UNIAXIAL COMPRESSIVE STRENGTH PREDICTION OF JET GROUTING COLUMNS USING SUPPORT VECTOR MACHINES

Joaquim Tinoco and António Gomes Correia
Departamento de Engenharia Civil/C-TAC
Universidade do Minho
Campus de Azurém, Guimarães,
Portugal
E-mail: {jabtinoco|agc}@civil.uminho.pt

Paulo Cortez
Centro Algoritmi, Departamento de Sistemas de Informação
Universidade do Minho
Campus de Azurém, Guimarães,
Portugal
E-mail: pcortez@dsi.uminho.pt

KEYWORDS
Soft soils, soil-cement mixtures, soil improvement, jet grouting, uniaxial compressive strength, regression, data mining, support vector machines, sensitivity analysis

ABSTRACT

Uniaxial compressive strength (UCS) is the mechanical properties currently used in geotechnical works design, namely in jet grouting (JG) treatments. However, when working with this soil improvement technology, due to its inherent geological complexity and high number of variables involved, such design is a hard, perhaps very hard task. To help in such task, a support vector machine (SVM), which is a data mining algorithm particularly adequate to explore high number of complex data, was trained to estimate UCS of JG samples extracted from real JG columns. In the present paper, the performance reached by SVM algorithm in UCS estimation is shown and discussed. Furthermore, the relation between mixture porosity and volumetric content of cement and the JG system were identified as key parameters by performing a 1-D sensitivity analysis. In addition, the effect and the interaction between the key variables in UCS estimation was measured and analyzed.

INTRODUCTION

Jet grouting (JG) technology is one of the most used soft-soil improvements methods (Falcão et al. 2000). According to JG technology, a high speed and pressure of grout (with or without other fluids) is injected into the subsoil, which cut and mixes the soil. At the end an improved mass of soil, often termed as *Soilcrete* is obtained. According to the number of fluids injected, three systems are conventionally in use: single, double and triple fluid system. Due to the heterogeneity of the soils, the constructive process of JG technology and nature of treatment fluid injected (normally water cement grout) there are many variables involved in treatment process (Nikbakhtan et al. 2010). Such conditions make the design of JG technology a complex geotechnical task. Nowadays, such design is almost performed based on empirical methods (Lee et al. 2005; Narendra et al. 1996), mainly in the initial project stages and in small scale geotechnical works where information is scarce. Therefore, and since these empirical methods are often too conservative and have a very limited applicability, the quality and the economy of the treatment can be compromised. Hence, and bearing in mind the high versatility of JG technology and its role in important geotechnical works, it is very important to develop rational models to estimate the effects of the different variables involved in JG process.

On the other hand, in the last few years some powerful tool, incorporating advanced statistic analysis, has been developed and are able to automatically extract important rules from vast and complex data. Such tools, usually known as data mining (DM) techniques, has been successfully applied in several scientific areas namely in Civil Engineering domain (Lai and Serra 1994; Rezania and Javadi 2007). One of the most interesting DM algorithms is the Support Vector Machines (SVM), which was used in the present work and has the particularity to be applied in both classification and regression problems. SVM is especially usefully to explore data with nonlinear relationships between several inputs and the target variable and had been successfully applied to solve geotechnical problems (Goh and Goh 2007; Tinoco et al. 2011b). The main criticism of "black box" DM techniques, such as SVM or artificial neural networks is the lack of explanatory power, i.e. the data-driven models are difficult to interpret by humans (Goh and Goh 2007). However, to overcome such drawback a sensitivity analysis (SA) procedure can be applied (Cortez and Embrechts 2011).

The performance reached by SVM algorithm trained with data collected directly from JG columns (JGS), with different JG parameters and *soilcrete* characteristics are shown and discussed in the present paper. Moreover, the key variables in UCS estimation are identified by applying a 1-D SA. Furthermore, the influence of the key variables in UCS estimation are quantified and discussed. In addition, and keep in mind a more realistic interpretation of the results a 2-D SA was performed to the first two key variables.

