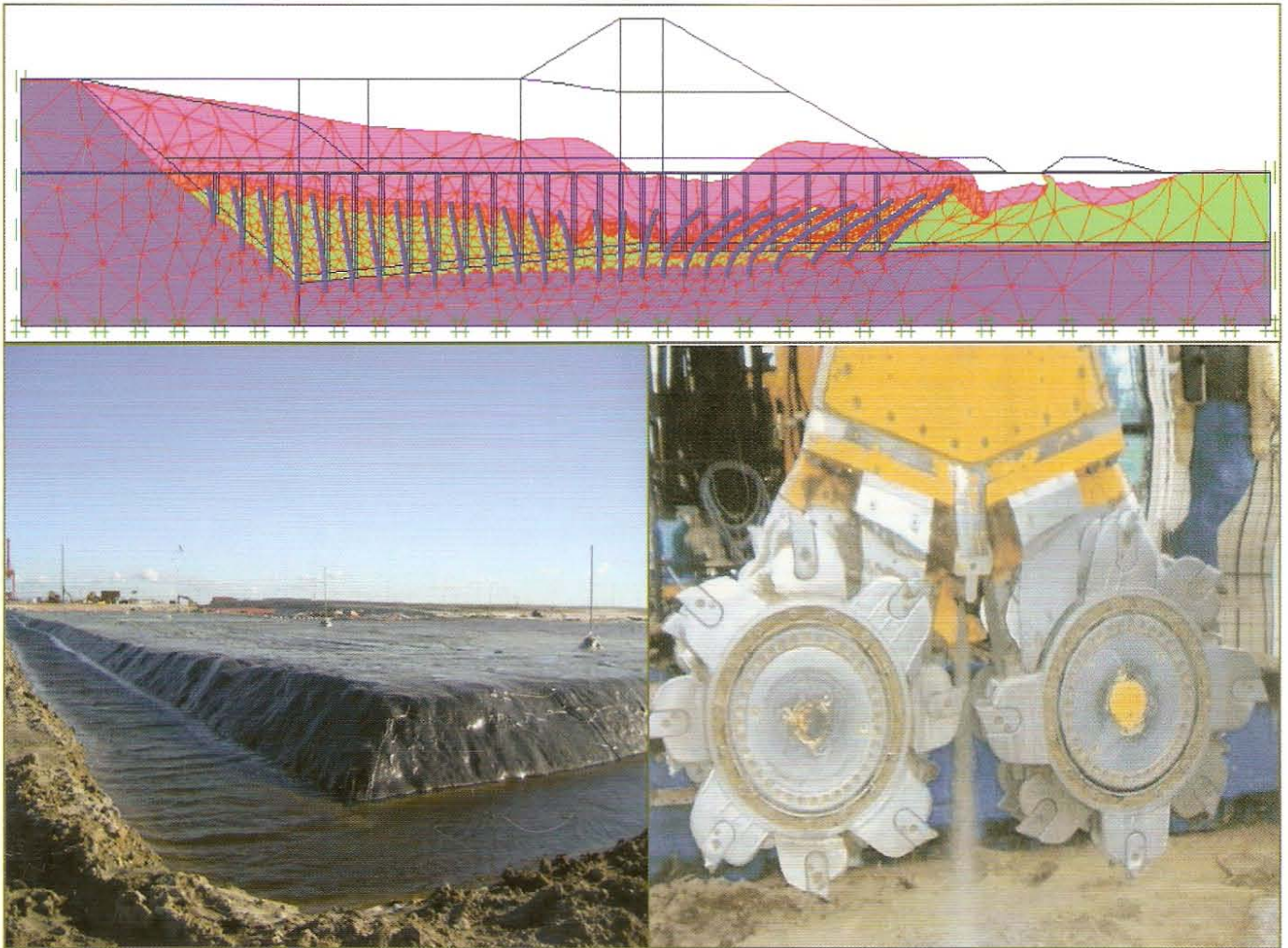


PROCEEDINGS OF THE INTERNATIONAL SYMPOSIUM

TC 211

IS-GI Brussels 2012



Recent Research, Advances & Execution Aspects of GROUND IMPROVEMENT WORKS

31 May – 1 June 2012, Brussels, BELGIUM

VOLUME IV

Organised by:

ISSMGE Technical Committee TC 211 Ground Improvement

Edited by:

Nicolas Denies & Noël Huybrechts

Belgische Groepering voor Grondmechanica en Geotechniek
Groupement Belge de Mécanique des Sols et de la Géotechnique

Comité Français de Mécanique des Sols

SIMSG ISSMGE



Application of a sensitivity analysis procedure to interpret uniaxial compressive strength prediction of jet grouting laboratory formulations performed by SVM model

Joaquim Tinoco, University of Minho, Portugal, jabinoco@civil.uminho.pt
António Gomes Correia, University of Minho, Portugal, agc@civil.uminho.pt
Paulo Cortez, University of Minho, Portugal, pcortez@dsi.uminho.pt

ABSTRACT

Jet Grouting (JG) technology is one of the most widely used soft-soil improvements methods around the world. When compared with other methods, JG versatility is highlighted, since it can be applied to several soil types, creates elements with different geometric shapes and, normally, is less expensive. However, due to inherent geological complexity of soil and high number of variables involved in JG process, JG design is a hard task. Nowadays, its design is essentially based on empirical rules that are often too conservatives, compromising the economy and the quality of the treatment. In the present study, data mining (DM) techniques, particularly the high learning capabilities of support vector machines (SVM) algorithm were used to predict uniaxial compressive strength (UCS) of JG laboratory formulations over time. Furthermore, and by performing a detailed sensitivity analysis, some important information was extracted based on the learned model. The high performance achieved by SVM algorithm in UCS prediction is summarized showing the high predictive accuracy reached ($R^2=0.93$). In addition, after apply a one- and two-dimensional sensitivity analysis, an important explanation of the model is given in terms of what are the key variables and its effect on UCS estimation, as well as the interaction level between input variables. Hence, it is shown that age of the mixtures, cement content and the relation between mixture porosity and volumetric content of cement have a high influence in strength behavior of JG laboratory formulations. Furthermore, was also possible to observe that water/cement ratio is the variable with higher interaction with age of the mixture and cement content.

