

ESTIMATIVA DA EVAPOTRANSPIRAÇÃO DE REFERÊNCIA EM ESTAÇÕES METEOROLÓGICAS DO ESPÍRITO SANTO COM TÉCNICAS DE MACHINE LEARNING

Iulo Pessotti Moro¹, Elaine Cordeiro dos Santos¹, Fernanda Lyra¹, Mariza Pereira de Oliveira Roza¹, Matheus Henrique da Silva¹, Juliana Oliveira Rodrigues¹.

¹Universidade Federal do Espírito Santo/Centro de Ciências Agrárias e Engenharias, Av. Governador Lindemberg, 316, Centro - 29550-000 - Jerônimo Monteiro – ES, Brasil.

Resumo

A evapotranspiração de referência (ET_o) é um elemento meteorológico importante para o manejo adequado dos recursos hídricos, sobretudo irrigação, obtida principalmente pelo método padrão Penman-Monteith FAO56. Objetivou-se modelar a ET_o em quatro estações meteorológicas do INMET no Espírito Santo (Vitória, Linhares, Alegre e Afonso Cláudio), considerando dados diários desde o início de operação até 2023. Para a modelagem, aplicou-se o algoritmo Random Forest, utilizando como variáveis de entrada a temperatura máxima (T_{max}), a amplitude térmica (AT) e a umidade relativa média (UR_{med}). Os dados foram divididos temporalmente em treino (70%) e teste (30%), normalizados com StandardScaler e avaliados por validação cruzada. O desempenho foi medido utilizando RMSE, MAE, R² e índice d'. Os resultados indicaram que a inclusão da UR_{med} contribuiu para maior precisão, com destaque para T_{max} e UR_{med} como variáveis mais influentes. Conclui-se que o Random Forest é eficaz na estimativa da ET_o e pode auxiliar em regiões com disponibilidade restrita de dados climáticos.

Palavras-chave: Aprendizado Supervisionado, Random Forest, Recursos Hídricos.

Área do Conhecimento: Engenharia Florestal.

Introdução

A evapotranspiração de referência (ET_o) é uma importante variável no planejamento econômico e sustentável dos recursos hídricos. Estimada em uma cultura padrão (gramado sem restrição hídrica), a partir da medição de elementos meteorológicos como temperatura do ar, umidade relativa, radiação solar e velocidade do vento (Moro et al., 2025). A ET_o também é usada para calcular a necessidade de água de cultivos agrícolas e florestais, por isso estimativas precisas são muito importantes para o planejamento da irrigação e a redução de perdas de água (Coutinho et al., 2020).

O método padrão para obter dados de ET_o é o Penman-Monteith FAO-56 descrito por Allen et al. (1998). Apesar da confiabilidade, sua aplicação é limitada em locais com dados meteorológicos incompletos. A sua fórmula também exige múltiplas variáveis e cálculos complexos. Modelos simplificados, baseados em regressão e Machine learning (ML), podem representar uma alternativa viável nessas condições (Alencar; Sediama; Mantovani, 2015; Ferreira, Cunha, Fernandes Filho, 2022).

