



Estrutura de redes neurais artificiais para estimar capacidade de campo em solos do Estado de Santa Catarina⁽¹⁾.

Diego Bortolini⁽²⁾; Jackson Adriano Albuquerque⁽³⁾; Cleber Rech⁽⁴⁾, Luana da Silva⁽²⁾; Rubia Borges Mendes Sponholz De Oliveira⁽⁵⁾.

⁽¹⁾ Trabalho executado com recursos do CNPq, FAPESC e UDESC.

⁽²⁾ Estudante de doutorado do Programa de Pós-graduação em Ciência do Solo, da Universidade do Estado de Santa Catarina; Lages, SC; diegobortanbortolini@gmail.com; ⁽³⁾ Professor; Universidade do Estado de Santa Catarina;

⁽⁴⁾ Apresentador, estudante de doutorado do Programa de Pós-graduação em Ciência do Solo, da Universidade do Estado de Santa Catarina; Lages, SC ⁽⁵⁾ Estudante de mestrado do Programa de Pós-graduação em Ciência do Solo; Universidade do Estado de Santa Catarina.

RESUMO: O uso de redes neurais artificiais (RNA) para estimativa da retenção e disponibilidade de água no solo se apresenta melhor do que o uso das regressões múltiplas lineares. Mas sua melhor estrutura e parâmetros não foram definidos para estudos em solos do Estado de Santa Catarina, e considerando que estes podem variar e resultar em melhorias na estimativa, este trabalho tem por objetivo o desenvolvimento e indicação de estrutura e parâmetro de redes neurais artificiais, com algoritmo *backpropagation* no software Weka para estimar capacidade de campo em solos do Estado de Santa Catarina. Foram testados para três modelos, distintas estruturas e parâmetros: número de variáveis de entrada, neurônios, taxa de aprendizagem, momento e tempo de treinamento para estimativa da capacidade de campo, ponto de murcha permanente e água disponível, sendo utilizadas 1184 amostras de solos coletadas em solos do Estado de Santa Catarina, sendo que as RNA foram desenvolvidas no software Weka. De maneira geral o aumento no número de variáveis de entrada melhora a estimativa da CC, e se recomenda o uso de seis neurônios na camada oculta, tempo de treinamento de 2000, taxa de aprendizagem de 0,05 e momento de 0,2, para as estimativas nos solos do Estado de Santa Catarina.

Termos de indexação: Weka, retenção de água, funções de pedotransferência.

INTRODUÇÃO

Propriedades dos solos têm sido utilizadas em diversos modelos matemáticos. Entretanto a determinação direta a campo ou no laboratório de propriedades hidráulicas é onerosa, o que dificulta sua obtenção. Para superar essas dificuldades, alguns pesquisadores propuseram modelos matemáticos para estimar de forma indireta a retenção de água no solo (Meng et al., 1987), que são conhecidos por equações ou funções de pedotransferência ou pedofunções (Bouma, 1989).

Para o desenvolvimento das funções de pedotransferência é necessária a escolha de um método matemático para construção dos modelos. O principal método é a regressão múltipla linear (RML), por ser de fácil obtenção e uso. No Estado de Santa Catarina Costa et al. (2013) utilizou RML para estimativa da água retida à capacidade de campo (CC), ponto de murcha permanente (PMP) e água disponível (AD), mas ainda sem resultados satisfatórios. Em vista disso Mendes (2014) usou a técnica das redes neurais artificiais (RNA) para estimar CC, PMP e AD e comparou os resultados aos das RML e encontrou melhores resultados com o uso das RNA.

No trabalho de Mendes (2014) foi utilizada uma configuração de RNA arbitrária, por não haverem trabalhos que testassem e indicassem estrutura e parâmetros a serem utilizados para estimar propriedades do solo. Nas RNA utilizadas, chamadas de *multilayers perceptrons* usando o algoritmo *backpropagation*, vários são as configurações e parâmetros que podem ser alterados para melhor eficiência das estimativas. Na estrutura das RNA pode ser alterado: número de entradas, camadas ocultas e neurônios; e nos parâmetros podem ser alterados: taxa de aprendizagem, momento, tempo treinamento, limite de validação e dobras para a validação cruzada.