1. INTRODUCTION

For constructions purposes the soil foundation should presents suitable geotechnical properties. However, due to increased urbanization and industrialization the availability of soils with such requirements is increasingly scarce. As a result, there is a growing demand for construction on soft soil, which is characterized by high plasticity, high fraction of fines and void ratio, low strength and high compressibility (Liu et al., 2008). In these cases, deep mixing methods are currently applied to improve soil properties, where powder cement or slurry cement is injected into the natural soil. Nowadays there are several soil improvements methods (Nikbakhtan et al., 2010) where Jet Grouting (JG) is highlighted. JG technology was first introduced by Yamakoda brothers in 60's decade and since them has been widely applied (Xanthakos et al., 1994). Such technology has aroused interest within the geotechnical community due to it great versatility. JG enable to improve mechanical and physical properties of different types of soil (ranging from coarse to fine-grained soils), obtain different geometries shapes (columns, panels), just require few equipments for its application and is economically attractive when compared with other similar methods.

The concept behind JG technology is to produce a soil-cement mixture, often termed as *soilcrete*, with an enhancement in terms of mechanical characteristics (strength and stiffness) and permeability. The new material is obtained by injecting grout, with or without other fluids (e.g. air or water) at high pressure and velocity into the sub soil. The fluids are injected through small nozzles placed at the end of a rod that, after introduced at the intend depth, is continually rotated and slowly removed up to the surface. Depending of the number of fluids injected, three systems are currently in use (Xanthakos et al., 1994). Single fluid system, where is only injected grout at high pressure and velocity; double fluid system is very similar to the single system but with the addition of an air shroud to the nozzle; and finally the triple fluid system, which differs from the single and double systems in that the erosion of the ground is carried out by a high pressure water jet shrouded with air with an additional low pressure grout line. In addition to these three systems there is also the Xjet system, also known as cross-jet or collided jetting that consists of a pair of intersecting air-shrouded water jets with separate grout jets and is designed to cut a nominal 2m diameter column in any ground (Shibazaki, 2004).

JG technology has been applied for different purposes, namely in the improvement of foundation soils, slope stabilization, and underpinning (Welsh and Burke, 1991). Furthermore, was also used as a method of performing block stabilization of contaminated soils as well as for the formation of barriers for the control of contaminant migration in the environmental field (Gazaway and Jasperse, 1992). In spite of being widely used, JG design lacks of rational approaches to predict its physical and mechanical properties. Indeed, even in large scale works, JG design is essentially based on empirically methods, as illustrated in Figure 1 and supported by knowhow of JG companies. The main reasons for such absence are related to the number of variables involved in JG process, which present complex or even nonlinear relationships, as well as due to inherent heterogeneity of the natural soil. As a result, and since such empirical rules are often too conservative, the quality and cost of the treatment can be compromised. Hence, and bearing in mind the high versatility of JG technology and its role in important geotechnical works, there is need to develop rational and reliable methods to estimate accurately JG mechanical and physical properties.

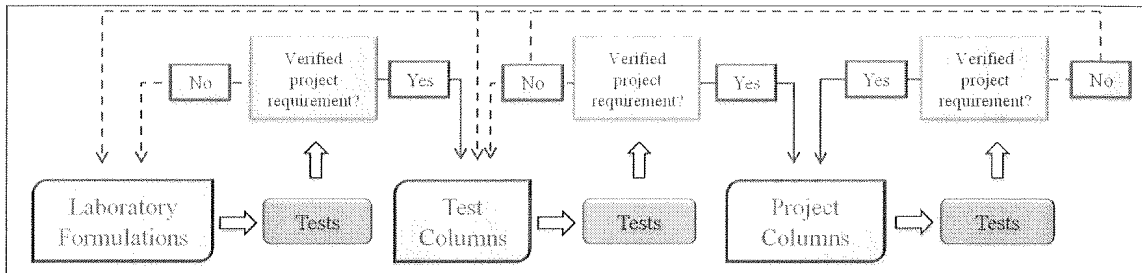


Figure 1: Empirical design of JG treatment.

One of the best ways to deal with high number of variables with complex relationship is to use an automate processes using artificial intelligent tools that analyze the raw data and extract useful knowledge. These tools can be seen as a good alternative to the traditional statistical approaches, which are unable to deal with such information, leading to poor results. Knowledge Discovery in Databases (KDD) (see Figure 2) is an iterative process that consists of several steps, where Data Mining (DM) is just one step (although crucial), aiming at the extraction of useful patterns from the observed data. According to CRISP-DM methodology, the following steps should be followed (Chapman et al., 2000): 1 - business understanding; 2 - data understanding; 3 - data preparation; 4 - modeling; 5 - evaluation; and 6 - deployment.

Two important DM tasks, classification and regression, require a supervised learning, where a model is adjusted to a dataset of examples that map I inputs (independent variables) into a given target (the dependent variable). There are several DM algorithms, each one with its own proposes and capabilities. Some of the most interesting supervised learning DM models are Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs), characterized by their flexibility and nonlinear learning capabilities. In the Geotechnics field, there are some successful applications of DM to different kinds of problems. Miranda et al. (2011) propose a new alternative regression models using ANNs for the analytical calculation of strength and deformability parameters of rock masses. Goh and Goh (2007) used SVMs to assess seismic liquefaction. Erzin (2007) studied the relationship between the swell pressure and soil suction behavior in specimens of Bentonite-Kaolinite clay mixtures with varying soil properties using ANNs. Tinoco et al. (2011b) applied SVMs and ANNs to predict deformability properties of JG laboratory formulations over time.

