

ABORDAGEM DATA MINING PARA A PREVISÃO DA RESISTÊNCIA À COMPRESSÃO UNIAXIAL DE MISTURAS LABORATORIAIS DE SOLO-CIMENTO

Data mining approach for unconfined compression strength prediction of laboratory soil cement mixtures

Joaquim Tinoco^a, António Alberto S. Correia^b, Paulo J. Venda Oliveira^c, A. Gomes Correia^d, Luís J. L. Lemos^c.

^a ISISE /ALGORITMI Research Center, Universidade do Minho, Portugal.

^b CIEPQPF, Universidade de Coimbra, Portugal.

^c ISISE, Universidade de Coimbra, Portugal.

^d ISISE, Universidade do Minho, Portugal.

RESUMO – A previsão da resistência à compressão uniaxial (q_u) de misturas solo-cimento é de elevada importância durante a fase de projeto. Para a sua quantificação são realizados ensaios laboratoriais, os quais consomem muito tempo e recursos. Neste artigo é apresentada uma nova abordagem para avaliação da q_u ao longo do tempo tirando proveito das elevadas capacidades de aprendizagem das técnicas de *Inteligência Artificial* (IA). Três algoritmos de IA, nomeadamente as Redes Neurais Artificiais (RNAs), as Máquinas de Vetores de Suporte (MVSs) e Regressões Múltiplas (RMs), foram treinados utilizando uma base de dados composta por 444 registos contemplando solos não coesivos, coesivos e orgânicos, assim como diferentes ligantes, condições de mistura e tempos de cura. Os resultados obtidos evidenciam um desempenho promissor na previsão da q_u de misturas laboratoriais de solo-cimento, sendo o melhor desempenho conseguido através da média das previsões obtidas pelas MVSs e RNAs ($R^2 = 0,95$). Estes modelos reproduzem eficazmente os principais efeitos das variáveis de entrada, nomeadamente da razão água/cimento, teor em cimento, teor em matéria orgânica e tempo de cura, as quais são conhecidas como preponderantes no comportamento de misturas solo-cimento.

ABSTRACT – The prediction of the uniaxial compression strength (q_u) of soil cement mixtures is of utmost importance for design purposes. This is done traditionally by laboratory tests which is time and resources consuming. In this paper it is presented a new approach to assess q_u over time based on the high learning capabilities of Artificial Intelligence (AI) techniques. A database of 444 records, encompassing cohesionless to cohesive and organic soils, different binder types, mixture conditions and curing time, were used to train three models based on Artificial Neural Networks (ANNs), Support Vector Machines (SVMs) and Multiple Regression (MRs). The results show a promising performance in q_u prediction of laboratory soil cement mixtures, being the best results achieved with an average of SVMs and ANNs model ($R^2 = 0.95$). These models catch very well the major effects of the input variables water/cement ratio, cement content, organic matter content and curing time, which are known as key parameters in soil cement mixtures behavior.

Palavras Chave – Misturas solo-cimento, resistência à compressão uniaxial, data mining, análise de sensibilidade.

Keywords – Soil-cement mixtures, unconfined compression strength, data mining approach, sensitivity analysis.

E-mails: jtinoco@civil.uminho.pt (J. Tinoco), aalberto@dec.uc.pt (A. Correia), pjvo@dec.uc.pt (P. Oliveira), agc@civil.uminho.pt (A. Correia), llemos@dec.uc.pt (L. Lemos)

1 – INTRODUÇÃO

O estudo das propriedades mecânicas de misturas solo-cimento é uma tarefa de elevada complexidade devido ao elevado número de parâmetros envolvidos. Ao longo das últimas décadas, diversos estudos foram realizados seguindo diferentes abordagens mas tendo em comum o objetivo de melhor compreender o comportamento de misturas solo-cimento ao longo do tempo.

No que diz respeito à resistência à compressão uniaxial (q_u), esta é medida em ambiente laboratorial com recurso a ensaios que consomem tempo e recursos, os quais geralmente são muito limitados. Por conseguinte, é cada vez mais importante reduzir o número de ensaios laboratoriais ao mínimo necessário sem contudo comprometer a segurança e a confiança nos dados obtidos. A prática comum consiste em preparar (antes de iniciar os trabalhos de construção) e ensaiar algumas amostras laboratoriais tendo como objetivo simular as condições *in situ* (EuroSoilStab, 2001; EN 14679:2005). Estas amostras, preparadas com o mesmo solo, ligante e água a utilizar *in situ*, permitem obter uma ideia sobre como a mistura se irá comportar *in situ*. No entanto, estas amostras laboratoriais também representam um importante custo para o projeto, devendo este por conseguinte ser minimizado.

Este cenário sublinha a necessidade de, pelo menos a nível de projeto, dispor de ferramentas capazes de definir o melhor conjunto de parâmetros a utilizar durante o tratamento *in situ*. No entanto, devido ao elevado número de parâmetros que afetam o comportamento mecânico de misturas solo-cimento, nomeadamente da respetiva q_u , o recurso a análises estatísticas tradicionais tem-se mostrado incapaz de resolver este problema eficazmente (Croce e Flora, 2000; Kirsch e Sondermann, 2001), existindo apenas algumas expressões empíricas e limitadas a um domínio de aplicação bastante restrito (Liu et al., 2008).

Tendo como objetivo ultrapassar esta limitação, uma primeira e bem-sucedida tentativa foi recentemente realizada, tirando vantagem da elevada capacidade de aprendizagem das técnicas de *Inteligência Artificial* (IA) (Tinoco et al., 2011; Tinoco et al., 2014; Gomes Correia et al., 2014; Tinoco et al., 2016), as quais têm sido aplicadas com sucesso no passado em diferentes domínios do conhecimento (Domingos, 2012; Inbarani et al., 2018; Deepa e Ganesan, 2016), incluindo na área da engenharia civil (Liao et al., 2012; Emamgholizadeh et al., 2017).

Embora anteriormente se tenha conseguido um muito bom desempenho na previsão da q_u de misturas laboratoriais de solo-cimento (ver Fig. 1), existem algumas limitações que ainda precisam de ser ultrapassadas (Tinoco et al., 2014). Em particular, a dependência dos modelos das propriedades da mistura, nomeadamente na respetiva porosidade, é uma das principais limitações. Como se pode observar na Fig. 2, a qual representa a importância relativa de cada variável de entrada na previsão da q_u , a porosidade da mistura (medida apenas após a preparação da mistura) tem um peso superior a 15%. Por outro lado, é ainda de sublinhar o facto de os modelos apresentados na Fig. 1 terem sido treinados utilizando uma base de dados relativa a misturas solo-cimento com elevadas dosagens de cimento (Gomes Correia et al., 2014). Nesse sentido, e tendo como principal objetivo eliminar a dependência dos modelos das propriedades finais das misturas, nomeadamente da respetiva porosidade, bem como alargar o domínio de aplicabilidade dos mesmos, foram desenvolvidos novos modelos para previsão da q_u ao longo do tempo através da aplicação de ferramentas de IA. Assim, os modelos aqui apresentados, além de não considerarem nenhuma propriedade da mistura solo-cimento contemplam um teor em cimento mais amplo.