SUPPORT VECTOR MACHINES

Support Vector Machines are very specific class of algorithms, which is characterized by use of kernels, absence of local minima, sparseness of the solution and capacity control obtained by acting on the margin, or on number of support vectors. When compared with other types of base learners, such as the famous multilayer perceptron, SVM represents a significant enhancement in functionality. The supremacy of SVM lies in their use of non-linear kernel functions that implicitly map inputs into high dimensional feature spaces. In this feature spaces linear operations may be possible that tray to find the best linear separating hyperplane ($y_i = \omega_0 + \sum_{i=1}^{m} \omega_i \emptyset_i(x)$), related to a set of support vector points, in the feature space. Thus, although

SVMs are linear learning machines with respect to feature spaces, they are in effect non-linear in the original input space. These attractive features and promising empirical performance are responsible for its gain of popularity.

SVM was initially proposed for classification problems by Vladimir Vapnik and his co-workers (Cortes and Vapnik 1995). Later, after the introduction of an alternative loss function proposed by Vapnik (Smola et al. 1996), called ε-insensitive loss function, was possible to apply SVM to a regression problems (Smola and Schölkopf 2004).

It is well known that SVM generalization performance (estimation accuracy) depends on a good setting of meta-parameters C, ε and the kernel parameters. The problem of optimal parameter selection is further complicated by the fact that SVM model complexity (and hence its generalization performance) depends on all three parameters. Parameter C controls the trade-off between complexity of the machine (flatness) and the number of non-separable data points and may be viewed as a "regularization" parameter (Goh and Goh 2007). Parameter ε controls the width of the ε-insensitive zone, used to fit the training data. The value of ε can affect the number of support vectors used to construct the regression function. Hence, both C and ε values affect model complexity (but in a different way). Selecting a particular kernel type and kernel function parameters is usually based on application-domain knowledge and should reflect distribution of input (x) values of the training data. In the present work was adopted the popular Gaussian kernel, since it presents less parameters than other kernels (e.g. polynomial):

$$K(x,x')) = \exp\left(-\gamma \cdot \|x - x'\|^2\right), \gamma > 0 \quad (1)$$

To reduce the search space, we adopt the heuristics proposed by Cherkassky and Ma (2004) to set the complexity penalty parameter, C=3 and the size of the insensitive tube $\varepsilon = \hat{\sigma}/\sqrt{N}$, where $\hat{\sigma} = 1.5/N \times \sum_{i=1}^{N}(y_i - \hat{y}_i)^2$, $\hat{y}_i$ is the value predicted by a 3-nearest neighbor algorithm and N the number of examples. The most important SVM parameter, the kernel parameter γ, was set using a grid search within $\{2^{-15}, 2^{-13}, \ldots, 2^{3}\}$, under an internal (i.e. applied over training data) 3-fold cross validation (Hastie et al. 2009).

All experiments were implemented in R tool (Team R 2009), using rminer library (Cortez 2010), which is particularly suitable for SVM training. Before fitting the SVM model, the data attributes were standardized to a zero mean and one standard deviation and before analyzing the predictions, the outputs post-processed with the inverse transformation (Hastie et al. 2009).

MODEL ASSESSMENT AND INTERPRETATION

In regression problem, the main goal is to induce a model that minimizes an error measurement between observed and predicted values considering N examples. For this purpose three common metrics were calculated (Tinoco et al. 2011a): Mean Absolute Deviation (MAD) Root Mean Squared Error (RMSE) and Coefficient of Correlation (R^2). The first two metrics should present lower values and R^2 should be close to the unit value. The regression error characteristic (REC) curve, which plots the error tolerance on the x-axis versus the percentage of points predicted within the tolerance on the y-axis (Bi and Bennett 2003), was also adopted during the analysis of the model performance.

To measure the generalization performance of the trained model were performed R runs under a cross validation approach. Under this scheme, the data are divided into k different subsets, being one used to test the model and the remains to fit it. At the end all data are used for training and testing. Yet, this method requires approximately k times more computation, because k models must be fitted. The final generalization estimate is evaluated by computing the MAD, RMSE and R^2 metrics for all N test samples.