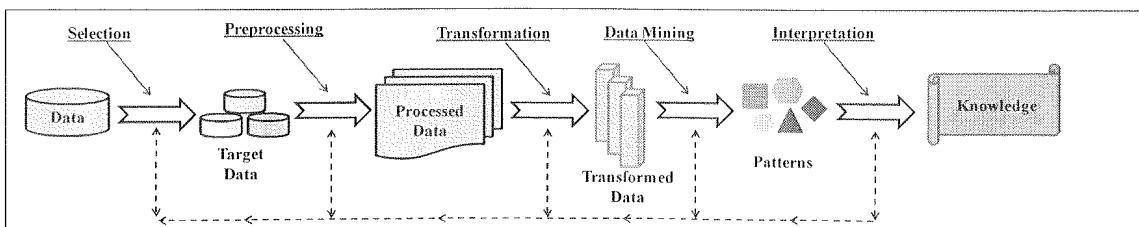


Figure 2: Steps of KDD process (adapted from Fayyad et al., 1996).

In the present work, a first step was made, by proposing a rational approach to predict Uniaxial Compressive Strength (UCS) of JG laboratory formulations (JGLF) over time, which is the mechanical properties currently used to measure the bearing capacity of foundations. For such purpose, an SVM was trained with UCS data collected from JGLF samples, in order to predict its strength over time. The main

results are here summarized and interpreted. Furthermore, an interpretability of SVM model is given by performing a detailed sensitivity analysis procedure. Such analysis allow us identify what are the key variables in UCS prediction of JGLF as well as a better understanding of the effect of each one in such prediction. In addition, was also measured the interaction between variables and interpreted its means and effect on UCS prediction.

2. METHODS AND DATA

2.1. Support Vector Machines

For regressions tasks there are several DM algorithms that can be applied, each one with its own advantage and limitations. The present study focuses on SVM algorithm that has been shown high learning capabilities even when working with complex data (Tinoco et al., 2011c). Support Vector Machines (Chen and Councill, 2003) are a very specific class of algorithms, based on statistic theory, characterized by use of kernels, absence of local minima during the learning phase 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 multilayer perceptron (the most popular neural network type), SVM represents a significant enhancement in functionality. The supremacy of SVM lies in their use of nonlinear kernel functions that implicitly map inputs into high dimensional feature spaces. In this feature spaces linear operations may be possible that try to find the best linear separating hyperplane ($y_i = \omega_0 + \sum_{i=1}^m \omega_i \phi_i(x)$), 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 (see Figure 3). These attractive features and promising empirical performance are responsible for its 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).

When working with SVM, 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 choosing a good parameter setting in a learning task is the so-called model selection. 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 (Cherkassky and Ma, 2004). In the present work was adopted the popular Gaussian kernel, since it presents less parameters than other kernels (e.g. polynomial):

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

For model selection, several approaches have been proposed (Huang et al., 2007; Momma and Bennett, 2011; Frohlich and Zell, 2005) in order to find the best set of parameters with less effort (time and computing consuming). In the present work, 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 $\epsilon = \hat{\sigma}/\sqrt{N}$, 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 $\{2^{-15}, 2^{-13}, \dots, 2^3\}$, under an internal (i.e. applied over training data) 10-fold cross validation (Hastie et al., 2009).

All experiments were implemented in R tool (R Development Core Team), using rminer library (Cortez, 2010), which is particularly suitable for SVM training through use advanced functions (particularly "mining" and "fit"). 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).

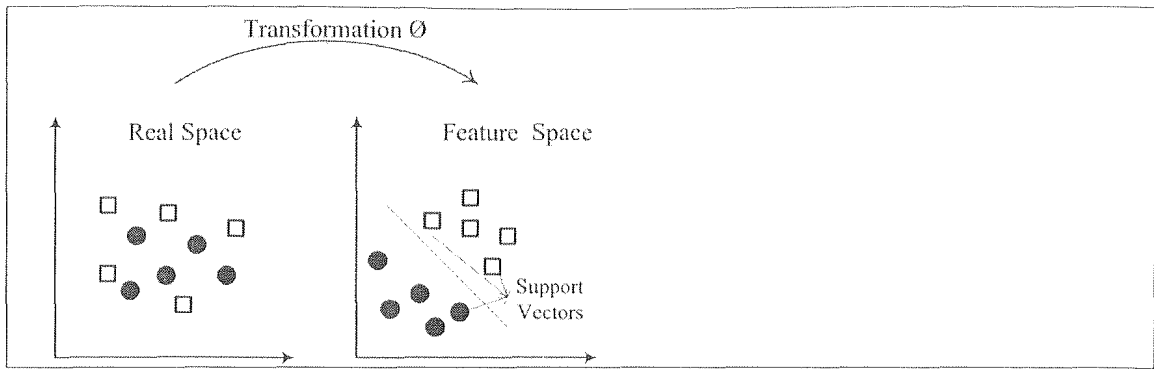


Figure 3: Example of a SVM transformation (adapted from (Cortez, 2010)).

