Universidade do Minho

Escola de Engenharia

Semana da Escola de Engenharia
October 24 - 27, 2011

SUPPORT VECTOR MACHINES IN MECHANICAL PROPERTIES PREDICTION OF JET GROUTING COLUMNS

Joaquim Tinoco[a], António Gomes Correia[a] and Paulo Cortez[b]

[a] C-TAC, Department of Civil Engineering, University of Minho

[b] Centre Algoritmi / Department of Information Systems, University of Minho

E-mail: jabtinoco@civil.uminho.pt

KEYWORDS

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

ABSTRACT

Strength and stiffness are 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, support vector machine (SVM), which is a data mining algorithm especially adequate to explore high number of complex data, can be used to learn the complex relationship between mechanical properties of JG samples extracted from real JG columns (JGS) and its contributing factors. In the present paper, the high capabilities of SVM in Uniaxial Compressive Strength (UCS) and Elastic Young Modulus estimation of JG laboratory formulations are summarized. After that, the performance reached by the same algorithm in the study of JGS are presented and discussed. It is shown, by performing a detailed sensitivity analysis, that the relation between mixture porosity and the volumetric content of cement, as well as the JG system are the key variables in UCS prediction of JGS. Furthermore, it is underlined the exponential effect of the age of the mixture in UCS estimation as well as the high iteration between these two key variables.

INTRODUCTION

Jet grouting (JG) technology is one of the most used soft-soil improvements methods to solve several geotechnical problems (Falcão et al. 2000; Gazzarrini et al. 2005). 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. The fluids are injected through small nozzles placed at the end of a rod that, after introduced at the intended depth, is continually rotated and slowly removed up to the surface. 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 a lot of variables involved in treatment process. Generally, the most important variables that affect the design of JG columns, namely its mechanical properties, are soil type, age of the mixture, mixture influx between soil and grout, exiting jet energy from nozzle, grout flow rate, rotating and lifting speed (Nikbakhtan et al. 2010b). All these variables 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; Nikbakhtan and Ahangari 2010a), 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.

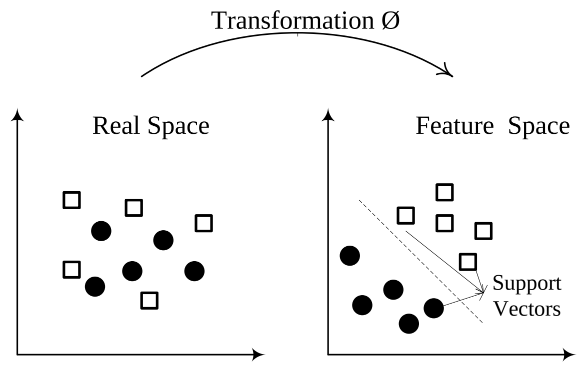
On the other hand, in the last few years some powerful tool, incorporating advanced statistic analysis, has been developed and are able to automatically extract important rules from vast and complex data. Such tools, usually known as data mining (DM) techniques, has been successfully applied in several scientific areas namely in Civil Engineering domain (Lai and Serra 1994; Rezania and Javadi 2007; Tinoco et al. 2011a).

One of the most interesting DM algorithms is the Support Vector Machines (SVM), which was used in the present work. This algorithm can be applied in both classification and regression problems. In a DM regression problem the main goal is to construct a regressor that works well on unseen examples, i.e. that generalises well. SVM is particularly useful to explore data with nonlinear relationships between several imputs and the target variable and had been successfully applied to solve geotechnical problems (Goh and Goh 2007; Tinoco et al. 2009; 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 procedure can be applied (Cortez and Embrechts 2011). Such procedure measures the response changes when a given input variable is varied through its domain, allowing understand how each variable affects the studied properties.

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 performing a one-dimensional sensitivity analysis. 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 two-dimensional sensitivity analysis was performed to the first two key variables. Previously, to these results, the high learning capabilities reached by SVM algorithm in UCS and Elastic Young Modulus at very small strain (E_0) of JG laboratory formulations (JGLF) are summarized.

SUPPORT VECTOR MACHINES

Support Vector Machines are a 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 (also known as backpropagation neural network), 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 (see Figure 1). 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.