Assim, os modelos propostos para a previsão da q_u de misturas laboratoriais de solo-cimento ao longo do tempo, desenvolvidos através da aplicação de três algoritmos de IA, nomeadamente as Redes Neurais Artificiais (RNAs), as Máquinas de Vetores de Suporte (MVSs) e Regressões Múltiplas (RMs), têm como atributos dez variáveis referenciadas como influentes no comportamento mecânico de misturas solo-cimento. Entre elas há a destacar o teor em cimento utilizado na preparação da mistura, a granulometria do solo e o tipo de ligante.

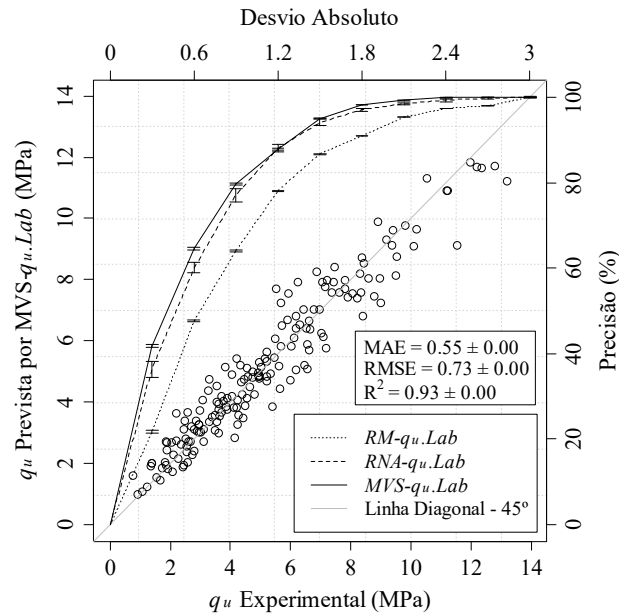
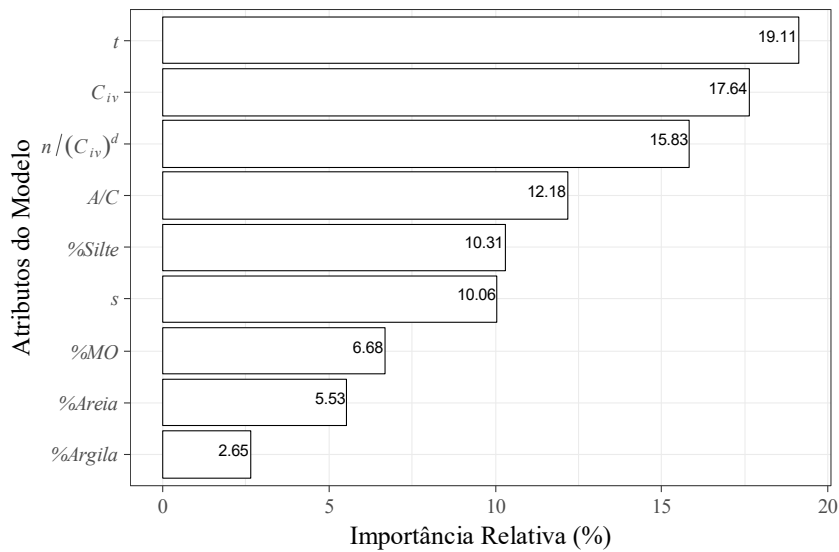


Fig. 1 – Desempenho dos modelos de IA na previsão da q_u (primeira tentativa) de misturas laboratoriais de solo-cimento (adaptado de Gomes Correia et al. (2014)).



t – tempo de cura (dias); C_{iv} – teor volumétrico de cimento; $n/(C_{iv})^d$ – relação entre a porosidade da mistura e o teor volumétrico de cimento; A/C – razão água/cimento; s – tipo de cimento; $\%Areia$ – teor em areia do solo; $\%Silte$ – teor em silte do solo; $\%Argila$ – teor em argila do solo; $\%MO$ – teor em matéria orgânica do solo

Fig. 2 – Importância relativa de cada variável de entrada na previsão da q_u de misturas laboratoriais de solo-cimento de acordo com o algoritmo MVSs (primeira tentativa) (adaptado de Tinoco et al. (2014)).

2 – METODOLOGIA

2.1 – Modelação

Para a modelação da q_u foram treinados três algoritmos de IA utilizando uma base previamente compilada e preparada, contendo resultados de testes à resistência à compressão uniaxial de misturas laboratoriais de solo-cimento bem como um conjunto de variáveis relacionadas com as características do solo e do ligante utilizados na preparação da mistura. Assim, foram treinados dois dos algoritmos de IA conhecidos pela sua elevada flexibilidade e capacidade de aprendizagem, nomeadamente as MVSs e as RNAs. Como base de comparação foi também treinada uma RM. Seguidamente é apresentada uma breve descrição dos algoritmos de IA aplicados, detalhando os respetivos parâmetros adotados.

Inicialmente desenvolvidas para problemas de classificação (Cortes e Vapnik, 1995), as MVSs foram posteriormente adaptadas a problemas de regressão através da introdução da função perda ϵ -insensitiva (Smola e Scholkopf, 2004). O principal objetivo das MVSs consiste em transformar os dados de entrada num espaço característico de elevada dimensão através de um mapeamento não-linear. As MVSs procuram posteriormente o melhor hiperplano de separação, relativo a um conjunto de vetores de suporte, nesse espaço característico. Esta transformação depende da função kernel adotada. O kernel Gaussiano é o mais popular por apresentar um menor número de parâmetros, tendo sido adaptado no presente estudo. Neste contexto, o desempenho das MVSs é afetado por três parâmetros: γ , o parâmetro do kernel; C , o parâmetro de penalização; e ϵ , a largura da zona ϵ -insensitiva (Gilan et al., 2012). Para auxiliar a escolha dos valores dos diferentes hiperparâmetros (γ , C e ϵ), foram adotadas as heurísticas propostas por Cherkassky e Ma (2004). Assim, para C foi adotado o valor de 3 e a largura da zona ϵ -insensitiva foi definida de acordo com $\epsilon = \hat{\sigma}/\sqrt{N}$, onde $\hat{\sigma} = 1,5/N \cdot \sum_{i=1}^N (y_i - \hat{y}_i)^2$, y_i é o valor observado, \hat{y}_i é o valor previsto pelo algoritmo dos 3-vizinhos próximos e N representa o número de registos da base de dados. O parâmetro kernel γ foi definido usando uma procura em grelha de $2^{\{-15; -11; -7; -3; 1\}}$, sob um esquema de validação cruzada interno.

