Under this setup, performance of the regression is affected by three parameters: γ - the parameter of the kernel, C - a penalty parameter, and ε - the width of a ε-insensitive zone. To reduce the search space, the first two values will be set using the heuristics of [2]: $C = 3$ and $\varepsilon = \hat{\sigma}/\sqrt{N}$, where $\hat{\sigma} = 1.5/N \cdot \sum_{i=1}^{N}(y_i - \hat{y}_i)^2$ and $\hat{y}_i$ is the value predicted by a 3-nearest neighbor algorithm. To optimize the kernel parameter γ, we adopted a grid search of $\{2^{-15}, 2^{-13}, ..., 2^{3}\}$.

FN are a general framework useful for solving a wide range of problems [1], where the functions of the neurons can be multivariate, multi-argument and it is also possible to use different learnable functions, instead of fixed functions. Moreover, there is no need to associate weights to connections between nodes, since the learning is achieved by the neural functions. When compared with ANN, there are some advantages [18]. For example, unlike ANN, FN can reproduce certain physical characteristics that lead to the corresponding network in a natural way. Also, the estimation of the network parameters can be obtained by resolving a linear system of equations, which returns a fast and unique solution, i.e. the global minimum is always achieved. These two types of networks have a similar structure, but they also have important differences. In FN the selection of the initial topology is normally based on the properties of the problem at hand and can be further simplified using functional equations and its neural functions (normally functions from a given family, such as polynomial or exponential) can be multidimensional and set during the learning phase. Furthermore, outputs neurons can be connected, which is not the case of standard ANN. In this work we use the FN to solve the following generic expression:

$$\hat{y} = \beta_0 \cdot \prod_{i=1}^{N} x_i^{\alpha_i} \tag{3}$$

where, $\{x_1, ..., x_i\}$ are the input parameters and $\{\beta_0, \alpha_1, ..., \alpha_i\}$ are the coefficients to be adjusted.

To learn the coefficients in equation (3), the following minimization problem was used:

$$Minimize Q = \sum_{S=1}^{S} \delta_s^2 = \sum_{S=1}^{S} \left(y_s - \beta_0 \cdot \prod_{i=1}^{N} x_i^{\alpha_i} \right)^2 \tag{4}$$

We also tested the classic multiple regression (using the R tool). Yet, the poor results achieved (when compared with ANN, SVM and FN) are not reported here due to space limitations.

The ANN and SVM models were training using rminer library [4], which facilitates the application of DM techniques in the R tool. The formulation and resolution of the FN was implemented in the free version of the GAMS [17].

3.2 Evaluation Measures

To assess and compare the performance of each predictive model, three evaluation measures were calculated: Mean Absolute Deviation - MAD; Root Mean Square Error - RMSE and the coefficient of determination - R^2:

$$MAD = \frac{\sum_{i=1}^{N} |y-\hat{y}|}{N}; RMSE = \sqrt{\frac{\sum_{i=1}^{N} (y-\hat{y})^2}{N}} \tag{5}$$

$$R^2 = \left(\frac{\sum_{i=1}^{N} (y-\bar{y}) \cdot (\hat{y}-\bar{\hat{y}})}{\sqrt{\sum_{i=1}^{N} (y-\bar{y})^2 \cdot \sum_{i=1}^{N} (\hat{y}-\bar{\hat{y}})^2}} \right)^2 \tag{6}$$

When compared with MAD, RMSE metric is more sensitivity to extreme errors. In a model with good performance, both MAD and RMSE should present lower values and R^2 should be close to unit value.

The Leave-One-Out scheme was adopted for measuring the predictive capability of each model, where sequentially one example is used to test the model and the remaining data is used for fitting the model. Under this scheme, there is need around N (the number of data samples) times more computation, since N models are fitted. The final generalization estimate is evaluated by computing the MAD, RMSE and R^2 metrics for all N test samples. To understand better the behavior of the JG material, the influence of each parameter was also quantified by applying a sensitivity analysis procedure [10]. This procedure determines the most important variables by successively holding all but one input constant and varying the other over its range of values to observe its effect on the system. A high variance observed in the outputs denotes a high input relevance.