O número de entradas representa o número de variáveis independentes. Na sequência é selecionado o número de camadas e de neurônios em cada camada. Taxa de aprendizagem corresponde ao quanto de erro que é repassado aos nós a cada iteração, e o momento permite aumentar a taxa de aprendizagem sem que ocorram oscilações, e atua no aumento da velocidade de convergência, para encontrar o mínimo de erro local que corresponde a melhor solução do problema (Galvão e Valença, 1999). O limite da validação e as dobras se referem a parâmetros da validação do modelo construído.

Considerando a variação destes parâmetros podem resultar em melhorias na estimativa, este trabalho tem por objetivo o desenvolvimento e



indicação de estrutura e parâmetro de RNA, com algoritmo *backpropagation* no software Weka para estimar a capacidade de campo em solos do Estado de Santa Catarina.

MATERIAL E MÉTODOS

Foram amostrados os horizontes pedogenéticos A, E, AB, BA e B e suas subdivisões de 70 perfis de solo no Estado de Santa Catarina. Isso totalizou um banco de dados com 296 horizontes pedogenéticos e 1.184 amostras com estrutura preservada, coletados com anel volumétrico (diâmetro de 6 cm e altura de 2,5 cm), e 296 amostras com estrutura alterada.

As amostras com estrutura preservada foram saturadas e submetidas a sucções de 1, 6 e 10 kPa em mesa de areia, e de 33, 100, 300, 500 e 1.500 kPa em câmaras de Richards. Foi considerada capacidade de campo (CC) a umidade volumétrica retida na sucção a 10 kPa. Após as amostras foram secas em estufa a 105°C para determinar a densidade do solo.

Na fração terra fina seca ao ar (TFSA) foi determinado: a distribuição granulométrica do solo (argila dispersa em hidróxido de sódio), os teores de argila total (0-0,002 mm), silte (0,002-0,053 mm), areia (0,053-2 mm) segundo Day (1965); o limite de plasticidade pelo método de Casagrande (Embrapa, 2011); o teor de carbono orgânico pelo método de Walkley & Black, modificado por Tedesco et al. (1995); e a densidade de partículas (Embrapa, 2011).

As RNA foram desenvolvidas utilizando software livre WEKA (Waikato Environment for Knowledge Analysis), versão 3.6.11 desenvolvido na University of Waikato Hamilton - New Zealand. Foi utilizada a função *Perceptron* de Camadas Múltiplas (*MultiLayerPerceptron* - MLP), e o algoritmo *backpropagation* para o treinamento das RNA's, com uma camada oculta, sendo que foram testados variações na estrutura e nos parâmetros da rede: número de variáveis de entrada, número de neurônios na camada oculta, taxa de aprendizagem, momento e tempo de treinamento. Foram testadas as seguintes variações:

- Número de variáveis de entrada:

3: teor de areia, teor de silte e teor de argila do solo;

4: teor de areia, teor de silte, teor de argila e teor de carbono orgânico do solo;

8: teor de areia, teor de silte, teor de argila, teor de carbono orgânico, densidade de partículas, densidade do solo, porosidade total e limite de

plasticidade do solo.

- Neurônios na camada oculta: 6; 10 e 20 neurônios;

- Taxa de aprendizagem: 0,01; 0,05 e 0,1;

- Momento: 0,1; 0,2 e 0,3;

- Tempo de treinamento: 1000 e 2000.

Juntamente, foi adotado o critério da validação cruzada em 10 dobras (*cross-validation*), com isso é possível estimar a precisão do método, onde o conjunto de dados é dividido aleatoriamente em 10 grupos (dobras) com tamanho e distribuição semelhantes. Uma dobra é separada como teste e as 9 dobras restantes são consideradas para treinamento da rede. Cada par (dobra de teste, dobra de treinamento) constitui uma iteração para a validação do algoritmo, totalizando 10 iterações. Após o treinamento, utilizando a dobra de treinamento, a dobra de teste é apresentada à rede em questão e o erro é calculado. Finalmente, a média das 10 dobras fornece uma estimativa do desempenho do algoritmo (Ferreira, 2005).

No programa WEKA os dados sofrem normalização automaticamente, sendo esta opção a ser selecionada no programa. Para o teste de configurações de estrutura de RNA's foi utilizado o banco de dados com 1184 amostras de horizontes A e B, sendo que os três modelos testados foram usados para estimar a retenção de água à CC.