Relativamente às RNAs, elas são um método de inteligência artificial que procura simular a estrutura do cérebro humano (ver Fig. 3) e do sistema nervoso através da sua arquitetura (Kenig et al., 2001). Este conceito foi inicialmente introduzido em 1943 por McCulloch and Pitts (1943), embora o seu uso tenha sido disseminado por Werbos (1974) através do desenvolvimento do algoritmo de *backpropagation*, o qual se veio a tornar numa ferramenta de referência no campo da previsão. As RNAs são uma técnica capaz de modelar problemas complexos e robustas a explorar dados com ruído. Neste estudo utilizou-se uma rede *multilayer perceptron* com conceções *feedforward* com uma única camada intermédia com H unidades de processamento. Uma vez que o desempenho da rede é sensível a H (balanço entre precisão e capacidade de generalização), adotou-se uma grelha de procura $\{0; 2; 4; 6; 8\}$ durante a fase de aprendizagem para encontrar o melhor valor de H . Esta grelha de procura utiliza apenas dados de treino, em que 70% são utilizados para propósitos de treino e 30% para validação. Depois de encontrar o melhor valor de H , a RNA é re-treinada com todos os dados de treino. Como função neuronal das unidades de processamento da camada intermédia foi utilizada a função logística $1/(1 + e^{-x})$.

Como base de comparação, neste trabalho foi também implementada uma RM. De acordo com o algoritmo de RMs as variáveis independentes são combinadas linearmente de forma a prever a variável dependente (Hastie et al., 2009). Esta natureza aditiva das RMs torna-as de fácil interpretação e amplamente utilizadas em tarefas de regressão. No entanto, uma das suas principais limitações prende-se com a sua ineficiência em modelar problemas de natureza não linear. Neste trabalho, foi adotada uma RM equivalente a uma RNA com zero camadas intermédias e uma função linear no nó de saída.

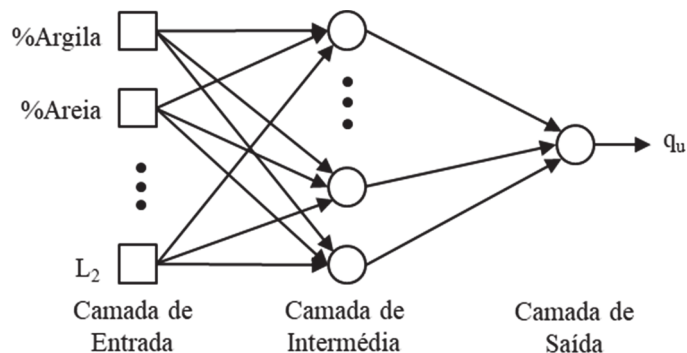


Fig. 3 – Arquitetura da RNA adotada no estudo da q_u de formulações laboratoriais de solo-cimento

O ambiente estatístico R (R Development Core Team, 2009), software livre e de código aberto, foi utilizado para conduzir todas as experiências realizadas. A biblioteca *rminer* (Cortez, 2010), a qual facilita a implementação de diversos algoritmos de IA, nomeadamente as MVSs e as RNAs, bem como diferentes abordagens de validação, em particular a validação cruzada implementada neste trabalho, foi também adotada neste estudo.

2.2 – Avaliação dos modelos

A avaliação dos modelos é uma etapa muito importante, permitindo medir o seu desempenho, quer em termos de precisão como também do ponto de vista de interpretação dos mesmos.

Para a comparação dos modelos e avaliação da respetiva precisão foram calculadas três métricas frequentemente utilizadas em problemas de regressão, nomeadamente (Hastie et al., 2009): Erro Médio Absoluto (MAE), Raiz do Erro Médio Quadrático (RMSE) e o coeficiente de determinação (R^2). Enquanto valores reduzidos de MAE e RMSE são indicadores de um modelo com elevado desempenho, o R^2 deverá aproximar-se tanto quanto possível do valor unitário. A principal diferença entre as métricas MAE e RMSE prende-se com o facto de a segunda ser mais sensível à existência de valores extremos uma vez que esta métrica utiliza o quadrado da diferença entre o valor experimental e o valor previsto pelo modelo. Quando comparado com a métrica MAE, o RMSE penaliza mais severamente modelos que apresentem maiores diferenças entre o valor experimental e o valor previsto, mesmo que isso aconteça apenas em alguns casos. Assim, estas duas métricas apresentam uma perspetiva diferente mas complementar sobre o desempenho dos modelos. Adicionalmente a estas três métricas, foram também calculadas as curvas REC (*Regression Errors Characteristic*) propostas por Bi and Bennett (2003). Estas curvas permitem uma avaliação visual e rápida do desempenho de diferentes modelos através da representação do erro tolerável (eixo dos xx) versus a percentagem de pontos calculados com um erro inferior ao tolerável (eixo dos yy).

A capacidade de generalização dos modelos é também um aspeto de elevada importância na avaliação de um modelo de IA. Nesse sentido, foi aplicada uma validação cruzada ($k - fold = 10$) (Hastie et al., 2009). Além disso todo o processo de treino foi repetido 5 vezes. De acordo com uma validação cruzada ($k - fold$) a totalidade dos dados são usados iterativamente para treino e para teste. Ou seja, a base de dados é dividida em k partes (onde k é um inteiro positivo), sendo o modelo treinado k vezes e ficando iterativamente uma das partes para propósitos de validação. No final, o melhor conjunto de parâmetros é utilizado para re-treinar o modelo utilizando a totalidade da base de dados.