4 Results of the Different Predictive Models

Similarly to the previous works, where DM techniques were used to study the behavior of JG laboratory formulation [15, 16], in the present study, a high performance was also reached in $E_{tg50\%}$ estimation. This performance is proved by lower values of MAD and RMSE metrics and the R^2 value close to the unit (see table 3).

Figure 1 shows the relation between $E_{tg50\%}$ measured versus predicted by SVM model. The same relation for the remaining models (ANN and FN) is very similar. As one can see in figure 1, all points are very close to the diagonal line that represents the perfect prediction. Despite of the better values of the metrics MAD, RMSE and R^2 in FN model, SVM is more consistent and interesting in terms of relative importance of each parameter. According to FN model the $E_{tg50\%}$ of JG laboratories formulations is almost only controlled by clay percentage of soil (55.92%), and we know that is not truth. Based on empirical knowledge, the mechanical properties of soil-cement mixtures are also affected by cement content.

Table 3 Comparison of the performance between the three models: ANN, SVM and FN, in $E_{tg50\%}$ estimation

	ANN	FN	SVM
MAD	0.27	0.18	0.19
RMSE	0.50	0.24	0.28
R^2	0.74	0.93	0.91

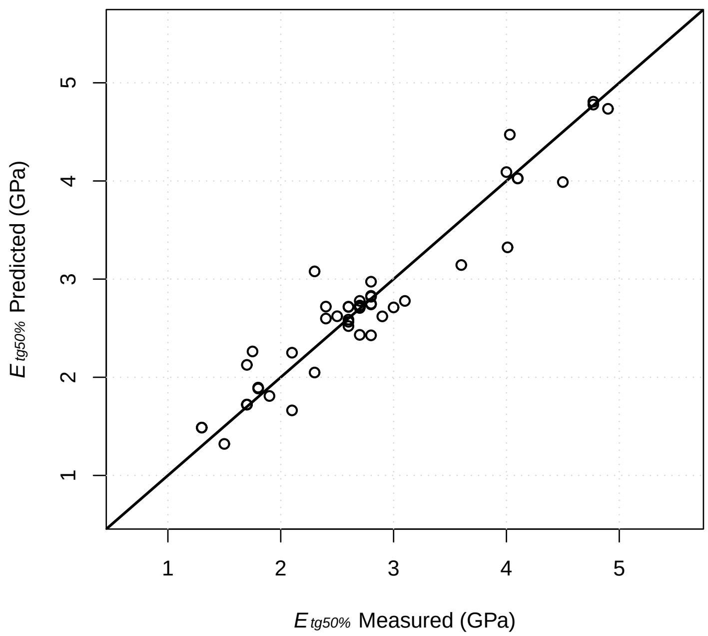

Fig. 1 Predicted versus desired $E_{tg50\%}$ of JG laboratory formulations using the SVM model

The performance reached by SVM model in $E_{tg50\%}$ prediction is very similar to the performance reached by the same model in Q_u [15] and E_0 [16] prediction. Table 4 summarizes the values of the coefficient correlation (R^2) of SVM model in Q_u, E_0 and $E_{tg50\%}$ estimation over time.