Figures 1: Example of a SVM transformation (adapted from (Cortez 2010))

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 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). For example, if C is too large (infinity), then the objective is to minimize the empirical risk only, without regard to model complexity part in the optimization formulation. This parameter is usually determined experimentally (trial and error) via the use of a training and test (validation) set. 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. The bigger ε, the fewer support vectors are selected. On the other hand, bigger ε-values results in more 'flat' estimates. 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 also should reflect distribution of input (x) values of the training data. The most kernel functions commonly used are the Gaussian and the Polynomial kernel.

In the present work was adopted the popular Gaussian kernel, since it presents less parameters than other kernels (e.g. polynomial):

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

To reduce the search space, we adopt the heuristics proposed in (Cherkassy and Ma 2004) to set the complexity penalty parameter, C=3 and the size of the insensitive tube $\varepsilon = \hat{\sigma}/\sqrt{N}$, where $\hat{\sigma} = 1.5/N \times \sum_{i=1}^{N}(y_i - \hat{y}_i)^2$ and $\hat{y}_i$is the value predicted by a 3-nearest neighbor algorithm. The most important SVM parameter, the kernel parameter γ (Hastie et al. 2008), 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. 2008).

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. 2008).

MODEL ASSESSMENT AND INTERPRETATION

Evaluation Measures

In a 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 main difference between RMSE and MAD is that former is more sensitivity to extreme values. 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.

Generalization Capacity

The overall generalization performance of the trained model was accessed by using 20 runs under 10-fold cross validation approach (Hastie et al. 2008). Under this scheme, the data are divided in 10 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 10 times more computation, because 10 models must be fitted. The final generalization estimate is evaluated by computing the MAD, RMSE and R^2 metrics for all N test samples.

Sensitivity Analysis

Besides obtaining a high predictive performance it is also important to extract human understandable knowledge from the data-driven model. For such purpose, a sensitivity analysis procedure (Cortez and Embrechts 2011) was performed. Such procedure is applied after the training phase and analyzes the model responses when a given input is changed. This procedure can be applied to any supervised DM model and allows to quantify the relative importance of each input parameter and also to measure the average effect of a given input on the target variable. Such quantification is determined by successively holding all inputs at a given baseline (e.g. 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. Higher response changes indicate a more relevant input. In particular, following the results achieved in (Cortez and Embrechts 2011), we adopt the gradient measure (S_a) to access the input relevance (R_a) of the attribute x_a (the higher the gradient, the higher is the input importance):

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

For more input influence details, the Variable Effect Characteristic (VEC) curve was plotted (Cortez and

Embrechts 2011). For a given input variable, the VEC curve plots the attribute L level values (x-axis) versus the sensitivity analysis responses (y-axis). In this paper, we set L=12. Furthermore, aiming to achieve a more realistic interpretation of the models a two-dimensional sensitivity analysis 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 (Cortez and Embrechts 2011).

JET GROUTING DATA

To train and test SVM algorithm was used a dataset composed by 288 results. 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 was measured in unconfined compression tests with on sample strain instrumentation (Gomes Correia et al. 2009). The input variables were selected 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 soil-cement mixture (e); cement content ($\%C$); water content (ω, %) and water/cement ratio (W/C). The main statistics of the input variables are shown in Table 1.

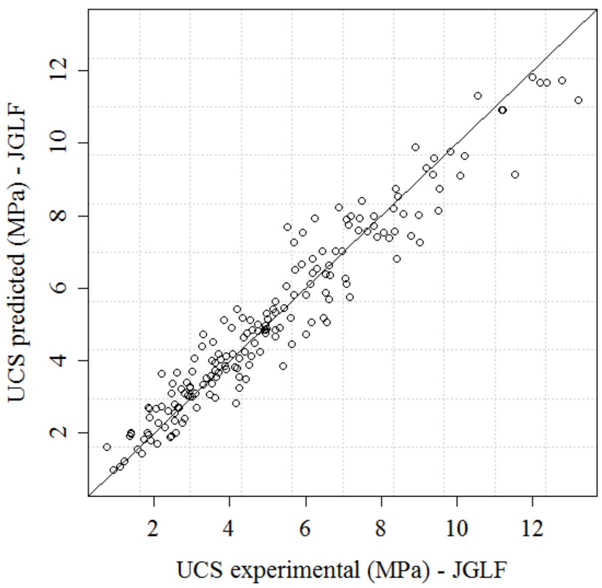
Table 1: Synopsis of the main statistics of the input variables

Variable	Min	Max	$\bar{u}$	σ
$n/(C_{iv})^d$	37.88	78.61	58.14	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	0.83
$\%C$	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

As previously referred all samples were collected from different columns but constructed under the same soil type. After some laboratory tests the soil under treatment was classified as lean clay - CL, which is composed by 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.

