



UNIVERSIDADE ESTADUAL PAULISTA  
"JÚLIO DE MESQUITA FILHO"  
Campus de Botucatu



**LUANA POSSARI MAZIERO COSTA**

**PREVISÃO DA EVAPOTRANSPIRAÇÃO DE REFERÊNCIA USANDO  
INFERÊNCIA NEURO-FUZZY**

**Botucatu**

**2021**



**LUANA POSSARI MAZIERO COSTA**

**PREVISÃO DA EVAPOTRANSPIRAÇÃO DE REFERÊNCIA USANDO  
INFERÊNCIA NEURO-FUZZY**

Tese apresentada à Faculdade de Ciências Agronômicas da Unesp, Campus de Botucatu, para obtenção do título de Doutora em Agronomia (Irrigação e Drenagem).

Orientador: Prof. Dr. Luís Roberto Almeida Gabriel  
Filho

Coorientador: Prof. Dr. Alexandre Dal Pai

**Botucatu**

**2021**

C837p

Costa, Luana Possari Maziero

Previsão da evapotranspiração de referência usando inferência neuro-fuzzy / Luana Possari Maziero Costa. -- Botucatu, 2021

92 p. : il., tabs., mapas

Tese (doutorado) - Universidade Estadual Paulista (Unesp), Faculdade de Ciências Agrônômicas, Botucatu

Orientador: Luís Roberto Almeida Gabriel Filho

Coorientador: Alexandre Dal Pai

1. Evapotranspiração de referência. 2. Algoritmos difusos. 3. Modelos matemáticos. 4. Análise de séries temporais. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Ciências Agrônômicas, Botucatu. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

## CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: PREVISÃO DA EVAPOTRANSPIRAÇÃO DE REFERÊNCIA USANDO INFERÊNCIA NEURO-FUZZY

**AUTORA: LUANA POSSARI MAZIERO COSTA**

**ORIENTADOR: LUÍS ROBERTO ALMEIDA GABRIEL FILHO**

**COORIENTADOR: ALEXANDRE DAL PAI**

Aprovada como parte das exigências para obtenção do Título de Doutora em AGRONOMIA (IRRIGAÇÃO E DRENAGEM), pela Comissão Examinadora:

*Luís R. A. Gabriel Filho*

Prof. Dr. LUÍS ROBERTO ALMEIDA GABRIEL FILHO (Participação Virtual)  
Gestão, Desenvolvimento e Tecnologia / Faculdade de Ciências e Engenharia de Tupã - UNESP

*pl Luís R. A. Gabriel Filho*

Prof. Dr. SERGIO AUGUSTO RODRIGUES (Participação Virtual)  
Bioprocessos e Biotecnologia / Faculdade de Ciências Agronômicas de Botucatu - UNESP

*pl Luís R. A. Gabriel Filho*

Prof. Dr. RODRIGO LILLA MANZIONE (Participação Virtual)  
FCE / UNESP/Tupã (SP)

*pl Luís R. A. Gabriel Filho*

Prof. Dr. ANTONIO ADOLFO MARTINS RAPOSO (Participação Virtual)  
Campus Tupã / Instituto Federal de São Paulo

*pl Luís R. A. Gabriel Filho*

Prof. Dr. ALEXSANDRO OLIVEIRA DA SILVA (Participação Virtual)  
Engenharia Agrícola / Universidade Federal do Ceará

Botucatu, 16 de dezembro de 2021



***Dedico,***

*A minha família, especialmente, ao meu filho Enrico.*



## **AGRADECIMENTOS**

Agradeço, primeiramente, a Deus pelo dom da vida.

À minha família, que me amparou e aconselhou nesses anos de doutorado.

Ao Prof. Dr. Luís Roberto Almeida Gabriel Filho que me orienta academicamente a mais de oito anos e tornou-se grande incentivador da minha carreira. Agradeço por cada conselho e toda a compreensão que teve comigo durante todos esses anos.

Ao meu coorientador Prof. Dr. Alexandre Dal Pai e ao Prof. Dr. Sérgio Augusto Rodrigues pelas grandes contribuições para a construção desta tese.

À UNESP, por proporcionar minha formação, gratuita e de qualidade, na graduação, mestrado e doutorado, em especial à Faculdade de Ciências Agrônômicas do Campus de Botucatu por proporcionar as condições necessárias para o meu aprendizado e desenvolvimento.

Aos docentes do Programa de Pós-Graduação em Agronomia (Irrigação e Drenagem) por me acolherem na pós-graduação e compartilharem seus conhecimentos para minha formação, mesmo quando tivemos que nos adaptar ao ensino remoto que a pandemia nos impôs.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa de estudos concedida.



“Conhecimento não é aquilo que você sabe, mas o que você faz com aquilo que você sabe”.

Aldous Huxley



## RESUMO

A evapotranspiração de referência é uma variável fundamental para o planejamento de sistemas de irrigação mais eficientes, uma vez que associada ao coeficiente da cultura permite estimar a necessidade hídrica da planta. Desta maneira, o objetivo do presente trabalho é estabelecer modelos de previsão de séries temporais da evapotranspiração de referência (Penman-Monteith/FAO 56) utilizando modelo neuro-fuzzy (ANFIS - *Adaptive Neuro-Fuzzy Inference System*), uma vez que este método apresenta resultados superiores quando comparados com outros métodos estocásticos na previsão de valores de uma série temporal. Para isso foi realizado a imputação de dados faltantes utilizando o método Correspondência Média Preditiva Ponderada (*midastouch*) das variáveis radiação solar global, velocidade do vento, temperatura do ar, temperatura máxima, temperatura mínima, umidade relativa do ar, umidade relativa do ar máxima e umidade relativa do ar mínima. A partir das séries históricas completas foram elaborados modelos de previsão de séries temporais que relacionam  $n$  valores de  $ET_o$  em tempos consecutivos e imediatamente anteriores ao tempo da observação a ser prevista. Os modelos neuro-fuzzy (ANFIS) elaborados utilizam o método de inferência Takagi-Sugeno, onde os dados são ordenados de acordo com a data de coleta, do mais antigo para o mais recente e os primeiros 85% dos dados sendo selecionados para treinamento e os demais 15% para validação. A partir da comparação dos modelos neuro-fuzzy elaborados, é possível verificar maior viabilidade na utilização de modelos gerais, que consideraram três anos de dados para o treinamento da rede, uma vez que apresentam erro percentual absoluto de 0,25% para Botucatu – SP e 0,26% para Tupã – SP. Portanto, a previsão da evapotranspiração de referência utilizando os modelos elaborados oferece estimativas confiáveis dos valores, que quando associados ao coeficiente da cultura, permitem gestão e planejamento de sistemas de irrigação de maneira eficiente e sustentável.

**Palavras-chave:** série histórica, Penman-Monteith, irrigação, ANFIS, dados ausentes.



## ABSTRACT

The reference evapotranspiration is a fundamental variable for the planning of more efficient irrigation systems, once associated with the crop coefficient, it allows the estimation of the plant's water requirement. Thus, the objective of the present work is to establish time series prediction models of reference evapotranspiration (Penman-Monteith/FAO 56) using the neuro-fuzzy model (ANFIS - Adaptive Neuro-Fuzzy Inference System), since this method presents superior results when compared with other stochastic methods in predicting values of a time series. For this, missing data was imputed using the Weighted Average Predictive Correspondence (midastouch) method of the variables global solar radiation, wind speed, air temperature, maximum temperature, minimum temperature, relative air humidity, maximum relative humidity and minimum relative humidity. From the complete historical series, time series prediction models were developed that relate  $n$  values of  $ET_o$  in consecutive times and immediately prior to the time of the observation to be forecast. The neuro-fuzzy models (ANFIS) elaborated use the Takagi-Sugeno inference method, where the data are sorted according to the collection date, from the oldest to the most recent and the first 85% of the data being selected for training and the remaining 15% for validation. From the comparison of the neuro-fuzzy models elaborated, it is possible to verify greater viability in the use of general models, which considered three years of data for the training of the network, since they present an absolute percentage error of 0.25% for Botucatu - SP and 0.26% for Tupã – SP. Therefore, the reference evapotranspiration forecast using the models developed offers reliable estimates of the values, which when associated with the crop coefficient, allow the management and planning of irrigation systems in an efficient and sustainable way.

**Keywords:** historical series, Penman-Monteith, irrigation, ANFIS, missing data.



## LISTA DE FIGURAS

Figura 1 – Localização das estações meteorológicas de coleta de dados no estado de São Paulo.....	34
Figura 2 - Quantidade de dados faltantes, por variável, para o banco de dados do município de (A) Botucatu – SP e (B) Tupã – SP.....	37
Figura 3 – Gráficos de densidade de probabilidade dos dados observados e imputados pelo método (A) Correspondência Média Preditiva ( $\rho_{mm}$ ) e (B) Correspondência Média Preditiva Ponderada ( $midastouch$ ) do município de Botucatu – SP.....	38
Figura 4 - Gráficos de densidade de probabilidade dos dados observados e imputados pelo método (A) Correspondência Média Preditiva ( $\rho_{mm}$ ) e (B) Correspondência Média Preditiva Ponderada ( $midastouch$ ) do município de Tupã – SP.....	39
Figura 5 – Gráfico de pontos para a distribuição dos dados imputados pelo método (A) Correspondência Média Preditiva ( $\rho_{mm}$ ) e (B) Correspondência Média Preditiva Ponderada ( $midastouch$ ) para o município de Botucatu – SP. ...	40
Figura 6 – Gráfico de pontos para a distribuição dos dados imputados pelo método (A) Correspondência Média Preditiva ( $\rho_{mm}$ ) e (B) Correspondência Média Preditiva Ponderada ( $midastouch$ ) para o município de Tupã – SP. ....	41
Figura 7 – Gráfico <i>box-plot</i> para a distribuição dos imputados pelo método (A) Correspondência Média Preditiva ( $\rho_{mm}$ ) e (B) Correspondência Média Preditiva Ponderada ( $midastouch$ ) para o município de Botucatu – SP. ...	42
Figura 8 - Gráfico <i>box-plot</i> para a distribuição dos imputados pelo método (A) Correspondência Média Preditiva ( $\rho_{mm}$ ) e (B) Correspondência Média Preditiva Ponderada ( $midastouch$ ) para o município de Tupã – SP. ....	43
Figura 9 - Arquitetura de um modelo <i>ANFIS</i> . ....	61
Figura 10 - Sistema neuro- <i>fuzzy</i> de previsão da evapotranspiração de referência utilizando a evapotranspiração calculada ( $ET_o$ ).....	64

Figura 11 – (A) Evapotranspiração de referência calculada pelo método de Penman-Monteith/FAO 56, (B) distribuição de frequências dos valores da série temporal e (C) densidade da probabilidade de obtenção da evapotranspiração de referência diária para Botucatu – SP.....	67
Figura 12 – Funções de autocorrelação e autocorrelação parcial das séries temporais das estações (A) outono, (B) inverno, (C) primavera, (D) verão e (E) geral do município de Botucatu – SP.....	69
Figura 13 – Erros de treinamento do modelo ANFIS Geral de Botucatu – SP para (A) um ano e (B) três anos de dados.....	72
Figura 14 – Gráficos de contorno utilizando série temporal de três anos das estações (A) Outono, (B) Inverno, (C) Primavera, (D) Verão e (E) Geral para o município de Botucatu – SP.....	73
Figura 15 – Arquitetura do modelo ANFIS geral utilizando série temporal de três anos para o município de Botucatu – SP.....	74
Figura 16 – Funções de pertinência das variáveis de entrada $ET_{o(t-1)}$ , $ET_{o(t-2)}$ e $ET_{o(t-3)}$ do modelo ANFIS de Botucatu – SP. ....	75
Figura 17 – Gráficos de previsão de $ET_o$ a partir da variação da $ET_{o(t-1)}$ e $ET_{o(t-2)}$ , com $ET_{o(t-3)}$ fixa para Botucatu – SP. ....	76
Figura 18 – (A) Evapotranspiração de referência calculada pelo método de Penman-Monteith/FAO 56, (B) distribuição de frequências dos valores da série temporal e (C) densidade da probabilidade de obtenção da evapotranspiração de referência diária para Tupã – SP.....	77
Figura 19 – Funções de autocorrelação e autocorrelação parcial das séries temporais das estações (A) outono, (B) inverno, (C) primavera, (D) verão e (E) geral do município de Tupã – SP.....	78
Figura 20 – Erros de treinamento do modelo ANFIS Geral de Tupã – SP para (A) um ano e (B) três anos de dados.....	81
Figura 21 – Gráficos de contorno utilizando série temporal de três anos das estações (A) Outono, (B) Inverno, (C) Primavera, (D) Verão e (E) Geral para o município de Tupã – SP.....	82

Figura 22 – Arquitetura do modelo ANFIS geral utilizando série temporal de três anos para o município de Tupã – SP. ....	83
Figura 23 – Funções de pertinência das variáveis de entrada $ET_{o(t-1)}$ e $ET_{o(t-2)}$ do modelo ANFIS de Tupã – SP. ....	84
Figura 24 – Gráficos de previsão de $ET_o$ a partir da variação da $ET_{o(t-1)}$ e $ET_{o(t-2)}$ , para o município de Tupã – SP. ....	85



## LISTA DE TABELAS

Tabela 1 – Resultado das medidas estatísticas NRMSE, MAPE e SMAPE calculados para os métodos <i>pmm</i> e <i>midastouch</i> dos municípios de Botucatu – SP e Tupã – SP. ....	45
Tabela 2 – Períodos de início e término dos subconjuntos de dados de cada estação meteorológica.....	56
Tabela 3 – Definição dos parâmetros dos modelos neuro- <i>fuzzy</i> de menor erro estatístico U de Theil para o município de Botucatu – SP.....	70
Tabela 4 – Erros U, RMSE e MAPE para os modelos Naive e ANFIS para o município de Botucatu – SP. ....	71
Tabela 5 – Definição dos parâmetros dos modelos neuro- <i>fuzzy</i> de menor erro estatístico U de Theil do município de Tupã – SP.....	79
Tabela 6 – Erros U, RMSE e MAPE para os modelos Naive e ANFIS para o município de Tupã – SP. ....	80



## LISTA DE ABREVIATURAS E SIGLAS

ANFIS	Adaptive Neuro-Fuzzy Inference System
$ET_0$	Evapotranspiração de Referência
INMET	Instituto Nacional de Meteorologia
MAPE	Erro Absoluto Médio Percentual
<i>midastouch</i>	Correspondência Média Preditiva Ponderada
NRMSE	Raiz Quadrada do Erro Quadrático Médio normalizado
<i>pmm</i>	Correspondência Média Preditiva
Rad	Radiação Solar Global
SMAPE	Erro Percentual Absoluto Médio Simétrico
T	Temperatura do Ar
Tmax	Temperatura do Ar Máxima
Tmin	Temperatura do Ar Mínima
UR	Umidade Relativa do Ar
Urmax	Umidade Relativa do ar Máxima
Urmin	Umidade Relativa do Ar Mínima
VV	Velocidade do Vento



## SUMÁRIO

<b>INTRODUÇÃO GERAL</b> .....	<b>25</b>
<b>Justificativa</b> .....	<b>27</b>
<b>Objetivos</b> .....	<b>28</b>
<b>CAPÍTULO 1 - IMPUTAÇÃO MÚLTIPLA PMM E MIDASTOUCH EM SÉRIES HISTÓRICAS DE DADOS METEOROLÓGICOS</b> .....	<b>29</b>
1.1 Introdução.....	31
1.2 Material e Métodos .....	33
1.3 Resultados e Discussão .....	36
1.4 Conclusões.....	46
<b>CAPÍTULO 2 – MODELAGEM NEURO-FUZZY PARA A PREVISÃO DA EVAPOTRANSPIRAÇÃO DE REFERÊNCIA ESTIMADA PELO MÉTODO DE PENMAN-MONTEITH</b> .....	<b>49</b>
2.1 Introdução.....	52
2.2 Material e Métodos .....	54
2.2.1 Séries Temporais.....	54
2.2.2 Cálculo da Evapotranspiração de Referência.....	56
2.2.3 Método de elaboração dos sistemas neuro- <i>fuzzy</i> .....	60
2.2.4 Avaliação e validação dos modelos.....	64
2.3 Resultados e Discussão .....	66
2.3.1 Serie temporal e modelagem ANFIS para o município de Botucatu – SP .....	66
2.3.2 Serie temporal e modelagem ANFIS para o município de Tupã – SP .....	76
2.4 Conclusões.....	85
<b>CONSIDERAÇÕES FINAIS</b> .....	<b>90</b>
<b>REFERÊNCIAS</b> .....	<b>91</b>



## INTRODUÇÃO GERAL

Diante da crescente preocupação com a escassez de recursos hídricos para as atividades agrícolas em algumas regiões do planeta, é fundamental que se conheça a evapotranspiração das culturas para o eficiente manejo de sistemas de irrigação, sendo que este parâmetro agrometeorológico é estimado a partir da evapotranspiração de referência ( $ET_o$ ) e do coeficiente de cultura ( $Kc$ ) (CARVALHO et al., 2011; JIANG et al., 2016; TEJERO et al., 2018).

A evapotranspiração é o resultado da evaporação das superfícies e da transpiração das plantas, de acordo com Allen et al. (1998), constituindo um dos componentes principais do balanço hídrico, dependendo do saldo da radiação e do déficit de pressão de vapor para a perda de água para a atmosfera (CARDOSO; JUSTINO, 2014; ADNAN; ULLAH; AHMED, 2020). Assim, o monitoramento de tais variáveis admite planejar programas de irrigação com maior eficiência, permitindo estimativas mais precisas de produção agrícola (STAN et al., 2016).

Estudos feitos por Klar et al. (2006) e Rossato et al. (2017) mostram que as condições climáticas e a umidade do solo influenciam diretamente a produção e a produtividade de culturas. Prando et al. (2015) evidenciam que o manejo correto da irrigação deve permitir o uso sustentável da água a fim de obter a maior produção possível por unidade de água utilizada. Já Tejero et al. (2018) afirmam que o manejo adequado de irrigação além de permitir a economia de água pode induzir a regulação da fisiologia da planta.

Além disso, Toureiro et al. (2017) afirmam que a necessidade de água da cultura é revelada por meio da evapotranspiração da mesma, podendo ser prevista a quantidade de água de irrigação a ser alocada para o planejamento de projetos de irrigação e de safra. Silva et al. (2015) ressaltam que além da estimativa da necessidade hídrica, a ( $ET_o$ ) permite o dimensionamento dos sistemas de irrigação (bombeamento, adução e distribuição de água). Desta maneira, a previsão de valores de evapotranspiração de referência devem ser feitos a partir de boa e imparcial estimativa de séries temporais de evapotranspiração de referência, conforme afirmam Tomas-Burguera et al. (2017).

O método de Penman-Monteith, adotado pela FAO (*Food and Agriculture Organization*) - Penman-Monteith/FAO 56, requer dados de temperatura média do

ar, velocidade do vento, umidade relativa média do ar, fluxo de calor no solo e saldo de radiação global (dado pela diferença entre o saldo de radiação de ondas curtas e o saldo de radiação de ondas longas) frequentemente disponíveis, com exceção da radiação de ondas longas, na maioria das estações meteorológicas do planeta, uma vez que tais dados são obtidos por meio de sensores que coletam informações em intervalos de tempos variados, disponibilizando grande volume de dados, conforme afirmam Baba, Vaz e Costa (2014).

Porém, Ventura et al. (2016) afirmam que problemas ocorridos nos instrumentos de coleta como desligamentos, manutenções, avarias, limitações físicas ou fenômenos climáticos ou mesmo na transmissão dos dados causam falhas nas séries de dados. Tais falhas também dificultam a aplicação de métodos, como é o caso do método padrão para o cálculo da evapotranspiração de referência de Penman-Monteith, sendo necessário a estimativa dos valores faltantes de algumas variáveis para que seja possível sua utilização.

Diante de tal necessidade, é comum a utilização de modelos matemáticos que representem o objeto ou fenômeno a ser estudado para que seja possível a estimativa ou mesmo a previsão de valores, fato também ressaltado por Macarthur, Stumpf e Oreffo (2018), já que a modelagem matemática clássica permite a otimização de processos e explicação de sistemas e fenômenos por meio de equações ou sistemas de equações.

Contudo, quanto maior a complexidade do fenômeno estudado mais interrelações de sistemas isolados são necessárias para sua representação, já que o aumento do número de variáveis dependentes e independentes envolvidas tornam as equações e sistemas de equações mais complexos, o que reduz a precisão oferecida para descrever o comportamento estudado (ZADEH, 1971; BASSANEZI, 2002; LABABIDI; BAKER, 2006).