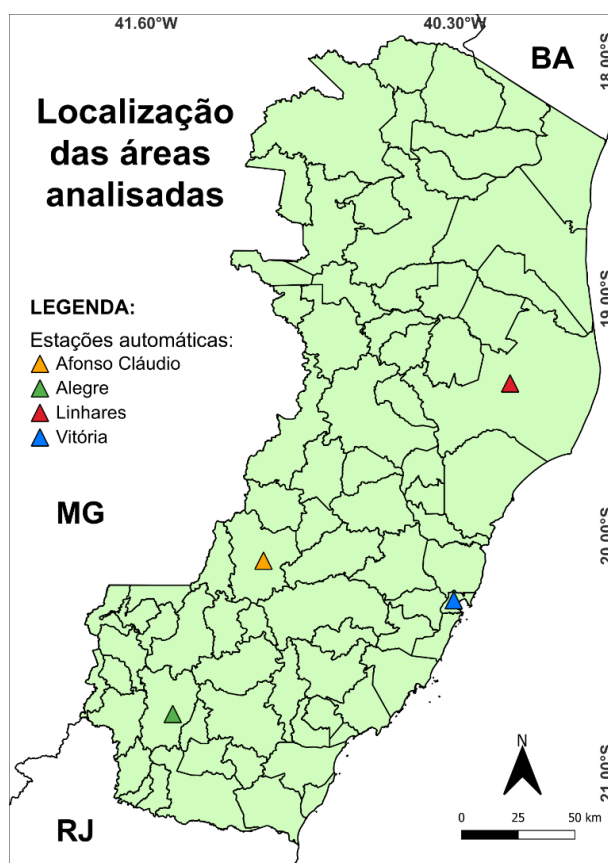
Técnicas de ML utilizando o algoritmo Random Forest (RF) na previsão da ET_o, tem apresentado bons resultados na literatura, em conjunto de dados complexos e não lineares (Ferreira, Cunha, Fernandes Filho, 2022). O RF é um modelo supervisionado que constrói múltiplas árvores de decisão a partir dos dados de treino e combina suas previsões para gerar estimativas em dados de teste (Breiman, 2001).

Este estudo teve como objetivo a aplicação da modelagem com machine learning e avaliação do algoritmo Random Forest em quatro estações meteorológicas do Espírito Santo, nos municípios de Alegre, Afonso Cláudio, Linhares e Vitória, analisando se um conjunto reduzido de variáveis meteorológicas é suficiente para estimar a ET_o com precisão, oferecendo uma alternativa viável ao método padrão FAO-56.

Metodologia

Foram utilizados dados diários de quatro estações meteorológicas do Instituto Nacional de Meteorologia (INMET) localizadas no estado do Espírito Santo: Vitória (A612), Linhares (A614), Alegre (A617) e Afonso Cláudio (A657) (Figura 1). Os dados compreendem o período desde o início de operação de cada estação meteorológica até dezembro de 2023. As quatro estações foram selecionadas por representarem diferentes regiões climáticas do Espírito Santo, contemplando tanto áreas litorâneas quanto interioranas, o que possibilita avaliar o desempenho do modelo de predição em cenários distintos de variabilidade climática.

Figura 1 – Estações meteorológicas



Fonte: Os autores (2025).

A evapotranspiração de referência foi estimada pelo algoritmo *Random Forest*, a partir dos dados de temperatura máxima (Tmax), amplitude térmica (AT) e umidade relativa média (URmed). Para comparar, utilizou-se a ETo obtida pelo método de Penman-Monteith ET_{OPM} , conforme a equação padrão recomendada pela FAO-56 (Allen *et al.*, 1998).

As etapas da modelagem foram: (i) remoção de valores nulos e registros incompletos; (ii) análise exploratória dos dados com aplicação do teste de Shapiro-Wilk para verificação da normalidade; (iii) cálculo da matriz de correlação de Spearman, uma vez que os dados não apresentaram distribuição normal; (iv) divisão temporal dos dados em conjunto de treino (70%) e teste (30%); (v) normalização dos dados com o método StandardScaler aplicado apenas ao treino; (vi) validação cruzada com TimeSeriesSplit (n = 5 divisões); (vii) cálculo das métricas estatísticas de desempenho (Quadro 1). Por fim, os resultados foram apresentados em gráficos de dispersão, comparando os valores de ETo previstos pelo modelo de RF com os observados pelo método de referência ET_{OPM} .