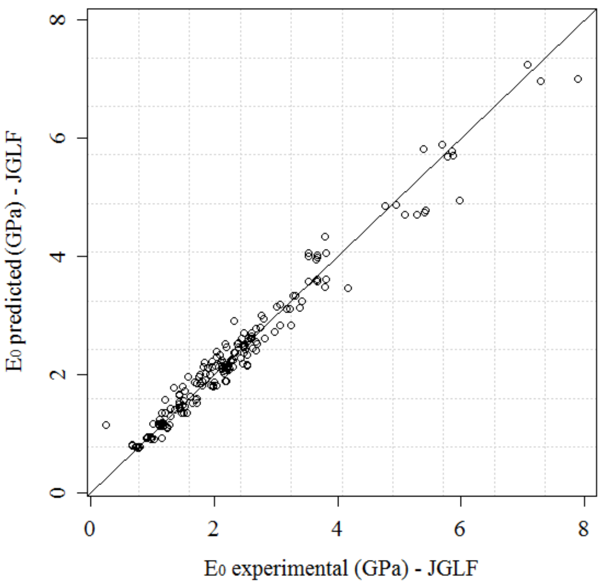
RESULTS AND DISCUSSION

Before analyze and discusses the results obtained in the study o UCS of JGS, we will summarize the main ones achieved in previous works where SVM was used to predict the mechanical properties i.e., UCS and E_0, of JGLF. When working with this kind of material, SVM is able to predict both UCS (Tinoco et al. 2011a) and E_0 (Tinoco et al. 2011b) with high accuracy. Figures 2 versus 3 show the excellent relation between the observed and predicted values of each example, which is assessed by proximity of all points to the diagonal line that represents the perfect prediction, respectively for UCS and E_0.

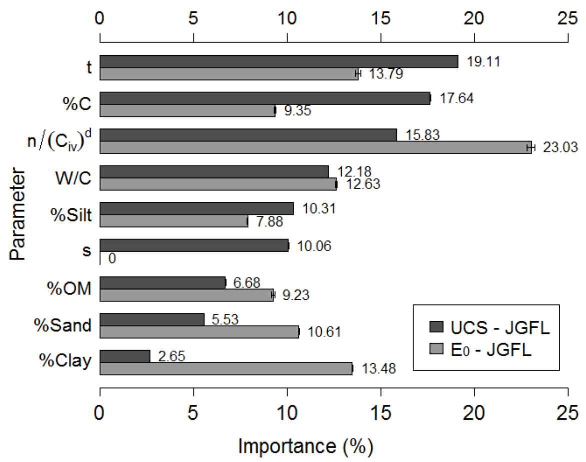


Figures 2: Relationship between experimental and predicted (SVM model) UCS values from JGLF samples (adapted from Tinoco et al. 2011a)

Performing a one-dimensional sensitivity analysis, the age of the mixture and the relation between the mixture porosity and volumetric content of cement were identified as key parameters in both mechanical properties estimation of JGLF (see Figure 3). In addition, to reliably predict UCS over time of JGLF, the cement content should be included. In the other hand, in E_0 estimation the soil properties, namely its clay content, should also be considered.

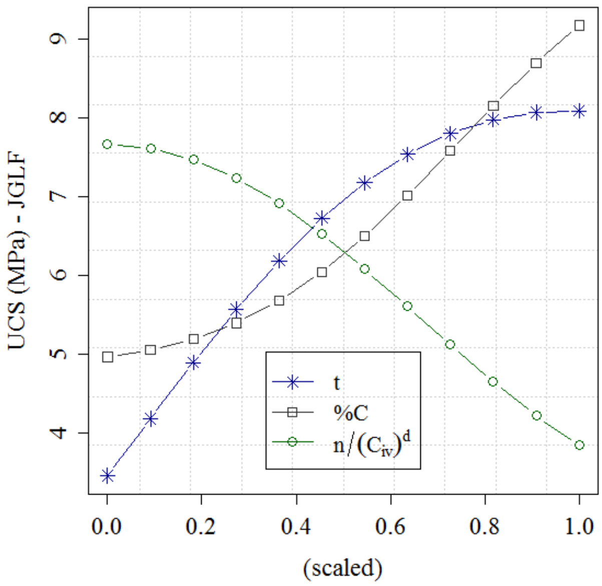


Figures 3: Relationship between experimental and predicted (SVM model) E_0 values from JGLF samples (adapted from Tinoco et al. 2011b)



