

## FUNÇÃO DE PEDOTRANSFERÊNCIA PSEUDO-CONTÍNUA PARA ESTIMAÇÃO DA UMIDADE DO SOLO

GABRIEL ALMEIDA GOMES<sup>1</sup>; MIGUEL DAVID FUENTES-GUEVARA<sup>2</sup>; ULISSES  
BRISOLARA CORRÊA<sup>3</sup>; STEFAN DOMIGUES NACHTIGALL<sup>4</sup>; RICARDO  
MATSUMURA ARAUJO<sup>5</sup>; LUÍS CARLOS TIMM<sup>6</sup>

<sup>1</sup>Universidade Federal de Pelotas – [gagomes@inf.ufpel.edu.br](mailto:gagomes@inf.ufpel.edu.br)

<sup>2</sup>Universidade Federal de Pelotas – [miguelfuge@hotmail.com](mailto:miguelfuge@hotmail.com)

<sup>3</sup>Instituto Federal Sul-Rio-Grandense – [ulissescorrea@ifsul.edu.br](mailto:ulissescorrea@ifsul.edu.br)

<sup>4</sup>Universidade Federal de Pelotas – [stefan.tefo@gmail.com](mailto:stefan.tefo@gmail.com)

<sup>5</sup>Universidade Federal de Pelotas – [ricardo@inf.ufpel.edu.br](mailto:ricardo@inf.ufpel.edu.br)

<sup>6</sup>Universidade Federal de Pelotas – [lctimm@ufpel.edu.br](mailto:lctimm@ufpel.edu.br)

### 1. INTRODUÇÃO

Com o objetivo de tomar boas decisões nas práticas agrícolas em relação ao manejo da água e do solo, é necessário o conhecimento do comportamento dos atributos físico-hídricos do solo, como o conteúdo de água no solo (D'EMILIO et al. 2018).

O conteúdo de água no solo é uma variável de difícil obtenção, devido ao tempo gasto na mensuração e ao uso de equipamentos caros na determinação. Por tanto, o uso de métodos alternativos para a sua determinação torna-se necessário, como é o caso das Funções de Pedotransferência (FPT). As FPT são muito utilizadas pois são ferramentas muito úteis na determinação de variáveis de difícil obtenção a partir de outras variáveis mais comuns e de fácil aquisição (GUNARATHNA et al. 2019).

Uma das formas de se obter as FPT é por meio de técnicas de inteligência artificial, como, por exemplo, as redes neurais artificiais (RNA). As RNA demonstram uma estimativa e acurácia melhor do que outros métodos mais corriqueiros devido a sua alta capacidade de resolver problemas complexos.

As FPT são principalmente utilizadas em escalas de vertentes, áreas experimentais ou continentais, e pouco frequentes em escala de bacia hidrográfica, unidade a qual é importante para o planejamento e gestão de recursos hídricos. Para a estimativa do conteúdo de água no solo em grande escala são necessários atributos do solo que comumente são obtidos a partir de levantamentos pedológicos ou de bancos de dados nacionais ou internacionais. Assim, a exploração e geração de métodos que tenham o potencial em estimar o conteúdo de água no solo em escalas maiores, como o caso das bacias hidrográficas, torna-se necessário (NEBEL et al. 2010). Frente ao anteriormente exposto, o objetivo do presente estudo foi desenvolver uma função de pedotransferência pseudo-contínua mediante RNA para a estimativa do conteúdo de água no solo ao longo de uma bacia hidrográfica.

### 2. METODOLOGIA

As áreas de estudos consideradas neste estudo foram a bacia hidrográfica Arroio Pelotas (BHAP) e a sua sub-bacia de cabeceira Sanga Ellert (SBHSE). As quais foram selecionadas devido às suas grandes participações na economia e importância social para a região Sul do Rio Grande do Sul. Mais especificamente nos municípios de Pelotas e Canguçu, onde predomina a agricultura familiar. Trabalhou-se a partir de informações de dados de solo previamente coletados e

analisados nessas bacias hidrográficas. A partir disso, um banco de dados de solo foi criado, contendo 14 atributos de solo de fácil aquisição e 1 variável de difícil determinação (conteúdo de água no solo).

