

## **DATA MINING no GUIDE TERRASSEMENTS ROUTIERS**

### **DATA MINING in the GUIDE TERRASSEMENTS ROUTIERS**

Marques, Rui, *Universidade do Minho, Guimarães, Portugal, marquesmnc@sapo.pt*  
Gomes Correia, António, *Universidade do Minho, Guimarães, Portugal, agc@civil.uminho.pt*  
Cortez, Paulo, *Universidade do Minho, Guimarães, Portugal, pcortez@dsi.uminho.pt*

#### **RESUMO**

Neste artigo descreve-se o processo de *Data Mining*, que visa a extração automática de conhecimento a partir de dados em bruto, sobre o “Guia Francês de Terraplenagens Rodoviárias” (*GTR*), com vista à reprodução das condições de compactação neste definidas. Para tal, utilizou-se uma base de dados constituída sobretudo pelas “Tabelas de compactação” do *GTR*, considerando variáveis qualitativas (material, compactador e nível de energia) e quantitativas (parâmetro Q/S, espessura (e), velocidade (V), número de aplicações de carga (N) e rendimento teórico (Q/L)). O suporte computacional foi composto pela ferramenta estatística *R* e pela biblioteca *RMiner*. A alta fiabilidade dos modelos baseados em redes neuronais, aliada aos critérios de arredondamento, permite automatizar a informação das condições de compactação com uma eficiência elevada.

#### **ABSTRACT**

This paper describes a Data Mining process, which intends to extract high-level knowledge from raw data, that was applied to the French Guide for Road Earthworks (*GTR*), with the intention of replicating the compaction conditions. For this purpose, the database consisted mainly of the “Tables of Compaction”, part of *GTR*, which encompass qualitative variables (material, cylinder and energy level) and quantitative variables (Q/S parameter, thickness (e), speed (V), load applications number (N) and theoretical compaction capacity (Q/L)). The computational support was provided by the *R* statistical tool and the *RMiner* library. The high performance of the models based in neural networks, coupled with the rounding criteria, allows an automation of the compaction conditions information with a high efficiency.

#### **1. INTRODUÇÃO**

A tarefa de compactação tem um papel determinante na resposta do sistema estrutura-fundação das obras em geral, e dos pavimentos e vias férreas em particular. A garantia de uma compactação adequada pode geralmente obter-se por meio de duas filosofias distintas, nomeadamente, através do “controlo por produto acabado” e do “controlo por procedimento” (Ubarri et al. 1989).

Em Portugal, o controlo da qualidade das obras em aterro é tradicionalmente feito através da especificação de características para o material compactado, como sejam o grau de compactação e o teor em água. Trata-se de um controlo por produto acabado. Pelo contrário, o controlo por procedimento consiste em estabelecer a metodologia segundo a qual deve efectuar-se a compactação fixando, segundo as características do material e o tipo de equipamento a empregar, a espessura da camada e o número de passagens. São exemplo desta filosofia de controlo as recomendações francesas do Guia de Terraplenagens Rodoviárias (*Guide Terrassements Routiers*), *GTR* (SETRA e LCPC 1992).

Certos inconvenientes identificados para os procedimentos do controlo por produto acabado, tais como a falta de fiabilidade dos ensaios, a necessidade de serem efectuados uma vez terminada a obra, a morosidade dos mesmos, os seus elevados custos, e sobretudo a dificuldade de correcção “à posteriori” das partes da obra com deficiente qualidade, declinaram a balança a favor das metodologias de controlo por procedimento.

Por outro lado, encontram-se no passado inúmeros estudos na área da compactação recorrendo a técnicas de Inteligência Artificial, os quais visaram a obtenção de modelos essencialmente para previsão dos parâmetros contemplados pelo controlo por produto acabado. Por conseguinte, neste estudo recorreu-se ao processo de *Data Mining* sobre a base de dados *GTR* (Secção 3.1), tendo em vista a obtenção de modelos envolvendo os parâmetros contemplados pela filosofia de controlo por procedimento.