Além disso, é comum a redução do volume de dados numéricos disponíveis à medida que a complexidade do fenômeno aumenta, além da existência de informações ambíguas e imprecisas acerca do fenômeno estudado. Neste caso, o raciocínio *fuzzy* permite descrever o comportamento de sistemas, admitindo o mapeamento das observações de entrada e saída (ZADEH, 1988; LABABIDI; BAKER, 2006). Aydin e Kayisli (2016) reforçam que o sistema de inferência neuro-*fuzzy* adaptável (ANFIS - *Adaptive Neuro-Fuzzy Inference System*) é eficiente para

estimar parâmetros, principalmente, de modelos matemáticos complexos e multivariados.

O estudo feito por Goyal et al. (2014), mostrou que os modelos de aprendizado superam os métodos empíricos tradicionais na modelagem de problemas, o que foi reforçado por Naghdi, Ghajar e Tsioras (2016), que obtiveram como resultado uma precisão relativamente mais alta no modelo ANFIS em comparação com o modelo de regressão linear múltipla. Neste contexto, o desenvolvimento de sistemas *fuzzy* tem se tornado uma alternativa bastante efetiva diante dos métodos estocásticos, quando empregados à avaliação de fenômenos agrários, biológicos e administrativos (GABRIEL FILHO et al., 2012).

Portanto, dada a complexidade envolvida na previsão de valores de uma série temporal, os sistemas de previsão de valores da evapotranspiração de referência podem ser estabelecidos de forma mais eficiente por meio de modelos que utilizam as redes neurais ao invés da utilização de métodos estatísticos. Essa metodologia aprimorada é denominada sistemas neuro-*fuzzy* ou ANFIS (*Adaptive Neuro-Fuzzy Inference System*). Ressalta-se que a metodologia ANFIS utiliza o método de inferência de Takagi-Sugeno, enquanto que sistemas *fuzzy* convencionais (os que não fazem uso das redes neurais) fazem uso do método de inferência de Mandani.

## **Justificativa**

Considerando: (1) a evapotranspiração de referência, uma medida fundamental para o cálculo da necessidade hídrica das plantas, (2) o manejo correto da irrigação, para que se tenha aumento de produtividade sem desperdício de recursos hídricos; e (3) os modelos neuro-*fuzzy*, que são capazes de descrever fenômenos complexos, destaca-se a importância de estudos visando melhorar o planejamento de sistemas de irrigação por meio da estimativa de séries temporais da evapotranspiração de referência utilizando dados meteorológicos disponíveis.

Para isso é necessário que os produtores rurais e pesquisadores tenham ferramentas que os auxiliem na tomada de decisão para planejamento dos recursos empregados na produção agrícola, para que garantam eficiência em produtividade sem desperdícios de recursos físicos (como a água) e financeiros.

Atualmente, é corriqueiro tomar decisões frente às recomendações de aplicação de lâminas de irrigação já pré-fixadas em experimentos agrônômicos, uma vez que

análises estatísticas convencionais fazem acertadamente as devidas comparações nos níveis pré-estabelecidos. Todavia, é de grande importância que produtores e/ou pesquisadores tenham acesso a uma informação geral, que possa, posteriormente, ser utilizada de maneira personalizada para a melhoria do planejamento de safra. Portanto, a importância do presente trabalho é caracterizada pelo alargamento do poder de decisão e avanço na gestão produtiva de produtores, oferecendo maior eficiência nos tratamentos culturais, através de uma metodologia que possibilita a obtenção de resultados superiores a outros comumente utilizados.

## **Objetivos**

O objetivo deste trabalho é estabelecer modelos de previsão de séries temporais da evapotranspiração de referência (Penman-Monteith/FAO 56) utilizando modelo neuro-*fuzzy* (ANFIS).

Especificamente, objetiva-se elaborar e avaliar o desempenho de um modelo que utilize observações anteriores para a estimação de um valor posterior, utilizando técnicas de inteligência artificial para relacionar valores passados da evapotranspiração de referência ( $ET_0$ ) com valores futuros.

Para alcançar o objetivo proposto, a tese foi dividida em dois capítulos, intitulados: (1) “Imputação múltipla *pmm* e *midastouch* em séries históricas de dados meteorológicos” e (2) “Modelagem neuro-*fuzzy* para a previsão da evapotranspiração de referência estimada pelo método de Penman-Monteith”.

## CAPÍTULO 1 - IMPUTAÇÃO MÚLTIPLA PMM E MIDASTOUCH EM SÉRIES HISTÓRICAS DE DADOS METEOROLÓGICOS

### Resumo

A presença de dados faltantes (*missings*) em séries de dados é comum e deve ser tratada de maneira adequada para não comprometer a aplicação de alguns métodos estatísticos e, conseqüentemente, minimizar os vieses que podem causar nos resultados e conclusões dos estudos. Para isso, existem diferentes métodos possíveis de serem aplicados de acordo com as características do conjunto de dados e o volume de dados faltantes. Assim, o objetivo deste capítulo é testar os procedimentos de imputação múltipla *pmm* (Correspondência Média Preditiva) e *midastouch* (Correspondência Média Preditiva Ponderada) para um conjunto de variáveis meteorológicas coletadas nos municípios de Botucatu – SP e Tupã – SP. Para isso, foram utilizados dados de radiação solar global, velocidade do vento, temperatura do ar, temperatura do ar máxima, temperatura do ar mínima, umidade relativa do ar, umidade relativa do ar máxima e umidade relativa do ar mínima do período de 20 de março de 2018 a 19 de março de 2021 da estação meteorológica da UNESP (Botucatu – SP) e do INMET (Tupã – SP). A série temporal de Botucatu – SP apresentou 1,4% de dados ausentes enquanto a de Tupã – SP possuía 7%, sendo que as reconstruções utilizando os métodos *pmm* e *midastouch* apresentaram resultados semelhantes nos gráficos de densidade de probabilidade dos dados observados e imputados e de distribuição dos dados por meio de pontos e *box-plot*. Ao comparar os métodos de imputação por meio dos indicadores estatísticos de erros, o banco de dados de Botucatu – SP apresentou diferenças mais significativas do que o de Tupã - SP, causado pelo menor volume de informações no banco de dados. Apesar de tais diferenças, o NRMSE dos bancos de dados apresentou proximidade, sendo de 1,398 para o método *pmm* e 1,403 para o *midastouch*, para Botucatu – SP e, para Tupã – SP, o método *pmm* apresentou NRMSE de 1,335 e o *midastouch* de 1,330. Portanto, o método *midastouch* apresentou melhor desempenho para séries maiores de dados meteorológicos.

**Palavras-chave:** dados ausentes, série temporal, reconstrução de base de dados.

## Abstract

The presence of missing data (missings) in data series is common and must be treated properly so as not to compromise the application of some statistical methods and, consequently, minimize the biases that may cause the results and conclusions of the studies. For this, there are different possible methods to be applied according to the characteristics of the dataset and the volume of missing data. Thus, the objective of this chapter is to test the procedures of multiple imputation pmm (Predictive Mean Correspondence) and midastouch (Weighted Mean Predictive Correspondence) for a set of meteorological variables collected in the cities of Botucatu – SP and Tupã – SP. For this, data on global solar radiation, wind speed, air temperature, maximum air temperature, minimum air temperature, relative air humidity, maximum relative air humidity and minimum relative air humidity for the period of 20 December were used. March 2018 to March 19, 2021 from the meteorological station of UNESP (Botucatu – SP) and INMET (Tupã – SP). The time series of Botucatu - SP presented 1.4% of missing data while that of Tupã - SP had 7%, and the reconstructions using the pmm and midastouch methods showed similar results in the probability density graphs of the observed and imputed data and distribution of data through points and box-plot. When comparing the imputation methods through statistical error indicators, the Botucatu - SP database showed more significant differences than the Tupã - SP database, caused by the lower volume of information in the database. Despite such differences, the NRMSE of the databases showed proximity, being 1.398 for the pmm method and 1.403 for the midastouch, for Botucatu - SP and, for Tupã - SP, the pmm method presented an NRMSE of 1.335 and the midastouch of 1.330. Therefore, the midastouch method performed better for larger series of meteorological data.

**Keywords:** missing data, time series, database reconstruction.

## 1.1 Introdução

É comum a presença de dados faltantes (*missings*) em diversos estudos científicos, sendo que tais ausências são percebidas em uma análise exploratória preliminar dos dados da pesquisa (MIOT, 2019; ALERYANI; WANG; DE LA IGLESIA, 2020). Além disso, a presença de dados faltantes deve ser avaliada com cuidado em muitos procedimentos estatísticos, já que podem causar vieses nos resultados e conclusões (AUDIGIER; HUSSON; JOSSE, 2014; CARVALHO, 2017). Desta forma, o tratamento adequado ou a imputação de valores ausentes torna-se uma etapa de pré-processamento essencial para a análise de dados, já que seu objetivo é minimizar os efeitos de conjuntos de dados incompletos (ARMINA *et al.*, 2017).

De acordo com Austin *et al.* (2021) o tratamento de dados ausentes deve ser escolhido entre cinco grandes categorias: *Listwise Deletion*, *Pairwise Deletion*, Imputação Única ou Simples, Máxima Verossimilhança e Imputação Múltipla. O método de *Listwise Deletion* prevê a exclusão de registros que possuem valor faltante realizando a análise apenas com os dados completos. Apesar da simplicidade, o método reduz o tamanho da amostra e o poder estatístico dos testes de hipótese (ANDRADE *et al.*, 2019). Diferentemente do método *Pairwise Deletion*, que exclui apenas o dado faltante mantendo as observações das outras variáveis do referido registro, assim, é possível sua utilização em análises onde as variáveis que possuam correlação com a variável de interesse apresentem dados completos (GUPTA; GUPTA, 2018). A imputação única ou simples prevê a estimação do valor ausente apenas uma vez, utilizando, entre diversos métodos, a média, moda, mediana, regressão logística e regressão linear para a imputação do valor, não sendo possível a captação da variabilidade e incerteza associada ao real valor perdido (RUBIN, 1987; ANDRADE *et al.*, 2019). A fim de superar tal desvantagem, foi criada a imputação múltipla que combina os resultados de imputação de  $m$  não respostas e  $m$  análises de dados completos, contabilizando a imprecisão de cada imputação (RUBIN, 1987; RIBEIRO, 2015). O método da máxima verossimilhança possui procedimentos matemáticos complexos, o qual a partir de uma função que contenha a informação do conjunto de dados (função de verossimilhança), busca-se maximizá-la por meio das estimativas dos parâmetros relacionados à distribuição de

probabilidade aderente aos dados observados (RIBEIRO, 2015; ANDRADE et al., 2019).

Estudos sobre o tratamento de dados faltantes evidenciam que o método de imputação múltipla é comumente utilizado para tal fim, visto ser uma maneira prática e eficiente de lidar com ausência de dados, já que o método substitui o valor ausente por um conjunto de valores que refletem a incerteza sobre o verdadeiro valor perdido, obtendo erros padrão e intervalos de confiança menores (CHHABRA; RANJAN, 2017; AUSTIN et al., 2021).

O tratamento de dados ausentes deve considerar ainda, o tipo e a quantidade de dados faltantes. Rubin (1976) classificou em três categorias (MCAR, MAR, MNAR) o problema de dados faltantes. A primeira MCAR (*missing complete at random*) considera a probabilidade de falta de dados ser totalmente ao acaso, independente da própria variável ou outra variável do banco de dados (AFRIFA-YAMOAHA et al., 2020). Os dados do tipo MAR (*missing at random*) ocorrem quando a probabilidade de perda de dados depende do valor de outras variáveis e os dados NMAR (*not missing at random*) surge quando o valor ausente ocorre de forma não aleatória, dependendo do próprio valor perdido (AIEB et al., 2019).

A respeito do percentual de dados faltantes, Harrell (2016) indica que o método de imputação única ou apenas a análise dos dados completos (eliminando os dados perdidos) seja utilizado para percentual de dados faltantes igual ou inferior a 5%. Se este índice estiver entre 5 e 15% a imputação única pode ser utilizada, porém indica-se a imputação múltipla e, para banco de dados com valor igual ou superior a 15% de dados perdidos recomenda-se o uso da imputação múltipla. Além disso, existem diferentes procedimentos de imputação de acordo com o tipo de variável, sendo indicado para dados contínuos os procedimentos de correspondência média preditiva (*pmm*) ou o algoritmo de florestas aleatória; para dados binários pode-se utilizar a regressão logística; para dados categóricos não ordenados pode-se optar pela regressão logística politômica ou análise discriminante linear; e para dados categóricos ordenados podem ser implementados os modelos de *odds* proporcionais (SILVA et al., 2019).

Vale ressaltar que, além destes procedimentos existe outros que podem ser ainda mais eficientes a determinados conjuntos de dados, além das diversas revisões metodológicas que resultam na proposição de novos procedimentos. Este é o caso de Gaffert, Meinfelder e Bosch (2018) que ao estudarem o método

Correspondência Média Preditiva (*pmm*), procedimento bastante popular para variáveis contínuas, verificam suas fragilidades e propõe um novo procedimento denominado Correspondência Média Preditiva Ponderada (*midastouch*), o qual apresentou resultado superior a outros nos estudos de simulação. Esse procedimento foi implementado no ambiente de análise de dados R (R CORE TEAM, 2021) via pacote *mice* (BUUREN; GROOTHUIS-OUDSHOORN, 2011).

Entre os autores que utilizaram estes métodos para imputação de dados oriundos de estações meteorológicas pode-se citar Silva et al. (2019) que imputaram dados de radiação solar global horária com o método *midastouch* para o município de Arapiraca – AL. Já o método *pmm* foi utilizado para imputar dados de precipitação diária por Costa et al. (2012) no estado da Paraíba, Carvalho et al. (2017) utilizaram a técnica em duas zonas homogêneas de precipitação compreendidas pela região semiárida do nordeste brasileiro e outra área compreendida pelo noroeste do estado de São Paulo e parte do estado do Mato Grosso do Sul, e Alves e Gomes (2020) aplicaram a técnica em séries temporais de precipitação da região da Bacia Hidrográfica do Médio São Francisco.

Assim, considerando (1) que as variáveis deste estudo (radiação solar global, velocidade do vento, temperatura do ar, temperatura do ar máxima, temperatura do ar mínima, umidade relativa do ar, umidade relativa do ar máxima e umidade relativa do ar mínima) possuem caráter quantitativo e contínuo; (2) a ampla utilização do método de Correspondência Média Preditiva (*pmm*) para imputação de dados de variáveis contínuas e (3) o melhor desempenho do método Correspondência Média Preditiva Ponderado (*midastouch*).

O objetivo deste trabalho é testar os métodos de imputação múltipla internos *pmm* e *midastouch* para dois conjuntos de variáveis meteorológicas, sendo o primeiro coletado no município de Botucatu – SP e o outro coletado no município de Tupã – SP.

## 1.2 Material e Métodos

Foram coletados dados horários de radiação solar global (Rad), velocidade do vento (VV), temperatura do ar (T), temperatura do ar máxima (Tmax), temperatura do ar mínima (Tmin), umidade relativa do ar (UR), umidade relativa do ar máxima (Urmax) e umidade relativa do ar mínima (Urmin) aferidas por duas estações

meteorológicas. A primeira localiza-se no município de Botucatu – SP, nas coordenadas geográficas 22° 50' Latitude Sul e 48° 25' de Longitude Oeste com altitude média de 786 metros, sob responsabilidade do Departamento de Solos e Recursos Ambientais da UNESP, Faculdade Ciências Agrônômicas, Fazenda Experimental Lageado. A segunda refere-se à estação automática do Instituto Nacional de Meteorologia (INMET) - Estação Tupã-A768, localizada no município de Tupã – SP nas coordenadas geográficas 21° 93' Latitude Sul e 50° 49' de Longitude Oeste com altitude média de 498 metros (Figura 1).

Figura 1 – Localização das estações meteorológicas de coleta de dados no estado de São Paulo.



Ressalta-se que os dados fornecidos pela estação de Botucatu – SP são diários e os dados de Tupã – SP são horários. Assim, os dados coletados no período de 20 de março de 2018 a 19 de março de 2021, totalizaram 1.096 informações diárias de cada variável para o município de Botucatu – SP e 26.304 informações horárias para cada variável para o município de Tupã – SP. Ao organizar os dados cronologicamente, do mais antigo para o mais recente, verificou-se a ausência de diversas informações em ambos os bancos de dados, as quais foram tratadas utilizando o ambiente R de análise de dados, versão 4.1.1 (R CORE TEAM, 2021).

Primeiramente, procedeu-se a avaliação dos bancos de dados a fim de verificar a quantidade de dados ausentes (*missing*), por meio dos pacotes *naniar* (TIERNEY

et al., 2020) e *VIM* (KOWARIK; TEMPL, 2016), obtendo a quantidade de dados faltantes para que se pudesse definir o método de imputação de dados mais adequado, considerando a recomendação de Harrell (2016) e o tipo de dados faltantes.

O método de imputação múltipla foi implementado por meio do pacote *mice* (BUUREN; GROOTHUIS-OUDSHOORN, 2011), o qual estabelece um processo iterativo com  $m$  imputações realizadas, sendo que neste trabalho foi estabelecido 5 imputações, que são agregadas para a contabilização da imprecisão e estimativa do dado faltante (ANDRADE et al., 2019).

Além da definição da quantidade de imputações que irão interagir para a estimativa dos dados faltantes, é preciso definir o procedimento a ser utilizado. Para este trabalho comparou-se o método Correspondência Média Preditiva (*pmm*) ao Correspondência Média Preditiva Poderada (*midastouch*), sendo que as imputações pelo método *pmm* são baseadas na distribuição preditiva posterior dos dados ausentes e o método *midastouch* na distribuição posterior dos parâmetros (Gaffert, Meinfelder e Bosch, 2018).

Para verificação da qualidade das imputações (*pmm* e *midastouch*) foram utilizados gráficos de densidade e pontos. Já para comparação dos desempenhos dos dois métodos utilizou-se as seguintes medidas estatísticas de erro: Raiz Quadrada do Erro Quadrático Médio normalizado (NRMSE), Erro Absoluto Médio Percentual (MAPE) e Erro Percentual Absoluto Médio Simétrico (SMAPE).

A medida NRMSE foi calculada por meio da Equação (1) implementada através pacote *imputeR* (FENG et al., 2020).

$$NRMSE_j = \sqrt{\frac{1}{n_j} \sum_{i=1}^n (y_{ij} - \widehat{y}_{ij})^2 / Var(\widehat{y}_{ij})} \quad (1)$$

onde  $y_{ij}$  é o  $i$ -ésimo dado observado da  $j$ -ésima variável,  $\widehat{y}_{ij}$  é o  $i$ -ésimo dado imputados da  $j$ -ésima variável,  $n_j$  o número de dados faltantes da variável  $j$  e  $Var(\widehat{y}_{ij})$  a variância dos valores imputados (CHO et al., 2020). O NRMSE global foi obtido pela média dos NRMSE de cada variável.

O Erro Absoluto Médio Percentual (MAPE), obtido pela Equação (2), expressa a precisão das imputações como uma porcentagem do erro (LEWIS, 1997).

$$MAPE = \frac{1}{n_j} \sum_{i=1}^n \frac{|y_{ij} - \widehat{y}_{ij}|}{y_{ij}} * 100 \quad (2)$$

onde  $y_{ij}$  é o  $i$ -ésimo dado observado da  $j$ -ésima variável,  $\widehat{y}_{ij}$  é o  $i$ -ésimo dado imputados da  $j$ -ésima variável,  $n_j$  o número de dados faltantes da variável  $j$ . Já o Erro Percentual Absoluto Médio Simétrico (SMAPE) dado pela Equação (3) foi calculado por meio do pacote *DescTools do R* (ANDRI et al., 2021).

$$SMAPE_j = \frac{100\%}{n_j} \sum_{t=1}^n \frac{|y_{ij} - \widehat{y}_{ij}|}{|y_{ij}| + |\widehat{y}_{ij}|} \quad (3)$$

sendo  $y_{ij}$  é o  $i$ -ésimo dado observado da  $j$ -ésima variável,  $\widehat{y}_{ij}$  é o  $i$ -ésimo dado imputados da  $j$ -ésima variável,  $n_j$  o número de dados faltantes da variável  $j$  (FLORES, 1986).