Figures 4: Comparison of the relative importance of each variable in UCS and E_0 of JGLF (SVM model), by performing one-dimensional sensitivity analysis

Analyzing the effect of key variables in UCS prediction (see Figure 5) of JGLF, it was possible to observe that its strength increase exponentially with age of the mixture. An opposite effect was observed for $n/(C_{iv})^d$, where UCS decrease when such relation increase (Tinoco et al. 2011b). The effect of these two variables is very similar in E_0 estimation of JGLF. Furthermore, and as expected, the %C has a positive impact in UCS estimation, mainly for cement content high than 40% (see Figure 5). In the other hand, JGLF stiffness decreases when clay content of the treated soil increases.

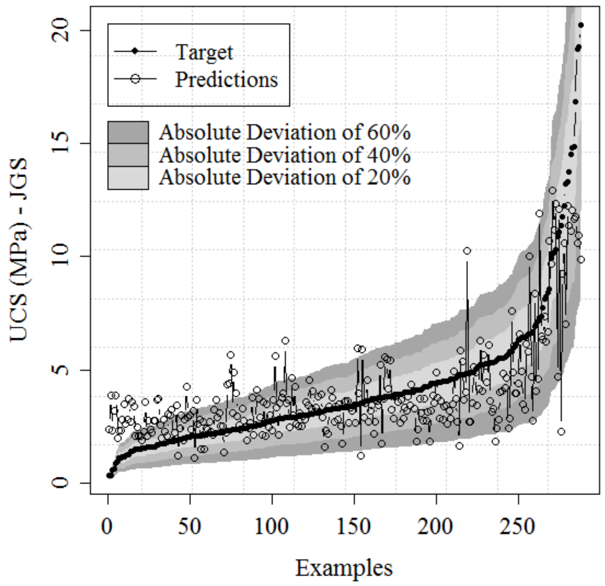


Figures 5: VEC curves of three key variables in UCS prediction of JGLF (SVM model), by performing one-dimensional sensitivity analysis (adapted from Tinoco et al. 2011b)

After these previous findings the same framework was applied for the prediction of UCS of JGS, described and discussed herewith. Then, the key input variables are identified by performing a one-dimensional sensitivity analysis procedure. The effect of each key variable in UCS prediction of JGS is analyzed and discussed. In addition, a two-dimension sensitivity analysis procedure was applied in order make a more realistic interpretation of key variables effect and the iteration between them.

On Figure 6 we compare the UCS observed 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

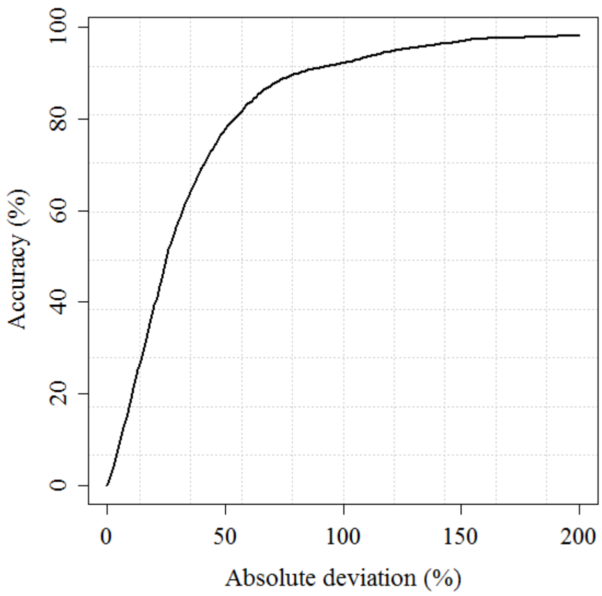
(0.63±0.01) and the values for MAD and RMSE metrics are 1.29±0.01 MPa and 1.87±0.02 MPa respectively. Observing the REC curve on Figure 7, which shows the accuracy obtained for a given absolute deviation (in percentage) we can observe a fast improvement on model accuracy. For example, to guaranty that the model is able to predict successfully 80% of the examples an error of 60% should be tolerated.