## 2. ESTADO DA ARTE

Encontram-se no passado inúmeros estudos na área da Geotecnia recorrendo a técnicas de Inteligência Artificial (IA). Toll (1996a, 1996b) e Shahin et al. (2001) apresentam vários exemplos de sistemas desenvolvidos no âmbito das mais diversas aplicações geotécnicas. A Figura 1 ilustra o número de estudos de IA em aplicações geotécnicas até ao ano de 1996.

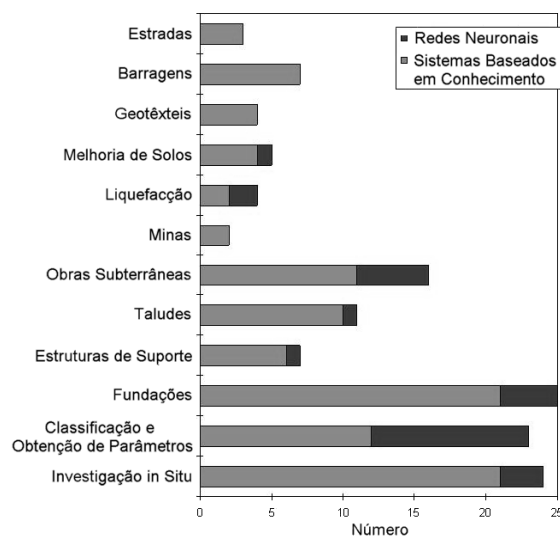


Figura 1 – Número de estudos de IA aplicados em Geotecnia até 1996 (Toll 1996a)

Por focarem a tarefa específica de compactação, objecto do presente trabalho, referem-se seguidamente três estudos desenvolvidos no domínio.

O *PACT* (Kotdawala e Hossain 1994) é um sistema baseado em conhecimento para a compactação de solos, desenvolvido tendo em vista a definição da espessura e do teor em água a adoptar para a compactação. Possui conhecimento sobre diferentes estaleiros de compactação, proveniente da aprendizagem sobre problemas associados à compactação de certos solos.

Basheer e Najjar (1995) e Najjar et al. (1996) desenvolveram um estudo baseado numa aproximação através de uma rede neuronal para a compactação de solos. Este pretendia ser usado para a previsão do teor em água óptimo e da baridade seca máxima do solo compactado. Os dados de entrada para a rede neuronal seriam o tipo de solo, as suas características granulométricas e os seus limites de consistência. Para solos no seu estado natural, a previsão era baseada apenas em três variáveis de entrada: os limites de liquidez e de plasticidade e ainda a massa volúmica.

Mais recentemente, Basheer (2001) efectuou um estudo visando a modelação empírica de curvas de compactação para solos coesivos. Nesse trabalho, foram desenvolvidos módulos a partir de modelos empíricos para simular as curvas de compactação dos solos coesivos a partir das propriedades fundamentais dos solos e da energia de compactação, por meio de regressão estatística assim como por técnicas baseadas em redes neuronais artificiais (RNAs). Reuniu-se um grande número de curvas de compactação cobrindo uma larga gama de solos finos, que foram utilizadas na modelação. Qualquer dos módulos mostrou prever as curvas de compactação dos solos com uma boa precisão, mas o módulo baseado em RNAs ofereceu melhor desempenho do que o modelo análogo baseado na estatística.

Por outra parte, na Universidade do Minho têm vindo a ser desenvolvidos distintos estudos empregando técnicas de IA visando diversas aplicações da Engenharia Civil. Exemplos destes estudos são o emprego de RNAs para identificação de áreas urbanas usando imagens de satélite (Lourenço et al. 2005) na área do Planeamento Urbano, a utilização de técnicas de optimização para idealização de sistemas aquíferos (Ferreira da Silva 2005) na Hidráulica, o uso do *Data Mining* para avaliação da estabilidade de pontes metálicas (Cruz et al. 2006) no domínio das Estruturas, e particularmente na Geotecnia, o emprego de técnicas de IA na previsão de parâmetros geomecânicos para modelação de obras subterrâneas (Miranda et al. 2006) e na análise da estabilidade de taludes (Martins et al. 2008). Potencia-se, desta forma, a aplicação de técnicas de IA a outras matérias da Engenharia Civil, e particularmente da Geotecnia.