Um outro aspeto muito importante num projeto de IA é conseguir perceber a forma como o modelo capturou o problema em estudo. Uma vez que os algoritmos de IA, em particular as RNAs

e as MVSs assentam em análises estatísticas complexas e são frequentemente designados por “black boxes”, são matematicamente muito complexos, havendo portanto a necessidade de “abrir” estes modelos de forma a facilitar a sua compreensão. Nesse sentido, Cortez e Embrechts (2013) propuseram uma nova abordagem de visualização através da realização de Análises de Sensibilidade (AS), as quais foram aplicadas neste trabalho. Uma AS é um método simples aplicado após a fase de treino medindo a resposta do modelo face à variação de um dos atributos do mesmo. Desta forma é possível quantificar a importância relativa de cada variável bem como o respetivo efeito médio na variável em estudo. Em particular, neste trabalho foi aplicado o método de Análise de Sensibilidade Global (ASG) (Cortez e Embrechts, 2013), o qual permite identificar interações entre as variáveis de entrada. Isto é conseguido através da variação em simultâneo de F atributos. Cada atributo toma diferentes valores (L níveis) dentro do respetivo intervalo de variação, e as restantes variáveis tomam um valor fixo de referência. Neste trabalho adotou-se o valor médio de cada variável como valor de referência e definiu-se $L=12$, o qual permite obter um nível de detalhe razoável dentro de um esforço computacional aceitável.

Utilizando os resultados da ASG é possível desenvolver diferentes técnicas de visualização. O gráfico de importância dos atributos mostra a influência relativa (R_a) de cada variável de entrada do modelo (de 0% a 100%). De acordo com a ASG, quanto maior for a variação produzida na variável de saída maior será a importância do respetivo atributo. Para medir este efeito, primeiro calculou-se a métrica gradiente (g_a) para cada um dos atributos. Depois, a importância relativa (R_a) foi calculada utilizando a seguinte equação:

$$R_a = \frac{g_a}{\sum_{i=1}^I g_i} \cdot 100(\%), \quad \text{onde } g_a = \frac{\sum_{j=2}^L |\hat{y}_{a,j} - \hat{y}_{a,j-1}|}{L-1} \quad (1)$$

onde a é a variável de entrada em análise, I é o número de variáveis, $\hat{y}_{a,j}$ é a resposta da análise de sensibilidade para $x_{a,j}$.

2.3 – Base de dados

Para fins de treino e teste dos modelos, foi compilada e organizada uma base de dados com 444 registos. Estas amostras fazem parte de diferentes estudos laboratoriais desenvolvidos na Universidade do Minho e na Universidade de Coimbra, tendo dois objetivos principais. Por um lado, analisar a influência de diferentes variáveis no comportamento mecânico de misturas laboratoriais de “jet grouting” e/ou “cutter soil mixing” (Tinoco et al., 2011). Por outro lado, definir a dosagem adequada que permita a melhor estabilização do ponto de vista técnico, económico e ambiental (Correia, 2011; Correia et al., 2015; Venda Oliveira et al., 2012, 2013, 2014).

Os solos usados na preparação das amostras laboratoriais foram recolhidos de oito locais diferentes. Um dos locais é a zona do Baixo Mondego, entre Coimbra e Figueira da Foz (Portugal), variando desde solos coesivos a não-coesivos e orgânicos a não-orgânicos, apresentando diferentes propriedades geotécnicas. Foram testados catorze ligantes distintos, incluindo cimento Portland, escória, cinzas volantes, cal e sílica ativa, aplicados individualmente ou combinados (Correia, 2011). Relativamente aos restantes sete locais, todos os solos são de natureza argilosa, contendo diferente percentagens de areia, silte, argila e matéria orgânica (Tinoco et al., 2011). Estas amostras foram preparadas com cimento do tipo CEM I 42,5R (cimento Portland com 100% de clínquer) e CEM II 42,5R (cimento Portland composto com $\geq 65\%$ clínquer). Foram ainda preparadas algumas amostras com cimento pozolânico (CEM IV/A 35,5R com $\geq 20\%$ clínquer).

Como atributo dos modelos foi selecionado um conjunto de 10 variáveis. A escolha deste conjunto de variáveis teve em consideração o conhecimento empírico relacionado com o comportamento de misturas solo-cimento, nomeadamente no que diz respeito à evolução da q_u ao longo do tempo (Sariosseiri e Muhunthan, 2009; Lee et al., 2005; Lorenzo e Bergado, 2004; Chen e Wang, 2006). Foi ainda tido em consideração o feedback obtido durante o processo de aprendizagem

dos modelos. Seguidamente são apresentadas as 10 variáveis selecionadas neste estudo para a previsão da q_u .

- Teor em argila do solo (%) - %Argila
- Teor em areia do solo (%) - %Areia
- Teor em silte do solo (%) - %Silte
- Teor em matéria orgânica do solo (%) - %MO
- Teor em água (%) - ω_0
- Teor em cimento (%) - a_w
- Razão Água/Cimento - A/C
- Idade da mistura (dias) - t
- Coeficiente relacionado com o tipo de ligante - C_s
- Coeficiente relacionado com o tipo de ligante secundário - L_2

O Quadro 1 resume as principais estatísticas das 10 variáveis de entrada, assim como da q_u , evidenciando o amplo intervalo dos valores do teor em cimento e da q_u . Na Fig. 4 é apresentada a matriz de correlação para todas as variáveis consideradas no estudo da q_u . Na diagonal são apresentados os histogramas para cada variável, abaixo da diagonal apresenta-se os gráficos de dispersão e na diagonal superior o respetivo coeficiente de correlação de *Pearson*. Da sua análise é perceptível a baixa correlação entre pares de variáveis, sendo a correlação mais significativa entre a %Silte e o ω_0 .

Quadro 1 – Resumo das principais estatísticas das variáveis de entrada e da q_u .

Variável	Mínimo	Máximo	Média	Desvio Padrão
%Argila	0,00	45,00	19,84	14,31
%Areia	0,00	99,00	22,97	22,10
%Silte	1,00	79,00	57,17	18,15
%MO	0,00	19,40	5,87	4,64
ω_0	7,17	113,05	64,96	24,48
a_w	3,00	284,32	55,42	69,21
A/C	0,63	10,91	3,30	2,05
t	3,00	90,00	25,71	15,42
C_s	0,20	0,31	0,22	0,06
L_2	0,00	1,00	0,61	0,49
q_u	0,10	13,19	2,77	2,72

3 – ANÁLISE E DISCUSSÃO DOS RESULTADOS

Nesta secção são apresentados os principais resultados do estudo realizado, visando o desenvolvimento de um modelo para a previsão da q_u de formulações laboratoriais de solo-cimento, através da aplicação de ferramentas estatísticas avançadas, usualmente conhecidas por IA.

O valor médio dos hyper-parâmetros e tempos de cálculo (acompanhados do respetivo intervalo de confiança a 95% de acordo com uma distribuição *t-student*) dos três modelos de IA treinados para a previsão da q_u de formulações laboratoriais de solo-cimento, daqui em diante designados por RNA.Lab, MVS.Lab e RM.Lab, são apresentados no Quadro 2.