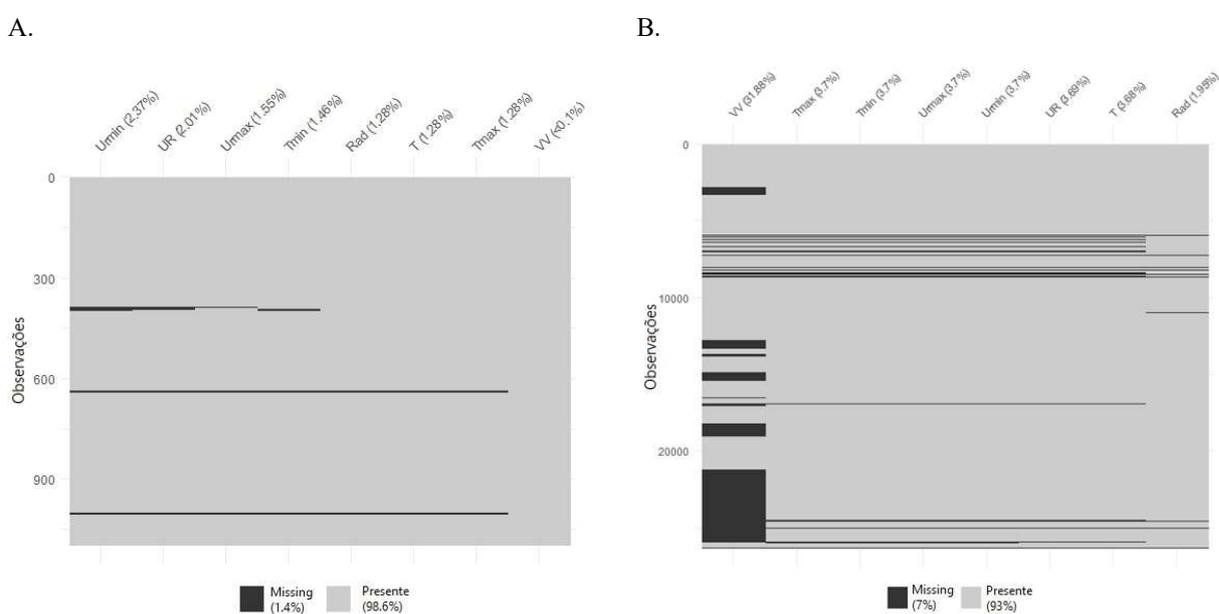
Devido ao conjunto de dados original já possuir dados ausentes, foi necessário simular os dados faltantes para o cálculo das medidas estatísticas de erro (NMRSE, MAPE e SMAPE). Para isso, os dados ausentes foram retirados do conjunto de dados original e posteriormente selecionado aleatoriamente 8% das linhas para a inserção de dados faltantes. A partir desse novo banco de dados o método de imputação foi executado a fim de preencher as falhas, possibilitando o cálculo das referidas medidas estatísticas de erro.

### 1.3 Resultados e Discussão

A avaliação dos bancos de dados mostra que as ausências são significativamente menores para o município de Botucatu – SP, uma vez que seus dados estão organizados de forma diária. Assim, tal município (Figura 2A) apresentou 2,37% de dados ausentes para a variável Urmin, 2,01% para UR, 1,55% para Urmax, 1,46% para Tmin, 1,28% para Rad, T e Tmax e menos de 0,1% para VV, totalizando 124 dados faltantes (1,4%).

Já o município de Tupã – SP (Figura 2B) possuía 14.744 dados ausentes (7%), sendo que o maior volume de dados faltantes ocorreu para a variável VV (31,88%) frente aos 3,7% para as variáveis Tmax, Tmin, Urmx, Urmin, UR e T e 1,95% para Rad. Vale ressaltar, que para a variável radiação solar global (Rad) os valores ausentes existentes no período das 19h às 6h foram preenchidos com o valor zero, visto que se trata do período noturno e não há incidência de radiação.

Figura 2 - Quantidade de dados faltantes, por variável, para o banco de dados do município de (A) Botucatu – SP e (B) Tupã – SP.



Assim, considerando que os dados ausentes são do tipo MAR, o percentual de dados ausentes de cada banco de dados e as recomendações de Harrell (2016), optou-se pela utilização do método de imputação múltipla, uma vez que tal método possibilita a compreensão da variabilidade e incerteza associada ao valor perdido para ambos os bancos de dados analisados. Desta forma, verificou-se a qualidade de imputação de dois de seus procedimentos: Correspondência Média Preditiva (*pmm*) e o Correspondência Média Preditiva Poderada (*midastouch*).

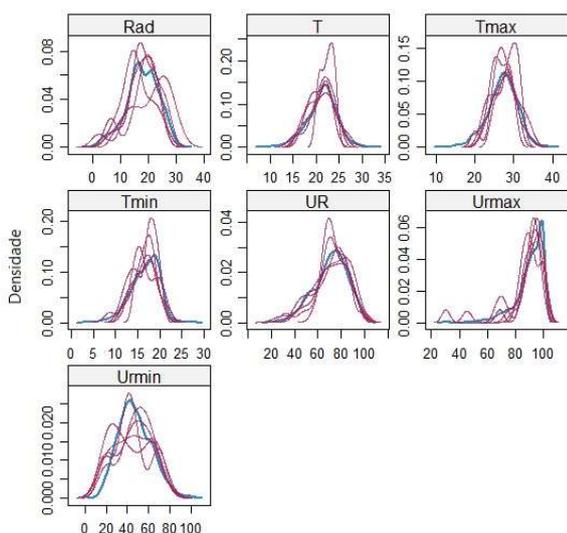
As séries temporais reconstruídas por ambos os métodos de imputação múltipla permitiram a manutenção da variabilidade temporal em escala horária para Tupã – SP e diária para Botucatu – SP, sendo este fato evidenciado pelos gráficos de densidade (Figura 3 e Figura 4), onde o eixo das abscissas (conhecido como eixo  $x$ )

de cada gráfico representa os valores observados para cada variável analisada e o eixo das ordenadas (conhecido como eixo  $y$ ) evidencia a densidade.

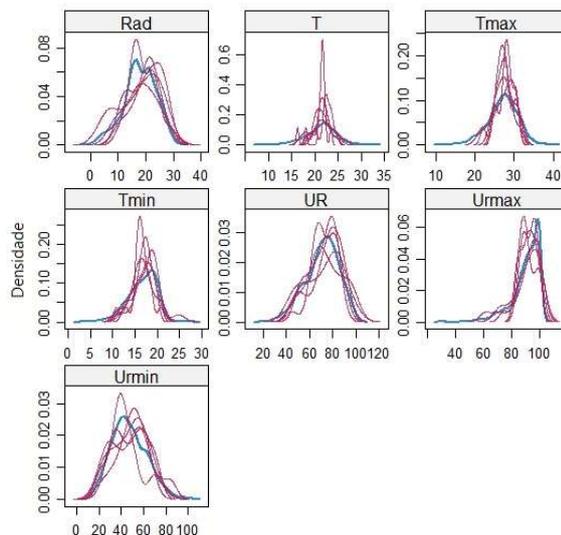
Os gráficos mostram a densidade de probabilidade dos dados originais por meio da linha azul e dos dados imputados pelas linhas vermelhas, sendo estas os resultados obtidos em cada uma das cinco interações do método de imputação de dados ausentes. Verifica-se pela análise dos resultados obtidos para o município de Botucatu – SP, que apenas as curvas de imputação do método Correspondência Média Preditiva (*pmm*) (Figura 3A) das variáveis T e UR apresentaram maior proximidade com a curva de densidade dos dados originais quando comparado às curvas de imputação do método Correspondência Média Preditiva Ponderada (*midastouch*) (Figura 3B), sendo que para as demais variáveis a proximidade maior com os dados originais ocorreu com o método *midastouch*. Vale ressaltar, que a variável velocidade do vento não apresenta gráfico de densidade de probabilidade, uma vez que este banco de dados apresenta apenas um dado faltante para esta variável.

Figura 3 – Gráficos de densidade de probabilidade dos dados observados e imputados pelo método (A) Correspondência Média Preditiva (*pmm*) e (B) Correspondência Média Preditiva Ponderada (*midastouch*) do município de Botucatu – SP.

A.



B.

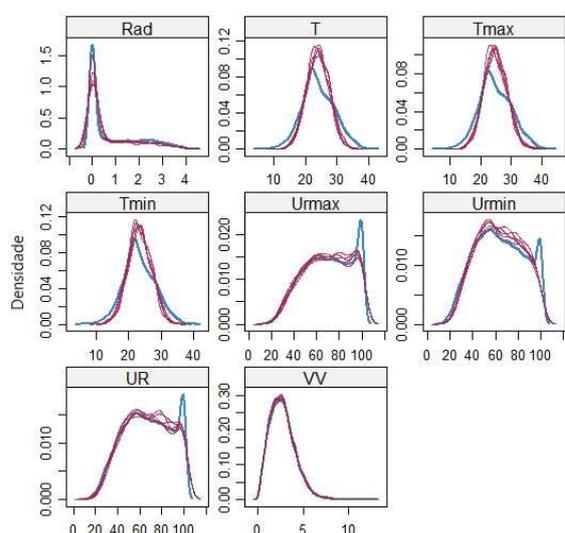


Legenda: Rad – Radiação Solar Global; T – Temperatura do Ar; Tmax – Temperatura do Ar Máxima; Tmin – Temperatura do Ar Mínima; UR – Umidade Relativa do Ar; Urmx - Umidade Relativa do Ar Máxima; Urmin - Umidade Relativa do Ar Mínima.

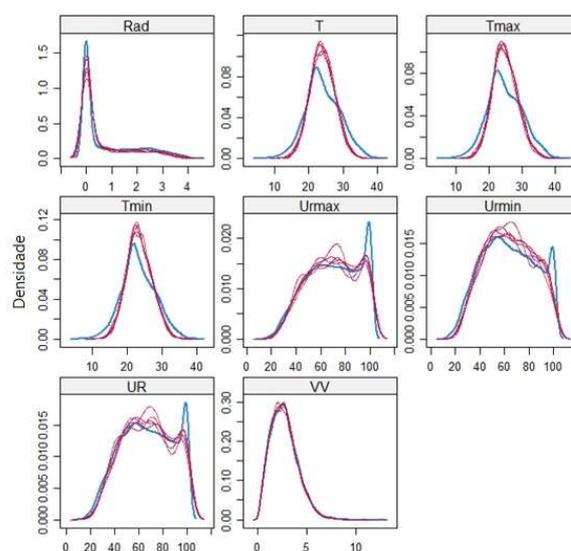
Ao observar as linhas de iteração obtidas pelos métodos *pmm* (Figura 4A) e *midastouch* (Figura 4B) para o município de Tupã – SP, verifica-se um melhor ajuste aos dados originais, sendo causada pelo maior detalhamento do banco de dados (informações horárias). Além disso, destaca-se comportamento bastante semelhante das linhas de imputação aos dados originais nas variáveis Rad e VV para ambos os métodos, sendo que para as demais variáveis há uma satisfatória aproximação das linhas de iteração ao comportamento dos dados originais.

Figura 4 - Gráficos de densidade de probabilidade dos dados observados e imputados pelo método (A) Correspondência Média Preditiva (*pmm*) e (B) Correspondência Média Preditiva Ponderada (*midastouch*) do município de Tupã – SP.

A.



B.



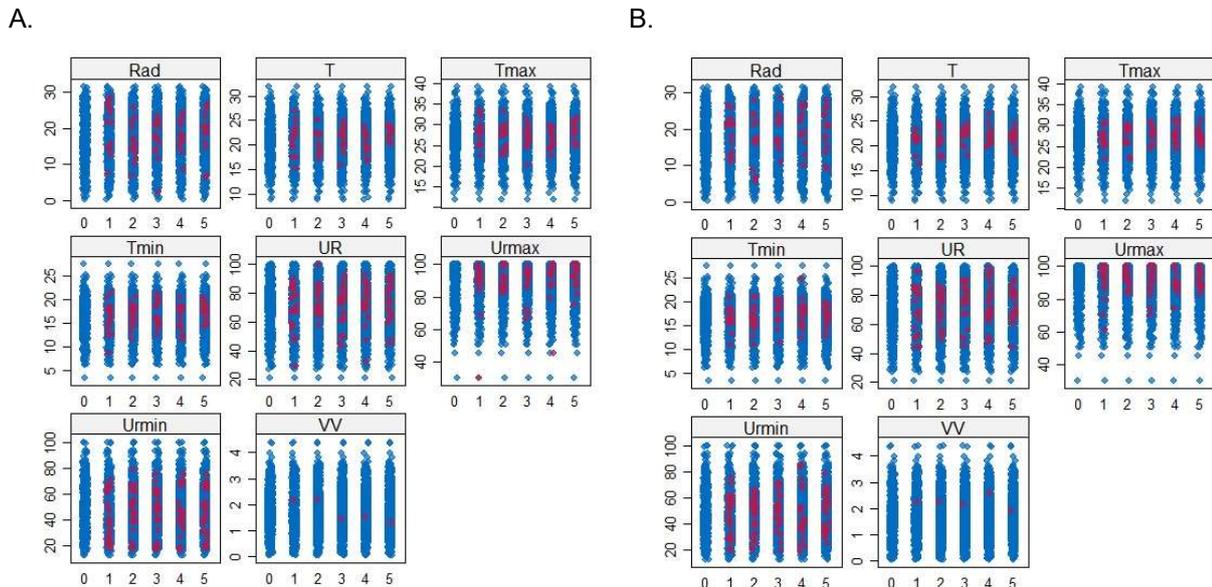
Legenda: Rad – Radiação Solar Global; T – Temperatura do Ar; Tmax – Temperatura do Ar Máxima; Tmin – Temperatura do Ar Mínima; UR – Umidade Relativa do Ar; Urmx - Umidade Relativa do Ar Máxima; Urmin - Umidade Relativa do Ar Mínima.

A qualidade das imputações também pode ser avaliada pelo gráfico de pontos, os quais mostram para cada variável analisada os valores imputados (vermelhos) e os valores observados (azuis), sendo que os valores dispostos na linha zero referem-se aos dados originais e as linhas de 1 a 5 mostram os valores atribuídos nas referidas iterações dos métodos de imputação.

Para o município de Botucatu – SP verifica-se que ambos os métodos (Figura 5) apresentam distribuição bastante semelhante, sendo influenciado pela baixa

quantidade de dados imputados, que possibilitou aos métodos a extração das informações sobre o comportamento dos dados ao longo do tempo em seu máximo.

Figura 5 – Gráfico de pontos para a distribuição dos dados imputados pelo método (A) Correspondência Média Preditiva (*pmm*) e (B) Correspondência Média Preditiva Ponderada (*midastouch*) para o município de Botucatu – SP.



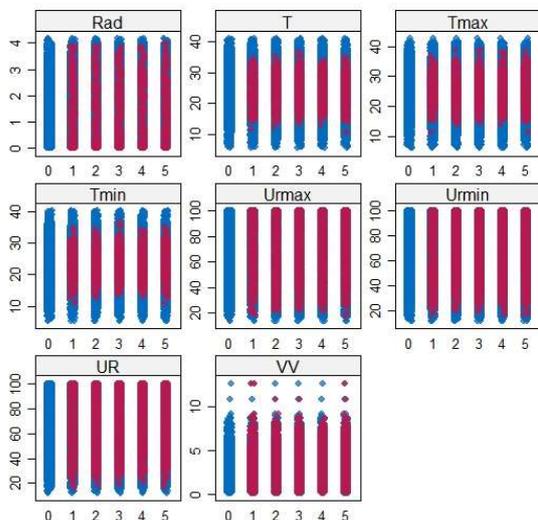
Legenda: Rad – Radiação Solar Global; T – Temperatura do Ar; Tmax – Temperatura do Ar Máxima; Tmin – Temperatura do Ar Mínima; UR – Umidade Relativa do Ar; Urmx - Umidade Relativa do Ar Máxima; Urmin - Umidade Relativa do Ar Mínima.

Já para o município de Tupã – SP o qual 7% do banco de dados possuía dados faltantes, verifica-se que os valores imputados estão contidos na amplitude dos dados originais, sendo que o método *pmm* (Figura 6A) apresenta maior centralidade dos dados imputados enquanto o método *midastouch* (Figura 6B) mostra tendência de que os valores imputados ocupem a amplitude observada nos dados originais.

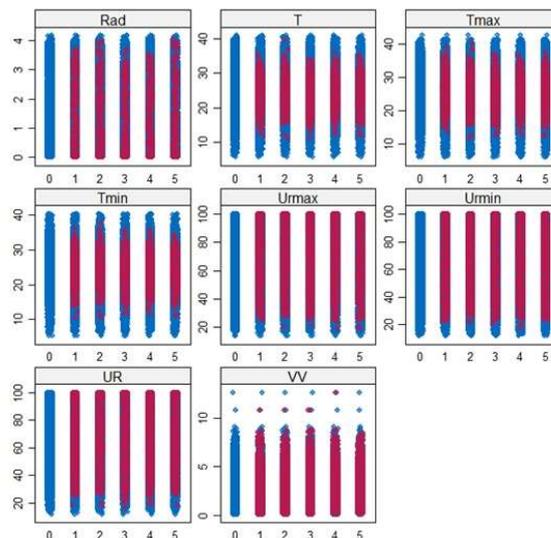
Este comportamento é explicado pela atenuação do viés de estimativa da variância total do método *pmm*, já que as imputações extraem amostras de uma distribuição preditiva posterior dos dados ausentes enquanto o método *midastouch* desenha as distribuições posteriores dos parâmetros para que a incerteza do conjunto de dados se propague (GAFFERT; MEINFELDER; BOSCH, 2018).

Figura 6 – Gráfico de pontos para a distribuição dos dados imputados pelo método (A) Correspondência Média Preditiva (*pmm*) e (B) Correspondência Média Preditiva Ponderada (*midastouch*) para o município de Tupã – SP.

A.



B.

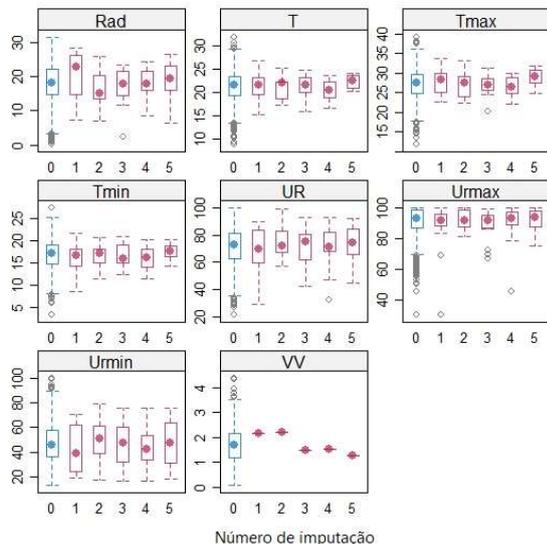


Legenda: Rad – Radiação Solar Global; T – Temperatura do Ar; Tmax – Temperatura do Ar Máxima; Tmin – Temperatura do Ar Mínima; UR – Umidade Relativa do Ar; Urmáx - Umidade Relativa do Ar Máxima; Urmín - Umidade Relativa do Ar Mínima.

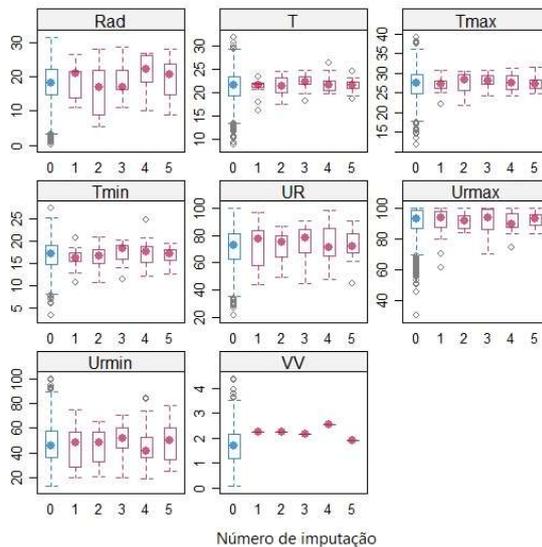
A distribuição dos dados originais e das cinco interações do processo de imputação do banco de dados do município de Botucatu – SP podem ser vista nos gráficos *box-plot* (Figura 7), onde verifica-se a presença de *outliers* (valores extremos) no banco de dados original (iteração 0), percebe-se ainda, que a imputação de dados tenta reproduzir as medidas estatísticas descritivas que formam o gráfico *box-plot* (mínimo, máximo, primeiro quartil, segundo quartil ou mediana e terceiro quartil) do banco de dados original. Assim, algumas interações apresentam *outliers* e outras sofrem alterações na média ou limites superior e inferior dos gráficos em ambos os métodos de imputação, a fim de testar o melhor ajuste para a atribuição dos valores. Vale ressaltar, que a variável velocidade do vento (VV) apresentou comportamento diferente das demais variáveis, pois havia apenas um valor faltante em sua série temporal. Resultados semelhantes foram obtidos por Oliveira (2019), permitindo concluir que a imputação de dados faltantes foi eficiente, já que a distribuição dos dados das interações foram semelhantes ao banco de dados original.