### 3. DATA MINING

#### 3.1. Base de dados *GTR*

O processo de *Data Mining* consiste na transformação de dados no estado bruto em conhecimento. Neste estudo, a base de dados a utilizar no processo de *Data Mining* constituiu-se sobretudo pelas “Tabelas de compactação” do “*Guide Terrassements Routiers*” (*GTR*), as quais são o resultado da experiência recolhida em França na tarefa de compactação, por diversos especialistas, num conjunto bastante representativo de obras. Esta experiência foi passada para o formato das ditas tabelas, permitindo, com base no conhecimento do duo material-compactador e na qualidade requerida para a compactação, determinar as condições aconselháveis de compactação: a espessura a adoptar para as camadas ( $e$ ), a velocidade de operação do compactador ( $V$ ), o número de aplicações de carga ( $N$ ), com as quais obter-se-á determinado rendimento por unidade de comprimento do eixo do compactador ( $Q/L$ ). A Figura 2 mostra uma matriz gráfica que representa, através de pontos, as associações de valores para o quarteto de variáveis material-compactador-energia- $Q/S$ , de acordo com as modalidades constantes nas “Tabelas de compactação” para aterros. Conclui-se, desta figura, a não linearidade do problema.

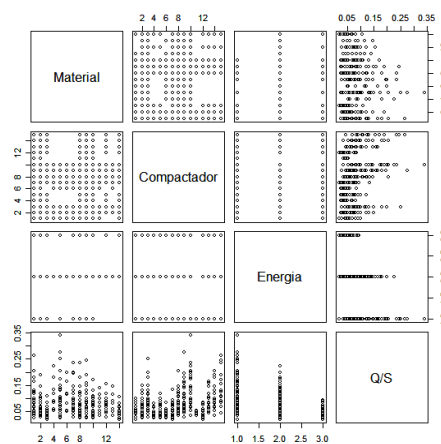


Figura 2 – Matriz de relacionamento entre algumas variáveis da base de dados *GTR*

### 3.2. Técnicas utilizadas

No presente estudo foram testadas distintas técnicas de *Data Mining*, desde a tradicional técnica de regressão múltipla, passando pelos métodos não paramétricos de árvores de decisão e de k-vizinhos próximos, até às técnicas não lineares baseadas em redes neuronais e em máquinas de vectores de suporte. A técnica baseada em redes neuronais foi aquela que permitiu uma maior fiabilidade, e por isso se apresenta seguidamente.

As RNAs pretendem reproduzir o modo de funcionamento do cérebro humano. Estas redes são constituídas por unidades de processamento (nodos) interligadas segundo uma dada configuração, sendo o *Perceptrão Multi-camada* (Figura 3) a mais popular (Haykin 1999). Os nodos constituem-se por: um conjunto de conexões ( $w_{ij}$ ), cada uma etiquetada por um peso, que tem um efeito excitatório para valores positivos e inibitório para valores negativos; um integrador ( $g$ ), que reduz os  $n$  argumentos de entrada (estímulos) a um único valor; e por uma função de activação ( $f$ ), que pode condicionar o sinal de saída, introduzindo uma componente de não linearidade no processo computacional. No presente estudo, os pesos da rede são inicialmente gerados aleatoriamente na gama  $[-0,7; +0,7]$  e a função de activação utilizada é a *logística* ( $1/(1+exp(-x))$ ). Depois, o algoritmo de treino é aplicado ajustando sucessivamente os pesos, até que o declive do erro seja aproximadamente nulo ou até que seja atingido um número máximo de iterações. O modelo de previsão para uma rede neuronal é dado pela seguinte expressão (Hastie et al. 2001):