Figures 6: Relationship between experimental and predicted (SVM model) UCS values from JGS

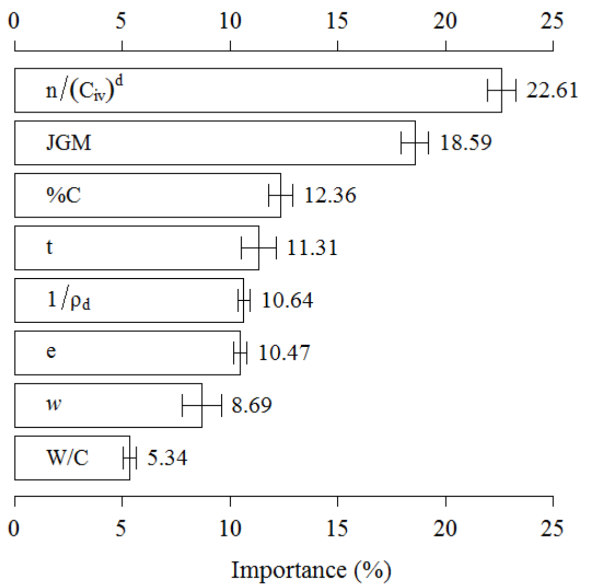
When evaluating a DM model, we should consider not only predictive accuracy but also model interpretability. In this paper, such interpretability is based on measuring which are the key input variables and how these affect the predicted output. To do so a one and two-dimensional sensitivity analysis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

Figure 8 give us an idea of the relative importance of each variable after applies a one-dimensional sensitivity analysis, 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 more relevant one, 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 according to SVM model is in agreement with the empirical knowledge. Among these four key variables, we can identify one that is related to JG process 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.

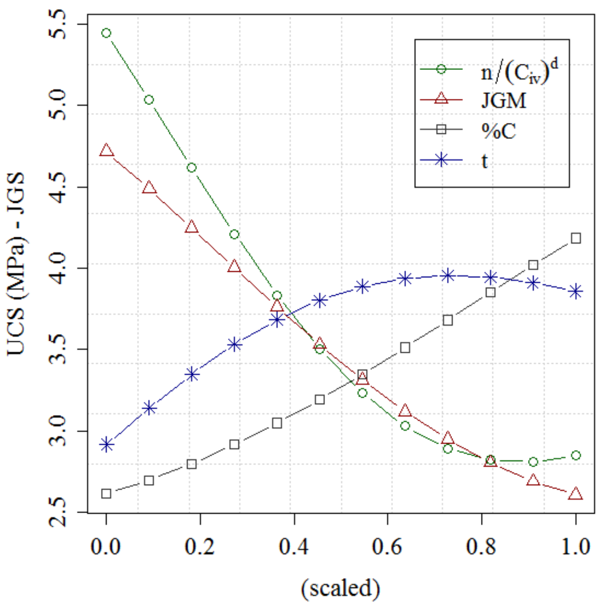


Figures 7: REC curve for SVM model in UCS prediction of JGS

Based also on one-dimensional sensitivity analysis, the effect of each variable on UCS prediction was quantified. The VEC curves of the four key variables previously identified are shown in Figure 9. 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 increase quickly in the early ages and after that tend to stabilize. This is the normal effect of the age in 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.



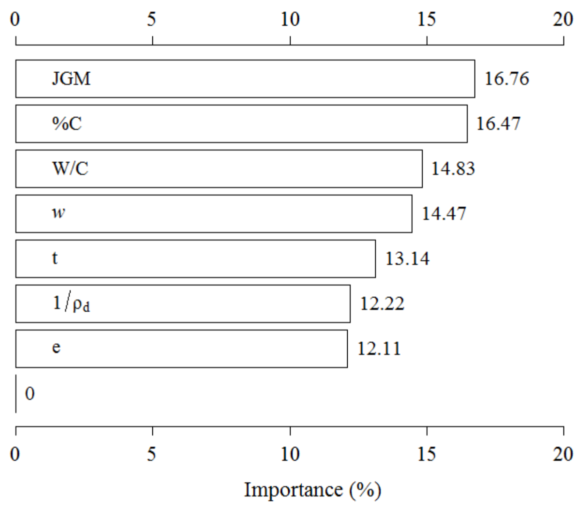
Figures 8: Relative importance of each variable in UCS prediction of JGS (SVM model,), quantified by one-dimensional sensitivity analysis