Besides to the performance reached by a DM model it should be also possible to extract human understandable knowledge from the data. To do it a SA procedure (Cortez and Embrechts 2011) was applied. This procedure (1-D SA), which is applied after the training phase and analyzes the model responses when a given input is changed, allowing to quantify the relative importance of each variable. Such quantification is determined by successively holding all inputs at their average values, except one input attribute that is varied through its range of values $(x_a \in \{x_1, \ldots, x_I\})$, with $(j \in \{1, \ldots, L\})$ levels. The obtained responses $(\hat{y}_{a,j})$ are stored and if there is a high gradient (S_a) observed, then this denotes a high input relevance (R_a), which is calculated by:

$$R_a = \frac{S_a}{\sum_{i=1}^{I} S_i \times 100(\%)},\ where\ S_a = \sum_{j=2}^{L}\left|\hat{y}_{a,j} - \bar{\hat{y}}_{a,j-1}\right| / (L-1) \quad (2)$$

For more input influence details, the Variable Effect Characteristic (VEC) curve was plotted. For a given input variable, the VEC curve plots the attribute L level values (x-axis) versus the sensitivity analysis responses (y-axis). Furthermore, aiming to achieve a more realistic interpretation of the models a 2-D SA was performed. Here, two variables are changed simultaneously and the response is measured. With the stored values it is possible to plot the VEC surface or VEC contour (see Cortez and Embrechts (2011)).

JET GROUTING DATA

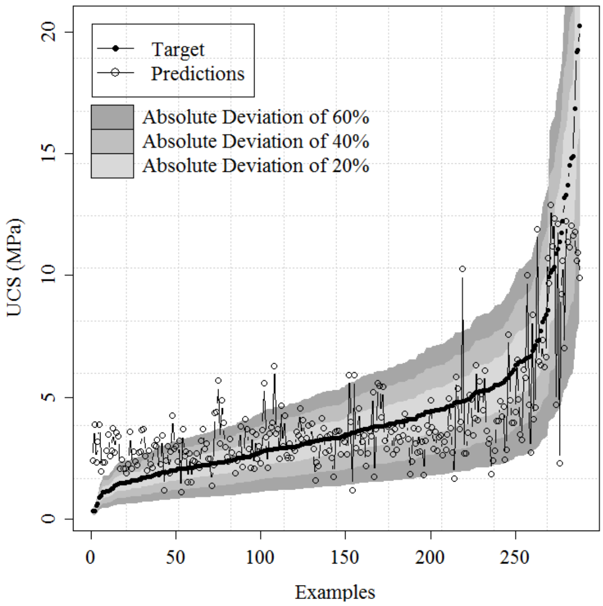
To train and test SVM algorithm was used a dataset composed by 288 records. The tested samples were collected from different columns constructed under the same soil type, at different times and were kept inside a box until be tested in order to keep the water content and not be damaged. UCS (in MPa) was measured in unconfined compression tests with on sample strain instrumentation (Correia et al. 2009). The input variables were chosen based on the expert knowledge about soil-cement mixtures (Shibazaki 2004) and with the experience reached by authors in laboratory formulations studies (Tinoco et al. 2011a). Thus, the following set of eight input variables were chosen: relation between the mixture porosity and the volumetric content of cement ($n/(C_{iv})^d$); age of the mixture (t, days); JG method (JGM); inverse of dry density of the soil-cement mixture ($1/\rho_d$, $m^3 \cdot kg^{-1}$; void ratio of the mixture (e); cement content ($\%C$); water content (ω, %) and water/cement ratio (W/C). The main statistics of both input and output variables are shown in Table 1. The soil under treatment was classified as lean clay (CL), with 39% of sand, 33% of silt, 27% of clay and 8.3% of organic matter. All columns were constructed with cement type CEM I 42.5 R.

Table 1: Summary of the input and output variables in UCS prediction of JGS

Variable	Min	Max	$\bar{u}$	σ
$n/(C_{iv})^d$	37.88	78.61	58.10	7.38
t	9.00	181.00	41.11	39.11
JGM	1.00	3.00	2.04	0.49
$1/\rho_d$	5.63E^{-4}	1.40E^{-3}	8.20E^{-4}	1.21E^{-4}
e	0.56	2.85	1.25	0.33
$\%C$	0.14	0.28	0.22	0.04
ω	2.50	96.80	36.88	12.98
W/C	0.83	1.00	0.89	0.06
UCS	0.32	20.27	3.33	3.07