A raiz do erro médio quadrático (REQM) e o coeficiente de determinação de Pearson (R^2) foram utilizados para avaliar o desempenho das FPT geradas a partir desta técnica. A indicação da melhor estrutura e configuração dos parâmetros da RNA foi realizada levando em consideração o maior valor de R^2 e menor REQM, buscando uma estrutura com menor número de neurônios, para maior facilidade nos cálculos das estimativas posteriormente da CC.

RESULTADOS E DISCUSSÃO

Inicialmente serão apresentados resultados variando o número de variáveis de entrada, tendo os resultados apresentados na tabela 1.

Para a CC com estrutura de três variáveis de entrada, os valores de R^2 variaram entre 0,60 e 0,63, mas a REQM foi constante, de 0,06 cm³ cm⁻³. O tempo de treinamento de 2000 foi melhor para todos os números de neurônios, momentos e taxas de aprendizagens ($R^2=0,63$). Para a taxa de aprendizagem de 0,01 não houve variação no R^2 para o número de neurônio na camada oculta, tempo de treinamento e momento, com R^2 de 0,63. Entretanto, o R^2 aumentou com a taxa de



aprendizagem e com o aumento do tempo de treinamento, e redução do momento e maior número de neurônios. Na taxa de aprendizagem de 0,2 o resultado de R^2 foi inferior independente do número de camadas ocultas, tempo de treinamento e momento. Portanto, para três variáveis de entrada o mais indicado foram taxas de aprendizado de 0,05 e 0,01 com tempo de treinamento de 2000, independente do número de neurônios e do momento.

Para a estrutura de rede com quatro variáveis de entrada, os valores de R^2 variaram entre 0,70 e 0,74, e o REQM variou de 0,05 a 0,06 $\text{cm}^3 \text{cm}^{-3}$. O tempo de treinamento de 2000 apresentou os melhores resultados em todos os números de neurônios, momentos e taxas de aprendizagem testadas. Valores de $R^2=0,73$ com $\text{REQM}=0,05$ foram observados para todos os números neurônios na camada oculta, nas taxas de aprendizagens de 0,05 e 0,1, tempo de treinamento de 2000, e em todos os momentos testados, portanto estas configurações podem ser indicadas para o modelo com quatro variáveis na camada de entrada.

Para a estrutura de rede com oito variáveis de entrada, os valores de R^2 variaram entre 0,78 e 0,87, e REQM variou entre 0,04 e 0,06 $\text{cm}^3 \text{cm}^{-3}$. Os resultados com maior R^2 ($\geq 0,86$) e com menor REQM foram observados no tempo de treinamento de 2000. Também nas taxas de aprendizagem de 0,05; 0,1 e 0,2 em todos os momentos com redes de 10 e 20 neurônios, e RNA mais simples, com seis neurônios, o melhor resultado está no uso da taxa de aprendizagem de 0,05, com momento de 0,2; e taxa de aprendizagem de 0,1 com momento de 0,01 e 0,05.

Observando-se os resultados gerais para os distintos números de variáveis de entrada das RNA, se verificou que o tempo de treinamento de 2000 melhorou os resultados das RNA. O número de camadas ocultas influenciou pouco os resultados. Assim, uma estrutura para construção de RNA's com bons resultados na estimativa da CC, PMP e AD para todos os números de variáveis de entrada é sugerido: seis neurônios na camada oculta; tempo de treinamento de 2000; taxa de aprendizagem de 0,05; momento de 0,2.

Quanto à estrutura, se observou que o aumento no número de variáveis na camada de entrada contribui mais para a melhoria das estimativas do que o aumento do número de neurônios na camada oculta, corroborando ao resultado de Soares et al. (2014). Mas o aumento do número de camadas de entrada gera maior custo e tempo para alimentação do modelo pela necessidade de mais análises.

CONCLUSÕES

O aumento no número de variáveis na camada de entrada melhora a estimativa da retenção e disponibilidade de água.

Indica-se para estimativa da CC, em solos do Estado de Santa Catarina, independente do número de entradas o uso de seis neurônios na camada oculta, tempo de treinamento de 2000, taxa de aprendizagem de 0,05 e momento de 0,2.

AGRADECIMENTOS

Agradecemos a UDESC, FAPESC e CNPq pelo auxílio financeiro através de bolsa de estudo e financiamento à pesquisa.

REFERÊNCIAS

BOUMA, J. Using soil survey data for quantitative land evaluation. *Advances Soil Science*, 9:177-213. 1989.

COSTA, A. da et al. Pedotransfer functions to estimate retention and availability of water in soils of the State of Santa Catarina, Brazil. *Revista Brasileira de Ciência do Solo*, 37:889-910, 2013.