$$\hat{y} = w_{o,0} + \sum_{j=I+1}^{o-1} f\left(\sum_{i=1}^I x_i w_{j,i} + w_{j,0}\right) w_{o,i} \quad (1)$$

onde  $w_{i,j}$  denota o peso da conexão desde o neurónio  $j$  até à unidade  $i$ ,  $o$  corresponde à unidade de saída,  $f$  à função de activação, e  $I$  ao número de neurónios de entrada.

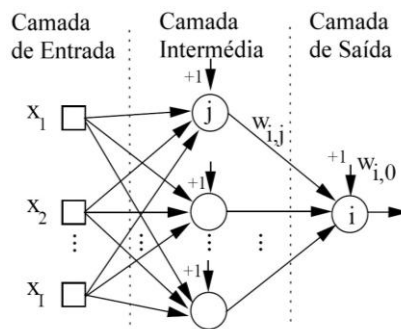


Figura 3 – *Perceptrão Multi-camada*

### 3.3. Processo de *Data Mining*

O processo de *Data Mining* (*DM*) foi levado a cabo no ambiente da ferramenta estatística *R* (R Development Core Team 2006). Nesta ferramenta utilizou-se a biblioteca *RMiner* criada por Cortez (2007), a qual possui um conjunto coerente de funções que facilita o uso de distintas técnicas de *DM*, recorrendo a pacotes disponíveis no programa *R*.

#### 3.3.1. Modelo de classes para a energia de compactação

Mostra-se primeiro um modelo para previsão do nível de energia de compactação no caso das camadas em aterro, em função ( $\sim$ ) de: a classe do material, o seu estado hídrico, a situação meteorológica e a eventual solução de tratamento para o material (Equação 2). Esse modelo foi

ajustado com os registos das “Tabelas de utilização dos materiais em aterro” do guia *GTR*. Para determinados registos especifica-se que não é aconselhável compactar (NC), e para outros casos identifica-se a necessidade de reclassificar (R) o material. Desta forma, para além dos possíveis níveis de energia Leve, Média e Pesada, introduziram-se as etiquetas “NC” e “R”.

$$\text{Energia} \sim \text{Material} + \text{Estado} + \text{Meteorologia} + \text{Solução Construtiva}. \quad (2)$$

Para a avaliação deste tipo de modelos é usualmente utilizado o conceito da matriz de confusão, a qual apresenta na diagonal o número de avaliações correctas em cada classe, quantificando nas restantes casas o número de avaliações erradas. O Quadro 1 mostra uma matriz de confusão para o ajuste de (2) através de uma rede neuronal, treinada e testada com todos os (503) registos das “Tabelas de utilização dos materiais em aterro”. Na totalidade dos registos observa-se, que a referida técnica apenas erra numa ocasião, ao classificar uma energia pesada como sendo média.

Quadro 1 – Matriz de confusão para o ajuste de (2) através de uma rede neuronal

	Leve	Média	Pesada	NC	R
Leve	23	0	0	0	0
Média	0	190	0	0	0
Pesada	0	1	154	0	0
NC	0	0	0	128	0
R	0	0	0	0	7

NC: Não é aconselhável Compactar; R: é necessário Reclassificar o geomaterial

Uma forma simples de medir o desempenho de um modelo na tarefa de classificação é calcular o rácio de acertos em relação ao número total de registos, podendo-se designar esta medida por *acuidade*. Além disso, caso a técnica de *DM* seja de parametrização aleatória, será conveniente utilizar uma estratégia de múltiplas execuções. Assim, o modelo (2) foi ajustado através de 20 redes neuronais, obtendo-se uma acuidade média de  $99.75 \pm 0.04$  %, sendo o intervalo de confiança determinado via um teste *t-student* para um nível de confiança de 95 %.