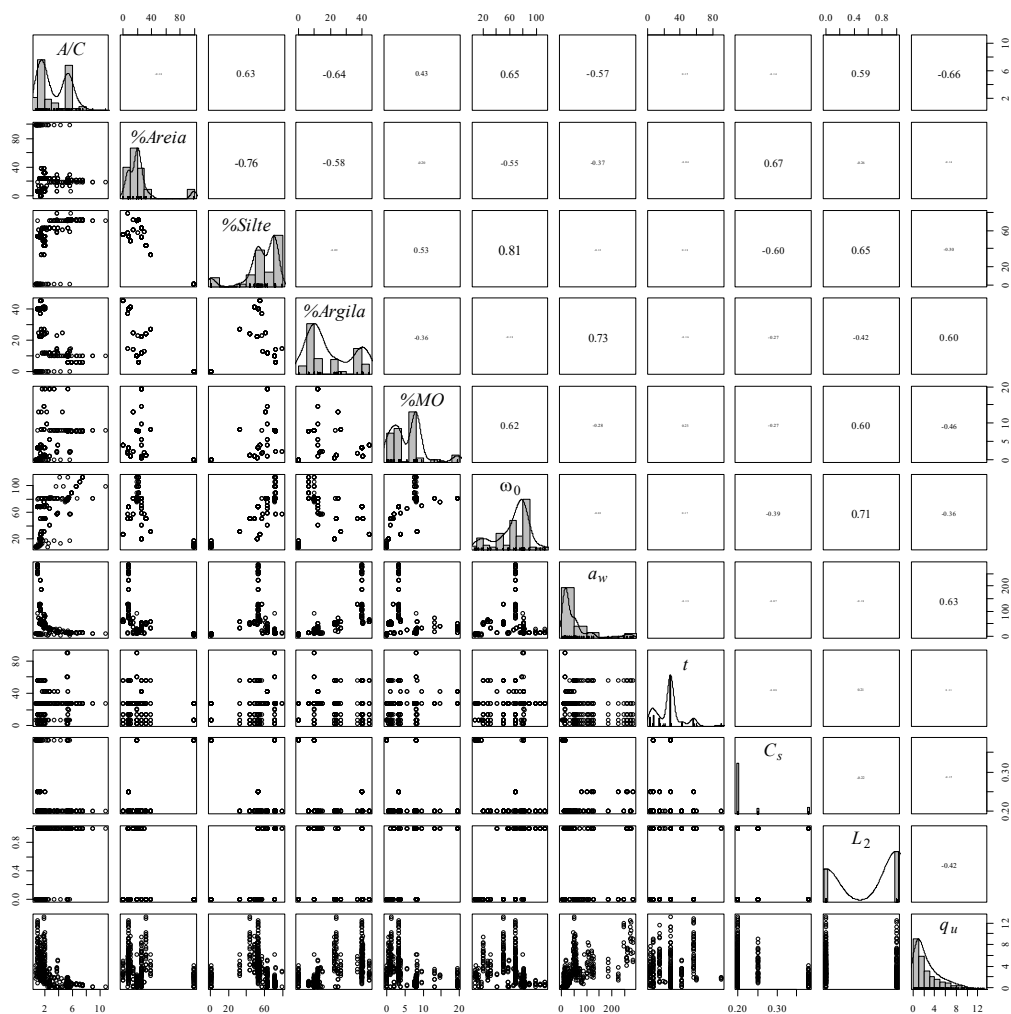


Fig. 4 – Matriz de correlação para todas as variáveis consideradas no estudo da q_u de formulações laboratoriais de solo-cimento

Quadro 2 – Hiper-parâmetros e tempos de cálculo de cada um dos modelos treinados (valor médio e respetivo intervalo de confiança a 95% de acordo com uma distribuição *t-student*).

Modelo	Hyper-parâmetro	Tempo (s)
<i>RM.Lab</i>	-	0,41±0,03
<i>RNA.Lab</i>	$H = 7 \pm 1$	17,18±0,45
<i>MVS.Lab</i>	$\gamma = 0,21 \pm 0,04; C = 3; \epsilon = -5,57 \pm 0,50$	9,63±0,28

Da análise do Quadro 3 observa-se um desempenho muito promissor na previsão da q_u de misturas laboratoriais de solo-cimento tendo por base o conjunto de variáveis seleccionadas, o qual não contempla nenhuma informação sobre as propriedades da mistura. De facto, ambos os algoritmos RNAs e MVSs foram capazes de prever q_u com elevada precisão, conseguindo um $R^2 =$

0,94. Tendo como referência as métricas MAE ou RMSE, é possível observar que o modelo MVS.Lab é capaz de prever q_u com uma precisão ligeiramente superior ao modelo RNA.Lab. Como espectável, o modelo RM.Lab apresenta o pior desempenho com um $R^2 = 0,68$, o que é significativamente inferior em comparação com o MVS.Lab ou RNA.Lab.

Quadro 3 – Comparação do desempenho dos modelos tendo por base as métricas MAE, RMSE e R^2 (valor médio e respetivo intervalo de confiança a 95% de acordo com uma distribuição *t-student*, melhores valores a negrito).

Modelo	MAE	RMSE	R^2
<i>RM.Lab</i>	1,08±0,01	1,53±0,01	0,68±0,00
<i>RNA.Lab</i>	0,46±0,02	0,69±0,05	0,94±0,01
<i>MVS.Lab</i>	0,43±0,01	0,67±0,03	0,94±0,01
<i>MVS&RNA.Lab</i> (média)	0,41±0,01	0,61±0,02	0,95±0,00

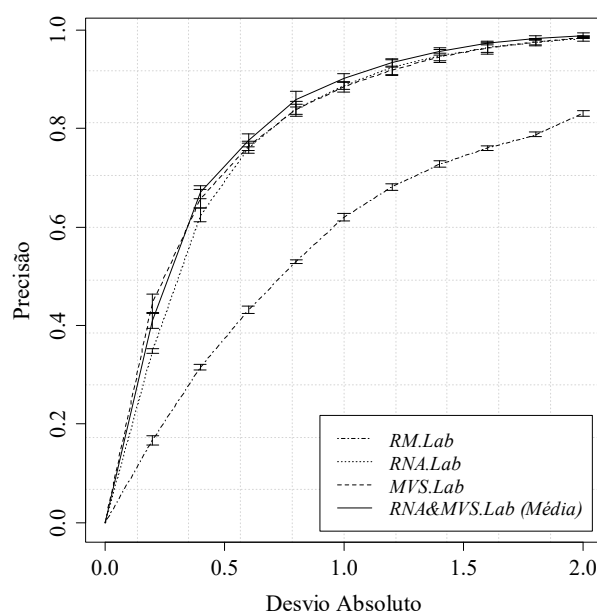


Fig. 5 – Comparação do desempenho dos modelos *MVS.Lab*, *RNA.Lab*, *RM.Lab* e *RNA&MVS.Lab Média* na previsão da q_u de misturas laboratoriais de solo-cimento com base nas curvas REC.