Figures 9: VEC curves for the more relevant variables in UCS prediction of JGS (SVM model), measured by one-dimensional sensitivity analysis

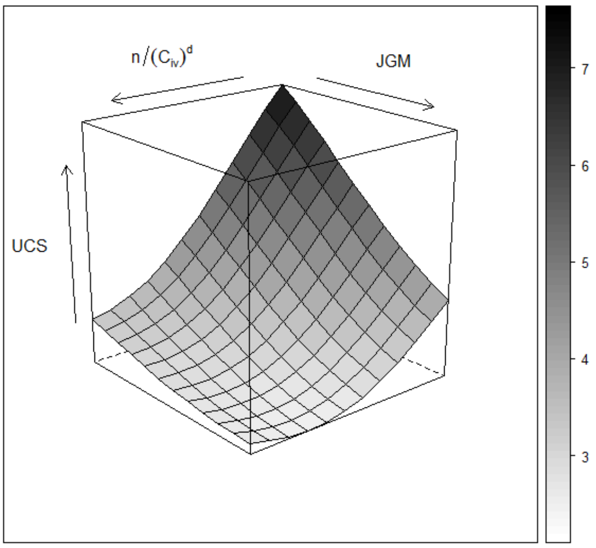
All observation previously exposed based on a one-dimensional sensitivity analysis, such as the relative importance of each variable or VEC curves, are very usefully to understand the behavior of JG mixtures. However, in this kind of analysis all variables are fixed in its mean value except one that is ranged from its minimum to maximum values. In real works this normally never happen. Therefore, in order to perform a more realistic analysis, a two-dimensional sensitivity analysis was carried out. In the next lines, the results of such analysis for $n/(C_{iv})^d$ and *JGM* (the first two key variables) are shown and discussed.

On Figure 10 we can see that *JGM* is the variable with the biggest iteration with $n/(C_{iv})^d$ in UCS prediction. Plotting of UCS prediction by SVM model when these two variables are changed simultaneously (see VEC surface on Figure 11), 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.

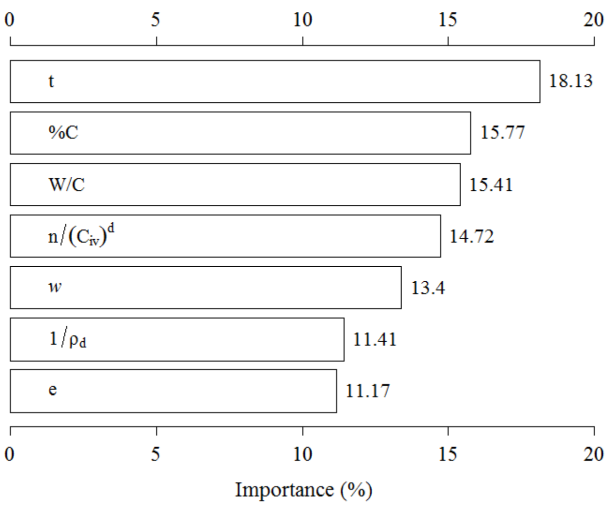


Figures 10: 2-D iteration with $n/(C_{iv})^d$ variable in UCS prediction of JGS (SVM model), quantified by two-dimensional sensitivity analysis