RESULTS AND DISCUSSION

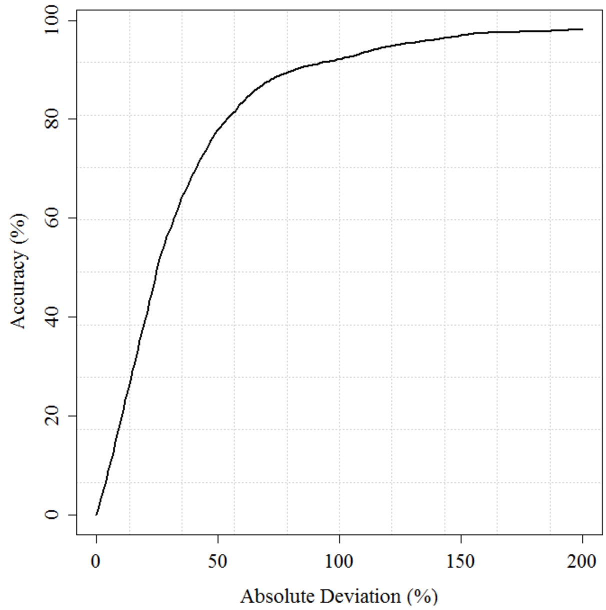
On figure 1, we compare the UCS of JGS measured with those predicted by SVM model for all 20 runs performed. In addition, it is also identified the areas for a prediction with an absolute deviation of 20%, 40% and 60%. As we can see, the accuracy reached by SVM model is relatively lower. Indeed, R^2 value is relatively worse (0.63±0.01) and the values for MAD and RMSE metrics are 1.29±0.01 MPa and 1.87±0.02 MPa respectively. However, observing the REC curve on figure 2, which shows the accuracy obtained for a given absolute deviation (in percentage), we can see a fast improvement on model accuracy. For example, to guaranty that the model is able to predict successfully 80% of the examples an error of 60% should be tolerated.



Figures 1: Relation between UCS JG samples observed versus predicted by SVM model

When evaluating a DM model, we should consider not only predictive accuracy but also model interpretability. In this paper, such interpretability is based on measuring which are the key input variables and how these affect the predicted output. To do so a 1-D and 2-D SA were applied.

Figure 3 give us an idea of the relative importance of each variable after applies a 1-D SA, with the correspondent t-student 95% confidence intervals for all 20 runs performed. As we can see, the relation $n/(C_{iv})^d$ is the input variable more relevant in UCS of JGS, with a weight around 23%. The *JGM*, *%C* and *t* are the next three variables more influents. From point of view of what is the known empirically in JG domain, the relative input importance of SVM model is in agreement with such knowledge. Among these four key variables we can identify one that is related to JG process (*JGM*) and three related with the mixture properties, namely its age, porosity and cement content. It is known that the cement content, mixture porosity and the age of the mixture should be taken in account in soil cement behavior study. Furthermore, if we are talking about soft-soils improvement using JG technology, make sense that JG system used should be considered. Depending of the JG system used the energy applied, the impact of the jet against the soil or the amount of cement inject will be different.