Os atributos do solo de fácil aquisição no banco de dados e considerados como variáveis de entrada foram: teor de argila, silte, areia, densidade do solo, macroporosidade, microporosidade, carbono orgânico do solo, elevação do terreno, declividade, aspecto do terreno, curvatura do terreno, curvatura do plano, curvatura do perfil e o potencial matricial de retenção de água no solo. Como variável de saída foi considerado o conteúdo de água no solo ( $\theta$ ).

Para a escolha do algoritmo de Aprendizado de Máquina (AM) a ser utilizado para a estimativa do  $\theta$ , foram realizados diversos testes no software WEKA (EIBE FRANK et al. 2016), explorando diversos tipos de algoritmos para selecionar o mais bem adaptável para o problema. Dentre os algoritmos proporcionados pelo WEKA, a RNA perceptron multicamadas demonstrou ser o algoritmo com maior potencial para estimar  $\theta$ .

Para a estimativa do conteúdo de água no solo, foi utilizada a metodologia desenvolvida por HAGHVERDI et al. (2012), com adaptações para desenvolver uma FPT pseudo-contínua através da técnica de RNA. Esta técnica utiliza o potencial matricial como um parâmetro de entrada na RNA, possibilitando assim a estimativa do conteúdo de água no solo em qualquer potencial matricial utilizado durante o treinamento da RNA. Conseqüentemente, a RNA foi estabelecida com um único parâmetro de saída referente ao  $\theta$ , estimando o  $\theta$  no potencial matricial que foi considerado como parâmetro de entrada. Assim, uma ampla faixa de potenciais matriciais foi usada como parâmetros de entrada (0; -1; -6, -10; -33, -100 e -1,500 kPa), permitindo a estimativa das correspondentes faixas de umidade que deram origem a uma curva pseudo-contínua.

O banco de dados estabelecido foi dividido em dois subconjuntos, um contendo 80% dos dados de entrada e saída (conjunto de treinamento) e o outro contendo o 20% restante dos dados (conjunto de teste). Essa separação foi necessária para treinar o modelo e posteriormente avaliá-lo. Depois dessa separação dos dados, houve outra divisão no conjunto de treinamento, sendo separados 10% dos dados dentro do algoritmo, que posteriormente seria utilizado como conjunto de validação durante o treinamento da RNA, teste esse necessário para poder verificar-se a presença de sobreajuste no treinamento, e também para permitir o uso da técnica de parada antecipada do treinamento.

A partir disso, a arquitetura da RNA MLP(Perceptron Multicamadas -- do inglês multilayer perceptron) foi desenvolvida na linguagem de programação Python, devido à grande quantidade de bibliotecas que a linguagem possui, facilitando a criação de RNAs. A RNA inicial desenvolvida consistiu de 3 camadas principais, sendo elas, uma camada de entrada, uma camada oculta e uma camada de saída, entre essas camadas foi utilizada a função de ativação *Rectified Linear Unit* (ReLU), assim como uma função de normalização de *Batch* com o objetivo de normalizar as entradas da rede. Na otimização da rede, foi utilizado o otimizador Gradiente Descendente Estocástico(SGD), ademais de um agendador para a variação do parâmetro da taxa de aprendizado à medida em que o treinamento da rede fosse realizado.

Um dos objetivos desta pesquisa foi a seleção das variáveis de entrada mais significativas para a estimativa do  $\theta$ . Para isso, o treinamento com diferentes subconjuntos de dados de entrada foram necessários nesta etapa, sendo necessária a escolha de um método de procura entre o método da Busca em Grade e a Busca Aleatória, porém, como o método da Busca em Grade seria

muito custoso computacionalmente a ser realizado, optou-se por utilizar a Busca Aleatória, pois mesmo que não garantisse o resultado ótimo, poder-se-ia explorar mais subconjuntos de variáveis de entrada de uma forma aleatória.