Em conclusão, o modelo (2) ajustado com uma rede neuronal, permite obter uma especificação para o nível de energia com uma fiabilidade próxima dos 100 %. Note-se que este modelo permite uma extrapolação para condições distintas das catalogadas nas recomendações francesas. A validade desta extrapolação carece ainda de maior aprofundamento.

### 3.3.2. Ajuste da rede neuronal para regressão

Pretendendo a replicação dos dados contidos nas “Tabelas de compactação”, as distintas técnicas de *DM* foram avaliadas quanto à sua capacidade de reprodução dos (422) registos para o caso das camadas em aterro e dos (127) registos para o caso das camadas de leito do pavimento. A técnica baseada em redes neuronais foi aquela que permitiu uma maior fiabilidade na reprodução dos registos.

Tendo em conta que a técnica baseada em RNAs tem uma parametrização inicial aleatória, com vista a obter uma maior confiança estatística, ajustaram-se 20 modelos distintos, utilizando todos os dados quer na fase de treino quer na fase de teste. De entre estes modelos, seleccionou-se aquele com menor *Desvio Absoluto (DA)*, calculado da seguinte forma:

$$DA = \sum_{i=1}^N |y_i - \hat{y}_i| \quad (3)$$

onde *N* designa o número de exemplos,  $y_i$  o valor desejado e  $\hat{y}_i$  o valor estimado pelo modelo.

### 3.3.3. Reprodução das condições de compactação

Partindo do conhecimento das variáveis a montante, particularmente o tipo de material, a classe do compactador, e igualmente a energia de compactação no caso das camadas em aterro, usa-se o modelo de rede neuronal para reproduzir as condições de compactação (Figura 4). Mostram-se nas Figura 5 e 6, respectivamente para as camadas em aterro e de leito do pavimento, os gráficos de dispersão de pontos dos valores estimados *versus* os tabelados no guia *GTR*, para o número de passagens (N) e para o rendimento teórico (Q/L).

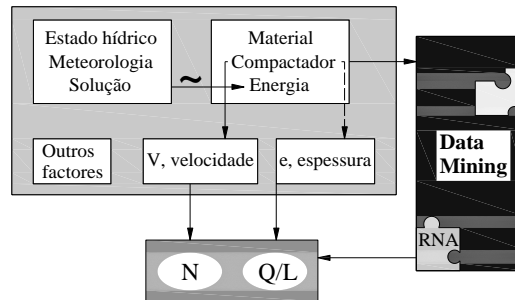


Figura 4 – Fluxograma para determinação das condições de compactação no caso das camadas em aterro

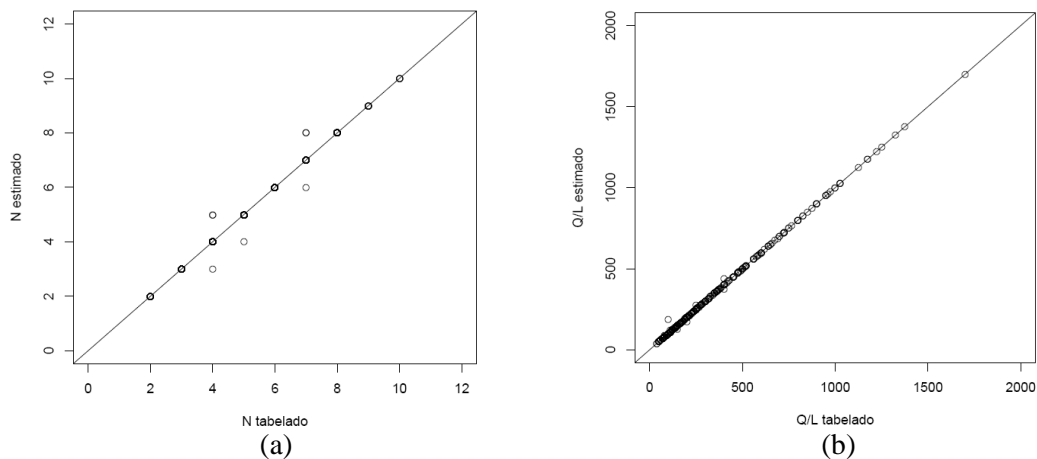


Figura 5 – Gráficos de dispersão de pontos dos valores de N e Q/L estimados *versus* os tabelados no guia *GTR*, para o caso das camadas em aterro

