

DESENVOLVIMENTO DE FUNÇÃO DE PEDOTRANSFERÊNCIA PARA ESTIMAR A CONDUTIVIDADE HIDRÁULICA SATURADA DO SOLO POR MEIO DE INTELIGÊNCIA ARTIFICIAL

LUIS FERNANDO DA SILVA MARTINEZ¹; GABRIEL BORGES DOS SANTOS²;
OTTONI MARQUES MOURA DE LEON³; LUÍS CARLOS TIMM⁴; DANIELLE
BRESSIANI⁵

¹Universidade Federal de Pelotas – nandoonly4work@gmail.com

²Universidade Federal de Pelotas – gabrielqwsantos@gmail.com

³Universidade Federal de Pelotas – ottonibaixo@gmail.com

⁴Universidade Federal de Pelotas – luisctimm@gmail.com

⁵Universidade Federal de Pelotas – daniebressiani@gmail.com

1. INTRODUÇÃO

O solo é um sistema complexo e fundamental para o ecossistema terrestre (LEHMANN, 2020), e obter informações mais precisas sobre os seus parâmetros traz benefícios para diversas áreas. A condutividade hidráulica saturada do solo (Ksat) é uma informação que traz diversos benefícios para a agricultura, como por exemplo, a potencialização de medidas corretivas para o aproveitamento de micronutrientes (SUCHITHRA; PAI, 2020), no auxílio adequado do manejo da irrigação (USOWICZ; LIPIEC, 2021) e é também um parâmetro chave para prever o comportamento hidráulico do solo (ZHANG; SCHAAP, 2019).

Resumidamente, a Ksat é uma propriedade chave para compreender o movimento da água no solo (BLANCO-CANQUI et al., 2017), porém, é uma propriedade na qual a medição direta é onerosa em termos de tempo e custo financeiro, inviabilizando medições referentes a grandes áreas (ZHANG; SCHAAP, 2019). Para suprir tal lacuna são utilizadas as funções de pedotransferência (FPTs), que são um método para estimar parâmetros do solo de difícil medição, a partir de parâmetros do solo de fácil medição e amplamente disponíveis em bancos de dados (ADHIKARY, 2008).

Muitas FPTs vêm sendo desenvolvidas para estimar propriedades hidráulicas do solo e diferentes técnicas, como estatísticas multivariadas avançadas, são utilizadas na construção do modelo de previsão (BEVINGTON, et al., 2016). Com isso, este trabalho tem por objetivo utilizar um modelo de aprendizado de máquina conhecido como redes neurais artificiais, para estimar a Ksat por meio da utilização de um banco de dados de solos subtropicais brasileiro.

2. METODOLOGIA

O banco de dados de solos deste estudo abrange diferentes áreas e bacias hidrográficas que se encontram, conforme classificação de Köppen-Geiger, sob influência do clima subtropical úmido (Cfa) (Figura 1), que se caracteriza por possuir uma regularidade na distribuição anual da pluviosidade (FAO, 2012). As áreas e bacias hidrográficas são: Bacia Hidrográfica do Arroio Fragata (BHAF), Bacia Hidrográfica Santa Rita (BHSR), Bacia hidrográfica de Sanga Ellert (BHSE), Bacia Hidrográfica do Arroio Pelotas (BHAP), Área de Estudo Cerro Largo (AECR) e Área de Estudo de Umuarama (AEU). O banco é composto por 552 amostras de solo e 11 variáveis: Textura (areia %, silte% e argila%), densidade do solo (g/cm³), densidade da partícula (g/cm³), porosidade total (cm³/cm³), porosidade efetiva (cm³/cm³), macroporosidade (cm³/cm³), microporosidade (cm³/cm³), teor de matéria orgânica (cm³/cm³), e condutividade hidráulica saturada (variável alvo).

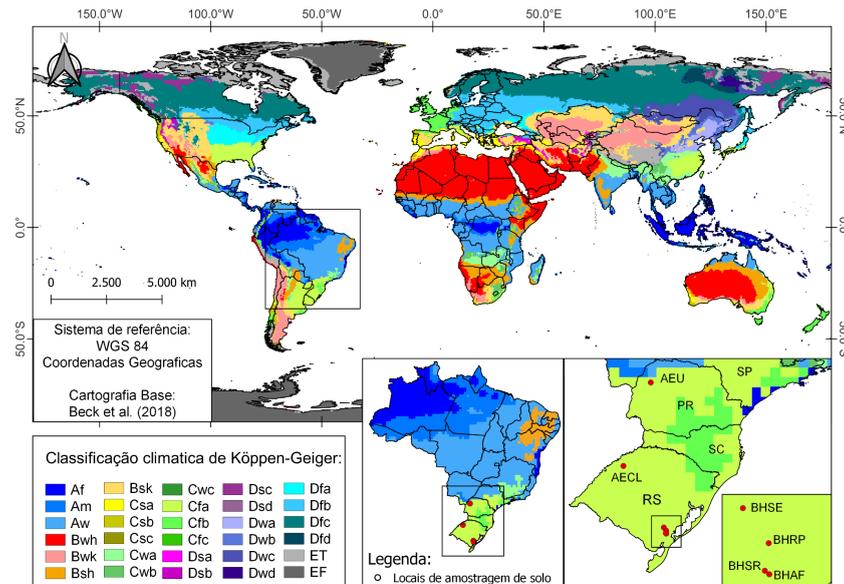


Figura 1 – Localização das áreas amostradas.