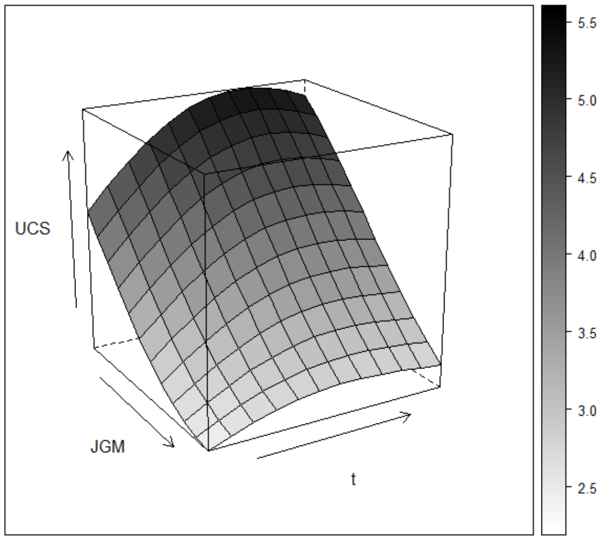
Now, if a similar analysis with *JGM* (second most relevant input variable) were performed, Figure 12 shows that the age of the mixture is the variable with the strongest iteration with *JGM* (18%). Observing the VEC surface for *JGM* and *t* iteration (Figure 13) we can see, that the highest values of UCS are reached when single fluid system is applied and *t* is high. Still, it is possible observe that for double fluid system and mainly for triple fluid system UCS increase slightly with the age of the mixture.



Figures 11: VEC surface for $n/(C_{iv})^d$ and *JGM* in UCS prediction of JGS (SVM model), measured by two-dimensional sensitivity analysis



Figures 12: 2-D iteration with *JGM* variable in UCS prediction of JGS (SVM model), quantified by two-dimensional sensitivity analysis



Figures 13: VEC surface for *JGM* and *t* in UCS prediction of JGS (SVM model), measured by two-dimensional sensitivity analysis

FINAL REMARKS AND CONCLUSIOS

The prediction of uniaxial compressive strength (UCS) of jet grouting (JG) samples extracted directly from JG columns (JGS) is a complex task that can be simplified through application of data mining (DM) techniques. In the present study support vector machines (SVM) were used to explore JG data in order to predict its UCS of JG samples collected from real JG columns. Furthermore, the main results of previous studies, where SVM was trained with JG laboratory formulations (JGLF), were also summarized.

The prediction of mechanical properties, i.e., UCS and Elastic young Modulus at early stage (E_0), of JGLF, can be easily and accurately done by SVM algorithm. Age of the mixture (*t*) and the relation between the mixture porosity and the volumetric content of cement ($n/(C_{iv})^d$) were identified as key variables in both mechanical properties prediction. As expected, the former has a positive effect in such prediction following an exponential behavior. On the other hand, $n/(C_{iv})^d$ has a negative impact in either UCS or E_0 estimations.

In regards of JGS, SVM experienced some difficulties to accurately estimate UCS of JG mixtures over time. However, some conclusions can be drawn. By performing a one-dimensional sensitivity analysis, it is shown that $n/(C_{iv})^d$, JG method (*JGM*), cement content (*%C*) and *t*, play an important role in UCS estimation of JGS over time. As expected, *%C* and *t* has a positive impact in UCS prediction. The former shows a roughly linear relationship with UCS while the latter has an exponential behavior. 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 two-dimensional sensitivity analysis, where $n/(C_{iv})^d$ and *JGM* are changed simultaneously we observe that the highest UCS values are reached for single fluid system and lower values of $n/(C_{iv})^d$. A similar conclusion was stated by performing a two-dimensional sensitivity analysis with *JGM* and *t*. It was observed that UCS of JG samples increase 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 contribution to better understand the behavior of JG mixtures. On the one hand, the number of JGLF can be drastically reduced, and on the other hand, the knowledge about JGS behavior was also improved. As a result of this knowledge, the quality and the cost of JG treatment can be improved by better controlling all parameter involved in JG process.

As a future works, we will try to improve the predictive capability of SVM model trained with JGS by combining other variables or subdividing the data base on clusters. Furthermore, the model applicability will be extended to other types of soils. In addition, SVM algorithm and other DM techniques will be applied to define predictive models of JG columns diameters as well as its stiffness.

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.

REFERENCES

Bi, J. and K. Bennett. 2003. “Regression Error Characteristic Curves”. *In Proceedings of the Twentieth International Conference on Machine Learning (CD-ROM)* (WashingtonDC, USA, Aug.21-24), AAAI Press.

Cherkassy, V. and Y. Ma. 2004. “Practical Selection of SVM Parameters and Noise Estimation for SVM Regression.” *Neural Networks*, No.17(1), 113-126.

Cortes, C. and V. Vapnik. 1995. “Support-Vector Networks.” *Machine Learning*, No. 20(3), 273-297.