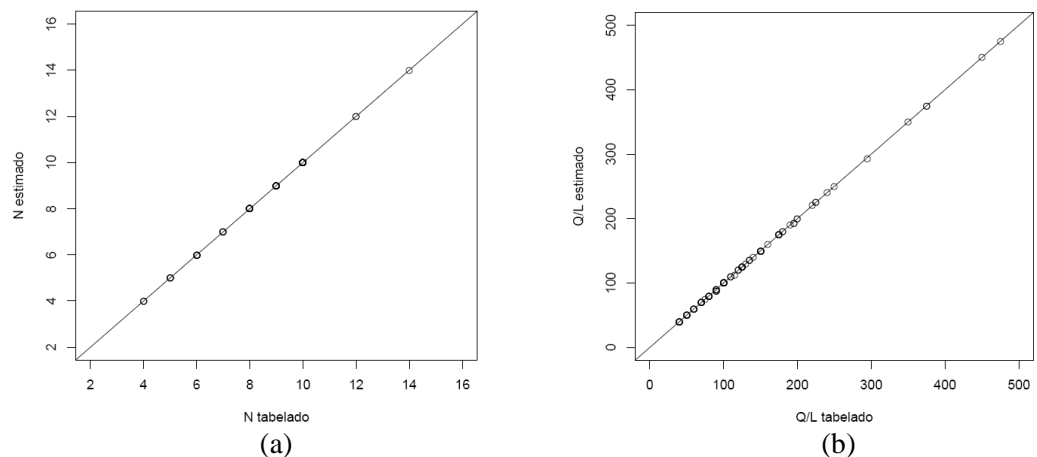


Figura 6 – Gráficos de dispersão de pontos dos valores de N e Q/L estimados *versus* os tabelados no guia *GTR*, para o caso das camadas de leito do pavimento

Como se pode observar, as previsões ajustam-se de modo quase perfeito aos valores tabelados, sobretudo para o caso do leito do pavimento. Ainda que a escassez de dados determine normalmente uma menor qualidade dos ajustamentos, a estratégia de validação adoptada, treinando e testando o modelo com todos os dados, e a menor possibilidade de erro associada a uma amostra mais pequena, permitem obter melhores resultados para o leito do pavimento.

#### 4. CONCLUSÕES

De um modo geral, a técnica de *Data Mining* que melhor se ajusta ao problema em estudo é a baseada em Redes Neurais Artificiais (RNAs). O elevado desempenho obtido com esta técnica confirma a não linearidade do problema. De realçar que os modelos de RNAs possibilitam uma reprodução fiel dos dados contidos nas “Tabelas de compactação”.

A actualização e adequação futuras da base de dados *GTR* poderão ter como suporte os modelos admitidos no presente estudo, com os quais poderá igualmente alimentar-se um processo de optimização de estaleiros de compactação.

#### AGRADECIMENTOS

Este estudo foi financiado pela Fundação para a Ciência e a Tecnologia no âmbito do Projecto POCI/ECM/61114/2004 – “Interaction soil-railway track for high speed trains”.