Para isso, os hiperparâmetros do modelo foram variados, assim como as variáveis de entrada, realizando diversos treinamentos e comparando as funções de custo dos modelos (função *Loss*), a qual permite avaliar a qualidade do treinamento da rede com base nos erros de estimativa nos conjuntos de treinamento e de validação. A função *Loss* escolhida para a verificação dos ajustes dos pesos da RNA foi a função Erro Quadrático Médio (MSE).

Outros ajustes de hiperparâmetros foram realizados na RNA como a técnica de parada antecipada do treinamento, uma técnica utilizada para a diminuição de sobreajuste no treinamento, visando que a rede obtenha uma melhoria no desempenho de estimação do  $\theta$  e na velocidade de treinamentos. A avaliação do desempenho da RNA foi realizada mediante análise estatística das estimativas de  $\theta$  em relação aos dados medidos, mediante os índices  $R^2$ , MAE e RMSE.

### 3. RESULTADOS E DISCUSSÃO

A RNA desenvolvida demonstrou uma boa capacidade de estimativa do  $\theta$ , utilizando como variáveis de entrada as variáveis, argila, densidade do solo, microporosidade, carbono orgânico do solo, curvatura do terreno, curvatura do plano.

A figura 1a demonstra o desempenho do melhor modelo de RNA encontrado até o momento, com as métricas estatísticas de  $R^2= 0,71$ , MAE= 0,06  $\text{cm}^3 \text{cm}^{-3}$  e RMSE=0,07  $\text{cm}^3 \text{cm}^{-3}$ , percebendo-se a proximidade da curva de predição à curva de dados reais.

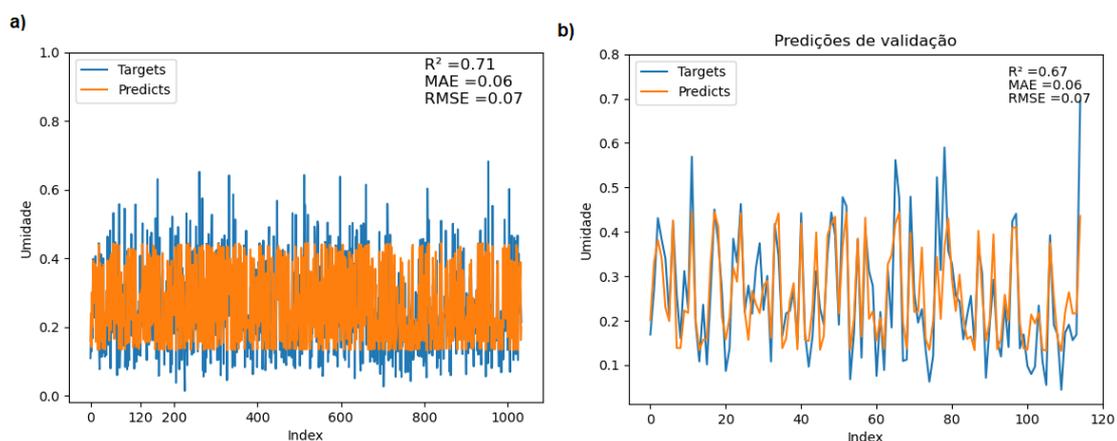


Figura 1. a) Comparação da predição de  $\theta$  com os dados do conjunto de treinamento. b) Comparação da predição de  $\theta$  com o conjunto de validação.

Na imagem 1b percebe-se a ótima qualidade do treino, já que podemos observar o acompanhamento da tendência dos dados medidos quando utilizado os dados do conjunto destinado à validação, demonstrando assim se há a presença de sobreajuste na RNA, sendo assim é possível afirmar que há pouca presença de sobreajuste.