Embora os modelos MVS.Lab e RNA.Lab apresentem um desempenho muito elevado, observou-se que o mesmo pode ainda ser melhorado calculando a média das previsões de MVS.Lab e RNA.Lab (*RNA&MVS.Lab*). Com esta combinação de modelos consegue-se um $R^2 = 0,95$ e um RMSE muito perto de 0,61 MPa (ver Quadro 3). A Fig. 5 apresenta as curvas REC de cada modelo, evidenciando o desempenho ligeiramente superior na previsão da q_u através do cálculo da média das previsões dos modelos MVS.Lab e RNA.Lab. A Fig. 5 ilustra ainda a diferença significativa entre o desempenho do modelo RM.Lab quando comparado com os modelos MVS.Lab e RNA.Lab na previsão da q_u de misturas laboratoriais de solo-cimento.

O desempenho promissor de ambos os modelos RNA.Lab e MVS.Lab na previsão da q_u está claramente evidenciado na Fig. 6 (Fig. 6a e Fig. 6b), a qual apresenta a relação entre os valores experimentais e previstos. Como se pode observar, todos os pontos se encontram muito próximos da linha diagonal, a qual representa um modelo ideal/perfeito. A Fig. 6c apresenta a mesma relação,

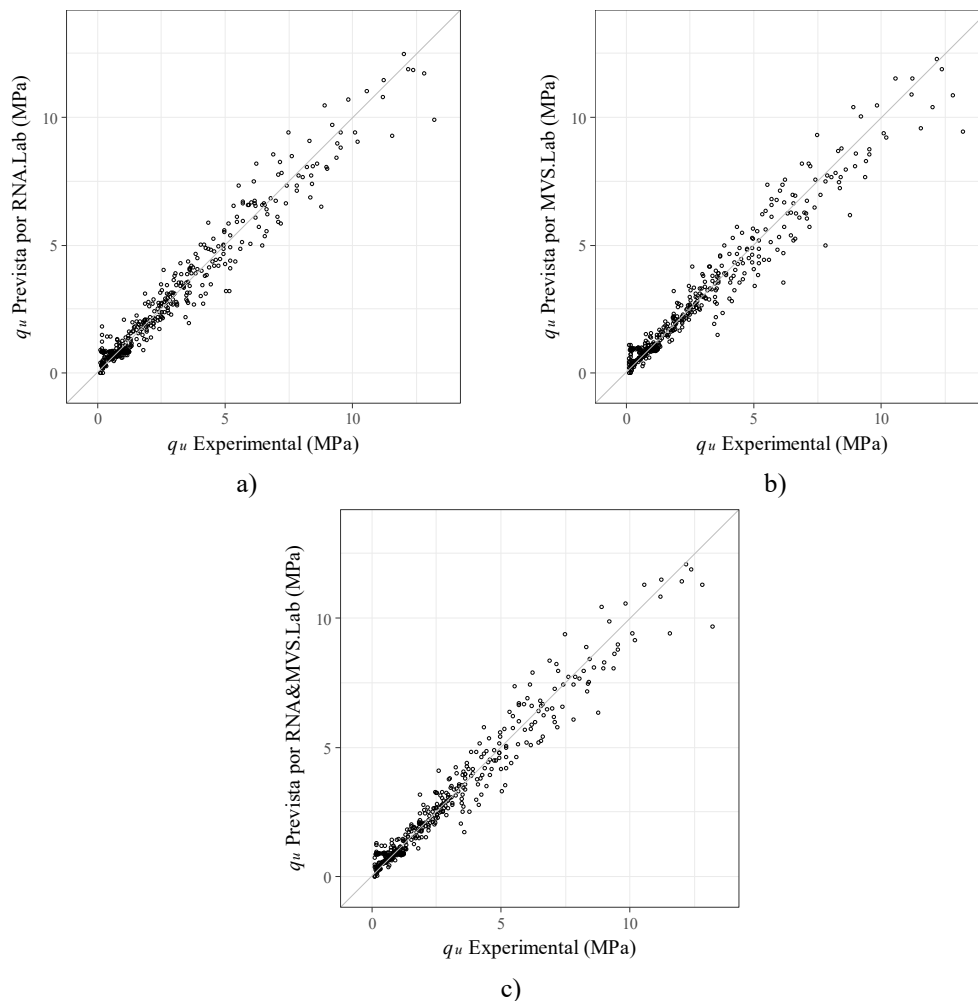


Fig. 6 – Relação entre os valores experimentais da q_u versus previstos pelos modelos: a) RNA.Lab; b) MVS.Lab; c) média de RNA.Lab e MVS.Lab.

mas considerando a média dos valores previstos pelos modelos RNA.Lab e MVS.Lab, ilustrando também neste caso o elevado desempenho conseguido.

Do ponto de vista de engenharia, complementarmente à precisão de um modelo é também muito importante perceber a forma como o mesmo modelou o problema, nomeadamente quando se está na presença de algoritmos de elevada complexidade matemática como as RNAs ou as MVSs. Nesse sentido, foi aplicada uma ASG (Cortez e Embrechts, 2013) sobre os modelos de forma a medir a influência dos atributos dos modelos na previsão da q_u . A Fig. 7 compara a importância relativa de cada variável de entrada dos modelos, mostrando que a relação A/C é a variável mais relevante na previsão da q_u de misturas laboratoriais de solo-cimento de acordo com os modelos RNA.Lab e MVS.Lab, com um peso relativo superior a 20%. As três variáveis chave seguintes são, de acordo com o modelo MVS.Lab, a_w , $\%MO$ e t . De acordo com o modelo RNA.Lab, o ranking é ligeiramente diferente, sendo ω_0 , $\%Silte$ e $\%Areia$ as três variáveis mais influentes logo a seguir à razão A/C . Comparando os dois modelos RNA.Lab e MVS.Lab, o último apresenta um ranking mais realista. De facto, entre as quatro variáveis mais relevantes, o modelo MVS.Lab inclui a influência do teor