Figura 7 – Gráfico *box-plot* para a distribuição dos imputados pelo método (A) Correspondência Média Preditiva (*pmm*) e (B) Correspondência Média Preditiva Ponderada (*midastouch*) para o município de Botucatu – SP.

A.



B.

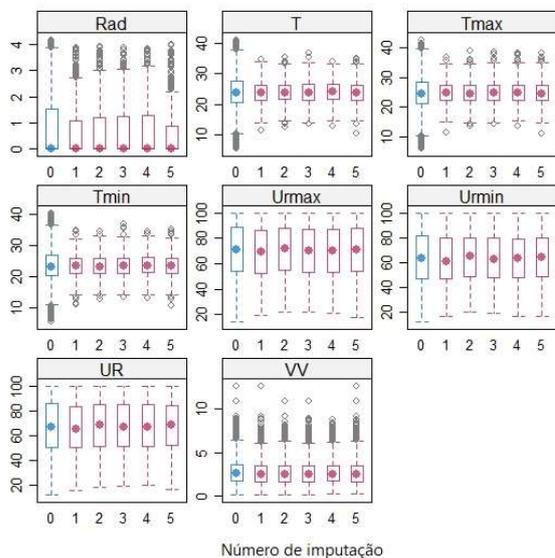


Legenda: Rad – Radiação Solar Global; T – Temperatura do Ar; Tmax – Temperatura do Ar Máxima; Tmin – Temperatura do Ar Mínima; UR – Umidade Relativa do Ar; Urmx - Umidade Relativa do Ar Máxima; Urmin - Umidade Relativa do Ar Mínima.

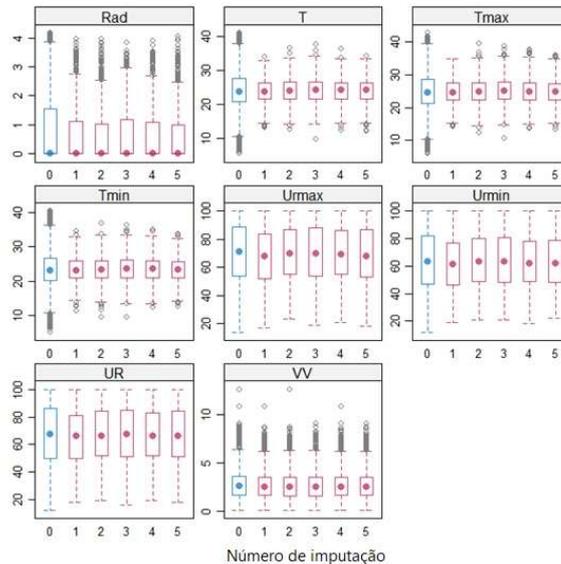
No caso do banco de dados de Tupã – SP (Figura 8), a presença de *outliers* é ainda maior para as variáveis Rad, T, Tmax, Tmin e VV, sendo que as variáveis UR, Urmx e Urmin não apresentam observações atípicas (*outliers*). Verifica-se também, a semelhança na distribuição dos dados das interações com os dados originais para ambos os métodos de imputação.

Figura 8 - Gráfico *box-plot* para a distribuição dos imputados pelo método (A) Correspondência Média Preditiva (*pmm*) e (B) Correspondência Média Preditiva Ponderada (*midastouch*) para o município de Tupã – SP.

A.



B.



Legenda: Rad – Radiação Solar Global; T – Temperatura do Ar; Tmax – Temperatura do Ar Máxima; Tmin – Temperatura do Ar Mínima; UR – Umidade Relativa do Ar; Urmx - Umidade Relativa do Ar Máxima; Urmin - Umidade Relativa do Ar Mínima.

A partir da simulação de dados faltantes, foi possível calcular o valor global médio da Raiz Quadrada do Erro Quadrático Médio normalizado (NRMSE), bem como para cada variável, além do Erro Absoluto Médio Percentual (MAPE) e o Erro Percentual Absoluto Médio Simétrico (SMAPE) para cada variável (Tabela 1). A análise de tais medidas evidencia que a imputação de valores faltantes por ambos os métodos apresentou resultados semelhantes, sendo realizadas previsões bastante precisas, já que as medidas estatísticas de erros apresentaram baixos valores. O NRMSE permite a comparação da acurácia da imputação entre as diferentes variáveis existentes no banco de dados, mostrando que apesar do método *pmm* apresentar melhor desempenho em algumas variáveis, os resultados globais médios ficaram muito próximos no banco de dados de Botucatu – SP (1,398 e 1,403 respectivamente) e no de Tupã – SP (1,335 e 1,330), sendo menor para *pmm* em Botucatu e maior em Tupã-SP.

O MAPE expressa o erro percentual médio em valores absolutos, assim quanto menor seu resultado, melhor foi a imputação de valores perdidos. Sua desvantagem

consiste em não obter respostas para séries que possuem valor zero, sendo por isso a não existência de valores para a variável Rad (possui valor zero nos horários noturnos) no banco de dados de Tupã – SP. Além disso, o MAPE é uma medida assimétrica, levando à penalização daquelas previsões que foram maiores que as reais. Assim, quando se compara os resultados de Urmin, o banco de dados de Botucatu – SP obteve 0,365 no método *pmm* e 0,427 no *midastouch* enquanto para Tupã – SP os resultados foram, respectivamente, 0,386 e 0,388, indicando que a diferença entre os valores obtidos pelos métodos no banco de dados de Botucatu – SP, é devido ao menor volume de informações para a extração dos parâmetros dos valores perdidos, uma vez que as informações estavam em escala diária. Desta forma, o erro SMAPE possibilita a correção desta assimetria, apresentando valores bastante próximos entre os métodos para as duas bases de dados. Com isso, a variável Urmin apresenta resultados de 0,311 e 0,365, respectivamente para os métodos *pmm* e *midastouch* para Botucatu – SP e 0,352 e 0,350 para os respectivos métodos do banco de dados de Tupã – SP.

Tabela 1 – Resultado das medidas estatísticas NRMSE, MAPE e SMAPE calculados para os métodos *pmm* e *midastouch* dos municípios de Botucatu – SP e Tupã – SP.

		Botucatu - SP		Tupã - SP	
		<i>pmm</i>	<i>midastouch</i>	<i>pmm</i>	<i>midastouch</i>
NRMSE	Global Médio	1,398	1,403	1,335	1,330
	Rad	1,370	1,435	1,375	1,380
	T	1,389	1,359	1,228	1,220
	Tmax	1,361	1,334	1,235	1,235
	Tmin	1,466	1,416	1,248	1,242
	UR	1,294	1,418	1,402	1,403
	Urmx	1,365	1,386	1,375	1,366
	Urmin	1,285	1,548	1,360	1,364
	VV	1,652	1,331	1,462	1,434
MAPE	Rad	0,575	0,694	-	-
	T	0,215	0,203	0,225	0,223
	Tmax	0,188	0,178	0,228	0,227
	Tmin	0,285	0,265	0,225	0,224
	UR	0,238	0,262	0,377	0,378
	Urmx	0,119	0,123	0,348	0,349
	Urmin	0,365	0,427	0,386	0,388
	VV	0,597	0,513	0,893	0,868
SMAPE	Rad	0,448	0,427	-	-
	T	0,202	0,191	0,210	0,209
	Tmax	0,179	0,166	0,212	0,212
	Tmin	0,259	0,243	0,210	0,210
	UR	0,206	0,226	0,346	0,344
	Urmx	0,112	0,115	0,319	0,316
	Urmin	0,311	0,365	0,352	0,350
	VV	0,525	0,430	0,584	0,593

Gaffert, Meinfelder e Bosch (2018) esclarecem que o algoritmo do método *midastouch* apresenta vantagem sobre os algoritmos do método *pmm* principalmente quando há grande incerteza sobre os parâmetros do modelo, ou seja, a diferença é maior à medida que a amostra é menor. Esta afirmação pode ser confirmada pelos resultados obtidos, já que para o banco de dados de Botucatu – SP (124 dados ausentes) houve diferenças maiores nos erros estatísticos do que para o banco de dados de Tupã – SP (14.744 *missings*). Tal efeito também foi constatado por Faquih et al. (2020) que obtiveram aumento do viés de estimativa e RMSE em tamanho de amostras maiores e mesmo percentual de dados ausentes.

## 1.4 Conclusões

Os resultados do estudo mostram que os métodos de imputação Correspondência Média Preditiva (*pmm*) e Correspondência Média Preditiva Ponderada (*midastouch*) apresentam resultados semelhantes oferecendo boas estimativas para dados faltantes de variáveis quantitativas contínuas. Desta forma, os métodos se mostram eficazes na imputação de dados faltantes independente da escala de tempo utilizada pela base de dados, oferecendo qualidade para as análises estatísticas, resultados e conclusões de estudos, já que minimizam possíveis distorções que poderiam ocorrer caso os dados ausentes do banco de dados original não fossem tratados.

Considerando a vantagem da propagação da incerteza do valor perdido apresentada pelo método *midastouch*, que foi evidenciada nos dados utilizados, recomenda-se a utilização deste método para a imputação de valores meteorológicos. Tais dados, que são fundamentais para a previsão da  $ET_o$ , constantemente apresentam falhas em suas séries temporais, devido a problemas técnicos em sensores ou armazenamento de dados, e o com a utilização do método *midastouch* de imputação múltipla é possível a correção dessas falhas com alta acurácia.

## Referências

- AFRIFA-YAMOA, E. et al. Missing data imputation of high-resolution temporal climate time series data. *Meteorological Applications*, v. 27, n. 1, 2020.
- AIEB, A. et al. A new approach for processing climate missing databases applied to daily rainfall data in Soummam watershed, Algeria. *Heliyon*, v. 5, n. 2, 2019.
- ALERYANI, A.; WANG, W.; DE LA IGLESIA, B. Multiple Imputation Ensembles (MIE) for Dealing with Missing Data. *SN Computer Science*, v. 1, n. 3, 2020.
- ALVES, L. E. R.; GOMES, H. B. Validação da Imputação Múltipla via Predictive Mean Matching para Preenchimento de Falhas nos Dados Pluviométricos da Bacia do Médio São Francisco. *Anuário do Instituto de Geociências*, v. 43, n. 1, p. 199–206, 2020.
- ANDRADE, D. F. DE et al. Técnicas de imputação de dados na análise de questionários contextuais. *In: Caderno de Pesquisa 1*. Brasília: Cebraspe, 2019, p. 537.
- ANDRI, S. et al. **DescTools: Tools for descriptive statistics**. R package version 0.99.44, 2021.
- ARMINA, R. et al. A Review On Missing Value Estimation Using Imputation

- Algorithm. **Journal of Physics: Conference Series**, v. 892, n. 1, p. 012004, 2017.
- AUDIGIER, V.; HUSSON, F.; JOSSE, J. A principal component method to impute missing values for mixed data. **Advances in Data Analysis and Classification**, v. 10, n. 1, p. 5–26, 2014.
- AUSTIN, P. C. et al. Missing Data in Clinical Research: A Tutorial on Multiple Imputation. **Canadian Journal of Cardiology**, v. 37, n. 9, p. 1322–1331, 2021.
- BUUREN, S. VAN; GROOTHUIS-OUDSHOORN, K. Mice: Multivariate Imputation by Chained Equations in R. **Journal of Statistical Software**, v. 45, n. 1, p. 1–67, 2011.
- CARVALHO, J. R. P. et al. Model for Multiple Imputation to Estimate Daily Rainfall Data and Filling of Faults. **Revista Brasileira de Meteorologia**, v. 32, n. 4, p. 575–583, 2017.
- CHHABRA, G.; RANJAN, V. V. AND J. A Comparison of Multiple Imputation Methods for Data with Missing Values. **Indian Journal of Science and Technology**, v. 10, n. 19, p. 1–7, 2017.
- CHO, B. et al. Effective Missing Value Imputation Methods for Building Monitoring Data. *In: 2020 IEEE International Conference on Big Data (Big Data)*. **Anais...** Atlanta, Georgia, USA, p. 2866–2875, 2020.
- COSTA, R. L. et al. Imputação Multivariada de Dados Diários de Precipitação e Análise de Índices de Extremos Climáticos (Imputation Multivariate of Precipitation Daily Data and Analysis of Climate Extremes Index). **Revista Brasileira de Geografia Física**, v. 5, n. 3, p. 661–675, 2012.
- FAQUIH, T. et al. A Workflow for Missing Values Imputation of Untargeted Metabolomics Data. **Metabolites**, v. 10, n. 12, p. 486, 2020.
- FENG, L. et al. **imputeR: A General Multivariate Imputation Framework**. R package version 2.2, 2020.
- FLORES, B. E. A pragmatic view of accuracy measurement in forecasting, **Omega**, v. 14, n. 2, p. 93–98, 1986.
- GAFFERT, P.; MEINFELDER, F.; BOSCH, V. Towards an MI-proper Predictive Mean Matching. *In: Proceedings of the Joint Statistical Meetings 2018 Survey Research Methods Section*. **Anais...** Vancouver, British Columbia, Canada, 2018.
- GUPTA, S.; GUPTA, D. M. K. A Survey on Different Techniques for Handling Missing Values in Dataset. **International Journal of Scientific Research in Computer Science, Engineering and Information Technology**, v. 4, n. 1, p. 295–301, 2018.
- HARRELL, J. F. E. **Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis**. 2 ed. New York: Springer International Publishing, 2016.
- KOWARIK, A.; TEMPL, M. Imputation with the R Package VIM. **Journal of Statistical Software**, v. 74, n. 1, p. 1–16, 2016.
- LEWIS, C. D. **Demand Forecasting and Inventory Control**. New York: Wiley, 1997.
- MIOT, H. A. Valores anômalos e dados faltantes em estudos clínicos e experimentais. **Jornal Vascular Brasileiro**, v. 18, 2019.
- OLIVEIRA, L. M. DE. **IMPUTAÇÃO MÚLTIPLA E FUNÇÕES DE PEDOTRANSFERÊNCIA PARA ESTIMATIVA DA DENSIDADE DE SOLOS DA**

**AMAZÔNIA ORIENTAL.** Tese (Doutorado em Agronomia) - Universidade Federal Rural da Amazônia, 2019.

R CORE TEAM. **R:** A language and environment for statistical computing. Vienna, Austria, 2021. Disponível em: <<http://cran.r-project.org/>>. Acesso em 19 mar. 2021.

RIBEIRO, E. A. **Imputação de dados faltantes via algoritmo EM e rede neural MLP com o método de estimativa de Máxima Verossimilhança para aumentar a acurácia das estimativas.** Dissertação (Mestrado em Ciência da Computação) – Universidade Federal do Sergipe, 2015.

RUBIN, D. B. Inference and Missing Data. **Biometrika**, v. 63, n. 3, p. 581–590, 1976.

RUBIN, D. B. **Multiple Imputation for Nonresponse in Surveys.** New York: Wiley, 1987.

SILVA, D. S. B. S. et al. IMPUTAÇÃO DE DADOS DIÁRIOS DE RADIAÇÃO SOLAR GLOBAL VIA AMBIENTE R. **Enciclopédia Biosfera**, v. 16, n. 29, p. 960, 2019.

TIERNEY, N. et al. **Naniar: Data Structures, Summaries, and Visualisations for Missing Data.** R package version 0.6.0, 2020.

## CAPÍTULO 2 – MODELAGEM NEURO-FUZZY PARA A PREVISÃO DA EVAPOTRANSPIRAÇÃO DE REFERÊNCIA ESTIMADA PELO MÉTODO DE PENMAN-MONTEITH

### Resumo

O planejamento de sistemas de irrigação eficientes requer o conhecimento da necessidade hídrica da planta, que pode ser estimada através da associação do coeficiente da cultura à evapotranspiração de referência ( $ET_o$ ). Assim, o objetivo deste capítulo é elaborar modelos de previsão da evapotranspiração de referência utilizando um sistema neuro-fuzzy que relacione valores de  $ET_o$  em tempos consecutivos e imediatamente anteriores ao tempo a ser previsto. Para isso foi utilizado séries temporais coletadas das estações meteorológicas de Botucatu – SP e Tupã – SP, no período de 20 de março de 2018 a 19 de março de 2021 para o cálculo da  $ET_o$  diária pelo método Penman-Monteith/FAO 56. Para a elaboração dos modelos de previsão baseado no sistema de inferência adaptativo neuro-fuzzy (ANFIS) foi considerado o período total coletado, além dos períodos correspondentes às estações do ano (outono, inverno, primavera e verão), sendo elaborados com três anos de dados e com um ano (20 de março de 2018 a 20 de março de 2019), a fim de verificar a capacidade de aprendizado do sistema neuro-fuzzy. Os dados foram ordenados do mais antigo ao mais recente, sendo os primeiros 85% dos dados utilizados para treinamento e os demais 15% para validação do modelo. Os resultados evidenciaram que os modelos de previsão de  $ET_o$  utilizando três anos de dados apresentaram menores erros de previsão, tendo em vista o maior número de informações para treinamento da rede, sendo que os modelos elaborados para as estações do ano apresentaram acurácia semelhantes ao modelo geral. Assim, o modelo geral com rede treinada em série temporal de três anos apresentou maior viabilidade em sua utilização. O modelo ajustado para Botucatu – SP caracterizou-se pela utilização de três informações anteriores, produzindo erro de 0,8194 com 47 épocas de treinamento, enquanto o modelo ajustado para Tupã – SP utilizou duas informações anteriores, obtendo erro de 0,990 em 81 épocas de treinamento. O valor da Raiz Quadrada do Erro Quadrático Médio (RMSE) obtido para Botucatu – SP foi de 1,06 enquanto para Tupã – SP foi de 1,30. Portanto, a utilização dos modelos elaborados para ambos os municípios permite a

previsão de valores de  $ET_o$  de maneira confiável, sendo possível a gestão e planejamento de sistemas de irrigação com maior eficiência e sustentabilidade.

**Palavras-chave:** ANFIS, estimativa de  $ET_o$ , séries temporais.

## Abstract

The planning of efficient irrigation systems requires knowledge of the plant's water requirement, which can be estimated by associating the crop coefficient with the reference evapotranspiration ( $ET_o$ ). Thus, the objective of this chapter is to develop reference evapotranspiration prediction models using a neuro-fuzzy system that relates  $ET_o$  values in consecutive times and immediately prior to the time to be predicted. For this, time series collected from the meteorological stations of Botucatu - SP and Tupã - SP, from March 20, 2018 to March 19, 2021, were used to calculate the daily  $ET_o$  by the Penman-Monteith/FAO 56 method. Forecasting models based on the neuro-fuzzy adaptive inference system (ANFIS) was considered the total period collected, in addition to the periods corresponding to the seasons (autumn, winter, spring and summer), being elaborated with three years of data and with one year (March 20, 2018 to March 20, 2019), in order to verify the learning capacity of the neuro-fuzzy system. Data were ordered from oldest to most recent, with the first 85% of the data used for training and the remaining 15% for model validation. The results showed that the  $ET_o$  prediction models using three years of data presented lower prediction errors, given the greater amount of information for training the network, and the models developed for the seasons showed accuracy similar to the general model. Thus, the general model with a three-year time series trained network showed greater feasibility in its use. The adjusted model for Botucatu - SP was characterized by the use of three previous pieces of information, producing an error of 0.8194 with 47 training epochs, while the adjusted model for Tupã - SP used two previous pieces of information, obtaining an error of 0.990 in 81 training epochs. The value of the Square Root of the Mean Square Error (RMSE) obtained for Botucatu – SP was 1.06 while for Tupã – SP it was 1.30. Therefore, the use of models developed for both municipalities allows the prediction of  $ET_o$  values in a reliable way, making it possible to manage and plan irrigation systems with greater efficiency and sustainability.

**Keywords:** ANFIS,  $ET_o$  estimate, time series.

## 2.1 Introdução

A evapotranspiração de referência ( $ET_o$ ) é um dos principais componentes do ciclo hidrológico, sendo que a utilização de valores precisos desta variável é essencial para a estimativa do balanço de água e rendimento das culturas, além do desenho e gestão de sistemas de irrigação (MANIKUMARI; MURUGAPPAN; VINODHINI, 2017). Desta maneira, Allen et al. (1998) destacam que além de fatores climáticos tais como radiação solar, temperatura do ar, déficit de pressão de vapor e velocidade do vento, outros fatores inerentes à cultura também afetam a evapotranspiração, sendo eles: tipo da cultura, densidade, variedade e estágio de crescimento. Isso ocorre, pois os aspectos como altura, cobertura do solo e resistência estomática da cultura influenciam as condições de solo e clima.