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: Mean Absolute Deviation (MAD), Root Mean Squared Error (RMSE) and Coefficient of Correlation (R^2). These metrics are defined as:

$$MAD = \frac{\sum_{i=1}^N |y - \hat{y}|}{N}; \quad RMSE = \sqrt{\frac{\sum_{i=1}^N (y - \hat{y})^2}{N}}; \quad R^2 = \left(\frac{\sum_{i=1}^N (y - \bar{y}) \times (\hat{y} - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^N (y - \bar{y})^2 \times \sum_{i=1}^N (\hat{y} - \bar{\hat{y}})^2}} \right)^2 \quad (2)$$

If the first two metrics show lower values and R^2 is close to the unit value, this means high predictive capacity of the model. The main difference between RMSE and MAD is that former is less sensitive 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 model performance.

The overall generalization performance of the trained model was accessed by using 20 runs under Leave-One-Out approach (Hastie et al., 2009), which is special suitable for small databases (e.g. lower than 100 examples). Under this scheme, one example at a time is used to test the model, and the remaining data is used to fit the model. At the end all data are used for training and testing. Yet, this method requires approximately N times more computation, because N 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 obtaining a high predictive performance it is also important to extract human understandable knowledge from the data-driven model. This is precisely the main drawback of such models. One of the best ways to “open” a DM model is to apply a sensitivity analysis procedure (Cortez and Embrechts, 2011). Such procedure is carried out 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 ($x_a \in \{x_1, \dots, x_l\}$), with ($j \in \{1, \dots, L\}$) 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 is the input importance):

$$R_a = S_a / \sum_{i=1}^l S_i \times 100(\%), \quad \text{where } S_a = \sum_{j=2}^L |\hat{y}_{a,j} - \bar{\hat{y}}_{a,j-1}| / (L - 1) \quad (3)$$

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 model 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).

2.2. Jet Grouting Data

During the learning phase of SVM algorithm was used a dataset with 175 records assembled from a laboratory experimental program, carried out at University of Minho, with the purpose of analyzing the influence of several variables in mechanical behavior of JG laboratory mixtures. Should be stressed that the acquisition of each data example requires considerable cost and amount of time, demanding laboratory work and material (e.g. cement). All records can be grouped in 9 different formulations characterized by its values of Water/Cement ratio (W/C), cement content (mass of cement/(mass of

soil+mass of cement)) (C), cement type (s) and soil properties, i.e., percentage of sand ($\%Sand$), percentage of silt ($\%Silt$), percentage of clay ($\%Clay$) and percentage of organic matter ($\%OM$). The input variables were chosen based on expert knowledge related to soil-cement mixtures (Shibazaki, 2004) and authors' experience (Tinoco et al., 2011a). Thus, additionally to the seven variables previously enumerated, the relation between mixture porosity and volumetric content of cement ($n/(C_v)^d$) as well as the age of the mixture (t , days) were also chosen as input variables of the model. The main statistics of the input and target variables of UCS dataset are shown in Table 1. For a physical characterization of the natural soil, some samples were collected and tested in laboratory. Although all soils were of clayed nature, they have different percentages of sand, silt, clay and organic matter. A detailed description of the distinct soil types can be found in Table 2. In the table, the first column denotes the construction site, while the third column shows the number of records that contain such soil. All laboratory formulations were prepared with cement type CEM I 42.5R ($s=0.20$), CEM II 42.5R ($s=0.20$) and CEM IV/A 35.5R ($s=0.25$).

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

Variable	Maximum	Minimum	Mean	Standard Deviation
t	56	3.00	21.60	19.24
C	0.75	0.24	0.47	0.15
$n/(C_v)^d$	74.26	48.83	62.59	7.26
W/C	1.12	0.68	0.88	0.16
s	0.25	0.20	0.21	0.02
$\%Sand$	39.00	0.00	13.57	11.53
$\%Silt$	57.00	33.00	50.49	5.49
$\%Clay$	45.00	22.50	35.89	7.74
$\%OM$	8.30	0.40	2.71	1.81
UCS	13.19	0.76	5.20	2.73

Table 2: Soil types present in the collected data

Site	Soil type	Frequency	$\%Sand$	$\%Silt$	$\%Clay$	$\%OM$
A	Lean clay (CL)	10	39.0	33.0	27.0	8.3
B	Organic lean clay (OL)	5	6.0	57.0	37.0	1.8
C	Fat clay (CH)	85	7.0	53.0	40.0	3.2
D	Silt Clay (CL-ML)	20	25.0	52.5	22.5	0.4
E	Lean clay (CL)	15	0.0	55.0	45.0	3.9
F	Silt Clay (CL-ML)	20	32.5	43.5	24.0	1.2
G	Lean clay (CL)	20	10.5	48.5	41.0	1.0

3. RESULTS AND DISCUSSION

SVM algorithm was able to learn with high accuracy the complex relationships between UCS of JGLF and its contributing factors. Figure 4 shows visually the excellent performance achieved by SVM model in UCS estimation of JGLF over time (all points are very close to the diagonal line), which is proven (mean value and 95% confidence intervals) by metrics $R^2=0.93\pm 0.00$, $MDA=0.55\pm 0.00$ and $RMSE=0.73\pm 0.00$. Furthermore, REC curve (dotted line) show a fast improvement on model accuracy. For example, SVM model can predict 90% of the records with an absolute deviation lower than 1.5MPa.

When evaluating a DM model, we should consider not only predictive accuracy but also try to understand how model works. In the present paper, model interpretability was measured by quantifying what are the key input variables in UCS estimation and its effect in such prediction. For such purpose was applied a detailed sensitivity analysis.