Table 4 Performance of each SVM predictive model in Q_u, E_0 and $E_{tg50\%}$

	SVM-Q_u	SVM-E_0	SVM-$E_{tg50\%}$
R^2	0.93	0.96	0.91

In spite of the high performance, assessed by metric MAD, RMSE and R^2, as well as the high relation between $E_{tg50\%}$ measured versus predicted, it is also important to quantify and analyze the relative relevance of each parameter in the model. Observing figure 2, which shows the importance of each input parameter in SVM predictive model of Q_u, E_0 and $E_{tg50\%}$, we can see that the relation between porosity and the volumetric content of cement $(\mathrm{n}/(\mathrm{C_{iv}})^{\mathrm{d}})$, the water/cement ratio (W/C) and the soil properties, namely the percentage of clay (%Clay), are the key parameters in $E_{tg50\%}$ prediction. Moreover, in Q_u estimation the age of the mixture and the percentage of cement should be included. In the other hand, it is interesting to observe that the soil properties are more relevant in deformability properties estimation (E_0 and $E_{tg50\%}$) than in strength prediction (Q_u). This observation makes some sense if we take into account that for low deformations, the grain size is the responsible for the main resistance capacity of the material. After the grains broke, the cohesion is sustained by soil-cement matrix. So, from this time, the age of the mixture and the percentage of cement take the main role in the strength capacity of the soil-cement mixture.

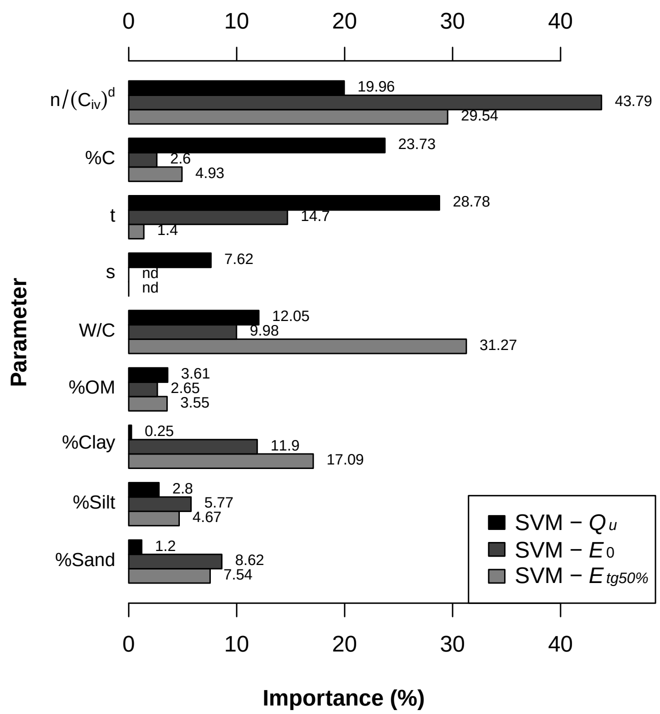

Fig. 2 Comparison of the relative importance of each parameter in each SVM predictive model of Q_u, E_0 and $E_{tg50\%}$

5 Conclusions and Future Works

A rational approach, based in soft computing tools, namely Artificial Neural Networks (ANN), Support Vector Machines (SVM) and Functional Networks (FN), was successfully applied to improve the knowledge about mechanical properties of Jet Grouting (JG) laboratory formulations. The uniaxial compressive strength (Q_u), the elastic Young modulus at very small strain (E_0) and the tangent elastic young modulus at 50% of the maximum applied stress ($E_{tg50\%}$) of Jet Grouting (JG) laboratory formulations can be estimated over time with high accuracy.

The high performance reached in mechanical properties prediction of JG laboratory formulations, allows us to conclude that Data Mining techniques are a useful tool (in particular, the SVM model) to better understand the behavior of soil-cement mixtures over time.

The sensitivity analysis carried out in this study allows the conclusion that the relation between porosity and the volumetric content of cement ($\mathrm{n}/(\mathrm{C_{iv}})^{\mathrm{d}}$), the water/cement ratio (W/C) and the soil properties, are the key parameters in deformability properties prediction. Furthermore, in Q_u estimation, the age of the mixture and the percentage of cement should be included. It was also possible to observe, that the soil properties are more relevant in deformability properties than in ultimate strength estimation of soil-cement mixtures.

The knowledge obtained from this study allows understanding better the mechanical behavior of soil-cement mixtures and will reduce the number of laboratory formulations carried out. As a result, the quality control process of JG columns is improved and the costs of laboratory material formulations are reduced. In future works, we intend to apply Data Mining techniques to develop predictive models for final diameter of real JG columns, as well as its mechanical properties. A sensitivity analysis procedure will be also applied in order to analyze the influence of each parameter.