A Ciência do NANO e seu impacto transformador no MACRO

Quadro 1 – Métricas utilizadas para avaliação de desempenho dos modelos

Métricas	Descrição
$\rho = 1 - \frac{6 \sum d_i^2}{N(N^2 - 1)} \quad (\text{Eq. 1})$	p: correlação de postos de Spearman, mede a associação monotônica entre as variáveis observadas e estimadas.; di: diferença entre os postos das variáveis correspondentes; N: é o número de observações.
$R^2 = 1 - \frac{\sum_{i=1}^N (o_i - P_i)^2}{\sum_{i=1}^N (o_i - \bar{o})^2} \quad (\text{Eq. 2})$	R ² : Coeficiente de Determinação; mede a proporção da variância explicada pelo modelo;
$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (P_i - O_i)^2}{N}} \quad (\text{Eq. 3})$	RMSE: Raiz do Erro Quadrático Médio, mede o erro médio ao quadrado, penalizando os erros maiores;
$\text{MAE} = N^{-1} \cdot \sum_{i=1}^N (P_i - O_i) \quad (\text{Eq. 4})$	MAE: Erro Médio Absoluto; mede o valor médio dos erros absolutos entre os valores estimados e observados;
$d' = 1 - \frac{\sum_{i=1}^N (P_i - O_i)^2}{\sum_{i=1}^N \left[\left(\left P_i - \bar{O} \right \right) + \left(\left O_i - \bar{O} \right \right) \right]^2} \quad (\text{Eq. 5})$	Em que: Pi: valor estimado ou previsto pelo modelo; Oi: é o valor observado ou valor medido e N: número total de observações. d': índice de concordância ajustado; Em que: Pi: valor de evapotranspiração de referência obtida pelo método testado; Oi: valor observado de evapotranspiração de referência mm d ⁻¹ ; $\bar{O}$: média dos valores obtidos pelo método padrão; e N: número de observações.

Fonte: Elaborado pelos autores (2025).

Resultados

A Tabela 1 apresenta o resultado da correlação de Spearman para cada estação meteorológica entre os elementos meteorológicos avaliados. Nota-se que Tmax está fortemente relacionado com EToPM. A Amplitude térmica (AT) apresenta uma correlação moderada. As correlações negativas ocorrem com umidade relativa média (URmed), associado à redução da ETo.

Tabela 1 – Valores de Correlação de Spearman para as variáveis meteorológicas escolhidas.

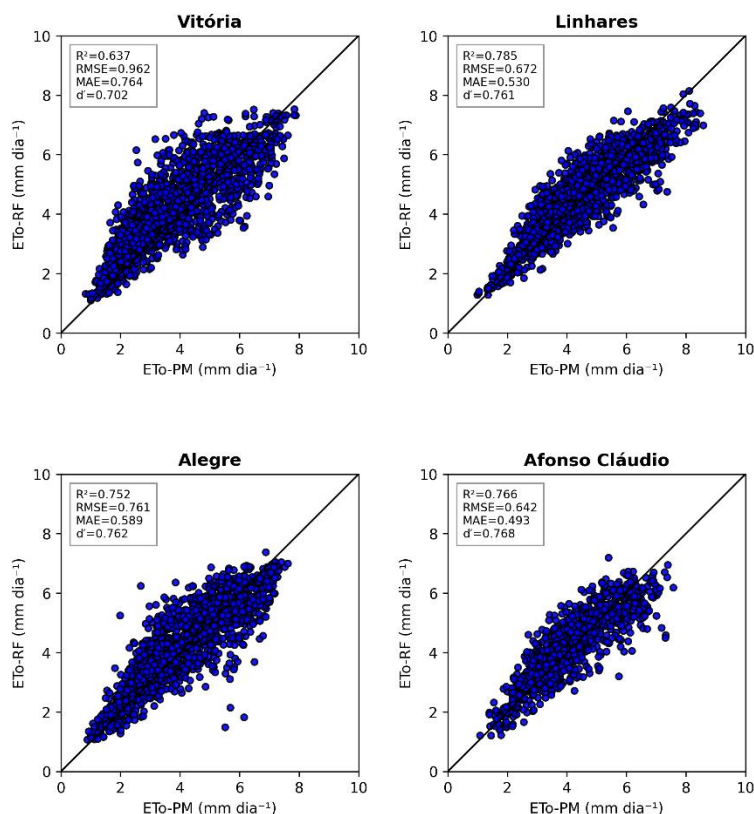
Nome da estação	Tmax- EToPM	AT- EToPM	URmed - EToPM
Alegre (A617)	0,878	0,589	-0,680
Afonso Cláudio (A657)	0,855	0,468	-0,635
Linhares (A614)	0,852	0,585	-0,692
Vitória (A612)	0,820	0,499	-0,654

Fonte: elaborado pelos autores (2025).