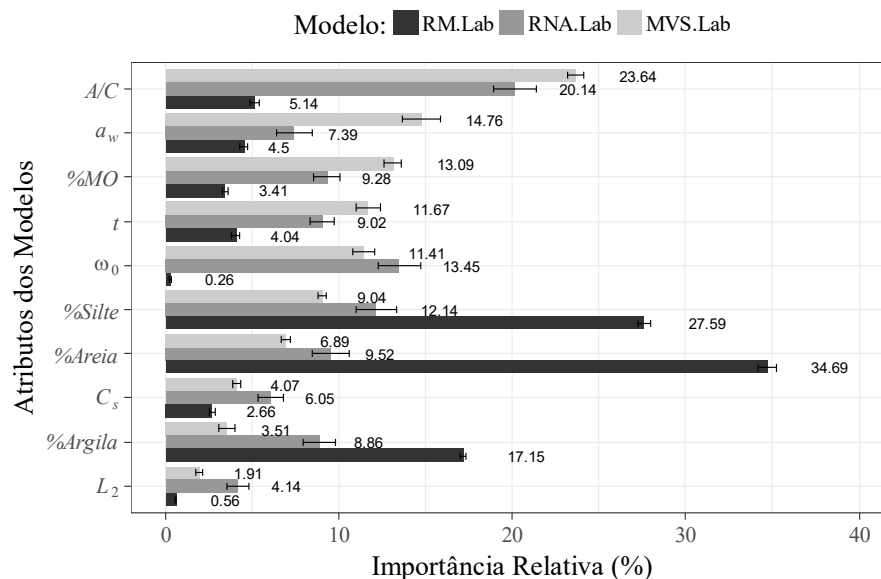


Fig. 7 – Comparação da importância relativa de cada variável de entrada dos modelos de acordo com uma ASG.

em água e do teor em cimento (A/C e a_w), o teor em matéria orgânica (%MO) e a idade da mistura (t), as quais são conhecidas por serem preponderantes no comportamento mecânico de misturas solo-cimento (Lorenzo e Bergado, 2004; Consoli et al., 2011; Horpibulsuk et al., 2011a, 2011b). De acordo com o modelo RNA.Lab, o efeito do teor em cimento é menos representativo (apenas presente na razão A/C) e o efeito do tempo de cura ocupa apenas sexta posição no ranking (menos de 10%). É sabido que a idade da mistura é uma das variáveis mais influentes no comportamento de misturas solo-cimento. Nesse sentido, considerando a precisão dos modelos bem como a importância relativa de cada variável, o modelo MVS.Lab parece ser a melhor escolha para a previsão da q_u ao longo do tempo de misturas laboratoriais de solo-cimento. Relativamente ao modelo RM.Lab, além do seu menor desempenho ($R^2 = 0,68$), a elevada influência das propriedades do solo e o reduzido efeito do tempo de cura e teor em cimento (cerca de 4%) parece muito pouco racional.

4 – CONSIDERAÇÕES FINAIS

Uma nova abordagem é proposta para a previsão da resistência à compressão uniaxial (q_u) de misturas laboratoriais de solo-cimento através da aplicação de ferramentas de *Inteligência Artificial* (IA). Os modelos propostos, suportados numa base de dados representativa constituída por 444 registos, são capazes de prever a q_u ao longo do tempo com uma precisão muito promissora ($R^2 = 0,95$). De sublinhar o facto de os modelos apenas considerarem informação disponível durante a fase de projeto, como por exemplo as propriedades dos solo, teor em cimento e água, etc. Isto ignifica que o projetista dispõe de uma ferramenta capaz de estimar a q_u em diferentes cenários (formulações), não tendo necessidade de preparar/testar qualquer amostra. Isto permite uma melhor otimização dos recursos disponíveis e, conseqüentemente, uma maior eficiência económica.

Através da aplicação de uma Análise de Sensibilidade Global (ASG) foi possível identificar as variáveis mais influentes na previsão da q_u ao longo do tempo. Assim, observou-se que a razão Água/Cimento (A/C) é a variável mais relevante, seguida pelo teor em cimento, teor em matéria

orgânica do solo e idade da mistura. Deste modo conseguiu-se uma melhor compreensão dos modelos propostos, os quais devido à sua natureza são de elevada complexidade matemática.

Como observação final, é de salientar o importante contributo das técnicas de IA para ajudar a resolver e melhor compreender problemas de elevada complexidade, em particular as máquinas de vetores de suporte e as redes neuronais artificiais.

5 – AGRADECIMENTOS

Este trabalho foi suportado pela FCT – Fundação para a Ciência e a Tecnologia, no âmbito do projeto ISISE com a referência UID/ECI/04029/2013, dos projetos CIEPQPF com as referências EQB/UI0102/2014 e UID/CEC/00319/2013, assim como pela bolsa de pós-doutoramento com a referência SFRH/BPD/94792/2013. Este trabalho foi também parcialmente financiado por fundos FEDER através do *Competitivity Factors Operational Programme* – COMPETE e por fundos nacionais através da FCT no âmbito dos projetos com as referências POCI-01-0145-FEDER-007633, POCI-01-0145-FEDER-007043 e POCI-01-0145-FEDER-028382. Agradece-se igualmente o apoio concedido pela ACIV – Associação para o Desenvolvimento da Engenharia Civil.

6 – REFERÊNCIAS BIBLIOGRÁFICAS

- Bi, J.; Bennett, K. (2003). *Regression error characteristic curves*. Proceedings of the Twentieth International Conference on Machine Learning, pp. 43-50, AAAI Press, Washington, DC, USA.
- Chen, H.; Wang, Q. (2006). *The behaviour of organic matter in the process of soft soil stabilization using cement*. Bulletin of Engineering Geology and the Environment, Vol. 65, nº 4, pp. 445-448.
- Cherkassky, V.; Ma, Y. (2004). *Practical selection of svm parameters and noise estimation for svm regression*. Neural Networks, Vol. 17, nº 1, pp. 113-126.
- Consoli, N.C.; Rosa, D.A.; Cruz, R.C.; Dalla Rosa, A. (2011). *Water content, porosity and cement content as parameters controlling strength of artificially cemented silty soil*. Engineering Geology, Vol. 122, nº 3-4, pp. 328-333.
- Correia, A. (2011). *Applicability of deep mixing technique to the soft soil of baixo Mondego*. Ph.D. thesis, School of Engineering, University of Coimbra, Coimbra, Portugal.
- Correia, A.A.; Oliveira, P.J.V.; Custódio, D.G. (2015). *Effect of polypropylene fibres on the compressive and tensile strength of a soft soil, artificially stabilised with binders*. Geotextiles and Geomembranes, Vol. 43, nº 2, pp. 97-106.
- Cortes, C.; Vapnik, V. (1995). *Support vector networks*. Machine Learning, Vol. 20, nº 3, pp. 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, pp. 572-583. LNAI 6171, Springer, Berlin, Germany.
- Cortez, P.; Embrechts, M. (2013). *Using sensitivity analysis and visualization techniques to open black box data mining models*. Information Sciences, Vol. 225, Mar, pp. 1-17.
- Croce, P.; Flora, A. (2000). *Analysis of single-fluid jet grouting*. Géotechnique, Vol. 50, Nº 6, pp. 739-48.