Acknowledgements The authors wish to thank to Portuguese Foundation for Science and Technology (FCT) the support given through the doctoral grant SFRH/BD/45781/2008. Also, the authors would like to thank the interest and financial support by Tecnasol-FGE and Tiago Valente for the dataset gathered from the laboratory formulations.

References

1. E. Castillo, A. Cobo, J. Gutirrez and R. Pruneda. (1999) Functional networks with applications: a neural-based paradigm, Springer.
2. V. Cherkassky and Y. Ma. Practical. (2004) Selection of SVM parameters and noise estimation for SVM regression, *Neural Networks* 17(1):113-126.
3. A. Gomes Correia, T. Valente, J. Tinoco, J. Falção, J. Barata, D. Cebola, and S. Coelho. (2009) Evaluation of mechanical properties of jet grouting columns using different test methods, in *Procedings of 17th International Conference on Soil Mechanics and Geotechnical Engineering* 2179-2171.

4. P. Cortez. (2010) Data Mining with Neural Networks and Support Vector Machines using the R/rminer Tool, in *P. Perner, editor, Advances in DataMining Applications and Theoretical Aspects, 10th Industrial Conference on Data Mining* 572-583, Berlin, Germany. LNAI 6171, Springer.
5. P. Croce and A. Flora. (2001) Analysis of single-fluid jet grouting, *Geotechnique* 50(6):739-748.
6. J. Falco, A. Pinto, and F. Pinto. (2000) Case histories of ground improvement solutions using jet grouting, Tecnasol FGE: Fundacoes e Geotecnia S.A.
7. A.T.C. Goh and SH Goh. (2007) Support vector machines: Their use in geotechnical engineering as illustrated using seismic liquefaction data, *Computers and Geotechnics* 34(5):410-421.
8. T. Hastie, R. Tibshirani, and J. Friedman. (2009) The Elements of Statistical Learning II: Data Mining, Inference, and Prediction, Springer.
9. S. Kenig, A. Ben-David, M. Omer, and A. Sadeh. (2001) Control of properties in injection molding by neural networks, *Engineering Applications of Artificial Intelligence* 14(6):819-823.
10. R.H. Kewley, M.J. Embrechts, and C. Breneman. (2000) Data strip mining for the virtual design of pharmaceuticals with neural networks, *Neural Networks* 11(3):668-679.
11. F. Kirsch and W. Sondermann. (2001) Ground improvement and its numerical analysis, in *Proceedings of the International Conference on Soil Mechanics and Geotechnical Engineering* 3:1775-1778.
12. A.B. Padura, J.B. Sevilla, J.G. Navarro, E.Y. Bustamante, and E.P. Crego. (2009) Study of the soil consolidation using reinforced jet grouting by geophysical and geotechnical techniques, *Construction and Building Materials* 23(3):1389-1400.
13. M. Shibazaki. (2004) State of practice of jet grouting, in *Grouting and Ground Treatment* 198-217.
14. A.J. Smola and B. Schölkopf. (2004) A tutorial on support vector regression, *Statistics and computing* 14(3):199-222.
15. J. Tinoco, A. Gomes Correia, and P. Cortez. (2009) A data mining approach for Jet Grouting Uniaxial Compressive Strength prediction, in *World Congress on Nature & Biologically Inspired Computing, NaBIC 2009* 553-558.
16. J. Tinoco, A. Gomes Correia, and P. Cortez. (2010) Applications of Data Mining Techniques in the estimations of Jet Grouting Laboratory Formulations over Time, in *1st international conference on information technology in Geo-Engineering, ICITG-Shanghai 2010* 92-100.
17. GAMS Development Corporation. Online Documentation, Welcome to the GAMS Home Page, http://www.gams.com/docs/document.htm (accessed May 17, 2010).
18. Y.Q. Zhou, D.X. He, and Z. Nong. (2005) Application of Functional Network to Solving Classification Problems, in *Proceedings of the World Academy of Science, Engineering and Technology* 7:390-393.