Na imagem 2a, é apresentada a curva da função *Loss* durante o treino, apresentando um comportamento satisfatório, sem grandes oscilações. Por outro lado na imagem 2b, é mostrada a representação gráfica da comparação entre o modelo estimando  $\theta$  e os valores reais do conjunto de dados para teste. Assim, pode-se verificar uma boa qualidade do modelo, com estatísticas de erro muito próximas do desempenho no treinamento e validação, permitindo uma boa estimativa do  $\theta$ .

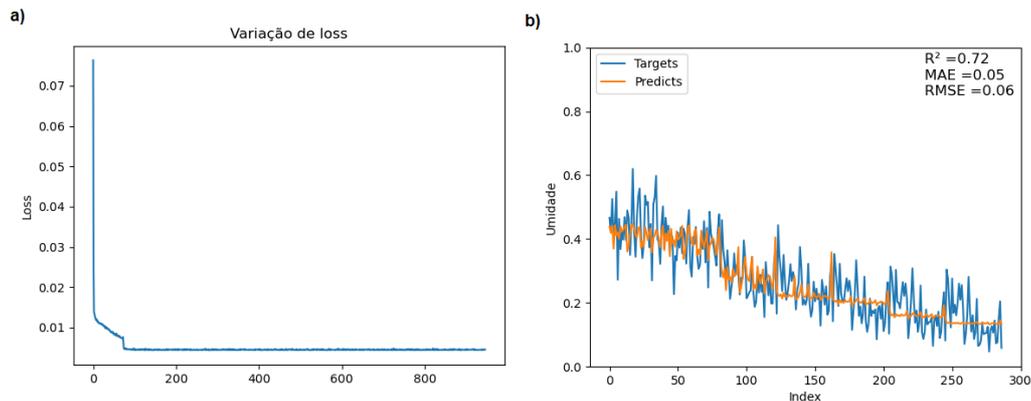


Figura 2. a) Curva demonstrando a variação da função *Loss* durante o treinamento. b) Comparação das previsões com os dados do conjunto teste.

#### 4. CONCLUSÕES

A RNA desenvolvida demonstrou potencial para a estimativa do  $\theta$ , acompanhando a tendência dos valores observados. Além disso, se quisermos aprimorar a qualidade do modelo, é possível continuar explorando diferentes hiperparâmetros e diferentes subconjuntos de entradas. Esta é uma solução válida para melhorar as estimativas e continuar aprimorando o desempenho da RNA, porém, com os resultados já apresentados, é possível afirmar que o modelo já possui um comportamento muito satisfatório para a estimativa do  $\theta$ .

#### 5. REFERÊNCIAS BIBLIOGRÁFICAS

- D'EMILIO, A.; AIELLO, R.; CONSOLI, S.; VANELLA, D.; IOVINO, M. Artificial neural networks for predicting the water retention curve of Sicilian agricultural soils. **Water**, v. 10, n.1431, p. 1-13, 2018.
- FRANK, E.; HALL, M.A.; WITTEN, I.H. **The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques"**, Morgan Kaufmann, Fourth Edition, 2016.
- GUNARATHNA, M; SAKAI, K; NAKANDAKARI, T; MOMII, K; KUMARI, M; AMARASEKARA, M. Pedotransfer functions to estimate hydraulic properties of tropical Sri Lankan soils. **Soil & Tillage Research**, v. 190, p. 109-119, 2019.
- HAGHVERDI, A.; CORNELIS, W.M.; GHAHRAMAN, B.A pseudo-continuous neural network approach for developing water retention pedotransfer functions with limited data. **Journal of Hydrology**, v. 442-443, p. 46-54, 2012.
- NEBEL, Á.L.C.; TIMM, L.C.; CORNELIS, W.; GABRIELS, D.; REICHARDT, K.; AQUINO, L.A.; PAULETTO, E.A.; REINERT, D.J. Pedotransfer functions related to spatial variability of water retention attributes for lowland soils. **Revista Brasileira de Ciência do Solo**, v. 34, n. 3, p. 669-680, 2010.