Devido a esta diversidade de fatores muitos métodos de estimação foram criados ao longo dos anos, sendo que o método desenvolvido por Penman-Monteith e, parametrizado pela FAO (*Food and Agriculture Organization*), conhecido como Penman-Monteith/FAO 56, foi admitido como padrão internacional para a estimativa de valores de  $ET_o$  devido a concordância dos resultados obtidos com os valores medidos por lisimetria ou balanço de energia (FERRONATO et al., 2016; LOPES et al., 2018). Tal método foi adotado neste trabalho para o cálculo da  $ET_o$ , já que de acordo com Carvalho et al. (2011), além de representar complexo sistema de transferência de água entre solo-planta-atmosfera para os diferentes climas presentes no mundo, utiliza em seu cálculo parâmetros meteorológicos observados em estações meteorológicas de superfície. Allen et al. (1998) salientam que o método se mostra consistente para os diferentes locais e climas, já que considera os parâmetros de trocas de energia e fluxo de calor latente, fundamentais para o processo. Assim, a  $ET_o$  apresenta-se como um processo complexo não-linear de grande importância para a melhoria da gestão de recursos hídricos e produtividade das lavouras, o que tem levado muitos pesquisadores a analisarem suas tendências (GOCIC; TRAJKOVIC, 2013; RIBEIRO; SIMEÃO; SANTOS, 2016).

A análise de tendência e obtenção de previsões requer a utilização de séries temporais, as quais são definidas por pontos de dados ordenados no domínio do tempo que, quando analisados, revelam a estrutura e o padrão dos dados possibilitando a elaboração de modelos de previsão (DEB et al., 2017; VLASENKO et al., 2018). Peng et al. (2015) explica que a previsão de valores futuros de séries

temporais é realizada com base nas observações anteriores de pontos temporais igualmente espaçados.

Desta forma, os modelos autorregressivos são constantemente utilizados para a modelagem de séries temporais, sendo o modelo Auto-Regressivo Integrado de Médias Móveis (ARIMA) um dos mais difundidos (CAMELO et al., 2017). O modelo ARMA (Modelos Autorregressivos de Médias Móveis) é uma combinação do modelo autorregressivo (AR), o qual explica a observação atual em função das  $n$  observações anteriores somada ao erro aleatório ( $p$ ) e o modelo de médias móveis (MA), que combina linearmente o erro da observação atual e das anteriores ( $q$ ). Assim, o modelo ARMA é indicado à series temporais estacionárias e não sazonais (CAMELO et al., 2017; LIMA; CASTRO; CARTAXO, 2019). A modelagem por meio do modelo ARIMA inclui os métodos AR e MA, porém devido a sua aplicação a séries não estacionárias, inclui o componente de integração ( $d$ ) que torna a série estacionária através da diferenciação dos dados. Para a utilização em séries que contenham sazonalidade, adiciona-se os termos sazonais  $(P, D, Q)_m$ , em que os parâmetros se referem aos  $m$  períodos sazonais contidos na série, tendo-se o modelo SARIMA (Modelo Auto-Regressivo Sazonal Integrado de Médias Móveis) (ALBUQUERQUE, 2019; MIGLIATO; PONTI, 2021).

Assim, modelos estatísticos têm sido amplamente utilizados para a realização de previsões, porém sua eficiência se dá em conjuntos de dados estacionários e lineares. Por isso, outro método de destaque são técnicas baseadas em redes neurais artificiais, a qual permite o relacionamento não-linear complexo entre as variáveis dependente e a independente, entre esses métodos estão os modelos neuro-*fuzzy* (CRONE; HIBON; NIKOLOPOULOS, 2011; VLASENKO et al., 2018).

De acordo com Vie, Kleinnijenhuis, Farmer (2020) a modelagem de espaços grandes e complexos é facilitada com a utilização de algoritmos genéticos, os quais podem ser implementados por meio da lógica *fuzzy*. Tal teoria, é capaz de reproduzir parte do raciocínio humano, o qual é composto pela indução e dedução, sendo que o primeiro pode ser desenhado matematicamente pela lógica clássica (bivalente), tendo seus erros transferidos para o raciocínio dedutivo, que tem a lógica *fuzzy* como ferramenta para a quantificação do grau de verdade dedutivo (ATHANASSOPOULOS; VOSKOGLOU, 2020). Isso ocorre devido a teoria introduzida por Lofti Asker Zadeh em 1965, possuir valores entre o “completamente verdadeiro” e “completamente falso”, mapeando o domínio de entrada e de saída.

Assim, a pertinência dos conjuntos *fuzzy* corresponde ao intervalo real contínuo  $[0,1]$ , sendo 0 a não pertinência e 1 a pertinência total, podendo-se acomodar infinitos valores entre os referidos pontos (SILVA et al., 2020).

Desta maneira, a teoria *fuzzy* apresenta resultados satisfatórios na aplicação de fenômenos agrônômicos, como nos estudos de Navarro-Hellín et al. (2016) que desenvolveram um sistema automático de apoio a decisão para a irrigação de plantações a partir de variáveis climáticas e de solo. Nourani e Komasi (2013) propuseram o modelo *fuzzy* para prever o escoamento superficial de bacias hidrográficas. Giusti e Marsili-Libelli (2015) confirmaram a economia de água na irrigação de culturas controladas por um sistema *fuzzy* de acionamento. Já Mamak et al. (2017) que obtiveram resultados superiores para o modelo ANFIS de previsão diária de  $ET_o$  quando comparados aos resultados de previsão obtidos pela fórmula empírica de Penman-Monteith/FAO 56, tendo como variáveis de entrada a umidade relativa, radiação solar, temperatura do ar e velocidade do vento.

Assim, o presente capítulo objetiva elaborar modelos de previsão da evapotranspiração de referência utilizando um sistema neuro-*fuzzy* que relacione os valores de  $ET_o$  de tempos consecutivos e imediatamente anteriores ao tempo da observação a ser prevista.

## **2.2 Material e Métodos**

### **2.2.1 Séries Temporais**

Para o cálculo da  $ET_o$  pela a equação de Penman-Monteith/FAO 56, foram utilizados dados coletados de duas estações meteorológicas. A primeira estação meteorológica está localizada no município de Botucatu, São Paulo, Brasil, nas coordenadas geográficas 22° 50' Latitude Sul e 48° 25' de Longitude Oeste com altitude média de 786 metros, sendo esta estação de responsabilidade do Departamento de Solos e Recursos Ambientais da UNESP, Faculdade Ciências Agrônômicas, Fazenda Experimental Lageado. A segunda estação meteorológica localiza-se no município de Tupã, São Paulo, Brasil, nas coordenadas geográficas 21° 93' Latitude Sul e 50° 49' de Longitude Oeste com altitude média de 498 metros, sendo esta estação automática (Estação Tupã-A768) de responsabilidade do Instituto Nacional de Meteorologia (INMET).

Os dados coletados das estações meteorológicas, necessários para obtenção dos valores de  $ET_o$ , foram radiação solar global, velocidade do vento, temperatura do ar, temperatura do ar máxima, temperatura do ar mínima, umidade relativa do ar, umidade relativa do ar máxima e umidade relativa do ar mínima, sendo necessário a estimativa de outras duas variáveis: saldo de radiação solar e fluxo de calor no solo, além do cálculo dos valores de coeficiente psicrométrico, pressão parcial de vapor, pressão de vapor de saturação e declividade da curva de pressão de vapor de saturação.

O período dos dados coletados compreende o intervalo entre 20 de março de 2018 (início do outono) a 19 de março de 2021 (fim do verão), sendo que os dados disponibilizados pela estação de Botucatu eram organizados de maneira diária (1.096 informações por variável) e a estação de Tupã – SP de maneira horária (26.304 informações de cada variável). Tendo em vista que a avaliação de desempenho do modelo seria mais adequada com a base de dados no mesmo formato, foi necessário a adequação de um dos conjuntos de dados coletados.

Considerando o estudo de Santos (2020) que aponta que a soma dos valores horários de  $ET_o$  possuem boa associação ao valor calculado de forma diária, desde que os dados climáticos disponíveis de forma horária estejam presente em quantidade e qualidade adequadas, e que a base de dados de Tupã – SP possuía 7% de dados ausentes e o de Botucatu 1,4%, foi necessária a imputação de valores conforme estabelecido pelo Capítulo 1, optando-se pela utilização da base de dados no formato diário. Com o valor da  $ET_o$  diária é possível o manejo eficiente da irrigação utilizando, por exemplo, seu funcionamento no horário noturno, proporcionando aumento da produtividade através do uso racional da água e da energia elétrica (ROSA; FIALHO; DOETZER, 2010).

Após a reconstrução do conjunto de dados foi realizado a somatória dos valores de radiação solar global registrados em cada dia, enquanto para as demais variáveis foi realizada a média dos valores, para posterior aplicação da equação de Penman-Monteith/FAO 56 a fim de calcular a  $ET_o$  diária e, a partir dela, prever a seu valor utilizando o sistema de inferência adaptativo neuro-*fuzzy* (ANFIS).

Desta forma, o presente capítulo utilizou duas classes de previsão de  $ET_o$  para cada estação meteorológica, sendo que na primeira a realização da previsão considerou um conjunto de dados com a totalidade dos dados obtidos no período para cada município e na segunda considerou quatro subconjuntos de dados, uma

para cada estação do ano, tendo os períodos definidos conforme (IAG USP, 2021) e evidenciados na Tabela 2.

Tabela 2 – Períodos de início e término dos subconjuntos de dados de cada estação meteorológica.

<b>Conjunto</b>	<b>2018</b>	<b>2019</b>	<b>2020</b>
<b>Outono</b>	20/03/2018 a	20/03/2019 a	20/03/2020 a
	20/06/2018	20/06/2019	19/06/2020
<b>Inverno</b>	21/06/2018 a	21/06/2019 a	20/06/2020 a
	21/09/2018	22/09/2019	21/09/2020
<b>Primavera</b>	22/09/2018 a	23/09/2019 a	22/09/2020 a
	20/12/2018	21/12/2019	20/12/2020
<b>Verão</b>	21/12/2018 a	22/12/2019 a	21/12/2020 a
	19/03/2019	19/03/2020	19/03/2021

Vale ressaltar, que foi realizada uma comparação entre os modelos estimados a partir de três anos de dados com modelos estimados usando apenas o primeiro ano de dados. O objetivo de tal comparação foi verificar se a capacidade de aprendizado do sistema neuro-*fuzzy* é melhor à medida que a série temporal torna-se maior e mais complexa.

Os estudos numéricos e experimentais realizados nesta tese foram desenvolvidos no Laboratório de Matemática Aplicada e Computacional (LabMAC) pertinente à Faculdade de Ciências e Engenharia da UNESP, Campus de Tupã, que possui licença do *software* Matlab®, que foi utilizado para a implementação dos modelos propostos.

### 2.2.2 Cálculo da Evapotranspiração de Referência

O cálculo da  $ET_o$  foi realizado utilizando o método Penman-Monteith/FAO 56, por meio da Equação (4), dada por:

$$ET_o = \frac{0,408 \Delta (R_n - G) + \frac{\gamma 900 u_2 (e_s - e_a)}{T + 273}}{\Delta + \gamma (1 + 0,34 U_2)} \quad (4)$$

sendo que:

- $ET_o$ : Evapotranspiração de Referência ( $\text{mm d}^{-1}$ );
- $R_n$ : saldo de radiação à superfície da cultura ( $\text{MJ m}^{-2} \text{d}^{-1}$ );
- $G$ : densidade do fluxo de calor do solo ( $\text{MJ m}^{-2} \text{d}^{-1}$ );
- $T$ : temperatura do ar a 2 m de altura ( $^{\circ}\text{C}$ );
- $U_2$ : velocidade de vento a 2 m de altura ( $\text{m s}^{-1}$ );
- $e_s$ : pressão de vapor de saturação (kPa);
- $e_a$ : pressão parcial de vapor (kPa);
- $\Delta$ : declividade da curva de pressão de vapor de saturação ( $\text{kPa } ^{\circ}\text{C}^{-1}$ ), e
- $\gamma$ : coeficiente psicrométrico ( $\text{kPa } ^{\circ}\text{C}^{-1}$ ).

Devido as estações meteorológicas utilizadas neste estudo não disponibilizarem as informações de densidade do fluxo de calor do solo ( $G$ ) e saldo de radiação global ( $R_n$ ), foi adotado o valor igual a zero para o fluxo de calor, já que de acordo com Allen et al. (1998), tal prática pode ser considerada, pois os valores que normalmente são registrados diariamente são baixos. Para os demais procedimentos analíticos (Equações 5 a(17) utilizados no cálculo da evapotranspiração utilizou-se a metodologia posta por EMBRAPA (2006). Assim, o saldo de radiação global ( $R_n$ ) é calculado por meio da Equação (5).

$$R_n = R_{ns} + R_{nl} \quad (5)$$

sendo:

$R_{ns}$ : saldo de radiação de ondas curtas ( $\text{MJ m}^{-2}\text{dia}^{-1}$ )

$R_{nl}$ : saldo de radiação de ondas longas ( $\text{MJ m}^{-2}\text{dia}^{-1}$ ).

Para o cálculo do saldo de radiação de ondas curtas utiliza-se a Equação (6), onde  $R_s$  é a radiação solar incidente ( $\text{MJ m}^{-2}\text{dia}^{-1}$ ), valor registrado pelas estações em questão, e  $\alpha$  é o coeficiente de reflexão da vegetação, assumindo valor 0,23 para a cultura de referência (grama).

$$R_{ns} = (1 - \alpha) R_s \quad (6)$$

O saldo de radiação de ondas longas é calculado por meio da Equação (7).

$$R_{nl} = -\sigma \left[ \frac{(T_{max} + 273,16)^4 + (T_{min} + 273,16)^4}{2} \right] (0,34 - 0,14\sqrt{e_a}) \left(1,35 \frac{R_s}{R_{so}} - 0,35\right) \quad (7)$$

em que:

- $\sigma$ : é a constante de Stefan-Boltzmann ( $4,90310 \cdot 10^{-9}$  MJ m<sup>-2</sup>dia<sup>-1</sup>);
- $T_{max}$  e  $T_{min}$ : temperatura máxima e mínima do dia, respectivamente (°C);
- $R_s$ : radiação solar incidente (MJ m<sup>-2</sup>dia<sup>-1</sup>);
- $R_{so}$ : radiação solar incidente na ausência de nuvens (MJ m<sup>-2</sup>dia<sup>-1</sup>);
- $e_a$ : pressão atual de vapor (kPa).

sendo que a pressão atual de vapor é calculada por meio da Equação (8) em que UR é a umidade relativa média do ar (%), informação fornecida por ambas as estações meteorológicas.

$$e_a = \frac{e_s UR}{100} \quad (8)$$

A pressão de saturação de vapor ( $e_s$ ) é obtida por meio pela Equação (9).

$$e_s = 0,6108 \exp\left(\frac{17,27 T}{T + 237,3}\right) \quad (9)$$

Já a radiação solar incidente na ausência de nuvens ( $R_{so}$ ) pode ser obtida com a Equação (10).

$$R_{so} = (0,75 + 2 \cdot 10^{-5} z) R_a \quad (10)$$

onde  $z$  é a altitude (m) e  $R_a$  refere-se à radiação solar no topo da atmosfera (MJ m<sup>-2</sup>dia<sup>-1</sup>), calculada pela Equação (11).

$$Ra = \frac{118,08}{\pi} dr[\omega_s \text{sen}(\varphi) \text{sen}(\delta) + \text{cos}(\varphi) \text{cos}(\delta) \text{sen}(\omega_s)] \quad (11)$$

sendo o valor de  $\varphi$  definido pela latitude do local em radianos. Para a conversão é necessário multiplicar a latitude (em graus) por  $\pi$  (3,14159) e dividir por 180. Após esse cálculo é aplicado a equação da distância inversa relativa entre a Terra e o Sol (rad), dada pela Equação (12).

$$dr = 1 + 0,033 \cos\left(\frac{2\pi}{365}J\right) \quad (12)$$

onde  $J$  corresponde o dia do ano, variando entre 1 e 365 ou 366 para anos bissextos.

A Equação (13) define o valor da declinação solar  $\delta$  (rad).

$$\delta = 0,409 \text{sen}\left(\frac{2\pi}{365}J - 1,39\right) \quad (13)$$

Já a Equação (14) fornece o ângulo horário ao nascer do sol  $\omega$  (rad).

$$\omega = \frac{\pi}{2} - \arctan\left(\frac{-\tan(\varphi) \tan(\delta)}{X^{0,5}}\right) \quad (14)$$

onde  $X = (1 - [\tan(\varphi)]^2 [\tan(\delta)]^2)$ , sendo que para  $X \leq 0$  assume-se valor igual a 0,00001.

Além dessas variáveis, ainda se faz necessário o cálculo dos valores de  $\Delta$ ,  $\gamma$ ,  $e_s$  e  $e_a$ . Assim, para o cálculo de  $\Delta$  utiliza-se a Equação (15).

$$\Delta = \frac{4098 \left[0,6108 \exp\left(\frac{17,27 T}{T+237,3}\right)\right]}{(T + 237,3)^2} \quad (15)$$

onde  $\exp$  expressa à base do logaritmo natural  $e$  (2,71828) elevada ao valor entre parêntesis.

O valor de  $\gamma$  pode ser definido através da Equação (16).

$$\gamma = 0,665 \cdot 10^{-3} P_{atm} \quad (16)$$

sendo  $P_{atm}$  a pressão atmosférica local (kPa), obtida com base na altitude do local  $z$  (m) através da Equação (17).

$$P_{atm} = 101,3 \left( \frac{293 - 0,0065 z}{293} \right)^{5,26} \quad (17)$$

### 2.2.3 Método de elaboração dos sistemas neuro-fuzzy

O modelo matemático proposto utilizou o sistema de inferência adaptativo neuro-fuzzy ANFIS (*Adaptive Neuro Fuzzy Inference System*), sendo que os dados obtidos foram ordenados de acordo com a data de coleta, do mais antigo para o mais recente, onde os 85% primeiros foram selecionados para treinamento e os demais 15% para a verificação do desempenho (validação) do modelo, assim como fizeram Khan e Khan (2019).

O sistema de inferência trata-se de uma modelagem de processos complexos, não lineares e vagos. Assim, o sistema neuro-fuzzy se concretiza com a utilização de redes neurais e do método de inferência de Takagi-Sugeno, sendo definido por uma série de regras do tipo:

$$\text{“se } x \in A \text{ e } y \in B, \text{ então } z = f(x, y)\text{”}$$

em que  $x$  e  $y$  são variáveis fuzzy,  $A$  e  $B$  são conjuntos fuzzy e  $f$  é uma combinação linear das variáveis de entradas e uma constante.

Vale ressaltar, que as variáveis fuzzy são os elementos envolvidos no fenômeno estudado, no caso deste estudo temperatura, umidade, velocidade do vento, saldo de radiação global e evapotranspiração de referência.

Desta forma, um sistema *fuzzy* que possua uma saída ( $z$ ) e duas entradas ( $x$  e  $y$ ) com 2 conjuntos *fuzzy* associado à cada uma delas, terá uma base de regras com quatro regras *fuzzy*, conforme Figura 9, sendo as regras estabelecidas como:

“Regra 1: Se  $x \in A_1$  e  $y \in B_1$ , então  $f_1 = p_1x + q_1y + r_1$ ”;

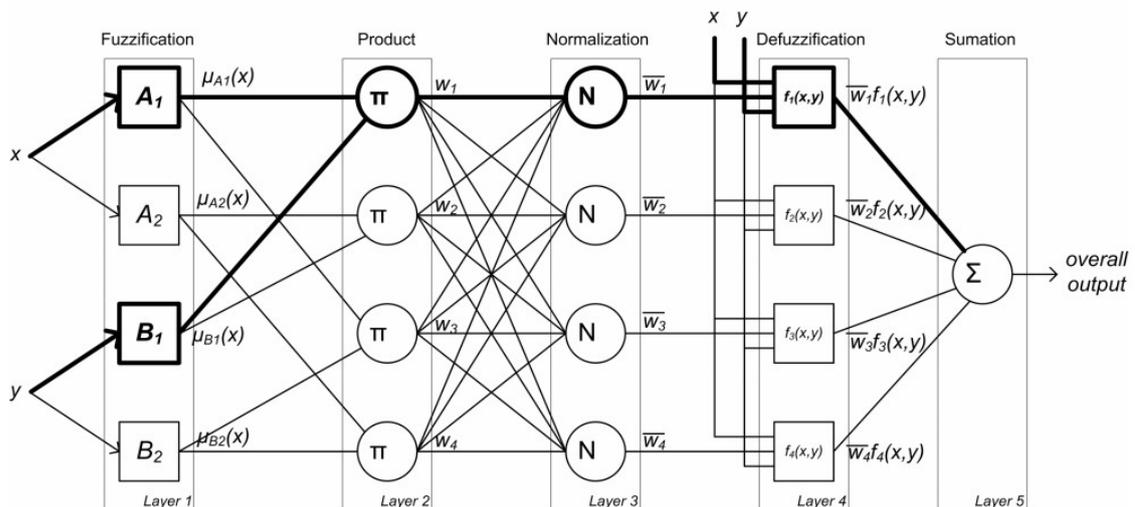
“Regra 2: Se  $x \in A_2$  e  $y \in B_1$ , então  $f_2 = p_2x + q_2y + r_2$ ”;

“Regra 3: Se  $x \in A_1$  e  $y \in B_2$ , então  $f_3 = p_3x + q_3y + r_3$ ”;

“Regra 4: Se  $x \in A_2$  e  $y \in B_2$ , então  $f_4 = p_4x + q_4y + r_4$ ”;

onde  $p$ ,  $q$  e  $r$  são os parâmetros consequentes da regra.