A figura 2, apresenta os resultados das métricas de desempenho e dispersão dos dados. O modelo *Random Forest* apresentou melhor desempenho nas estações de Linhares e de Afonso Cláudio, com maiores valores de R² e d' e menores erros (RMSE e MAE). Já em Vitória e Alegre, o desempenho foi inferior.

A Ciência do NANO e seu impacto transformador no MACRO

Figura 2 – Gráficos de dispersão entre os valores observados ETo_{PM} e os estimados por RF



Fonte: Os autores (2025).

Discussão

Conforme observado na Figura 2, o desempenho do modelo variou entre as estações. Em Vitória (A612), verificou-se maior ocorrência de superestimativas em valores baixos e subestimativas em valores altos de ETo, além de apresentar o menor R² (0,637) e o maior RMSE (0,962). Esse resultado pode estar relacionado à maior complexidade climática da região litorânea, marcada por elevada variabilidade atmosférica, possíveis inconsistências nos registros meteorológicos como também à ausência de variáveis de entrada capazes de representar essas condições (Ayaz *et al.*, 2021).

Nas estações de Linhares, Alegre e Afonso Cláudio, o ajuste foi mais consistente, com pontos mais próximos da linha 1:1, indicando boa capacidade de predição em relação ao método de referência de Penman-Monteith (ETo_{PM}). Os melhores desempenhos foram observados em Linhares e Afonso Cláudio, com R² de 0,785 e 0,766, e com maior concordância ($d \approx 0,76$) respectivamente, enquanto Alegre apresentou desempenho intermediário (R²=0,752). Esses resultados reforçam que a performance do modelo, em conjunto com as variáveis de entrada, é sensível a fatores locais.

Os resultados deste estudo estão em consonância com os de Stefanidis *et al.*, (2025), que ao estimar a evapotranspiração potencial na Grécia com Random Forest obtiveram R² de 0,917 e RMSE de 0,468 mm dia⁻¹ utilizando temperatura e radiação solar. A diferença observada em relação a esse desempenho decorre principalmente da ausência de variáveis mais sensíveis à ETo, como radiação e vento, evidenciando que o modelo mantém elevada eficiência mesmo sob condições de entrada limitadas.

O modelo com algoritmo Random Forest apresentou desempenho satisfatório na estimativa da evapotranspiração de referência (ETo) nas estações meteorológicas avaliadas. Baratto *et al.* (2022), apontaram o Random Forest como o método mais adequado para espacialização da ETo no Espírito Santo. Embora o desempenho variável entre estações, e restrição de variáveis, o RF manteve boa capacidade preditiva, corroborando sua aplicabilidade em contextos regionais.

A Ciência do NANO e seu impacto transformador no MACRO

De forma consistente, Sammen *et al.* (2023) destacaram o Random Forest como o algoritmo mais preciso no semiárido iraquiano, superando SVM e boosting, Já Dias *et al.* (2021) observaram o seu desempenho preditivo para ETo mensal no Brasil ($R^2 > 0,90$). No presente estudo, apesar de número reduzidos das variáveis de entrada para o modelo, optou-se também por elas estarem correlacionadas a ETo.

Entre as limitações, superestimativas e subestimativas observadas nos valores de ETo, estão associadas à característica do algoritmo Random Forest, que tende a estimar com maior acurácia os valores centrais em detrimento dos extremos (Breiman, 2001; Harrison, 2019). Dessa forma, estações com maior número de dados podem apresentar desvios nos valores extremos de ETo. Estratégias como a inclusão de variáveis adicionais (radiação solar e velocidade do vento) ou técnicas de avançadas no aprendizado do modelo podem contribuir para melhorar a predição (Sharafi; Mohammadi Ghalei, 2024).

