

DESEMPENHO DE REGRESSÃO LINEAR MÚLTIPLA E REDES NEURAIS ARTIFICIAIS NA CONSTRUÇÃO DE FUNÇÕES DE PEDOTRANSFERÊNCIA PARA A ESTIMATIVA DO CONTEÚDO DE ÁGUA NO SOLO

DANIEL BAPTISTA DE MEDEIROS¹; MAURÍCIO BRAGA DE PAULA²; LUÍS CARLOS TIMM³

¹Universidade Federal de Pelotas (UFPEL) – danielmedeiroseg@outlook.com

²Universidade Federal de Pelotas (UFPEL) – maubrapa@gmail.com

³ Universidade Federal de Pelotas (UFPEL) – luisctimm@gmail.com

1. INTRODUÇÃO

O conteúdo de água no solo é uma importante variável para as atividades ligadas à agricultura e ao manejo e conservação do solo e da água, uma vez que a disponibilidade hídrica implicará diretamente no sucesso da cultura e no controle de processos erosivos (VERECKEN et al., 2008; WANG et al., 2024). Porém, devido à necessidade de coletar-se observações de solo e posterior tratamento laboratorial, a determinação deste conteúdo torna-se onerosa e muitas vezes inviável. Isto faz com que haja a necessidade da utilização de funções matemáticas denominadas Funções de Pedotransferência (PTFs) (AINA; PERYASWAMI, 1985), que são funções que utilizam variáveis de solo de fácil obtenção, como textura, densidade do solo e tipo de horizonte, para determinar grandezas de obtenção dificultosa, como o conteúdo de água no solo, por exemplo (AL MAJOU, 2008).

Com o avanço da tecnologia, muitas técnicas computacionais têm sido utilizadas para a elaboração de PTFs, dentre elas as Redes Neurais Artificiais (RNAs), que consistem em estruturas complexas formadas por um conjunto de neurônios que se comunicam entre si a partir de variáveis de entrada para identificar um determinado padrão. O algoritmo da RNA irá determinar os pesos sinápticos de cada uma destas variáveis que serão processadas pelos neurônios nas camadas intermediárias (ocultas) e por fim resultará em uma camada de saída que apresentará o valor da variável resposta (GERÓN, 2019).

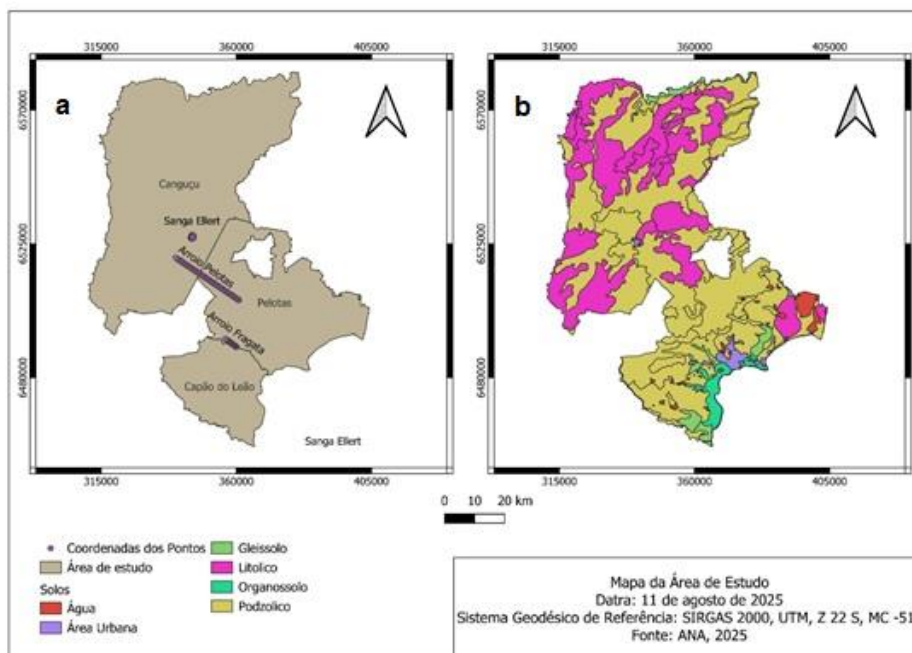
O objetivo do presente trabalho foi comparar o desempenho de PTFs geradas por dois modelos estatísticos diferentes quanto à estimativa do conteúdo de água no solo usando Regressão Linear Múltipla e Redes Neurais Artificiais.

2. METODOLOGIA

2.1 ÁREA DE ESTUDO

O presente estudo foi realizado nas seguintes bacias hidrográficas localizadas no sul do Rio Grande do Sul: Bacia do Arroio Pelotas, Bacia Sanga Ellert e Bacia do Arroio Fragata (Figura 1).

Figura 1: (a) Municípios, e (b) Tipos de Solo, da área de estudo



Fonte: Adaptado de ANA, 2025.

Foram utilizadas 300 amostras de solo coletadas na camada de solo de 0-20 cm na seguinte proporção: 169 pontos na Bacia Sanga Ellert, 100 pontos na Bacia do Arroio Pelotas e 31 pontos na Bacia do Arroio Fragata.

Em cada amostra, foram determinadas as frações texturais do solo (porcentagem de areia, de silte, de argila), carbono orgânico, densidade do solo e conteúdo de água no solo no potencial mátrico de -10 kPa. Foram construídas nove PTFs, uma por meio de Regressão Linear Múltipla e oito por meio de uma RNA com hiperparâmetros variados, conforme a Tabela 1.