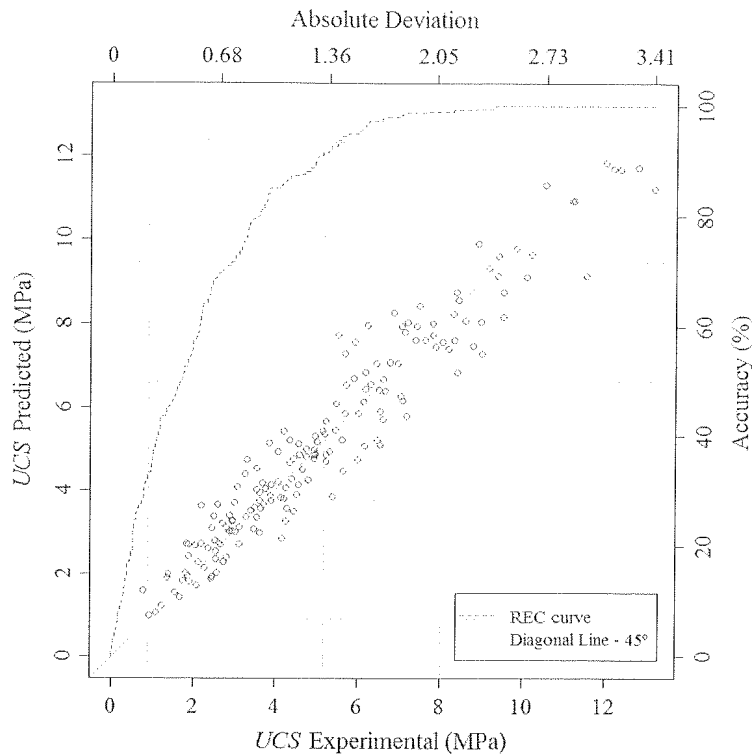


Figure 4: Relation between UCS experimental versus UCS predicted by SVM model (the dashed line represents the REC curve and should be read on top horizontal and right vertical axis).

The averaged relative importance of each input variable and the correspondent t-student 95% confidence interval for all 20 runs performed was quantified by one-dimensional sensitivity analysis. Figure 5 (left side) shows such results and allow to point out some important conclusions. One the one hand, the key variables that control the mechanical behavior of JGLF, in terms of strength, are the age of the mixture with an impact around 19%, followed by cement content (18%) and $n/(C_{iv})^d$ (16%). One the other hand, and based on t-student values, SVM algorithm is very consistent for all 20 runs performed. Furthermore, this relative importance ranking is in agreement with the empirical knowledge related with soil-cement mixtures. It is well known that t has a strong influence in such mixtures behavior, mainly if we take in account the range of t variable in the dataset used during the learning phase, i.e. $t \leq 56$ days time of cure. On the other hand, make all sense that C shows high impact when we are study soil-cement mixtures (Horpibulsuk et al., 2003). In addition, the mixture porosity it is also considered throughout $n/(C_{iv})^d$ variable. The W/C ratio and cement type have a less impact in strength behavior of JGLF, with an relative importance around 12% and 10% respectively. Finally, maybe seem strange the lower influence of soil properties. However, this observations can be explained if we take in account that all soils are very similar, i.e., all soil were classified as clayed nature, just differing on its sand, silt, clay and organic matter content.

Based also on one-dimension sensitivity analysis, the effect of the four key variables on UCS estimation was quantified and is depicted in Figure 5 (right side). The first observation we can drawn is that all four variables have a nonlinear effect on UCS behavior of JGLF. Second, and as expected, t and C have a positive impact in such prediction. On the other hand, increasing $n/(C_{iv})^d$ and/or W/C the strength of the mixture decrease. VEC curve of t shows a convex shape which means that age effect is more pronounced for early ages of cure and trend to stabilize for advanced ages (up to 45 days time of cure). It is also interesting to observe the shape of C VEC curve. In this case, C improves considerably UCS for values higher than 45%. Lastly, it is also possible to observe that $n/(C_{iv})^d$ and W/C VEC curves have a very similar effect (concave shape) on UCS prediction.

All previous results were obtained by performing a one-dimensional sensitivity analysis, i.e., holding all variables at its mean values and range only one at each time. As we know, such conditions rarely or even never happen in the real word. Hence, and keeping in mind a more realistic and detailed analysis, in the nest lines we present and discusses some important observation taken from a two dimensional sensitivity analysis, i.e., when two variables are changed simultaneously. Thus, was measured the interaction level of all variables with t and C and plotted the VEC surfaces for: t and W/C ; t and $n/(C_{iv})^d$; C and W/C and C and t .

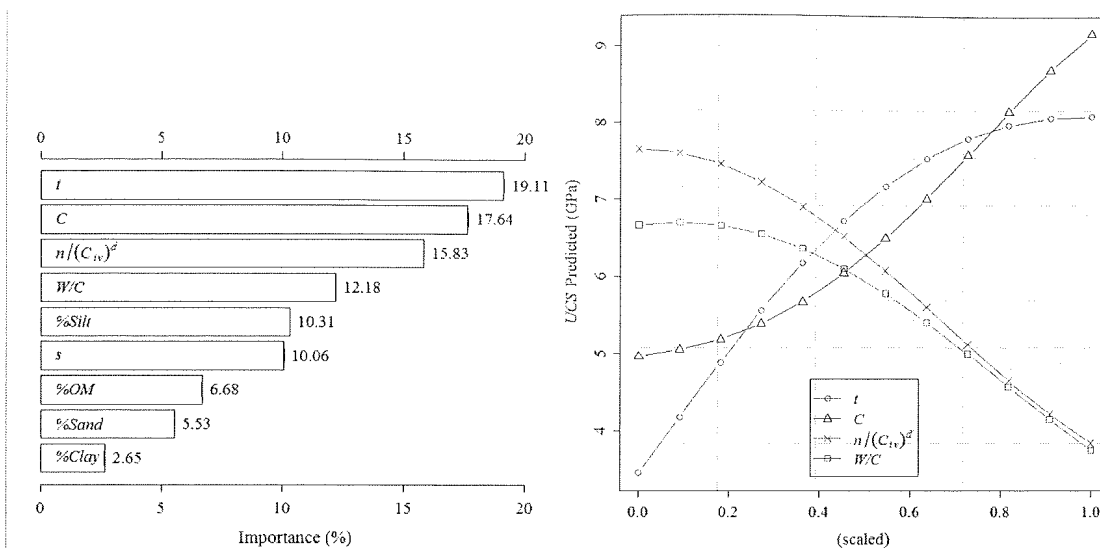


Figure 5: Relative importance of each input variable (left side) and VEC curves of t , C , $n/(C_v)^d$ and W/C variables, according to SVM mode, quantified by one-dimensional sensitivity analysis.