O método estrutural para a estimativa da condutividade hidráulica saturada (Ksat) foi realizado no Google Collaboratory (Colab) pelo Jupyter Notebook, com a linguagem Python, utilizando bibliotecas, como: Pandas, Numpy, Matplotlib.pyplot, e Seaborn, e seguiu um roteiro empírico de 8 passos, descrito abaixo:

1. Entender o problema - consta na observação de cada variável e entender seu significado de forma “filosófica” e criar hipóteses sobre a importância de cada variável para o modelo.

2. Estudo univariado - estudo da variável alvo, analisando sua distribuição e detecção de outliers. Referente a distribuição, observou-se que os dados não possuíam boa aderência à distribuição normal. Optou-se então pela transformação da variável utilizando a distribuição de Johnson (Johnson, 1939). Tal transformação ajuda a simplificar a análise estatística e modelagem, facilitando a interpretação e melhorando o desempenho do modelo.

Em relação a outliers, foi constatado 12% de valores acima da referência do limite superior. Neste caso todas as amostras foram mantidas, pois a variabilidade é intrínseca da natureza da ksat, subentendo que não houve qualquer erro no experimento ou erro de anotação.

3. Estudo multivariado - estudo da interrelação entre as variáveis preditoras, assim como, suas relações com a variável alvo. Foi utilizado a Correlação de Pearson, Correlação de Spearman e foram gerados gráficos de dispersão de todas as variáveis.

4. Limpeza Básica - variáveis preditoras que possuíam correlação maior que 0.95 foram observadas. Essas variáveis possuem basicamente o mesmo valor preditivo referente a variável alvo, não havendo nenhum ganho de informação, servindo apenas para aumentar o ruído amostral. Logo foi mantida apenas uma variável representante desse conjunto e o restante foi deletado.

5. Normalização das variáveis preditoras - a fim de diminuir a volatilidade, todas as variáveis foram padronizadas onde cada amostra foi diminuída da média global da variável e dividido pelo desvio-padrão.

6. Divisão em dados de treino e dados de teste - foi realizada através de validação cruzada, que consiste em dividir os dados em conjuntos de treinamento e teste múltiplos para obter uma avaliação mais robusta do modelo.

7. Aplicação do modelo - foi aplicada a rede neural artificial com ativação Relu;

8. Avaliação do modelo - foi realizada a partir do Coeficiente de Determinação (R^2) e Erro Absoluto Médio (MAE).

3. RESULTADOS E DISCUSSÃO

Com a aplicação do modelo, obteve-se um R^2 de 0,47 e um Erro Absoluto Médio de 19,55. A priori trata-se de um resultado bastante promissor frente a outros estudos da literatura, porém, precisa-se ressaltar que o modelo utilizou 10 variáveis preditoras enquanto as funções tradicionais na literatura são bem mais simples, utilizando em média de 1 a 5 variáveis (ZHAO et al., 2016; XU et al., 2017; ZOU; ZHA; 2021).

Conforme mostrado na Figura 2, para valores de Ksat entre 0 e 50, observa-se um erro relativamente baixo, porém quando o valor da variável de estudo aumenta sua magnitude, também aumenta significativamente o erro. O modelo não foi capaz de adaptar-se a valores extremos da variável.

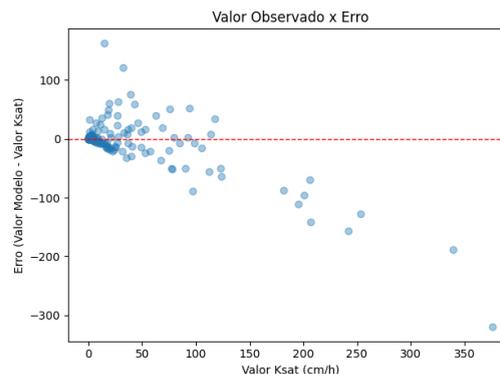


Figura 2 - Valor previsto pelo modelo - Valor Ksat

A rede neural foi aplicada de forma simples, sem qualquer tipo de ajuste de hiperparâmetros. Futuramente poderão ser realizadas as seguintes melhorias neste modelo, tais como: (a) Ajuste de hiperparâmetros - taxa de aprendizado, número de camadas e neurônios, funções de ativação, dentre outros; (b) Troca de algoritmo - pode-se admitir o uso de outro método, caso este mostre uma melhor aderência a este estudo; (c) Feature Engineering - criação de variáveis através da combinação com as variáveis já existentes, a fim de ajudar o modelo a captar melhor a variabilidade da variável de saída; (d) Coleta de mais dados - com um banco de dados maior, espera-se conseguir uma melhor representatividade; (e) Testar outras distribuições nas variáveis de entrada a na variável alvo; e (f) Usar stacking models - combina as previsões de vários modelos para obter uma previsão mais precisa e robusta, geralmente usando outro modelo para realizar essa combinação.