- Deepa, N.; Ganesan, K. (2016). *Multi-class classification using hybrid soft decision model for agriculture crop selection*. Neural Computing and Applications, pp. 1-14.
- Domingos, P. (2012). *A few useful things to know about machine learning*. Communications of the ACM, Vol. 55, n° 10, pp. 78-87.
- Emamgholizadeh, S.; Bahman, K.; Bateni, S.M.; Ghorbani, H.; Marofpoor, I.; Nielson, J.R. (2017). *Estimation of soil dispersivity using soft computing approaches*. Neural Computing and Applications, Vol. 28, n° 1, pp. 207-216.
- EN 14679:2005. *Execution of special geotechnical works - Deep mixing*. European Committee for Standardization.
- EuroSoilStab (2001). *Development of Design and Construction Methods to Stabilise Soft Organic Soils. Design Guide Soft Soil Stabilization*. Industrial & Materials Technologies Programme (BriteEuRam III), European Commission, Brussels, Belgium, CT97-0351, EC Project No. BE 96-3177.
- Gilan, S.; Bahrami Jovein, H.; Ramezani-pour, A. (2012). *Hybrid support vector regression – particle swarm optimization for prediction of compressive strength and rcpt of concretes containing metakaolin*. Construction and Building Materials, Vol. 34, Sep, pp. 321-329.
- Gomes Correia, A.; Tinoco, J.; Cortez, P. (2014). *Use of data mining in design of soil improvement by jet grouting*. In: D.T. et al. (ed.) Second International Conference on Information Technology in Geo-Engineering (ICITG 2014), pp. 43-63, IOS Press, Durham, UK.
- Hastie, T.; Tibshirani, R.; Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Second Edition, Springer-Verlag, New York.
- Horpibulsuk, S.; Rachan, R.; Suddeepong, A.; Chinkulkijniwat, A. (2011a): *Strength development in cement admixed bangkok clay: laboratory and field investigations*. Soils and Foundations, Vol. 51, n° 2, pp. 239-251.
- Horpibulsuk, S.; Rachan, R.; Suddeepong, A. (2011b). *Assessment of strength development in blended cement admixed bangkok clay*. Construction and Building Materials, Vol. 25, n° 4, pp. 1521-1531.
- Inbarani, H.H.; Kumar, S.U.; Azar, A.T.; Hassanien, A.E. (2018). *Hybrid rough-bijective soft set classification system*. Neural Computing and Applications, Vol. 29, n° 8, pp. 67-78.
- Kenig, S.; Ben-David, A.; Omer, M.; Sadeh, A. (2001). *Control of properties in injection molding by neural networks*. Engineering Applications of Artificial Intelligence, Vol. 14, n° 6, pp. 819-823.
- Kirsch, F.; Sondermann, W. (2001). *Ground improvement and its numerical analysis*. Proceedings of the international conference on soil mechanics and geotechnical engineering, Vol. 3. AA Balkema Publishers; pp. 1775–1778.
- Lee, F.H.; Lee, Y.; Chew, S.H.; Yong, K.Y. (2005). *Strength and modulus of marine clay-cement mixes*. Journal of geotechnical and geoenvironmental engineering, Vol. 131, n° 2, pp. 178-186.
- Liao, S.; Chu, P.; Hsiao, P. (2012). *Data mining techniques and applications. A decade review from 2000 to 2011*. Expert Systems with Applications, Vol.39, n° 12, pp. 11303-11311.
- Liu, Y.; Zhang, D.; Liu, Z.; Deng, Y. (2008). *Assessment of unconfined compressive strength of cement stabilized marine clay*. Marine Georesour Geotechnol; Vol.26, n° 1, pp. 19-35.

- Lorenzo, G.; Bergado, D. (2004). *Fundamental parameters of cement-admixed clay-new approach*. Journal of Geotechnical and Geoenvironmental Engineering, vol. 130, n° 10, pp. 1042-1050.
- McCulloch, W.S.; Pitts, W. (1943). *A logical calculus of the ideas immanent in nervous activity*. The bulletin of mathematical biophysics, Vol. 5, n° 4, pp. 115-133.
- R Development Core Team (2009). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Web site: <http://www.r-project.org/>
- Sariosseiri, F.; Muhunthan, B. (2009). *Effect of Cement Treatment on Geotechnical Properties of some Washington State Soils*. Engineering Geology, Vol. 104, n° 1-2, pp. 119-125.
- Smola, A.; Scholkopf, B. (2004). *A tutorial on support vector regression*. Statistics and Computing, Vol. 14, n° 3, pp. 199-222.
- Tinoco, J.; Alberto A.; Venda Oliveira, P.; Gomes Correia, A.; Lemos, L. (2016). *A data-driven approach for qu prediction of laboratory soil-cement mixtures*. Procedia Engineering, Vol. 143, pp. 566-573.
- Tinoco, J.; Gomes Correia, A.; Cortez, P. (2011). *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, n°3, pp. 1257-1262.
- Tinoco, J.; Gomes Correia, A.; Cortez, P. (2014). *A novel approach to predicting young's modulus of jet grouting laboratory formulations over time using data mining techniques*. Engineering Geology, Vol. 169, Feb, pp. 50-60.
- Venda Oliveira, P.J.; Correia, A.A.; Garcia, M.R. (2012). *Effect of organic matter content and curing conditions on the creep behavior of an artificially stabilized soil*. Journal of materials in civil engineering, Vol. 24, n° 7, pp. 868-875.
- Venda Oliveira, P.J.; Correia, A.A.; Garcia, M.R. (2013). *Effect of stress level and binder composition on secondary compression of an artificially stabilized soil*. Journal of Geotechnical and Geoenvironmental Engineering, Vol. 139, n° 5, pp. 810-820.
- Venda Oliveira, P.J.; Correia, A.A.; Lopes, T.J. (2014). *Effect of organic matter content and binder quantity on the uniaxial creep behavior of an artificially stabilized soil*. Journal of Geotechnical and Geoenvironmental Engineering, Vol. 140, n° 9, pp. 04014053.
- Werbos, P. (1974). *Regression: New tools for prediction and analysis in the behavioral sciences*. Ph.D. thesis, Harvard University, Harvard, USA.