Figura 9 - Arquitetura de um modelo ANFIS.



Fonte: DUKA (2015)

Verifica-se que a arquitetura de um sistema ANFIS é composta por cinco camadas, descritas a seguir:

- a) **Camada 1:** nesta camada ocorre a *fuzzyficação*, onde as variáveis de entrada são associadas a um conjunto *fuzzy* com grau de pertinência, sendo a saída desta camada, que pode ser representada por:

$$O_i^j = \mu_i^j(x_i)$$

onde,  $i$  é o número de variáveis de entrada,  $j$  é o número de regras de combinação e  $\mu$  o grau de pertinência associada ao conjunto *fuzzy*. Esta camada é adaptativa, uma vez que os parâmetros antecedentes são ajustados durante o processo de treinamento do sistema (MATHUR; GLESK; BUIS, 2016).

- b) Camada 2:** esta é uma camada não adaptativa, onde os nós implementam o conectivo  $e$  da regra, obtendo uma saída com peso  $w_i$ , que representa a força de disparo da regra, ou seja, o grau de pertinência de atendimento ao consequente da regra (KESHTEGAR et al., 2018a). Assim, a força de disparo da primeira regra da arquitetura apresentada é descrita por  $w_1 = \mu_{A_1}(x) \cdot \mu_{B_1}(y)$ . A implementação do conectivo é feita por meio do produto algébrico ou operação de T-norma e o operador mínimo, onde tem-se:

$$O_j^2 = w_j = \prod_{i=1}^2 \mu_i^i = \mu_1^i x \mu_2^i = \min(\mu_1^i; \mu_2^i); \quad j = 1,2,3,4.$$

- c) Camada 3:** Nesta camada ocorre a normalização e, portanto, os nós não são adaptativos (KESHTEGAR et al., 2018a). Assim, calcula-se a relação do peso de cada regra e a soma dos pesos das demais regras, já que cada nó desta camada recebe entradas da camada anterior, podendo ser definida por:

$$O_j^3 = \bar{w}_j = \frac{w_j}{\sum_{i=1}^4 w^i} = \frac{w_j}{w_1 + w_2 + w_3 + w_4}; \quad j = 1,2,3,4.$$

- d) Camada 4:** A camada da *defuzzyficação* apresenta nós adaptativos (que se adaptam durante o processo de aprendizado), que calculam o valor do consequente ponderado de cada regra do modelo, sendo que as saídas são calculadas pelo produto entre os níveis, podendo ser representadas por:

$$\bar{w}_j \cdot f_j(x, y) = (p_j x_1 + q_j x_2 + r_j); \quad j = 1,2,3,4$$

em que,  $\bar{w}_j$  representa o peso normalizado da regra  $j$ ,  $f_j(x, y)$  é a função de saída da camada 3 e  $p_j, q_j, r_j$  é o conjunto de parâmetros do consequente linear de cada regra (MATHUR; GLESK; BUIS, 2016).

- e) **Camada 5:** Esta camada é formada por um único nó, não adaptativo, que produz a saída global do sistema ANFIS, somando todos os sinais de entrada:

$$O_j^4 = f = \sum_{j=1}^2 f_j.$$

Para o ajuste automático dos parâmetros, o método ANFIS utiliza duas técnicas de aprendizado: o *backpropagation* (gradiente descendente) e o híbrido, que combina o *backpropagation* com o erro de mínimos quadrados (AL-DUNAINAWI; ABBOD; JIZANY, 2017). Neste estudo, foi utilizado o método de aprendizado *backpropagation*, que de acordo com Duka (2015), ajusta os parâmetros das funções de associação (antecedentes) a fim de minimizar o erro do modelo.

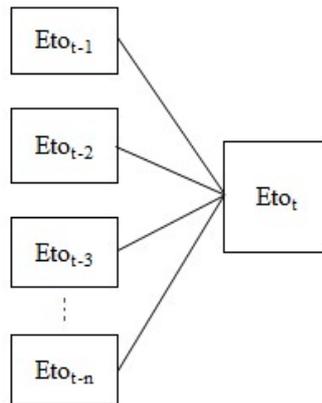
O modelo desenvolvido relaciona valores de  $ET_o$  em tempos consecutivos e imediatamente anteriores ao valor a ser previsto, descrito da forma  $y_t = F(y_{t-1}, y_{t-2}, y_{t-3}, \dots, y_{t-n})$ , no qual  $y_t$  é o valor de  $ET_o$  ( $y$ ) no tempo ( $t$ ) (tempo atual na qual realizar-se-á a previsão), e  $F$  é uma função não-linear que representa o modelo neuro-*fuzzy* criado a partir dos dados de treinamento.

Para a modelagem matemática de previsão da  $ET_o$  procederam-se simulações do modelo neuro-*fuzzy* ANFIS considerando a (1) variação da quantidade de variáveis de entrada de 2 a 4 informações/lags (tempos de atraso do tipo  $t - 1, t - 2, t - 3$  e  $t - 4$  onde  $t$  corresponde à  $ET_o$  no tempo  $t$ ); (2) variação da quantidade de funções de pertinência entre 2 e 3; e (3) variação da quantidade de épocas de treinamento entre 1 e 100 épocas, a partir do *Fuzzy Logic Toolbox* pertencente ao software MATLAB®.

Verifica-se por meio da Figura 10 que o sistema possui  $m$  variáveis de entrada, que são as informações de  $ET_o$  calculadas através da equação de Penman-Monteith/FAO 56, sendo utilizadas as informações  $t - 1, t - 2, t - 3, \dots, t - n$  onde  $t$  indica a  $ET_o$  no tempo  $t$ , que são combinados para fornecer a variável de saída, que refere-se à evapotranspiração prevista. Portanto, o modelo realiza a previsão da

evapotranspiração a partir da combinação de dados da própria  $ET_o$  que foram calculados previamente.

Figura 10 - Sistema neuro-fuzzy de previsão da evapotranspiração de referência utilizando a evapotranspiração calculada ( $ET_o$ ).



#### 2.2.4 Avaliação e validação dos modelos

Para a avaliação da tendência do modelo em superestimar ou subestimar a evapotranspiração de referência em relação à série histórica utilizada, foram utilizadas as seguintes medidas estatísticas de erro:

##### a) Erro absoluto médio percentual

Segundo Bayer e Belli (2016) mede a precisão de uma previsão por meio da porcentagem de seu erro, dada pela Equação (18):

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_{observado} - y_{fuzzy}|}{y_{observado}} * 100 \quad (18)$$

##### b) Raiz Quadrada do Erro Médio Quadrático

De acordo com Hallak e Pereira Filho (2011), expressa a acurácia dos resultados numéricos expressos na mesma dimensão da variável analisada, obtida pela Equação (19):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{observado} - y_{fuzzy})^2} \quad (19)$$

### c) U de Theil

Indica a qualidade das predições, calculado pela Equação (20), sendo que resultados mais próximos a zero indicam que o erro de previsão é menor que método ingênuo (Naive).

$$U = \frac{\sqrt{\sum_{i=1}^{n-1} (y_{fuzzy} + 1 - y_{observado} + 1)^2}}{\sqrt{\sum_{i=1}^{n-1} (y_{observado} + 1 - y_{observado})^2}} \quad (20)$$

onde:

- $y_{fuzzy}$  e  $y_{observado}$ : respectivamente, valor calculado pelo modelo *fuzzy* e valor observado
- $n$ : quantidade de pares de entrada-saída do vetor usado no treinamento.

Desta forma, dentre as combinações do modelo ANFIS simuladas, aquele com o menor valor no erro estatístico U de Theil será considerado o de melhor desempenho. Estes modelos são comparados aos modelos Naive, por meio dos erros estatísticos, para a verificação da eficiência nas previsões de  $ET_o$  para as séries geral e estações do ano. Assim, o modelo que apresentar o menor erro de previsão será considerado o mais eficiente. Vale ressaltar, que o modelo Naive pressupõe que o valor posterior é igual ao valor anterior, servindo como referência para comparações com a eficácia de outros modelos (WERNER; BISOGNIN; MENEGOTTO, 2017; LYNCH; GORE, 2021).

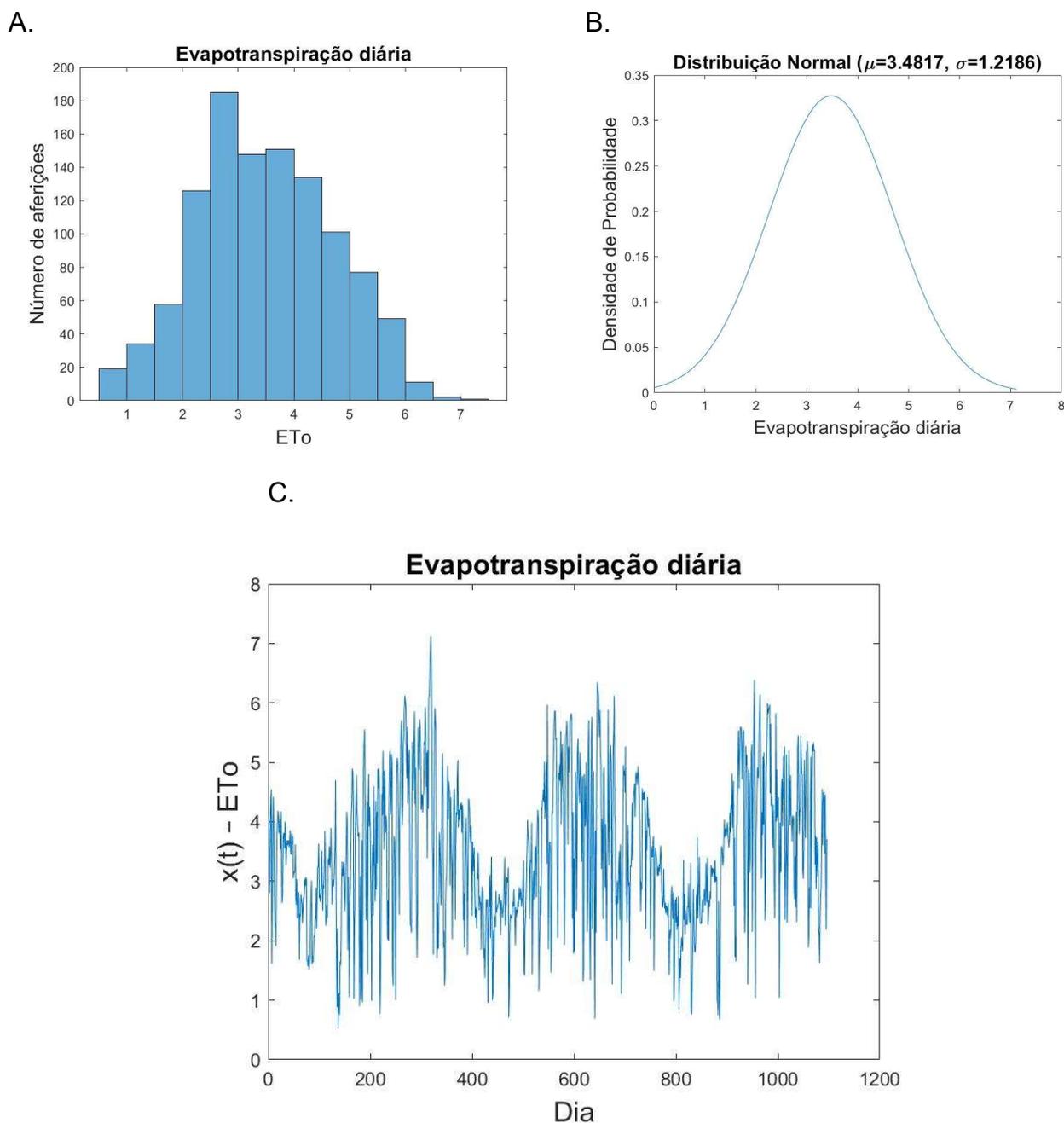
## 2.3 Resultados e Discussão

### 2.3.1 Serie temporal e modelagem ANFIS para o município de Botucatu – SP

A evapotranspiração de referência calculada por meio do método de Penman-Monteith/FAO 56 para o período de 3 anos possui variação mostrada na Figura 11, onde verifica-se a natureza estocástica da  $ET_o$  com oscilação entre 0,5 e 7 mm/dia<sup>-1</sup>, tendo maior frequência no intervalo entre 2,5 e 3 mm/dia<sup>-1</sup>, com 180 observações e menor frequência com valor acima de 7 mm/dia<sup>-1</sup> (Figura 11A).

Percebe-se ainda a sazonalidade da série histórica, visto que os valores de  $ET_o$  possuem tendência de queda nos períodos do outono (do dia 1 ao 93, 366 ao 458 e 732 ao 823) motivada pela chegada do inverno, que possui valores de temperatura e radiação solar global menores que as demais estações, representando os vales da série temporal. A tendência de elevação dos valores de  $ET_o$  ocorre nos períodos da primavera (do dia 187 a 276, 553 a 642 e 918 a 1007), haja visto a elevação da temperatura e radiação solar global causada pelo verão, que representa os picos da série (Figura 11C).

Figura 11 – (A) Evapotranspiração de referência calculada pelo método de Penman-Monteith/FAO 56, (B) distribuição de frequências dos valores da série temporal e (C) densidade da probabilidade de obtenção da evapotranspiração de referência diária para Botucatu – SP.



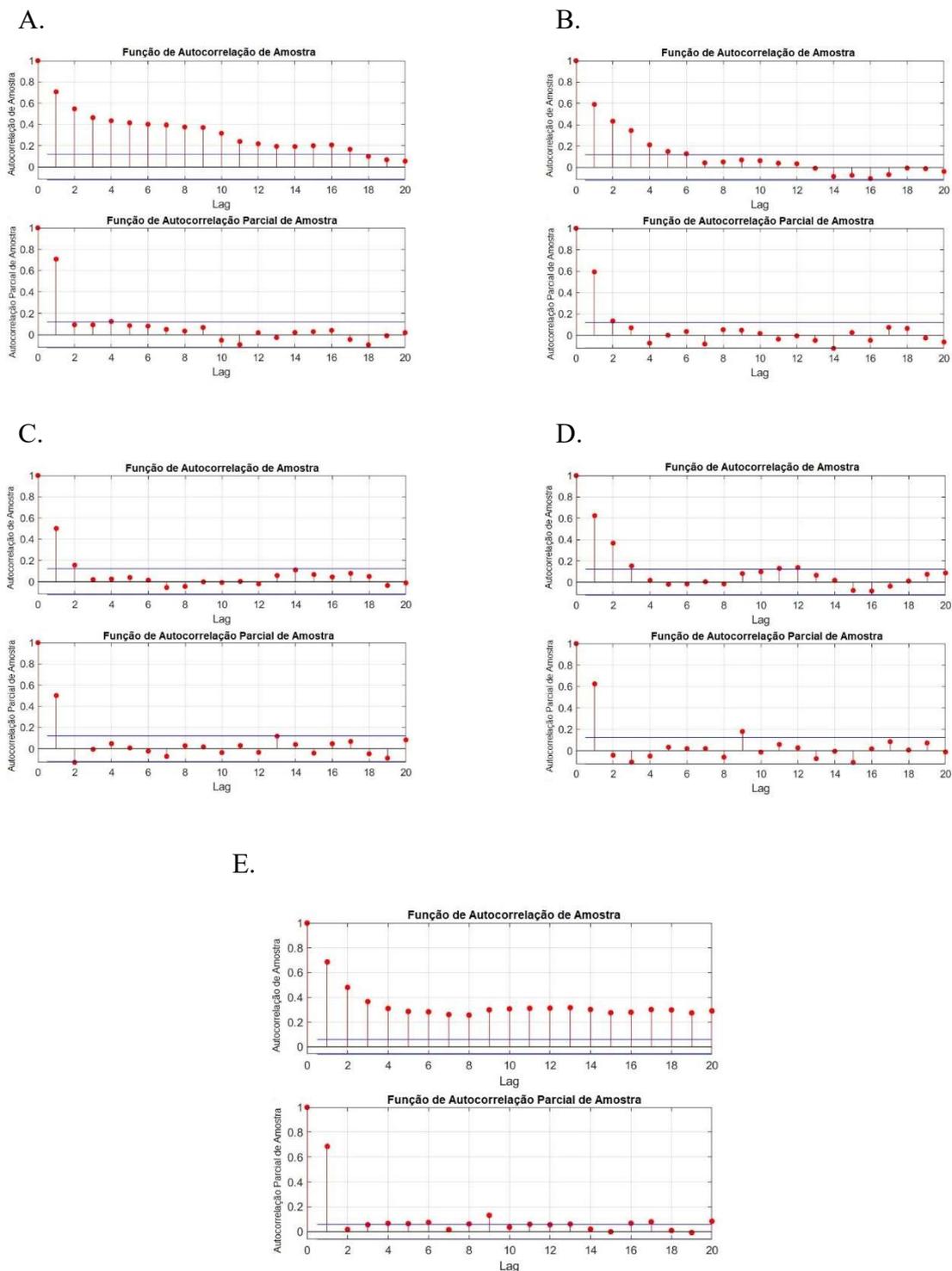
O gráfico de distribuição de frequências indica que o centro da distribuição está em torno de  $3,5 \text{ mm/dia}^{-1}$ , sugerindo uma distribuição normal dos dados (Figura 11B) com média de  $3,48 \text{ mm/dia}^{-1}$  e desvio padrão de  $1,21 \text{ mm/dia}^{-1}$ . Além disso, tem-se que a probabilidade dos valores de  $ET_o$  estarem entre  $1,06$  e  $5,9 \text{ mm/dia}^{-1}$  é de  $95,44\%$ .

A referida série temporal de valores de evapotranspiração de referência é utilizada para a elaboração dos modelos ANFIS geral, outono, inverno, primavera e verão, sendo que a função de autocorrelação foi utilizada para verificar a estrutura de dependências das observações e ajudar a definir o intervalo de um dos parâmetros a ser simulado (quantidade de variáveis de entrada).

De acordo com Nazaré, Castro e Gabriel Filho (2020) a função de autocorrelação permite compreender o tipo do modelo e quais os valores anteriores que possuem informações relevantes para a previsão dos futuros valores. Assim, cada *lag* representa o número de períodos de tempo que separa os dados de séries temporais, sendo que no gráfico da função de autocorrelação parcial a quantidade de *lags* representada é definida por  $n/4$  (sendo  $n$  o número de observações), conforme recomendação de Box e Jenkins (1976).

Desta forma, a Figura 12 evidencia, por meio da função de autocorrelação, que se trata de um modelo autorregressivo, uma vez que possui decaimento exponencial das correlações. Ao observar a função de auto correlação parcial verifica-se que se refere a um modelo autorregressivo de ordem 1 ( $AR_{(1)}$ ) para as séries do outono, verão e geral; e de ordem 2 ( $AR_{(2)}$ ) para as séries do inverno e primavera, uma vez que apenas o *lag* 1 e *lag* 2, respectivamente, apresentam correlação significativa para a previsão do próximo valor de  $ET_o$ .

Figura 12 – Funções de autocorrelação e autocorrelação parcial das séries temporais das estações (A) outono, (B) inverno, (C) primavera, (D) verão e (E) geral do município de Botucatu – SP.



A partir da análise das funções de autocorrelação e autocorrelação parcial é possível concluir que é possível realizar a previsão dos valores de evapotranspiração de referência, já que a observação atual é explicada em função de informações anteriores e do erro aleatório. Desta forma, os modelos neuro-fuzzy

foram elaborados considerando a variação da quantidade de *lags* (variáveis de entrada), entre outros fatores, que explicariam a informação a ser prevista.

Considerando a variação dos parâmetros testados para os modelos neuro-*fuzzy*, a combinação de variáveis de entrada, funções de pertinência e épocas de treinamento que obteve o menor valor do erro estatístico U de Theil, ou seja, aquele que apresentou maior eficiência na previsão dos valores, foi apresentada por modelos que possuem as características definidas conforme Tabela 3.

