

## PREENCHIMENTO DE FALHAS METEOROLÓGICAS DE Evaporação Diária VIA Random Forests

**DANIEL FERREIRA AFONSO 1<sup>1</sup>, MATHEUS DE PAULA FERREIRA 2<sup>2</sup>, MÁRCIO JOSÉ DE SANTANA3<sup>3</sup>, ÉDIMO FERNANDO ALVES MOREIRA 4<sup>4</sup>, EUSÍMIO FELISBINO FRAGA JÚNIOR 5<sup>5</sup>**

<sup>1</sup> Doutorando em Engenharia de Sistemas Agrícolas, ESALQ/ USP, e-mail: danielfafonso@usp.br

<sup>2</sup> Doutor em Bioestatística, UFV, email: matheusferreiraufv@gmail.com

<sup>3</sup> Doutor em Engenharia Agrícola, UFLA, email: marcosantana@iftm.edu.br

<sup>4</sup> Doutor em Bioestatística, UFV, email: edimo@iftm.edu.br

<sup>5</sup> Doutor em Engenharia de Sistemas Agrícolas, ESALQ/USP, email: eusimiofraga@ufu.br

Apresentado no  
LII Congresso Brasileiro de Engenharia Agrícola - CONBEA 2023  
18 a 21 de outubro de 2023 – Ribeirão Preto - SP, Brasil

**RESUMO:** As razões das perdas de dados meteorológicos são diversas e comprometem a consistência das séries históricas, prejudicando a sua aplicação. O objetivo do trabalho foi avaliar o uso da modelagem via *Random Forests* para o preenchimento de falhas meteorológicas no município de Uberaba, Minas Gerais, Brasil. Foram extraídos do INMET os dados de Evaporação Diária da estação meteorológica convencional do período de 19 anos. Para o ajuste dos modelos de aprendizado foram considerados, como variáveis de entrada, os valores de cinco dias anteriores e cinco dias posteriores a uma determinada observação  $y$  e esta foi utilizada como variável de saída. Os dados obtidos foram divididos, aleatoriamente, em duas partes: um conjunto para o treinamento e outro para o teste. Foi utilizada a raiz quadrada do erro quadrático médio (RMSE) como medida de qualidade de ajuste. Os modelos de *random forests* apresentaram fraco desempenho (RMSE > 20%) e assim são considerados ineficientes no preenchimento das falhas.

**PALAVRAS-CHAVE:** séries de dados meteorológicos, aprendizado estatístico, dados ausentes.

### FILLING IN METEOROLOGICAL FAILURES OF DAILY EVAPORATION VIA RANDOM FORESTS

**ABSTRACT:** The reasons for the loss of meteorological data are diverse and compromise the consistency of historical series, jeopardizing their application. The objective of this work was to evaluate the use of modeling via Random Forests to fill meteorological gaps in the municipality of Uberaba, Minas Gerais, Brazil. Data from the conventional meteorological station for the period of 19 years were extracted from INMET for the following characteristic Daily Evaporation. For the adjustment of the learning models, the values of five days before and five days after a certain observation  $y$  were considered as input variables, and this was used as an output variable. The data obtained were randomly divided into two parts: a set for training and another for testing. The square root mean square error (RMSE) was used as a measure of goodness of fit. Random forests models performed poorly (RMSE > 20%) and are therefore considered inefficient in filling gaps.

**KEYWORDS:** climate data series, statistical learning, missing data.

**INTRODUÇÃO:** As informações meteorológicas são de grande utilidade na agricultura. Os dados meteorológicos são úteis para as tomadas de decisão no plantio, na colheita, na aplicação de defensivos, entre outras. (ZHANG et al., 2021). Ainda que cuidados sejam tomados para preservar a qualidade e a precisão dos dados provenientes das estações, a maior parte desses é afetada por uma proporção de dados faltantes. As razões das perdas de dados, referidas como falhas, são diversas: reorganização das redes de estações, erro de transmissão, medição, mau funcionamento dos sensores, problemas de calibração, intempéries climáticas e vandalismo (CHINASHO et al., 2021). Há uma grande variedade de técnicas utilizadas para o preenchimento de falhas. Entre elas, podemos citar: o uso da regressão linear múltipla (WOLDESENBET et al., 2016), a ponderação regional; a interpolação do inverso da distância; o método da razão normal e o método tradicional do Reino Unido (KANDA et al., 2018). Mais recentemente, destacamos o uso de modelos de aprendizado estatístico, como as redes neurais artificiais (COUTINHO et al., 2018). Outra técnica de aprendizado estatístico em potencial para o preenchimento de falhas meteorológicas, mas que ainda é usada de maneira incipiente, é a modelagem via *Random Forest*. Um modelo de aprendizado estatístico envolve a construção de um modelo matemático para uma variável de saída, ou variável resposta, em função de uma ou mais variáveis de entradas, também chamadas de variáveis explicativas (HASTIE; TIBSHIRANI; FRIEDMAN, 2009). Em face da diversidade de métodos para preenchimentos de falhas utilizando tanto geo-estatística como outras técnicas de preenchimento muitas vezes complexos e de difícil aplicação a metodologia utilizada neste trabalho se mostra como uma alternativa inovadora. Frente a lacuna de trabalhos avaliando o potencial da metodologia *Random Forest* no preenchimento de falhas de banco de dados meteorológicos o objetivo foi avaliar o uso da modelagem via *Random Forests*, em um contexto de aprendizado estatístico, para o preenchimento de falhas meteorológicas no município de Uberaba, Minas Gerais, Brasil.