#### REFERÊNCIAS

- Basheer, I.A. e Najjar, Y.M. (1995). A neural-network for soil compaction. *Proc. 5th Int. Symp. Numerical Models in Geomechanics*, Davos, Switzerland (eds. Pande, G.N. e Pietruszczak, S.), Rotterdam: Balkema, 435-440.
- Basheer, I.A. (2001). Empirical modelling of the compaction curve of cohesive soils. *Can. Geotech. J. / Rev. can. geotech.* 38(1): 29-45.
- Cortez, P. (2007). RMiner: *Data Mining with Neural Networks and Support Vector Machines using R*, In Introduction to Advanced Scientific Softwares and Toolboxes, International Association for Engineering, In press.
- Cruz, P.J.S., Lourenço, L., Santos, M., Quintela, H. e Cortez, P. (2006). Strength of corroded tapered plate girders under pure shear. *Proceedings of the International Conference on Bridge Maintenance, Safety and Management*, Porto, Portugal. Taylor & Francis, ISBN 978-0-415-40315-3.
- Ferreira da Silva, J.F. (2005). Localização e gestão optimizadas de sistemas de captação e de controlo da intrusão marinha em aquíferos costeiros. *7.º Simpósio de Hidráulica e Recursos Hídricos dos Países de Língua Oficial Portuguesa (SILUSBA)*, Évora, Portugal.
- Hastie, T., Tibshirani, R. e Friedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer-Verlag, NY, USA.
- Haykin, S. (1999). *Neural Networks – A Comprehensive Foundation*. Prentice-Hall, New Jersey.
- Kotdawala, S.J. e Hossain, M. (1994). Knowledge and Data-Driven Expert-System for Soil Compaction. *Proc. 8th Int. Conf. Computer Methods and Advances in Geomechanics, Morgantown*, (eds. Siriwardane, H.J. e Zaman, M.M.), Rotterdam: Balkema, Vol. 1, 465-470.
- Lourenço, J., Ramos, L., Ramos, R., Santos, H. e Fernandes, D. (2005). Urban areas identification through clustering trials and the use of neural networks. *14th European Colloquium on Theoretical and Quantitative Geography*, Tomar, Portugal.

- Martins, F., Araújo, A. e Marques, R. (2008). Análise da estabilidade e optimização da geometria de taludes recorrendo a técnicas de inteligência artificial. *XI Congresso Nacional de Geotecnia*, Coimbra, Portugal.
- Miranda, T., Gomes Correia, A., Silva, C. e Ribeiro e Sousa, L. (2006). Desenvolvimento de Sistemas de Inteligência Artificial para Análise de Risco em Túneis. *Anais do III Congresso Luso-Brasileiro de Geotecnia*, Curitiba, Brasil, 155-160.
- Najjar, Y.M., Basheer, I.A. e Naouss, W.A. (1996). On the identification of compaction characteristics by neuronets. *Computers and Geotechnics*, 18(3): 167-187.
- R Development Core Team (2006). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL: <http://www.R-project.org>.
- SETRA e LCPC (1992). Service d'Etudes Techniques des Routes et Autoroutes e Laboratoire Central des Ponts et Chaussées. *Réalisation des remblais et des couches de forme*. France, Ministère de l'Equipement, du Logement et des Transports.
- Shahin, M.A., Jaksa, M.B. e Maier, H.R. (2001). Artificial neural network applications in geotechnical engineering. *Australian Geomechanics*, 36(1): 49-62.
- Toll, D.G. (1996a). Artificial Intelligence Systems for Geotechnical Engineering with specific reference to Ground Improvement. *10th European Young Geotechnical Engineers' Conference Izmir*, URL: <http://www.ins.itu.edu.tr/10ygec/ygec.html>.
- Toll, D.G. (1996b). Artificial Intelligence Applications in Geotechnical Engineering. *Electronic Journal of Geotechnical Engineering*, URL: <http://geotech.civen.okstate.edu/ejge/ppr9608/index.htm>.
- Ubarri, V.E., Cabrera, J.A.H. e Boccaleri, S.R. (1989). *Terraplenes y pedraplenes*. Madrid, MOPU - Ministerio de Obras Públicas y Urbanismo, Secretaría General Técnica, Centro de Publicaciones.