Figure 6 shows the relative interaction between all variables with t (left side) and C (right side), i.e., the first two key variables. In both situations the highest interaction is observed for W/C , with an relative importance around 14%. This observation shows that in spite of W/C being the fourth variable with more impact in UCS prediction, based on a one-dimensional sensitivity analysis, it should be taken in account in JGLF behavior since it has a strong interaction with other variables, namely with t and C . The highest interaction of W/C with t can be explained if we consider that the gain of strength is related to the decreasing of free water in the mixture (hardness process). This means that when is used a high W/C ratio we will need to wait more time to obtain the same strength than for a lower W/C ratio. Then, we can conclude that to obtain a fast hardness process, the mixture should be prepared with lower values of W/C ratio. The high interaction between C and W/C explain the cement content effect. Normally, mixtures with high C are prepared using grout with lower W/C ratio. Therefore, is clear that C and W/C has a strong interaction. Another interest observation from Figure 6 is related with the soil properties. Again this input shows low impact on UCS estimation of JGLF.

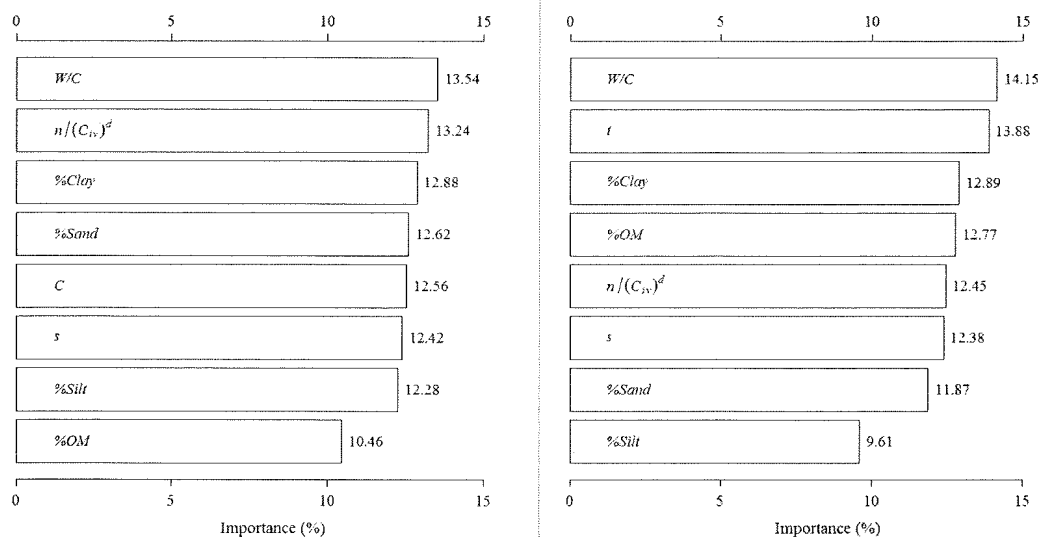


Figure 6: Interaction level of all variables with t (left side) and C (right side), according to SVM model, when a two-dimension sensitivity analysis is applied.

Plotting the interaction effect between t and W/C in UCS prediction, we obtain the VEC surface shown in Figure 7 (left side). This surface shows precisely the high effect of the interaction between these two variables, evidenced by the high range of UCS values for different combinations of t and W/C (since 2 MPa to 9MPa). Furthermore, it is also possible to observe that mixtures with high W/C ratios tend to

stabilize for ages more early. Based on VEC surface for t and $n/(C_{iv})^d$ (Figure 7, right side), we can see the high impact interaction that these two variables also have in UCS estimation of JGL (UCS range from 2MPa to 9MPa). In addition, the effect of t on UCS is more pronounced for lower values of $n/(C_{iv})^d$ than for higher values. This means that when we have a mixture with high porosity (or lower cement content) the UCS will just slight increase over time.