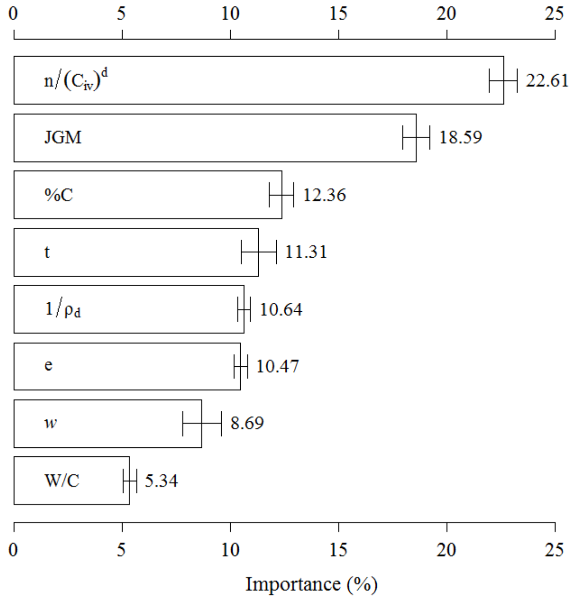


Figures 2: REC curve for SVM model

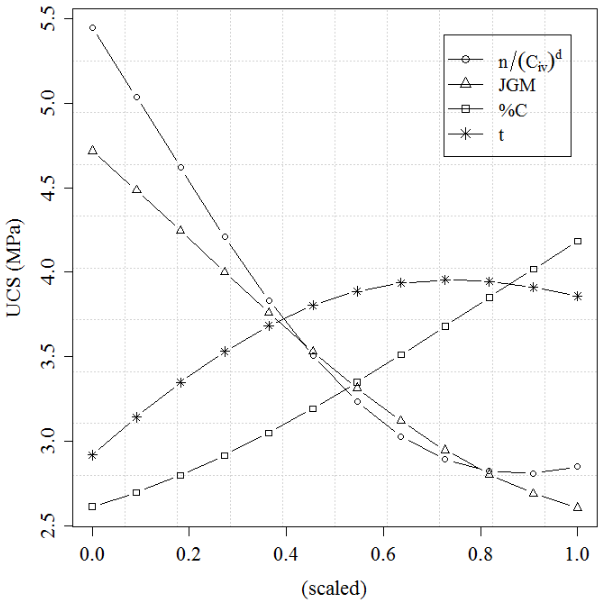
Based also on 1-D SA, the effect of each variable on UCS prediction was quantified. The VEC curves of the four key variables previously identified are shown in figure 4. The age of the mixture and the cement content has a positive impact in UCS prediction. However, its effect in UCS is different. The VEC curve of t shows a convex shape that means that UCS increases quickly in the early ages and after that tend to stabilize (typical behavior of cement mixtures). In the other hand, VEC curve of *%C* is almost linear. The remains two key variables have a similar effect on UCS prediction. UCS of JG samples decrease according to an exponential shape if $n/(C_{iv})^d$ increase, i.e., increasing mixture porosity or decreasing cement content. We also can observe that the UCS decrease almost linearly with the JG method. That means that the biggest strength is reached by application of single fluid system.

All observation previously exposed based on a 1-D SA, such as the relative importance of each variable or VEC curves are very usefully to understand the behavior of JG mixtures. However, in this kind of analysis all variables are fixed in its mean value except one that is ranged from its minimum to maximum values. In real works this normally never happen. thus, in order to do a more realistic analysis we carried out a

2-D SA. In the next lines, the results of a 2-D SA for $n/(C_{iv})^d$ and *JGM* (the first two key variables) are exposed and discussed.

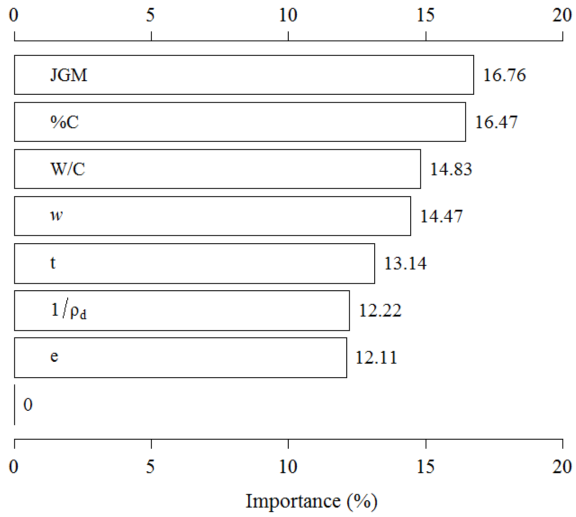


Figures 3: Relative importance of each variable according SVM model, quantified by a 1-D SA

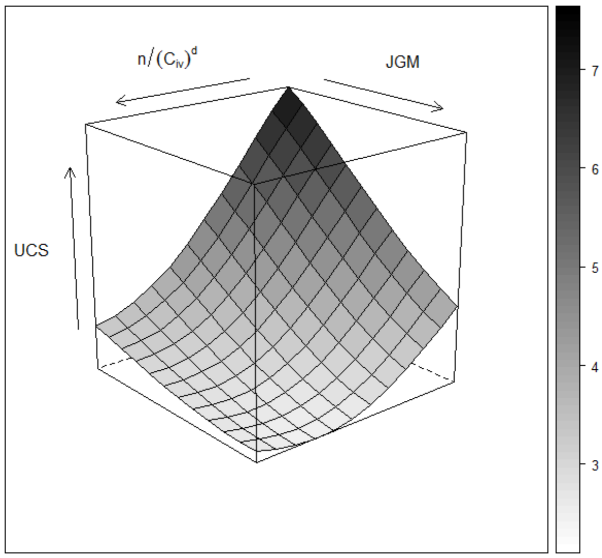


Figures 4: VEC curves for the more relevant variables, according to SVM model measured by 1-D SA