EMBRAPA. Serviço Nacional de Levantamento e Classificação dos Solos. Manual de métodos de análise de solo. 2. Ed rev. 2011. 230 p.

FERREIRA, E.J. Abordagem genética para seleção de um conjunto reduzido de características para construção de ensembles de redes neurais: aplicação à língua eletrônica. São Carlos, Escola de Engenharia de São Carlos, Universidade de São Paulo, 2005. p. (Dissertação de Mestrado).

GALVÃO, C.O.; VALENÇA, M.J.S. (org) Sistemas inteligentes: aplicações a recursos hídricos a ciências ambientais. Porto Alegre: Editora da Universidade, UFRGS, ABRH, 1999. 246p.

MENDES, R. B. Predição da retenção de água em solos de Santa Catarina através de redes neurais artificiais. Dissertação (Mestrado em Ciência do Solo) - Universidade do Estado de Santa Catarina. Lages, 2014.

MENG, T. P. et al. Models to predict water retention in semiarid sandy soil. *Soil Science Society of America Journal*, 51:1563-1565, 1987.

SOARES, F. C. et al. Redes neurais artificiais na estimativa da retenção de água do solo. *Ciência Rural*, Santa Maria, 44:293-300, 2014.

TEDESCO, M.J. et al. Análises de solo, plantas e outros materiais. 2. ed. Porto Alegre: Universidade Federal do Rio Grande do Sul, 1995. 174p.

Tabela 1 – Resultados de desempenho de RNA para estimativa da CC em solos de Estado de SC.