Cortez, P. 2010. “Data Mining with Neural Networks and Support Vector Machines Using the R/rminer Tool”. *In Advances in Data Mining - Applications and Theoretical Aspects, 10th Industrial Conference on Data Mining, Lecture Notes in Artificial Intelligence 6171* (Sydney, Autralia, Dec.14-17), IEEE, 572-583.

Cortez, P. and M. Embrechts. 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, Apr.11-15), IEEE, 341-348.

Falcão, J.; A. Pinto and F. Pinto. 2000. “Cases Histories of Ground Improvement Solutions using Jet-Grouting.” Research Report, Tecnasol FGE: Fundacoes e Geotecnia S.A.

Gazzarrini, P.; M. Kokan and S. Jungaro. 2005. “Case History of Jet Grouting in British Columbia. Underpinning of CN Rail Tunnel in North Vancouver.” Research Report, The Grout Line.

Goh, A. and S. Goh. 2007. “Support Vector Machines: Their use in Geotechnical Engineering as Illustrated using Seismic Liquefaction Data.” *Computers and Geotechnics*, No.34(5), 410-421.

Gomes Correia, A.; T. Valente; J. Tinoco; J. Falcão; J. Barata; D. Cebola and S. Coelho, “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, Oct.5-9), IOS Press, 2179-2171.

Hastie, T.; R. Tibshirani and J. Friedman. 2008. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer-Verlag.

Lai, S. and M. Serra. 1997. “Concrete Strength Prediction by Means of Neural Network.” *Construction and Building Materials*, No.11(2), 93-8.

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. and K. Ahangari. 2010a. “Field study of the influence of various jet grouting parameters on soilcrete unconfined compressive strength and its diameter.”

International Journal of Rock Mechanics and Mining Sciences, No.47, 685-689.

Nikbakhtan, B.; K. Ahangari and N. Rahmani. 2010b. "Estimation of Jet Grouting Parameters in Shahriar Dam, Iran." *Mining Science and Technology*, No.20(3), 472-477.

R Development Core Team. 2009. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Rezania, R. and A. Javadi. 2007. "A New Genetic Programming Model for Predicting Settlement of Shallow Foundations." *Canadian Geotechnical Journal*, No.44(12), 1462–73.

Shibazaki, M. 2004. "State of Practice of Jet Grouting, in Grouting and Ground Treatment". *Grouting and Ground Treatment*, 198-217.

Smola, A. 1996. "Regression Estimation with Support Vector Learning Machines." Research Report, Technische Universitat Munchen.

Smola, A.J. and B. Schölkopf. 2004. "A Tutorial on Support Vector Regression." *Statistics and Computing*, No.14(3), 199-222.

Tinoco, J.; A. Gomes Correia and P. Cortez. 2009. "A Data Mining Approach for Jet Grouting Uniaxial Compressive Strength Prediction". *In Proceedings of World Congress on Nature & Biologically Inspired Computing* (Coimbatore, India, Dec.9-11). IEEE, 553-558.

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

Tinoco, J.; A. Gomes Correia and P. Cortez. 2011b. "A Data Mining Approach for Predicting Jet Grouting Geomechanical Parameters". *In Proceedings of GeoHunan International Conference II: Emerging Technologies for Design, Construction, Rehabilitation, and Inspections of Transportation Infrastructures: Rood Material and New Innovations in Pavement Engineering, Geotechnical Special Publication No. 23* (Hunan, China, Jun.9-11). ASCE, 97-104.

AUTHORS' BIOGRAPHIES

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". He is interested to soft soil improvement methods, namely jet grouting technology, soil cement mixtures behavior and data mining techniques. 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. He is supervisor of the PhD thesis. E-mail address agc@civil.uminho.pt.

PAULO CORTEZ is Associate Professor at the Department of Information Systems of U. Minho researcher at centro Algoritmi, with interests in the fields of: business intelligence, data mining, neural networks, evolutionary computation and forecasting. Currently, he is an associate editor of the Neural Processing Letters journal and he participated in 7 R&D projects (principal investigator in 2). His research has appeared in Journal of Heuristics, Journal of Decision Systems, Artificial Intelligence in Medicine, Neurocomputing, Neural Processing Letters, and others. His e-mail address is pcortez@dsi.uminho.pt and his Web-page can be found at http://www3.dsi.uminho.pt/pcortez.