In figure 5 we can see that *JGM* is the variable with the biggest interaction with $n/(C_{iv})^d$ in UCS prediction. Plotting of UCS prediction by SVM model when these two variables are changed simultaneously (see Figure 6), keeping the remains at their means values it is possible to observe that the highest strength is reached when single fluid system is applied and is produced a soil cement mixture with lower values of $n/(C_{iv})^d$. These results coincide with those obtained from the interpretation of the VEC curves for these two variables.



Figures 5: 2-D interaction with $n/(C_{iv})^d$ variable in UCS prediction according to SVM model, quantified by a 2-D SA



Figures 6: VEC surface for $n/(C_{iv})^d$ and *JGM* in UCS prediction, according to SVM model measured by 2-D SA

When a similar procedure was carried out with *JGM* (second most relevant input variable), was possible observe that the age of the mixture is the variable with the strongest interaction with *JGM* (18%). In addition, it was possible to see that the highest values of UCS are reached when single fluid system is applied and *t* is high. Furthermore, it is possible observe that for double and mainly for triple fluid system UCS increase slightly with the age of the mixture.

FINAL REMARKS AND CONCLUSIONS

In the present study support vector machines (SVM) were used to explore jet grouting (JG) data, collected directly from JG columns (JGS), in order to predict its uniaxial compressive strength (UCS).

Although SVM model has experienced some difficulties to accurately estimate UCS of JGS over time, some important conclusions can be drawn. By performing a 1-D sensitivity analysis (SA), we have shown that the relation between the mixture porosity and the volumetric content of cement ($n/(C_{iv})^d$), JG method (*JGM*), cement content (*%C*) and age of the mixture (*t*), play an important role in UCS estimation over time. As expected, *%C* and *t* has a positive impact in UCS prediction, underlining the exponential relationship of the latter with UCS. In the other hand, $n/(C_{iv})^d$ and *JGM* have a negative impact in UCS estimation. It is appealing to observe that the *JGM* effect is approximately linear. Performing a 2-D SA, it was observed that the highest UCS values are reached for single fluid system and for lower values of $n/(C_{iv})^d$. In addition, UCS of JGS increases exponentially with *t* reaching the highest values for single fluid system. Moreover, when triple fluid system is applied UCS just slightly increases over time.

The knowledge obtained from the present study is a great contribute to understanding better the behavior of JG mixtures. As a result of this knowledge, the quality and the cost of JG treatment can be improved by controlling better parameters involved in JG process. As a future works, an attempt to improve the model predictive capability as well as its applicability will be done. In addition, SVM and other data mining algorithm will be applied to define predictive models of JG columns diameters as well as its stiffness.

ACKNOWLEDGMENTS

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.

REFERENCES

Bi, J. and Bennett, K. 2003. "Regression error characteristic curves". *In Proceedings of the Twentieth International Conference on Machine Learning* (Washington, DC, August 21-24). AAAI Press, 43-50.

Cherkassky, V. and Ma, Y. 2004. "Practical selection of SVM parameters and noise estimation for SVM regression." *Neural Networks*, Vol.17, No.1, 113-126.

Gomes Correia, A.; Valente, T.; Tinoco, J.; Falcão, J.; Barata, J.; Cebola, D. and Coelho, S. 2009. "Evaluation of mechanical properties of jet grouting columns using different test methods." *In Proceedings of 17th International Conference on Soil Mechanics and Geotechnical Engineering* (Alexandria, Egypt, October 5-9). IOS Press, 2169-2171.

Cortes, C. and Vapnik, V. 1995. "Support-vector networks." *Machine learning*, Vol.20, No.3, 273-297.