| Entrada | Número de neurônios | | 6 | | 10 | | 20 | | | |
|---------|-----------------------|---------|----------------------|----------------|------|----------------|------|----------------|------|------|
| | Tempo de aprendizagem | Momento | Tempo de treinamento | R ² | REQM | R ² | REQM | R ² | REQM | |
| 3 | 0,01 | 0,1 | 1000 | 0,63 | 0,06 | 0,63 | 0,06 | 0,63 | 0,06 | |
| | | | 2000 | 0,63 | 0,06 | 0,63 | 0,06 | 0,63 | 0,06 | |
| | | 0,2 | 1000 | 0,63 | 0,06 | 0,63 | 0,06 | 0,63 | 0,06 | |
| | | | 2000 | 0,63 | 0,06 | 0,63 | 0,06 | 0,63 | 0,06 | |
| | | 0,3 | 1000 | 0,63 | 0,06 | 0,63 | 0,06 | 0,63 | 0,06 | |
| | | | 2000 | 0,63 | 0,06 | 0,63 | 0,06 | 0,63 | 0,06 | |
| | 0,05 | 0,1 | 1000 | 0,62 | 0,06 | 0,63 | 0,06 | 0,63 | 0,06 | |
| | | | 2000 | 0,63 | 0,06 | 0,63 | 0,06 | 0,63 | 0,06 | |
| | | 0,2 | 1000 | 0,62 | 0,06 | 0,62 | 0,06 | 0,63 | 0,06 | |
| | | | 2000 | 0,63 | 0,06 | 0,63 | 0,06 | 0,63 | 0,06 | |
| | | 0,3 | 1000 | 0,62 | 0,06 | 0,62 | 0,06 | 0,63 | 0,06 | |
| | | | 2000 | 0,63 | 0,06 | 0,63 | 0,06 | 0,63 | 0,06 | |
| | 0,1 | 0,1 | 1000 | 0,62 | 0,06 | 0,62 | 0,06 | 0,63 | 0,06 | |
| | | | 2000 | 0,63 | 0,06 | 0,63 | 0,06 | 0,63 | 0,06 | |
| | | 0,2 | 1000 | 0,62 | 0,06 | 0,62 | 0,06 | 0,63 | 0,06 | |
| | | | 2000 | 0,62 | 0,06 | 0,63 | 0,06 | 0,63 | 0,06 | |
| | | 0,3 | 1000 | 0,62 | 0,06 | 0,62 | 0,06 | 0,62 | 0,06 | |
| | | | 2000 | 0,62 | 0,06 | 0,63 | 0,06 | 0,63 | 0,06 | |
| | 4 | 0,01 | 0,1 | 1000 | 0,71 | 0,05 | 0,71 | 0,05 | 0,70 | 0,05 |
| | | | | 2000 | 0,72 | 0,05 | 0,72 | 0,05 | 0,72 | 0,05 |
| | | | 0,2 | 1000 | 0,71 | 0,05 | 0,71 | 0,05 | 0,71 | 0,05 |
| | | | | 2000 | 0,72 | 0,05 | 0,72 | 0,05 | 0,72 | 0,05 |
| | | | 0,3 | 1000 | 0,71 | 0,05 | 0,71 | 0,05 | 0,71 | 0,05 |
| | | | | 2000 | 0,72 | 0,05 | 0,72 | 0,05 | 0,72 | 0,05 |
| 0,05 | | 0,1 | 1000 | 0,72 | 0,05 | 0,72 | 0,05 | 0,72 | 0,05 | |
| | | | 2000 | 0,73 | 0,05 | 0,74 | 0,05 | 0,73 | 0,05 | |
| | | 0,2 | 1000 | 0,72 | 0,05 | 0,72 | 0,05 | 0,72 | 0,05 | |
| | | | 2000 | 0,73 | 0,05 | 0,74 | 0,05 | 0,73 | 0,05 | |
| | | 0,3 | 1000 | 0,72 | 0,05 | 0,72 | 0,05 | 0,72 | 0,05 | |
| | | | 2000 | 0,72 | 0,05 | 0,74 | 0,05 | 0,73 | 0,05 | |
| 0,1 | | 0,1 | 1000 | 0,72 | 0,05 | 0,72 | 0,05 | 0,72 | 0,05 | |
| | | | 2000 | 0,73 | 0,05 | 0,73 | 0,05 | 0,73 | 0,05 | |
| | | 0,2 | 1000 | 0,72 | 0,05 | 0,72 | 0,05 | 0,72 | 0,05 | |
| | | | 2000 | 0,73 | 0,05 | 0,73 | 0,05 | 0,73 | 0,05 | |
| | | 0,3 | 1000 | 0,72 | 0,05 | 0,72 | 0,05 | 0,72 | 0,05 | |
| | | | 2000 | 0,73 | 0,05 | 0,73 | 0,05 | 0,73 | 0,05 | |
| 8 | | 0,01 | 0,1 | 1000 | 0,83 | 0,04 | 0,78 | 0,05 | 0,79 | 0,05 |
| | | | | 2000 | 0,85 | 0,04 | 0,80 | 0,05 | 0,80 | 0,04 |
| | | | 0,2 | 1000 | 0,84 | 0,04 | 0,79 | 0,04 | 0,79 | 0,05 |
| | | | | 2000 | 0,86 | 0,04 | 0,80 | 0,04 | 0,81 | 0,04 |
| | | | 0,3 | 1000 | 0,84 | 0,04 | 0,79 | 0,05 | 0,79 | 0,05 |
| | | | | 2000 | 0,85 | 0,04 | 0,81 | 0,04 | 0,81 | 0,04 |
| | 0,05 | 0,1 | 1000 | 0,83 | 0,04 | 0,84 | 0,04 | 0,83 | 0,04 | |
| | | | 2000 | 0,85 | 0,04 | 0,86 | 0,04 | 0,86 | 0,04 | |
| | | 0,2 | 1000 | 0,84 | 0,04 | 0,84 | 0,04 | 0,84 | 0,04 | |
| | | | 2000 | 0,86 | 0,04 | 0,86 | 0,04 | 0,86 | 0,04 | |
| | | 0,3 | 1000 | 0,84 | 0,04 | 0,84 | 0,04 | 0,84 | 0,04 | |
| | | | 2000 | 0,85 | 0,04 | 0,86 | 0,04 | 0,86 | 0,04 | |
| | 0,1 | 0,1 | 1000 | 0,84 | 0,04 | 0,85 | 0,04 | 0,85 | 0,04 | |
| | | | 2000 | 0,85 | 0,04 | 0,86 | 0,04 | 0,85 | 0,04 | |
| | | 0,2 | 1000 | 0,83 | 0,04 | 0,84 | 0,04 | 0,85 | 0,04 | |
| | | | 2000 | 0,84 | 0,04 | 0,86 | 0,04 | 0,85 | 0,04 | |
| | | 0,3 | 1000 | 0,84 | 0,04 | 0,84 | 0,04 | 0,85 | 0,04 | |
| | | | 2000 | 0,84 | 0,04 | 0,86 | 0,04 | 0,85 | 0,04 | |