**MATERIAL E MÉTODOS:** Os dados utilizados neste trabalho foram provenientes de uma estação meteorológica automática instalada no município de Uberaba, Minas Gerais (latitude 19°73' S, longitude 47°95' N e 737 m de altitude). O clima da região, de acordo com a classificação de Köppen, é caracterizado como Tropical Savannah Aw (BECK et al., 2018). O conjunto de dados meteorológicos foi obtido no banco de dados meteorológicos para ensino e pesquisa do Instituto Nacional de Meteorologia (INMET) e disponibilizado no Banco de Dados Meteorológicos para Ensino e Pesquisa (BDMEP). A estação meteorológica convencional é composta de sensores isolados que registram, continuamente, os dados meteorológicos (pressão atmosférica, temperatura, umidade relativa do ar, precipitação, radiação solar, direção e velocidade do vento, insolação diária e evaporação diária, etc.). Os principais instrumentos utilizados são: actinógrafo, anemômetro, barômetro, evaporímetro de piche, heliógrafo, higrógrafos, pluviógrafo, psicrômetro, termohigrógrafo, termômetro. Esses registros são coletados em três horários de observação às 12 UTC, 18 UTC e 24UTC, respectivamente 9h, 15h e 21h, no horário oficial de Brasília e anotados por um técnico observador, a cada intervalo. Foram extraídos os dados do período de 27/12/1999 a 30/08/2018. As seguintes informações meteorológicas foram coletadas em escala diária: Evaporação Diária (ED). Para o ajuste dos modelos de aprendizado, foram considerados, como variáveis de entrada, os valores de cinco dias anteriores e cinco dias posteriores a uma determinada observação  $y$ . Esta observação  $y$  foi utilizada como variável de saída. O mesmo procedimento foi adotado para todas as variáveis estudadas. O conjunto de dados obtido, foi dividido, aleatoriamente, em duas partes: um conjunto para o treinamento (80% dos valores) e outro para o teste (20%) dos valores. As observações de treinamento são aquelas usadas no ajuste dos modelos. As observações de teste são utilizadas para a avaliação dos modelos ajustados.

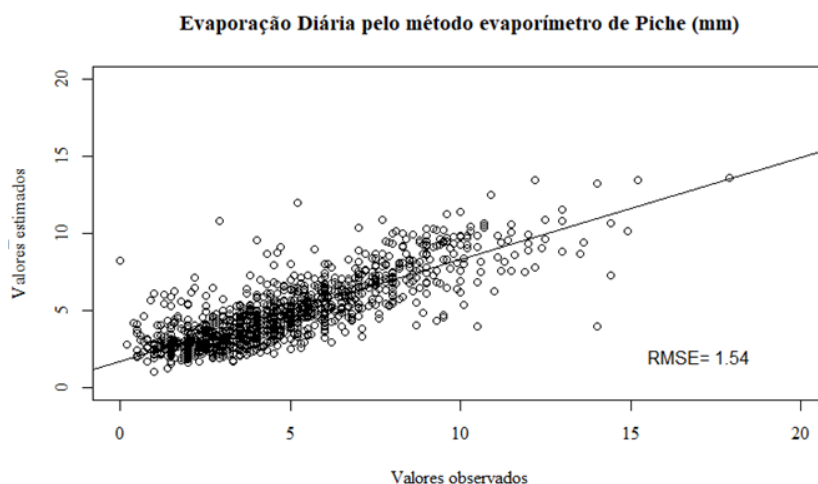
**RESULTADOS E DISCUSSÃO:** Previamente a modelagem via *random forests* foi realizada uma análise descritiva dos dados meteorológicos do período de estudo, 27/12/1999 a 30/08/2018 (Tabela 1).

**Tabela 1** – Análise descritiva dos valores observados de 27/12/1999 a 30/08/2018

	Méd	Mín.	Máx.	Var.	Des.	CV	K	Assim.
ED	4,87	0	19,5	7,03	2,65	54,39	1,05	0,92

\*Méd.=Média; Mín.=Mínimo; Máx.=Máximo; Var.=Variância; Des.=Desvio-padrão; CV=Coeficiente de Variação (%); K=Curtose e Assim.=Assimetria. ED=Evaporação Diária.