Cortez, P. 2010. "Data Mining with Neural Networks and Support Vector Machines using the R/rminer Tool." *In P. Perner (Ed.), Advances in Data Mining: Applications and Theoretical Aspects, 10th Industrial Conference on Data Mining. LNAI 6171* (Berlin, Germany, July 12-14). Springer, 572-583.

Cortez, P. and Embrechts, M. 2011. "Opening Black Box Data Mining Models Using Sensitivity Analysis." *In Proceedings of the 2011 IEEE Symposium on Computational Intelligence and Data Mining* (Paris, France, April 11-15). IEEE, 341-348.

Falcão, J.; Pinto, A. and Pinto F. 2000. "Case histories of ground improvement solutions using jet grouting." Geotechnical news, Tecnasol-FGE: Fundacões e Geotecnia S.A.

Goh, A. and Goh, S. 2007. "Support vector machines: Their use in geotechnical engineering as illustrated using seismic liquefaction data." *Computers and Geotechnics*, Vol.34, No.5, 410-421.

Hastie, T.; Tibshirani, R. and Friedman, J. 2009. *The elements of statistical learning: Data mining, inference, and prediction*. Springer-Verlag New York.

Lai, S. and Serra, M. 1997. "Concrete strength prediction by means of neural network." *Construction and Building Materials*, Vol.11, No.2, 93-98.

Lee, F.H.; Y. Lee; S.H. Chew and K.Y. Yong. 2005. "Strength and modulus of marine clay-cement mixes." Journal of geotechnical and geoenvironmental engineering, No.131(2), 178-186.

Narendra, B.S.; P.V. Sivapullaiah; S. Suresh and S.N. Omkar. 2006. "Prediction of unconfined compressive strength of soft grounds using computational intelligence techniques: A comparative study." Computers and Geotechnics, No.33(3), 196-208.

Nikbakhtan, B.; Ahangari, K. and Rahmani, N. 2010. "Estimation of jet grouting parameters in Shahriar dam, Iran." *Mining Science and Technology*, Vol.20, No.3, 472-477.

Rezania, M. and Javadi, A. 2007. "A new genetic programming model for predicting settlement of shallow foundations." *Canadian Geotechnical Journal*, Vol.44, No.12, 1462-1473.

Shibazaki, M. 2003. "State of practice of jet grouting." *In Proceedings of 3rd International Specialty Conference on Grouting and Ground Treatment* (New Orleans, Louisiana, February 10-12). ASCE, 198-217.

Smola, A. and Schölkopf B. 2004. "A tutorial on support vector regression." *Statistics and computing*, Vol.14, No.3, 199-222.

Smola, A. et al. 1996. "Regression estimation with sup- port vector learning machines". Master's thesis, Technische Universit at Munchen.

R Development Core Tem 2009. "R: A language and environment for statistical computing." R Foundation for Statistical Computing Vienna, Austria.

Tinoco, J.; Gomes Correia, A. and Cortez, P. 2011a. "Application of Data Mining Techniques in the Estimation of the Uniaxial Compressive Strength of Jet Grouting Columns over Time." *Construction and Building Materials*, Vol.25, No.3, 1257-1262.

Tinoco, J.; Gomes Correia, A. and Cortez, P. 2011b. "A Data Mining Approach for Predicting Jet Grouting Geomechanical Parameters." *In Proceedings of GeoHunan 2011: Rood Material and New Innovations in Pavement Engineering, Geotechnical Special Publication No. 23*, (Hunan, China, June 6-11). ASCE, 97-104.

BIOGRAPHY

JOAQUIM TINOCO was born in Braga, Portugal and went to the University of Minho, where he studied civil engineering and obtained his degree in 2008 (5-years graduation). Currently, he is developing the PhD works at University of Minho on "Application of Data Mining Techniques to Jet Grouting Columns Design". His e-mail address is jabtinoco@civil.uminho.pt.

ANTÓNIO GOMES CORREIA is full Professor at University of Minho and Director of Centre for Territory, Environment and Construction. His E-mail address is agc@civil.uminho.pt.

PAULO CORTEZ is Associate Professor at the Department of Information Systems of University of Minho and researcher at centro Algoritmi. His e-mail address is pcortez@dsi.uminho.pt and his Web page can be found at http://www3.dsi.uminho.pt/pcortez.