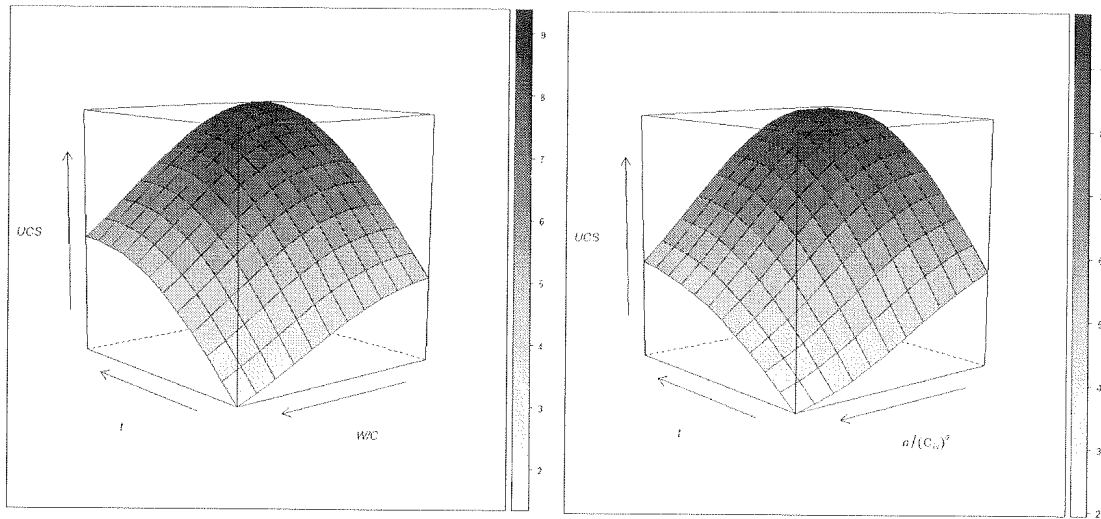


Figure 7 : VEC surface for t and W/C (right side) and t and $n/(C_{iv})^d$ (left side) in UCS prediction, according to SVM model, quantified by two-dimensional sensitivity analysis.

Figure 8 shows the VEC surfaces for C and W/C (left side) and C and t (right side). In these pictures we can see the high impact interaction of both W/C and t (UCS range from 3MPa to 11MPa). The VEC surface of C and W/C (left side of Figure 8) shows a fast increasing of UCS for higher values of C when W/C decrease. This behavior can be explained by the high amount of cement in such condition (high C and low W/C ratio). Observing the VEC surface of C and t (right side of Figure 8) we can see an almost linear effect of C for advanced ages.

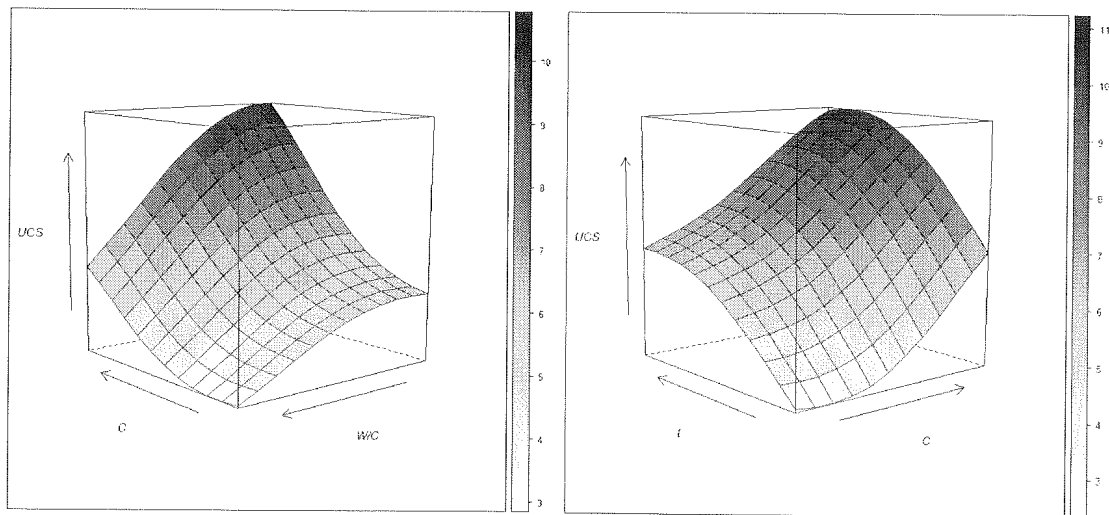


Figure 8: VEC surface for C and W/C (left side) and C and t (right side) in UCS prediction, according to SVM model, quantified by two-dimensional sensitivity analysis.

4. CONCLUSION

In the context of Jet Grouting (JG) technology the main drawback is related with the absence of rational design models to predict mechanical properties of JG mixtures, namely its Uniaxial Compressive Strength (UCS). Nowadays, this task is essentially performed considering the results of laboratory tests and empirical models. However, in the preliminary stages of design, mainly in smaller geotechnical works, the decision regarding the parameter values is based on limited information. Thus, the use of data from past projects can be seen as a good solution to overcome this problem. The application of Data Mining (DM)

techniques to well organized data gathered from large geotechnical works can provide a strong framework to the development of models that can be very useful in future projects.

In the present work, a rational model has been proposed to predict *UCS* of jet grouting laboratory formulations (JGLF) over time using DM techniques. The proposed model is able to accurately predict *UCS* of JGLF mixtures from 3 to 56 days in advance (i.e. the range values for variable age) prepared with clayed soils containing different percentage of sand, silt, clay and organic matter. Furthermore, some important observations were pointed out by performing a detailed sensitivity analysis. Age of the mixture (*t*), cement content (*C*) and the relation between mixture porosity and the volumetric content of cement ($n/(C_v)^4$) were identified as key variable in *UCS* prediction of JGLF. These three variables have a nonlinear effect on *UCS* estimation according to exponential laws. In addition, the Water/Cement (*W/C*) ratio shows a strong interaction with the first two key variables on *UCS* prediction.

Moreover, and since the proposed model are in agreement with the empirical knowledge related to JGLF, it can be conclude that such model can be used to describe rationally the actual empirical knowledge related to this material. All obtained knowledge can be seen as an important advance for geotechnical civil engineering works since contribute for a better understanding of JG mixtures behavior, which will allow to significantly reduce the number of laboratory formulations need to be prepared. At the end, the economy and the quality of the treatment will be improved.

As future works, the same methodology, i.e. DM techniques will be applied in order to develop rational models to estimate deformability properties of JGLF. In addition, the same tools will be also used to define predictive model for strength and stiffness of real JG columns as well as its diameter.