Para a evaporação diária (ED), o desempenho do modelo de *random forests* ajustado foi considerado ruim. Nesta variável o RMSE, em percentual, foi de 31,62% (1,54 em 4,87) (Figura 2, Tabela 1). Credita-se esse fraco desempenho a alta dispersão relativa dos dados, evidenciada pelo alto CV desta variável (Tabela 1). A utilização de um modelo de *random forests*, como realizado neste trabalho, para o preenchimento de falhas meteorológicas para esta variável não é recomendada, pois poderia levar a interpretações imprecisas.



**Figura 2** - Diagrama de dispersão e RMSE (root mean square error) entre valores observados e estimados via *random forests* para evaporação diária no período de 27/12/1999 a 30/08/2018

O processo de evaporação diária (ED) desempenha um papel crítico no sistema terrestre, conduzindo os principais processos da superfície terrestre nos ciclos de energia, água e carbono. Informações quantitativas de evaporação/ evapotranspiração são utilizadas para resolução de numerosos desafios que envolvem o manejo das águas, climatologia (variabilidade e mudanças climáticas), meteorologia (condições climáticas), hidrologia (ciclo da água), engenharia de irrigação (programação de irrigação), e na previsão das cheias e na construção e operação de reservatórios (GIOVANELLA et al., 2021). A evaporação diária é dependente de inúmeras outras variáveis, tais como, a umidade relativa do ar, temperatura da superfície e velocidade do vento. Tal fato faz com que essa variável tenha grandes flutuações nos seus valores, o que dificulta sua modelagem.

**CONCLUSÕES:** O modelo de *random forests* apresentou fraco desempenho para evaporação diária e assim é considerado ineficiente no preenchimento das falhas.

**AGRADECIMENTOS:** Ao Instituto Federal do Triângulo Mineiro – Campus Uberaba pelo espaço destinado ao desenvolvimento da Dissertação.

## REFERÊNCIAS:

- BARTON, M. et al. **Forecasting impacts of biological control under future climates: mechanistic modelling of an aphid pest and a parasitic wasp.** Ecological Modelling, Amsterdam, v. 457, 109679, oct. 2021. DOI: 10.1016/j.ecolmodel.2021.109679
- BECK, H. E. et al. **Present and future Köppen-Geiger climate classification maps at 1-km resolution.** Scientific Data, London, v. 5, 180214, oct. 2018. DOI: 10.1038/sdata.2018.214
- CHINASHO, A. et al. **Evaluation of seven gap-filling techniques for daily station-based rainfall datasets in south ethiopia.** Advances in Meteorology, London, v. 2021, 9657460, aug. 2021. DOI: 10.1155/2021/9657460
- COUTINHO, E. R. et al. **Application of Artificial Neural Networks (ANNs) in the gap filling of meteorological time series.** Revista Brasileira de Meteorologia, Fortaleza, v. 33, n. 2, p. 24-36, apr./jun. 2018. DOI: 10.1590/0102-778620140115
- ETEDALI, H. R.; AHMADI, M. **Evaluation of various meteorological datasets in estimation yield and actual evapotranspiration of wheat and maize (case study: Qazvin plain).** Agricultural Water Management, Amsterdam, v. 256, 107080, oct. 2021. DOI: 10.1016/j.agwat.2021.107080
- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The elements of statistical learning: data mining, inference, and prediction.** New York: Springer, 2009.
- INSTITUTO NACIONAL DE METEOROLOGIA. **Sistema Nacional de Meteorologia.** INMET, Brasília: INMET, 2014. Disponível em: <http://www.inmet.gov.br>. Acesso em: 25 jun. 2021.
- KANDA, N. et al. **Performance of various techniques in estimating missing climatological data over snowbound mountainous areas of Karakoram Himalaya.** Meteorological Applications, Oxford, v. 25, n. 3, p. 337-349, nov. 2017. DOI: 10.1002/met.1699
- LU, W. et al. **Effect of climate change induced agricultural risk on land use in Chinese small farms: implications for adaptation strategy.** Ecological Indicators, New York, v. 115, 106414, aug. 2020. DOI: 10.1016/j.ecolind.2020.106414
- REN, X. et al. **Deep learning-based weather prediction: a survey.** Big Data Research, Leiden, v. 23, 100178, feb. 2021. DOI: 10.1016/j.bdr.2020.100178
- WOLDESENBET, T. A. et al. **Gap filling and homogenization of climatological datasets in the headwater region of the Upper Blue Nile Basin, Ethiopia.** International Journal of Climatology, Chichester, v. 37, n. 4, p. 2122-2140, aug. 2016. DOI: 10.1002/joc.4839
- ZHANG, Y. et al. **Agricultural drought prediction in China based on drought propagation and large-scale drivers.** Agricultural Water Management, Amsterdam, v. 255, 107028, sep. 2021. DOI: 10.1016/j.agwat.2021.107028