Tabela 3 – Definição dos parâmetros dos modelos neuro-*fuzzy* de menor erro estatístico U de Theil para o município de Botucatu – SP.

Modelo	Variáveis de entrada ( <i>lags</i> )		Funções de pertinência		Épocas de Treinamento		Regras	
	1 ano	3 anos	1 ano	3 anos	1 ano	3 anos	1 ano	3 anos
Outono	2	3	2	3	1	99	4	27
Inverno	2	2	2	2	1	63	4	4
Primavera	3	2	2	2	1	66	8	4
Verão	2	2	2	2	93	1	4	4
Geral	2	3	2	2	67	47	4	8

Os referidos modelos foram avaliados considerando os erros U, RMSE e MAPE (Tabela 4), sendo que a análise dessas medidas estatísticas mostra que a previsão dos valores de  $ET_o$  calculados pelo modelo ANFIS possuem maior precisão do que o modelo Naive (previsão ingênua), visto que os valores das estatísticas de erro testadas são menores para o modelo ANFIS, evidenciando baixo erro de previsão dos valores de  $ET_o$ .

Pesquisas de Keshtegar et al. (2018b) apresentaram valores de RMSE semelhantes a este estudo, sendo que também obtiveram superioridade do modelo ANFIS em comparação com os demais métodos testados (Rede Neural Artificial e Modelo de Árvore M5).

Tabela 4 – Erros U, RMSE e MAPE para os modelos Naive e ANFIS para o município de Botucatu – SP.

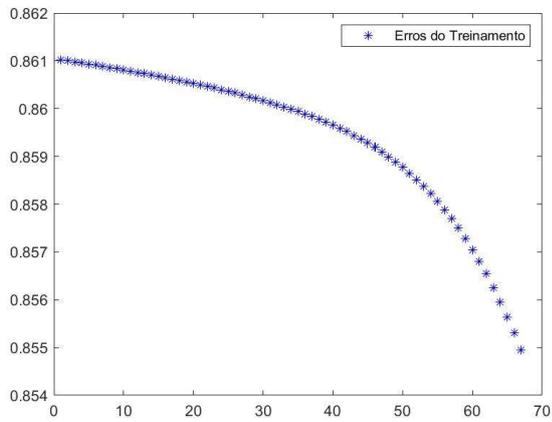
Estação	Modelo	U		RMSE		MAPE	
		1 ano	3 anos	1 ano	3 anos	1 ano	3 anos
Outono	Naive	1	1	0,48	0,63	0,30	0,46
	ANFIS	0,85	0,88	0,41	0,56	0,16	0,25
Inverno	Naive	1	1	1,35	0,73	0,92	0,46
	ANFIS	1,05	0,97	1,42	0,70	0,65	0,27
Primavera	Naive	1	1	0,64	1,14	0,49	0,90
	ANFIS	2,01	0,92	1,17	1,05	0,18	0,25
Verão	Naive	1	1	0,91	1,03	0,71	0,72
	ANFIS	0,95	0,91	0,86	0,93	0,23	0,24
Geral	Naive	1	1	1,08	1,17	0,86	0,88
	ANFIS	0,91	0,90	0,98	1,06	0,25	0,25

Além do melhor desempenho do modelo ANFIS frente ao Naive, verifica-se que os modelos elaborados utilizando três anos de dados apresentam menores valores de RMSE quando comparados com os modelos treinados com série temporal de um ano de dados, com exceção do modelo do outono, verão e geral. Apesar destes três modelos apresentarem maior valor do erro médio quadrático, o erro absoluto de sua previsão é inferior a 1%, o que mostra que o aprendizado da rede é melhor quando a série temporal possui mais informações.

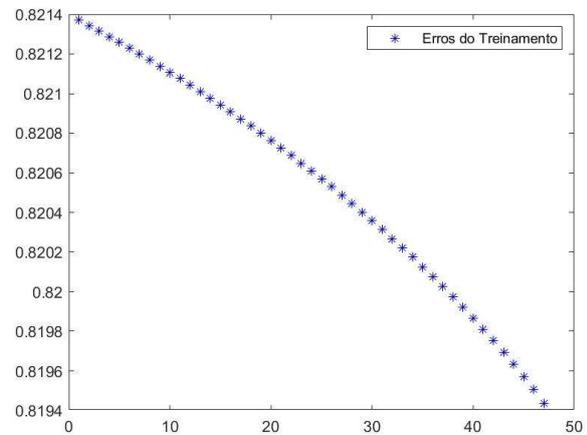
Quando se observa o erro de treinamento do modelo geral na combinação das variáveis de entrada e obtenção da previsão da  $ET_o$ , a superioridade do modelo que utiliza três anos de dados é reforçada, já que o menor erro de treinamento é de 0,8194 obtido na 47ª época de treinamento (Figura 13B), enquanto que com a série temporal reduzida, o menor erro foi de 0,855 obtido com 67 épocas de treinamento (Figura 13B). Assim, o aprendizado da rede, que ajusta os parâmetros nos nós das estruturas do modelo a fim de obter proximidade entre saídas e entradas, foi mais longo no primeiro ano devido a menor quantidade de dados da série (SPACCA, 2019).

Figura 13 – Erros de treinamento do modelo ANFIS Geral de Botucatu – SP para (A) um ano e (B) três anos de dados.

A.

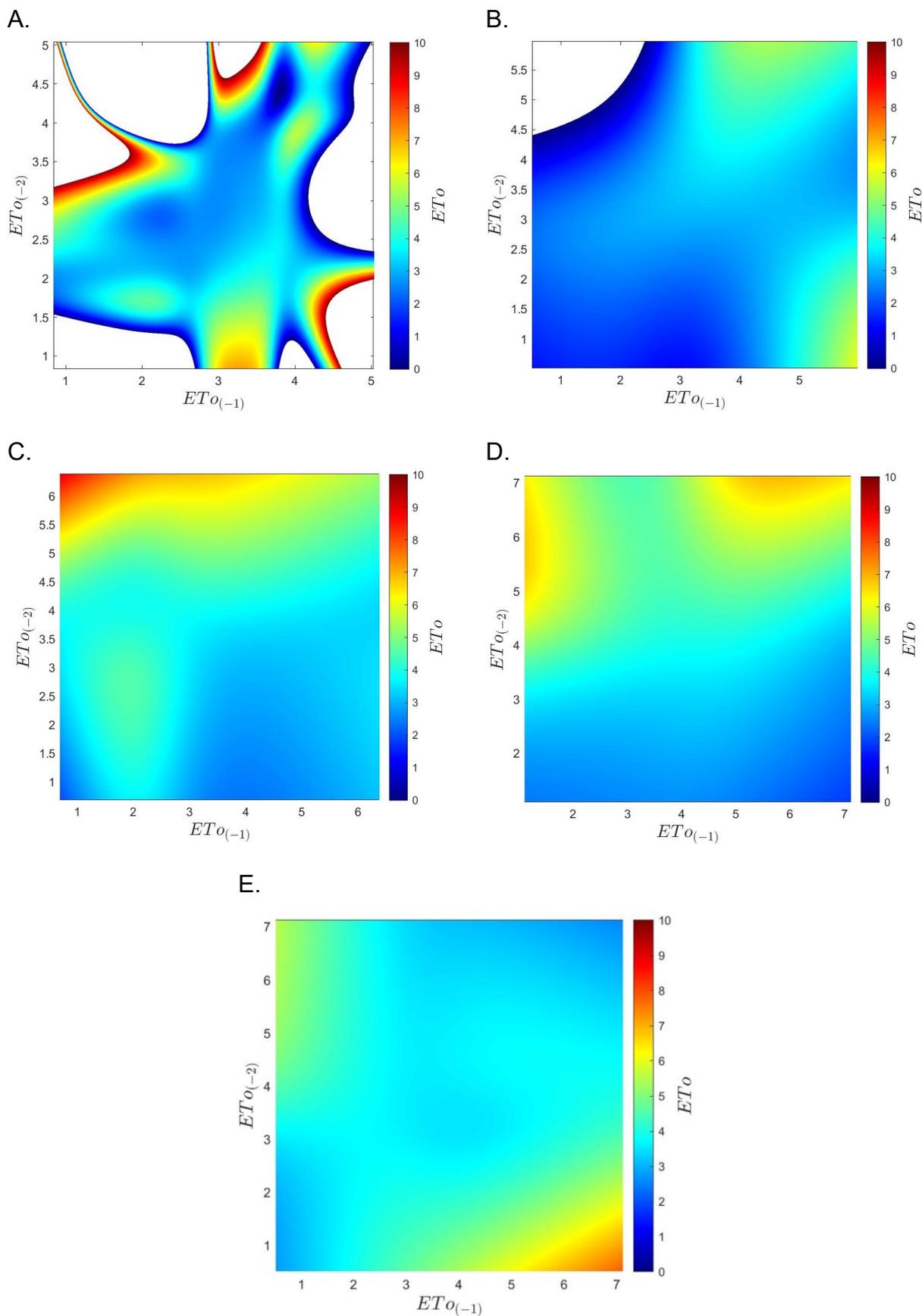


B.



Considerando que os modelos ANFIS que utilizam a série temporal com três anos de dados apresenta maior eficiência na previsão dos valores de  $ET_o$ , a Figura 14 apresenta os gráficos de contorno obtidos com os melhores modelos de cada estação do ano e o modelo geral com variação do *lag 1* ( $ET_{o(t-1)}$ ) e *lag 2* ( $ET_{o(t-2)}$ ). As demais variáveis de entrada, quando houverem, são fixadas automaticamente no valor médio da variável.

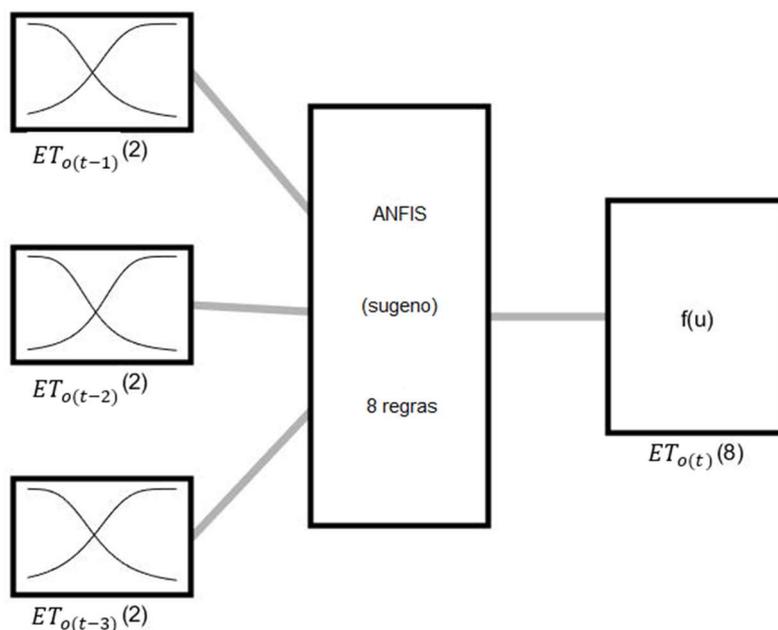
Figura 14 – Gráficos de contorno utilizando série temporal de três anos das estações (A) Outono, (B) Inverno, (C) Primavera, (D) Verão e (E) Geral para o município de Botucatu – SP.



Verifica-se que as Figura 14A e Figura 14B apresentam regiões sem o mapeamento dos valores de  $ET_o$  (regiões brancas), o que significa que não há previsão para estes casos de combinação, pois a série temporal em que a rede foi treinada não apresentava a ocorrência de tais casos, apesar da combinação gráfica ser possível. Para as Figura 14C, Figura 14D e Figura 14E o mapeamento está completo, evidenciando o treinamento da rede em todas as possibilidades gráficas. Desta forma, é possível verificar o valor da evapotranspiração de referência previsto pelos modelos elaborados, a partir da combinação dos valores de um e dois *lags* anteriores ao valor de interesse. Assim, as regiões preenchidas com cores frias (azul) apresentam os valores baixos de  $ET_o$  e as regiões de cores quentes (vermelho) os valores mais altos, sendo possível identificar a mudança gradativa nos valores previstos, conforme o estabelecido na Teoria dos Conjuntos *Fuzzy*.

Visando a praticidade na utilização de um único modelo que seja eficiente para todo o ano e considerando que as medidas estatísticas de erro são semelhantes para os modelos das estações do ano e a autocorrelação entre as variáveis da série temporal apresentou maior dependência para a série completa (geral), o modelo geral com três anos de dados torna-se mais interessante para ser usado para a previsão da  $ET_o$ , cuja arquitetura é mostrada na Figura 15.

Figura 15 – Arquitetura do modelo ANFIS geral utilizando série temporal de três anos para o município de Botucatu – SP.

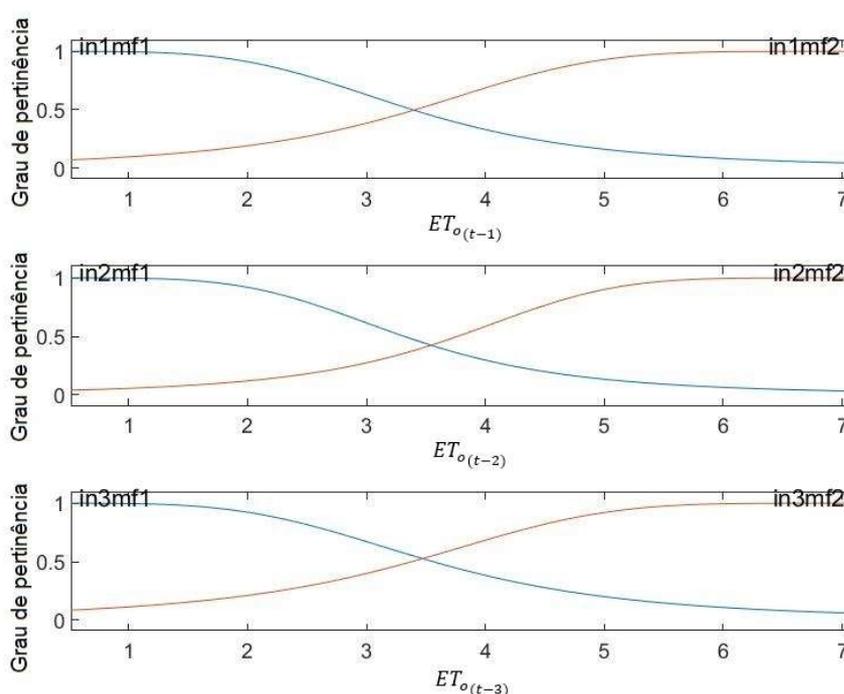


Tal arquitetura é composta por três variáveis de entrada ( $ET_{o(t-1)}$ ,  $ET_{o(t-2)}$  e  $ET_{o(t-3)}$ ), cuja as informações correspondem aos respectivos valores anteriores ao valor previsto na série temporal. Cada uma das variáveis de entrada possui duas funções de pertinência, ou conjuntos *fuzzy*, para a acomodação dos valores, sendo estas do tipo sino generalizada, que são definidas pela Equação 21.

$$gbellmf_{(x;a,b,c)} = \frac{1}{1 + \left| \frac{x-c}{a} \right|^{2b}} \quad (21)$$

A função de pertinência do tipo sino generalizada, representada na Figura 16, apresenta simetria em torno de um número real definido pelo parâmetro  $c$  da equação, o parâmetro  $a$  determina a largura da função de pertinência e a transição entre a pertinência total e a não pertinência é definida pelo parâmetro  $b$ . A análise da Figura 16 mostra que o grau de pertinência de 0,5 ocorre para valores de  $ET_o$  de, aproximadamente,  $3,5 \text{ mm/dia}^{-1}$  para as três variáveis de entrada, demonstrando semelhança no comportamento das funções e transição suave entre os dois conjuntos *fuzzy*.

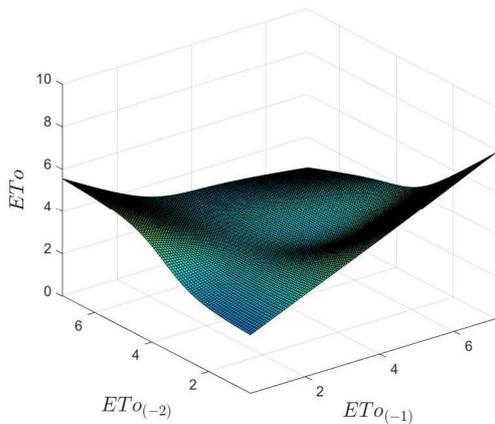
Figura 16 – Funções de pertinência das variáveis de entrada  $ET_{o(t-1)}$ ,  $ET_{o(t-2)}$  e  $ET_{o(t-3)}$  do modelo ANFIS de Botucatu – SP.



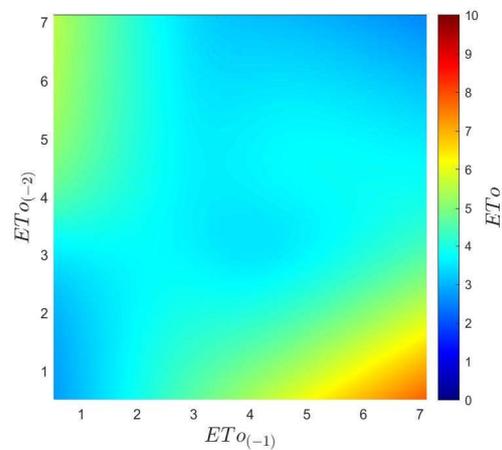
A quantidade de regras do sistema neuro-*fuzzy* são definidas pelo total de combinações possíveis entre as funções de pertinência e variáveis de entrada. Desta forma, o modelo geral de Botucatu – SP apresentou oito regras de combinação ( $2^3$ ). A partir dessas combinações são obtidos os resultados de previsão do valor de  $ET_o$ , que podem ser vistos por meio do gráfico de superfície (Figura 17A) e mapa de contorno (Figura 17B).

Figura 17 – Gráficos de previsão de  $ET_o$  a partir da variação da  $ET_o(t-1)$  e  $ET_o(t-2)$ , com  $ET_o(t-3)$  fixa para Botucatu – SP.

A.



B.



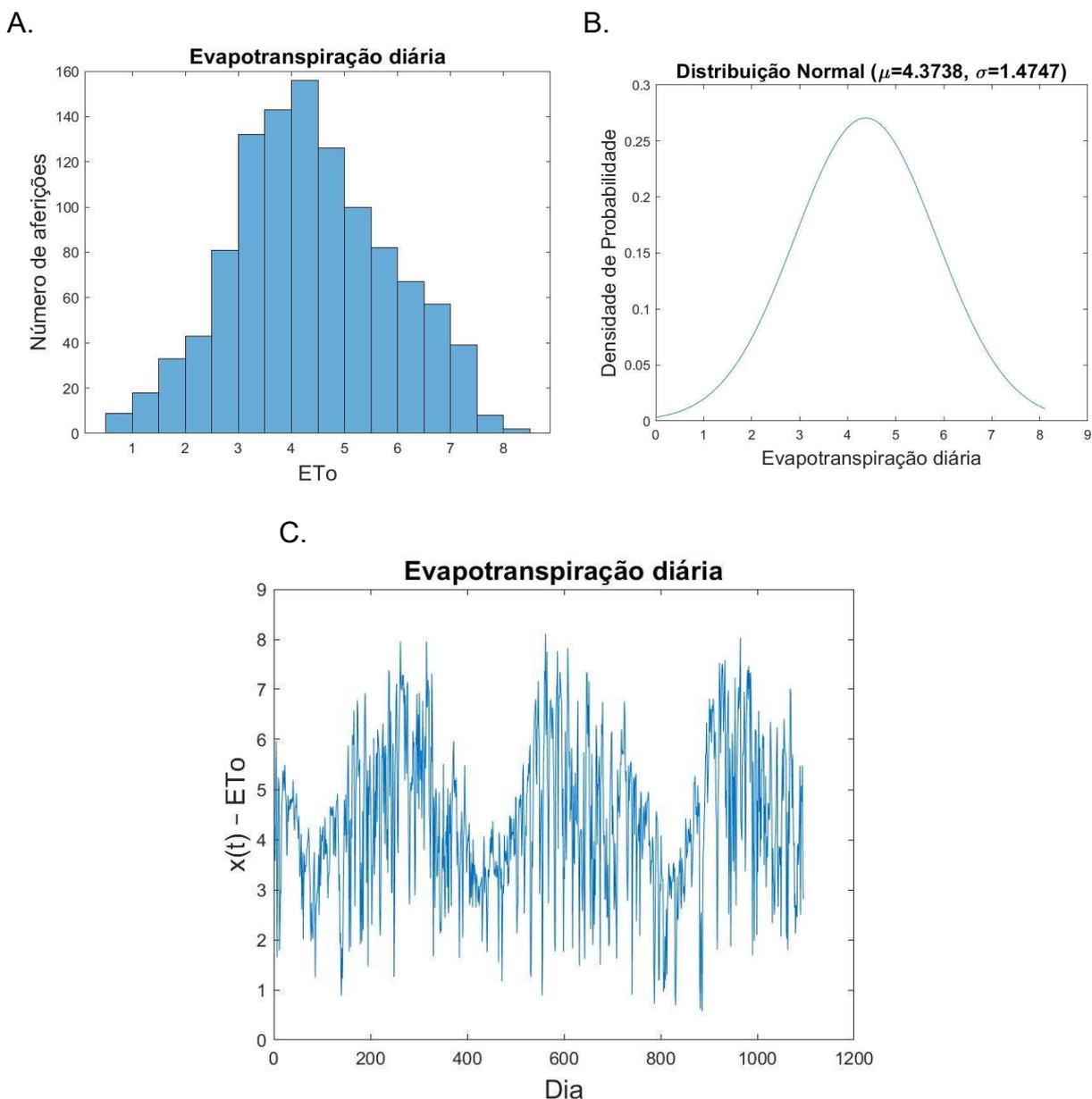
Os gráficos mostram o comportamento do valor de  $ET_o$  quando o valor  $ET_o(t-3)$  está fixo na média da variável e as demais variáveis de entrada assumem  $n$  valores definidos em seus eixos. Desta forma, tem-se que a maior frequência observada de  $ET_o$  está entre 3,5 e 4,5 mm/dia<sup>-1</sup> (como a distribuição de frequências da série temporal), obtida quando a  $ET_o$  do dia anterior (*lag* 1) está entre 0 e 3 mm/dia<sup>-1</sup> e *lag* 2 entre 0 e 3,5 mm/dia<sup>-1</sup> ou quando o *lag* 1 apresenta  $ET_o$  de 4 mm/dia<sup>-1</sup> com *lag* 2 acima de 3 mm/dia<sup>-1</sup> (região azul).