Os resultados confirmam o potencial do Random Forest como alternativa viável ao método padrão de Penman-Monteith, especialmente em regiões onde há limitações de monitoramento meteorológico. A utilização de apenas três variáveis de entrada (temperatura máxima, amplitude térmica e umidade relativa média) mostrou-se eficiente para gerar estimativas da ETo, o que amplia sua aplicabilidade em áreas com restrição de dados (Santos *et al.*, 2023). Tal característica é relevante no estado do Espírito Santo, onde a diversidade topográfica e climática exige métodos acessíveis para subsidiar o manejo de recursos hídricos e a agricultura irrigada.

Conclusão

Com base conjunto de dados das estações, o modelo Random Forest apresentou melhor desempenho nas estações de Linhares e Afonso Cláudio, com maior concordância em relação ao método de Penman-Monteith, enquanto Vitória registrou os menores ajustes. A inclusão de variáveis adicionais pode contribuir para reduzir desvios em valores extremos de ETo, sobretudo em regiões de maior diversidade climática.

O modelo mostrou-se aplicável em locais com restrições de monitoramento meteorológico e passível de ampliação em estudos futuros que envolvam um número maior de estações. Os resultados evidenciam o potencial do uso de algoritmos de aprendizado de máquina em estudos agrometeorológicos, ao mesmo tempo em que reforçam a necessidade de considerar as particularidades de cada estação e a qualidade das séries temporais utilizadas.

Referências

ALENCAR, L. P. DE; SEDIYAMA, G. C.; MANTOVANI, E. C. Estimativa da evapotranspiração de referência (ETo padrão FAO), para Minas Gerais, na ausência de alguns dados climáticos.

Engenharia Agrícola, vol. 35, no. 1, p. 39–50, Feb. 2015. DOI 10.1590/1809-4430-

Eng.Agric.v35n1p39-50/2015. Available at:

http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-69162015000100039&lng=pt&lng=pt.

ALLEN, R. G.; PEREIRA, L. S.; RAES, D.; SMITH, M. **Crop Evapotranspiration: guidelines for computing crop water requirements**, n. 56. Rome: [s. n.], 1998. Available at:

<https://www.fao.org/4/x0490e/x0490e00.htm>.

AYAZ, A.; RAJESH, M.; SINGH, S. K.; REHANA, S. Estimation of reference evapotranspiration using machine learning models with limited data. **AIMS Geosciences**, vol. 7, no. 3, p. 268–290, 2021. DOI 10.3934/geosci.2021016. Available at: <http://www.aimspress.com/article/doi/10.3934/geosci.2021016>.

BARATTO, P. F. B.; CECÍLIO, R. A.; DE SOUSA TEIXEIRA, D. B.; ZANETTI, S. S.; XAVIER, A. C. Random forest for spatialization of daily evapotranspiration (ET₀) in watersheds in the Atlantic Forest. **Environmental Monitoring and Assessment**, vol. 194, no. 6, p. 449, 23 Jun. 2022. DOI 10.1007/s10661-022-10110-y. Available at: <https://link.springer.com/10.1007/s10661-022-10110-y>.

BREIMAN, L. Random Forests. **Machine Learning**, vol. 45, no. 1, p. 5–32, Oct. 2001. DOI

A Ciência do NANO e seu impacto transformador no MACRO

10.1023/A:1010933404324. Available at: <https://link.springer.com/10.1023/A:1010933404324>.

COUTINHO, E. R.; MADEIRA, J. G. F.; SILVA, R. M. da; OLIVEIRA, E. M. de; DELGADO, A. R. S. Avaliação de Métodos de Estimativa da Evapotranspiração de Referência (ET_o) Diária Para Regiões dos Estados do Rio de Janeiro e Espírito Santo. **Revista Brasileira de Meteorologia**, vol. 35, no. 4, p. 649–657, Dec. 2020. DOI 10.1590/0102-77863540069. Available at: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-77862020000400649&lng=pt.