5. ACKNOWLEDGEMENTS

The authors wish to thank to “Fundação para a Ciência e a Tecnologia” (FCT) for the financial support under the strategic project PEst-OE/ECI/UI4047/2011 and 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., “Regression Error Characteristic Curves”, *Twentieth International Conference on Machine Learning (CD-ROM)*, AAAI Press, 40-50, 2003.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. and Wirth, R., “CRISPDM1.0: Step-by-Step Data Mining Guide”, *CRISP-DM consortium*, 2000.
- Chen, Y. and Councill, J., “An Introduction to Support Vector Machines: A Review”, *AI Magazine*, AAAI Press, 24(2):105-107, 2003.
- Cherkassky, V. and Ma, Y., “Practical Selection of SVM Parameters and Noise Estimation for SVM Regression”, *Neural Networks*, Elsevier, 17(1):113-126, 2004.
- Cortes, C. and Vapnik, V., “Support Vector Networks”, *Machine Learning*, Springer, 20(3):273-297, 1995.
- Cortez, P., “Data Mining with Neural Networks and Support Vector Machines Using the R/rminer Tool”, *Advances in Data Mining – Applications and Theoretical Aspects 10th Industrial Conference on Data Mining (ICDM 2010)*, *Lecture Notes in Artificial Intelligence 6171*, Springer, 572-583, 2010.
- Cortez, P. and Embrechts, M., “Opening Black Box Data Mining Models Using Sensitivity Analysis”, *2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM 2011)*, IEEE, 341-348, 2011.
- Erzin, Y., “Artificial Neural Networks Approach for Swell Pressure versus Soil Suction Behavior”, *Canadian Geotechnical Journal*, NRC Research Press, 44(10):1215-1223, 2007.
- Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P., “The KDD Process for Extracting Useful Knowledge from Volumes of Data”, *Communications of the ACM*, ACM, 39(11):27-34, 1996.
- Frohlich, H. and Zell, A., “Efficient Parameter Selection for Support Vector Machines in Classification and Regression via Model-Based Global Optimization”, *Neural Networks*, IEEE, 3:1431-1436, 2005.

Liu, S., Zhang, D., Liu, Z. and Deng, Y., "Assessment of Unconfined Compressive Strength of Cement Stabilized Marine Clay", *Marine Georesources & Geotechnology*, Taylor & Francis, 26(1):19-35, 2008.

Gazaway, H. and Jasperse, B., "Jet grouting in Contaminated Soils", *Grouting, Soil Improvement and Geosynthetics (GSP 30)*, ASCE, 206-214, 1992.

Goh, A. and Goh, S., "Support Vector Machines: Their use in Geotechnical Engineering as Illustrated using Seismic Liquefaction Data", *Computers and Geotechnics*, Elsevier, 34(5):410-421, 2007.

Hastie, T., Tibshirani, R. and Friedman, J., "The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer-Verlag, 2nd Edition, 2009.

Horpibulsuk, S., Miura, N. and Nagaraj, T., "Assessment of Strength Development in Cement-Admixed High Water Content Clays with Abrams' Law as a Basis", *Geotechnique*, Thomas Telford, 53(4): 439-444, 2003.

Huang, C., Lee, Y., Lin, D. and Huang, S., "Model Selection for Support Vector Machines via Uniform Design", *Computational Statistics & Data Analysis*, Elsevier, 52(1):335-346, 2007.

Miranda, T., Gomes Correia, A., Santos, M., Sousa, L. and Cortez, P., "New Models for Strength and Deformability Parameter Calculation in Rock Masses Using Data-Mining Techniques", *International Journal of Geomechanics*, ASCE, 11:44-58, 2011.

Momma, M. and Bennett, K., "A Pattern Search Method for Model Selection of Support Vector Regression", *Proceedings of the SIAM International Conference on Data Mining*, Philadelphia: SIAM, 261:274, 2011.

Nikbakhtan, B., Ahangari, K. and Rahmani, N., "Estimation of Jet Grouting Parameters in Shahriar Dam, Iran", *Mining Science and Technology*, Elsevier, 20(3):472-477, 2010.

R Development Core Team, "R: a Language and Environment for Statistical Computing", R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org>.

Smola, A. and others, "Regression Estimation with Support Vector Learning Machines", *Master's Thesis*, Technische Universit at Munchen, 1996.

Smola, A. and Schölkopf, B., "A Tutorial on Support Vector Regression", *Statistics and Computing*, Springer, 14(3):199-222, 2004.

Shibazaki, M., "State of practice of jet grouting", *Grouting and Ground Treatment: 3rd International Specialty Conference on Grouting and Ground Treatment*, ASCE, 198-217, 2004.

Tinoco, J., Gomes Correia, A. and Cortez, P., "Application of Data Mining Techniques in the Estimation of the Uniaxial Compressive Strength of Jet Grouting Columns over Time", *Constructions and Building Material*, Elsevier, 25(3):1257-1262, 2011a.

Tinoco, J., Gomes Correia, A. and Cortez, P., "Using Data Mining Techniques to Predict Deformability Properties of Jet Grouting Laboratory Formulations over Time", *Progress in Artificial Intelligence*, Springer, 491-505, 2011b.

Tinoco, J., Gomes Correia, A. and Cortez, P., "Uniaxial Compressive Strength Prediction of Jet Grouting Columns using Support Vector Machines", *The 2011 European Simulation and Modeling Conference, EUROSIS*, 326-330, 2011c.

Xanthakos, P., Abramson, L. and Bruce, D., "Ground Control and Improvement", Wiley-Interscience, 1994.

Welsch, J. and Burke, G., "Jet Grouting – Uses For Soil Improvement", *Geotechnical Engineering Congress (GSP No. 27)*, 334-345, 1991.