Tabela 1: Cenários estudados

	FA	CA	N	O
Regressão Linear Múltipla				
RNA Cenário 1	ReLU	1	10	lbfgs
RNA Cenário 2	Logistic	1	10	lbfgs
RNA Cenário 3	Logistic	2	10, 10	lbfgs
RNA Cenário 4	Logistic	1	10	adam
RNA Cenário 5	Tanh	2	10, 10	lbfgs
RNA Cenário 6	ReLU	3	20, 10, 5	lbfgs
RNA Cenário 7	Tanh	3	20, 10, 5	lbfgs
RNA Cenário 8	Logistic	3	20, 10, 6	lbfgs

Hiperparâmetros da RNA: FA: Função de Ativação; CA: Número de Camadas Ocultas; N: Número de Neurônios por Camada Oculta; e O: Otimizador.

As cinco primeiras variáveis citadas anteriormente foram utilizadas nos modelos como variáveis preditoras, enquanto o conteúdo de água no solo foi adotado como variável resposta.

O modelo de Regressão Linear Múltipla foi construído no software Microsoft Excel, enquanto o modelo de Rede Neural Artificial foi construído na linguagem Python dentro da plataforma Google Colab. Para a Rede Neural, o conjunto amostral foi

dividido em 80% para treinamento e 20% para teste do algoritmo. A taxa de aprendizado utilizada foi o valor considerado padrão para otimizadores adam e sgd (0,001).

Para avaliar o desempenho das PTFs, foram utilizadas as seguintes métricas:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$RMSE = \sqrt{MSE}$$

Em que:

\hat{y}_i : Valor estimado em cada ponto amostral;

y_i : Valor observado em cada ponto amostral;

\bar{y} : Valor médio;

MSE: Erro Médio Quadrático;

RMSE: Raiz Quadrada do Erro Médio Quadrático;

R^2 : Coeficiente de Determinação.

Os valores das métricas apresentadas nas equações anteriores foram comparados entre o modelo gerado pelo Excel e os oito cenários criados na RNA.

3. RESULTADOS E DISCUSSÃO

O modelo que apresentou melhor desempenho foi a RNA do Cenário 2, com uma camada oculta de 10 neurônios, função de ativação logistic e otimizador lbfgs, alcançando $R^2 = 0,523$ (Tabela 2). Embora esse valor seja superior ao obtido pela Regressão Linear Múltipla ($R^2 = 0,408$), o ganho é relativamente modesto e indica que a capacidade preditiva dos modelos ainda é limitada. Observa-se que a escolha da função de ativação foi determinante, já que a troca de logistic por ReLU (Cenário 1) reduziu substancialmente o desempenho, enquanto o aumento do número de camadas ocultas não trouxe benefícios, provavelmente em função do número reduzido de amostras disponíveis (300), insuficiente para explorar arquiteturas mais complexas. O pior resultado foi obtido no Cenário 4, em que apenas o otimizador foi alterado para adam, evidenciando a alta sensibilidade das RNAs a pequenas mudanças nos hiperparâmetros.

Tabela 2: Desempenho dos modelos testados

	MSE (cm ³ /cm ³)	RMSE (cm ³ /cm ³)	R ²
Regressão Linear Múltipla	0,00288	0,05363	0,408
RNA Cenário 1	0,00296	0,05440	0,326
RNA Cenário 2	0,00209	0,04580	0,523
RNA Cenário 3	0,00216	0,04650	0,508
RNA Cenário 4	0,00391	0,06250	0,110
RNA Cenário 5	0,00228	0,04780	0,479
RNA Cenário 6	0,00268	0,05180	0,388
RNA Cenário 7	0,00232	0,04800	0,472
RNA Cenário 8	0,00210	0,04580	0,521

4. CONCLUSÕES

As RNAs apresentam potencial para a construção de funções de pedotransferência, mas sua aplicação prática requer bases de dados extensas, inclusão de variáveis adicionais que capturem melhor os processos físico-hidráulicos do solo e uso de técnicas de regularização e validação mais robustas. Futuros trabalhos deverão explorar esses aspectos no intuito de aumentar o desempenho das RNAs na estimativa do conteúdo de água do solo.

5. REFERÊNCIAS BIBLIOGRÁFICAS

AINA, P. O.; PERIASWAMY, S. P. Estimating available water-holding capacity of western Nigerian soils from texture and bulk density, using core and sieved samples. **Soil Science**, 1985.

AL MAJOU, H.; BRUAND, A.; DUVAL, O.; LE BAS, C.; VAUTIER, A. Prediction of soil water retention properties after stratification by combining texture, bulk density and the type of horizon. **Soil Use and Management**, v. 24, p. 383–391, 2008.

ANA. **Catálogo de Metadados ANA**, 2025. Acessado em: 20 jul. 2025. Disponível em: <https://metadados.snirh.gov.br/geonetwork/srv/por/catalog.search#/home>

GÉRON, Aurélien. **Mãos à obra: aprendizado de máquina com ScikitLearn e Tensor Flow**, 2019.

VERECKEN, H.; HUISMAN, J. A.; BOGENA, H.; VANDERBORGHT, J. VRUGT, J. A.; HOPMANS, J. W. On the value of soil moisture measurements in vadose zone hydrology: A review. **Water Resources Research**, Estados Unidos, v. 44, p. 1-21, 2008.

WANG, Y.; LIANGSHENG, S.; HU, Y.; XIAOLONG, H.; WENXIANG, S.; WANG, L. A comprehensive study of deep learning for soil moisture prediction. **Hydrology and Earth System Sciences**, Munique, v. 28, p. 917-943, 2024.