DIAS, S. H. B.; FILGUEIRAS, R.; FERNANDES FILHO, E. I.; ARCANJO, G. S.; SILVA, G. H. da; MANTOVANI, E. C.; CUNHA, F. F. da. Reference evapotranspiration of Brazil modeled with machine learning techniques and remote sensing. **PLOS ONE**, vol. 16, no. 2, p. e0245834, 9 Feb. 2021. DOI 10.1371/journal.pone.0245834. Available at: <https://dx.plos.org/10.1371/journal.pone.0245834>.

FERREIRA, L. B.; DA CUNHA, F. F.; FERNANDES FILHO, E. I. Exploring machine learning and multi-task learning to estimate meteorological data and reference evapotranspiration across Brazil. **Agricultural Water Management**, vol. 259, p. 107281, Jan. 2022. DOI 10.1016/j.agwat.2021.107281. Available at: <https://linkinghub.elsevier.com/retrieve/pii/S0378377421005588>

HARRISON, M. **Machine Learning Pocket Reference: Working with Structured Data in Python**. 1st ed. [S. l.]: O'Reilly Media, 2019.

MORO, I. P.; ZANETTI, S. S.; CECÍLIO, R. A.; PEZZOPANE, J. E. M.; XAVIER, A. C. Influência dos Elementos Meteorológicos na Evapotranspiração de Referência Diária. **Revista Brasileira de Meteorologia**, v. 40, 2025. <https://doi.org/10.1590/0102-778640240043>

SAMMEN, S. S.; KISI, O.; AL-JANABI, A. M. S.; ELBELTAGI, A.; ZOUNEMAT-KERMANI, M. Estimation of Reference Evapotranspiration in Semi-Arid Region with Limited Climatic Inputs Using Metaheuristic Regression Methods. **Water**, vol. 15, no. 19, p. 3449, 30 Sep. 2023. DOI 10.3390/w15193449. Available at: <https://www.mdpi.com/2073-4441/15/19/3449>

SANTOS, P. A. B. dos; SCHWERZ, F.; CARVALHO, L. G. de; BAPTISTA, V. B. da S.; MARIN, D. B.; FERRAZ, G. A. e S.; ROSSI, G.; CONTI, L.; BAMBI, G. Machine Learning and Conventional Methods for Reference Evapotranspiration Estimation Using Limited-Climatic-Data Scenarios. **Agronomy**, vol. 13, no. 9, p. 2366, 12 Sep. 2023. DOI 10.3390/agronomy13092366. Available at: <https://www.mdpi.com/2073-4395/13/9/2366>.

SHARAFI, S.; MOHAMMADI GHALENI, M. Revealing accuracy in climate dynamics: enhancing evapotranspiration estimation using advanced quantile regression and machine learning models. **Applied Water Science**, vol. 14, no. 7, p. 162, 24 Jul. 2024. DOI 10.1007/s13201-024-02211-5. Available at: <https://link.springer.com/10.1007/s13201-024-02211-5>.

STEFANIDIS, S.; IOANNOU, K.; PROUTSOS, N.; KARMIRIS, I.; STEFANIDIS, P. Comparative Analysis of Machine Learning Algorithms for Potential Evapotranspiration Estimation Using Limited Data at a High-Altitude Mediterranean Forest. **Atmosphere**, vol. 16, no. 7, p. 851, 12 Jul. 2025. DOI 10.3390/atmos16070851. Available at: <https://www.mdpi.com/2073-4433/16/7/851>.

Agradecimentos

Agradecemos a Fundação de Amparo à Pesquisa e Inovação do Espírito Santo (FAPES) pelo apoio financeiro e incentivo à pesquisa, através do edital FAPES nº 14/2023 - PROCAP 2024 DOUTORADO, a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.