4. CONCLUSÕES

O presente estudo buscou mostrar que através de um número adequado de amostras e com técnicas intermediárias de manipulação de dados e machine learning, é possível criar modelos com poder preditivo maior do que o dos modelos existentes na literatura. O trabalho realizou metodologia de 8 passos, em

linguagem Python, no ambiente Google Colab, para criar um modelo de redes neurais artificiais, para estimar a Ksat por meio da utilização de um banco de dados de solos subtropicais brasileiro.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- ADHIKARY, P. P.; CHAKRABORTY, D.; KALRA, N.; SACHDEV, C. B.; PATRA, A. K.; KUMAR, S.; TOMAR, R. K.; CHANDNA, P.; RAGHAV, D.; AGRAWAL, K.; SEHGAL, M. Pedotransfer functions for predicting the hydraulic properties of Indian soils. **Soil Research**, v. 46, n. 5, p. 476-484, 2008.
- BECK, H. E.; ZIMMERMANN, N. E.; MCVICAR, T. R.; VERGOPOLAN, N.; BERG, A.; WOOD, E.F. Present and future Köppen-Geiger climate classification maps at 1-km resolution. **Nature Scientific Data**, v.5, n.180214, 2018.
- BEVINGTON, J.; PIRAGNOLO, D.; TEATINI, P.; VELLIDIS, G.; MORARI, F. On the spatial variability of soil hydraulic properties in a Holocene coastal farmland. **Geoderma**, v. 262, p. 294-305, 2016.
- BLANCO-CANQUI, H.; WIENHOLD, B. J.; JIN, V. L.; SCHMER, M. R.; KIBET, L. C. Long-term tillage impact on soil hydraulic properties. **Soil and Tillage Research**, v. 170, p. 38-42, 2017.
- FAO. Food and Agriculture Organization of the United Nations. 2012. **Global ecological zones for FAO forest reporting**: 2010 update. FAO Forestry Paper 179. Rome, Italy: FAO.
- JOHNSON, N. L. Systems of Frequency Curves Generated by Methods of Translation. **Biometrika**, v. 36, n. 1/2, p. 149-176, 1949.
- LEHMANN, J.; BOSSIO, D. A.; KÖGEL-KNABNER, I.; RILLIG, M. C. The concept and future prospects of soil health. **Nature Reviews Earth & Environment**, v. 1, n. 10, p. 544-553, 2020.
- SUCHITHRA, M. S.; PAI, M. L. Improving the prediction accuracy of soil nutrient classification by optimizing extreme learning machine parameters. **Information processing in Agriculture**, v. 7, n. 1, p. 72-82, 2020.
- TREJO-ALONSO, J.; FUENTES, C.; CHÁVEZ, C.; QUEVEDO, A.; GUTIERREZ-LOPEZ, A.; GONZÁLEZ-CORREA, B. Saturated hydraulic conductivity estimation using artificial neural networks. **Water**, v. 13, n. 5, p. 1-15, 2021.
- USOWICZ, B.; LIPIEC, J. Spatial variability of saturated hydraulic conductivity and its links with other soil properties at the regional scale. **Scientific Reports**, v. 11, n. 1, p. 1-12, 2021.
- XU, C.; XU, X.; LIU, M.; LIU, W.; YANG, J.; LUO, W.; ZHANG, R.; KIELY, G. Enhancing pedotransfer functions (PTFs) using soil spectral reflectance data for estimating saturated hydraulic conductivity in southwestern China. **Catena**, v. 158, p. 350-356, 2017.
- ZHANG, Y.; SCHAAP, M. G. Estimation of saturated hydraulic conductivity with pedotransfer functions: A review. **Journal of Hydrology**, v. 575, p. 1011-1030, 2019.
- ZHAO, C.; SHAO, M.; JIA, X.; NASIR, M.; ZHANG, C. Using pedotransfer functions to estimate soil hydraulic conductivity in the Loess Plateau of China. **Catena**, v. 143, p. 1-6, 2016.
- ZOU, B.; ZHA, Y.. Using machine learning methods as a pedotransfer function to estimate soil hydraulic parameters. **Journal of Irrigation and Drainage**, v. 40, n. 5, p. 81-87, 2021.