### 2.3.2 Série temporal e modelagem ANFIS para o município de Tupã – SP

A série temporal de evapotranspiração de referência calculada através do método de Penman-Monteith/FAO 56 para o município de Tupã – SP, também apresenta natureza estocástica com variação da  $ET_o$  entre 0,5 e 8 mm/dia<sup>-1</sup> (Figura 18), mostrando-se superiores aos valores observados no município de Botucatu –

SP, porém com comportamento semelhante de tendência, com queda nos períodos do outono e crescimento naqueles que representam os dias de primavera (Figura 18C).

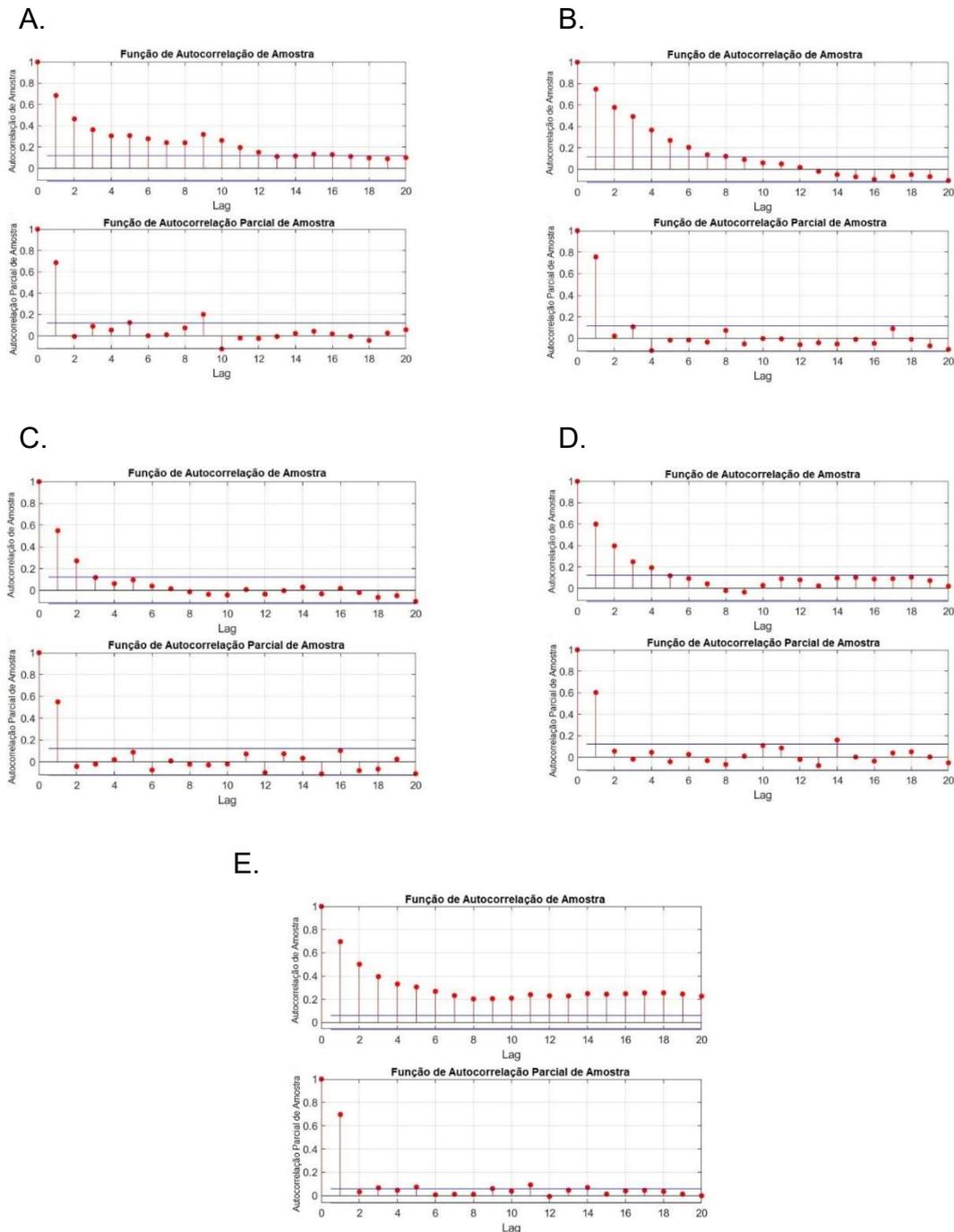
Figura 18 – (A) Evapotranspiração de referência calculada pelo método de Penman-Monteith/FAO 56, (B) distribuição de frequências dos valores da série temporal e (C) densidade da probabilidade de obtenção da evapotranspiração de referência diária para Tupã – SP.



O gráfico de distribuição de frequências mostra que a maior ocorrência de  $ET_0$  está entre 4 e 4,5 mm/dia<sup>-1</sup> (Figura 18A), com distribuição aproximadamente normal (Figura 18B) com média de 4,37 mm/dia<sup>-1</sup> e desvio padrão de 1,47 mm/dia<sup>-1</sup>, indicando que 95,44% dos valores de  $ET_0$  estão entre 1,43 e 7,31 mm/dia<sup>-1</sup>.

Com a série temporal de valores de evapotranspiração de referência foi elaborado os modelos neuro-*fuzzy* geral, outono, inverno, primavera e verão para o município de Tupã – SP, sendo que a função de autocorrelação mostra que refere-se a um modelo autorregressivo e a função de autocorrelação parcial define que há significância estatística apenas para 1 *lag*, tratando-se de um modelo autorregressivo de ordem 1 ( $AR_{(1)}$ ), conforme resultados apresentados na Figura 19.

Figura 19 – Funções de autocorrelação e autocorrelação parcial das séries temporais das estações (A) outono, (B) inverno, (C) primavera, (D) verão e (E) geral do município de Tupã – SP.



Considerando as funções de autocorrelação e autocorrelação parcial verifica-se a viabilidade da elaboração de modelos neuro-*fuzzy* para a previsão dos valores de  $ET_o$ , uma vez que o valor previsto pode ser explicado pela informação anterior a ele. Assim, as simulações do modelo neuro-*fuzzy* ANFIS para Tupã – SP que obtiveram o menor valor no erro estatístico U de Theil são caracterizados na Tabela 5, onde verifica-se que o número de épocas de treinamento utilizada para obtenção dos melhores resultados variou entre os modelos, evidenciando que a rede neural necessita de quantidade de treinamento diferente para cada conjunto de dados, sendo um parâmetro bastante particular de cada série temporal trabalhada.

Tabela 5 – Definição dos parâmetros dos modelos neuro-*fuzzy* de menor erro estatístico U de Theil do município de Tupã – SP.

Modelo	Variáveis de entrada ( <i>lags</i> )		Funções de pertinência		Épocas de Treinamento		Regras	
	1 ano	3 anos	1 ano	3 anos	1 ano	3 anos	1 ano	3 anos
Outono	2	2	2	2	100	1	4	4
Inverno	2	2	2	3	1	1	4	9
Primavera	4	2	3	2	100	81	81	4
Verão	2	2	2	2	96	1	4	4
Geral	2	2	2	2	59	81	4	4

Considerando os resultados da qualidade da predição mostrada pelo indicador U (Tabela 6), os modelos do inverno e primavera com um ano de dados e o outono e inverno com três anos de dados possuem desempenho inferior ao modelo de previsão ingênua (Naive), ou seja, inferir que o valor futuro é igual ao valor passado é mais explicativo do que considerar a incerteza na atribuição dos valores da série temporal. Porém, o indicador MAPE do modelo ANFIS das referidas estações apresenta-se inferior ao modelo Naive, indicando que a porcentagem do erro de previsão é menor quando se considera a incerteza (ANFIS).

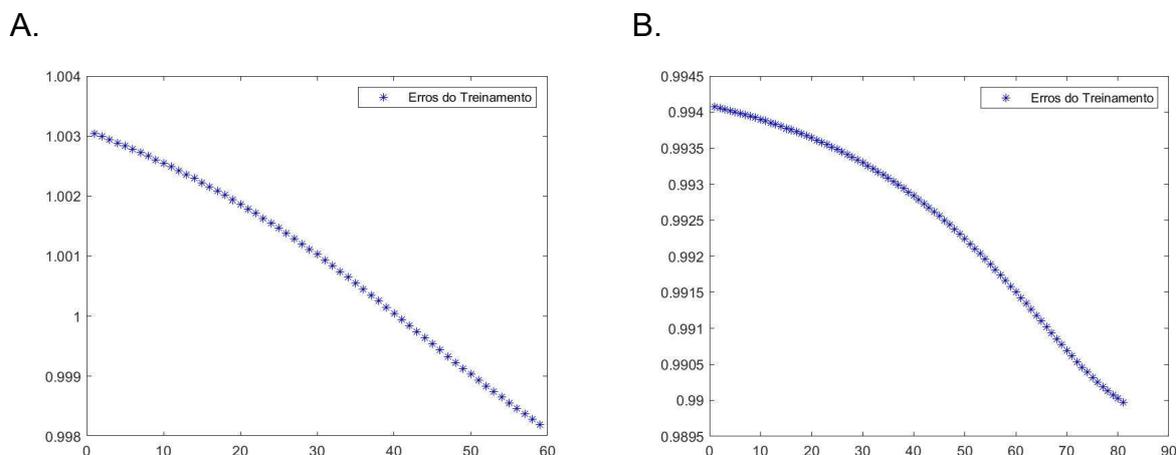
Tabela 6 – Erros U, RMSE e MAPE para os modelos Naive e ANFIS para o município de Tupã – SP.

Estação	Modelo	U		RMSE		MAPE	
		1 ano	3 anos	1 ano	3 anos	1 ano	3 anos
Outono	Naive	1	1	0,73	0,87	0,46	0,65
	ANFIS	0,96	1,07	0,69	0,93	0,26	0,45
Inverno	Naive	1	1	1,43	0,91	0,96	0,64
	ANFIS	1,15	1,06	1,64	0,95	0,45	0,26
Primavera	Naive	1	1	1,04	1,70	0,75	1,26
	ANFIS	1,22	0,84	1,42	1,43	0,18	0,30
Verão	Naive	1	1	1,57	1,28	1,20	0,94
	ANFIS	0,85	0,90	1,33	1,15	0,34	0,28
Geral	Naive	1	1	1,38	1,46	1,07	1,09
	ANFIS	0,90	0,89	1,24	1,30	0,28	0,26

Resultados de RMSE semelhantes a estes foram obtidos por Roy et al. (2020) quando testaram o modelo ANFIS para a previsão de  $ET_o$ , sendo que quando comparados os resultados obtidos por modelos que utilizam algoritmos de otimização de ajuste apresentam resultados inferiores, demonstrando boa acurácia na previsão de valores. Além disso, a verificação do erro absoluto (MAPE) mostra que os modelos que utilizam a série temporal completa (três anos) apresentaram valores inferiores ao modelo treinado com um ano de dados, com exceção do modelo obtido para o outono e primavera. Isso ocorre devido à ambos serem estações de transição, ocorrendo maior instabilidade atmosférica, principalmente na primavera (FERREIRA, 2016). Apesar disso, sua previsão apresenta erro absoluto inferior a 1%, evidenciando que uma série temporal maior e mais complexa, contribui para o aprendizado da rede.

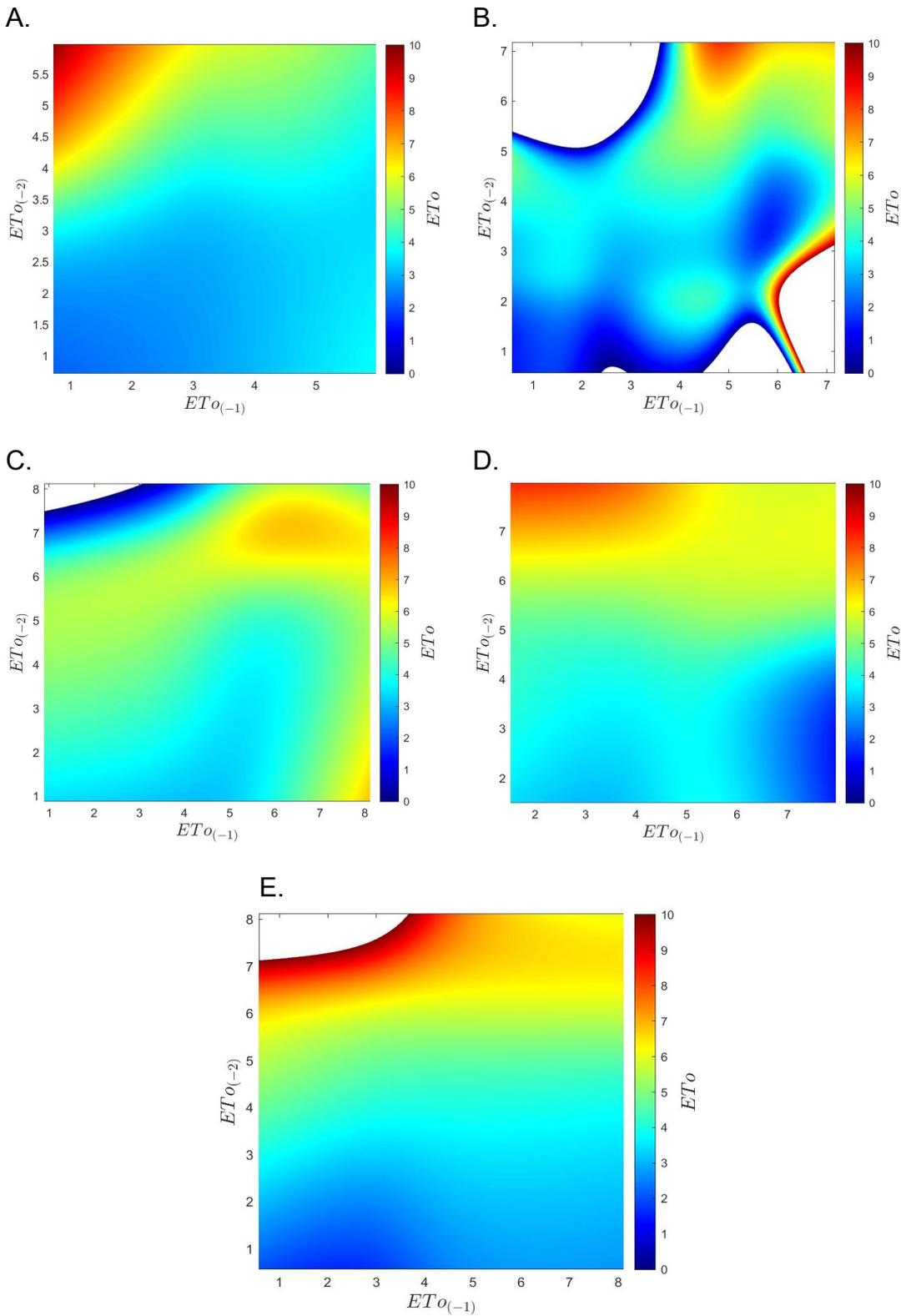
Os resultados de treinamento, estabelecidos na Figura 20A, mostram que o modelo ANFIS com série temporal de um ano apresentou na primeira época de treinamento erro de 1,003 com queda ao longo das 59 interações chegando a 0,998, aproximadamente. Já o modelo que utilizou três anos de dados (Figura 20B) necessitou de 81 épocas de treinamento para que o erro reduzisse de 0,994 para 0,990. Percebe-se que, em ambos os modelos a queda no erro de aprendizado foi pequena, porém devido ao maior volume de dados disponível para o treinamento da rede, o modelo ANFIS com três anos mostra maior acurácia na previsão.

Figura 20 – Erros de treinamento do modelo ANFIS Geral de Tupã – SP para (A) um ano e (B) três anos de dados.



Desta forma, considerando o melhor desempenho na previsão da evapotranspiração de referência dos modelos ANFIS que utilizam a série temporal com três anos de dados, a Figura 21 apresenta os gráficos de contorno das estações e do modelo geral, evidenciando o resultado da  $ET_o$  a partir da combinação dos valores dos *lags* 1 e 2, já que os melhores modelos obtidos para o município de Tupã – SP apresentam duas variáveis de entrada ( $ET_{o(t-1)}$  e  $ET_{o(t-2)}$ ). Vale ressaltar, que o preenchimento com cores frias (azul) apresentam os menores valores de  $ET_o$  e com cores quentes (vermelho) os maiores valores, havendo a mudança gradativa de cores, que representam a progressiva mudança de pertencimento dos valores aos conjuntos *fuzzy*.

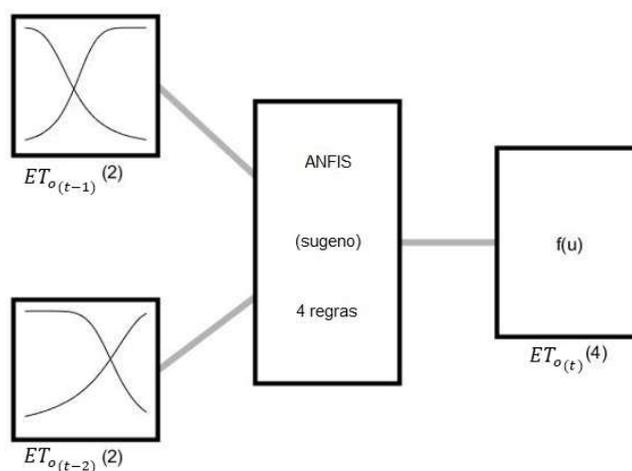
Figura 21 – Gráficos de contorno utilizando série temporal de três anos das estações (A) Outono, (B) Inverno, (C) Primavera, (D) Verão e (E) Geral para o município de Tupã – SP.



A partir da análise gráfica, verifica-se que os modelos inverno (Figura 21B), primavera (Figura 21C) e geral (Figura 21E) possuem regiões sem o mapeamento

dos valores de  $ET_o$  (regiões brancas), evidenciando que, apesar da combinação gráfica ser possível, a série temporal de evapotranspiração de referência não apresentou casos de combinação daqueles valores. Apesar dos modelos do outono (Figura 21A) e verão (Figura 21D) apresentarem mapeamento completo, a viabilidade de utilização do modelo geral para três anos é maior, já que se tem apenas um mapeamento de previsão, aplicável para todo o ano. Além disso, a combinação dos valores que são apresentados no outono e verão estão contidos no modelo geral. Além da viabilidade na utilização do modelo ANFIS treinado na série temporal de três anos de dados, o erro estatístico MAPE apresenta o menor valor quando comparado com os modelos das estações do ano, evidenciando que sua previsão de  $ET_o$  para o município de Tupã – SP é eficiente. Assim, a arquitetura de tal modelo é mostrada na Figura 22.

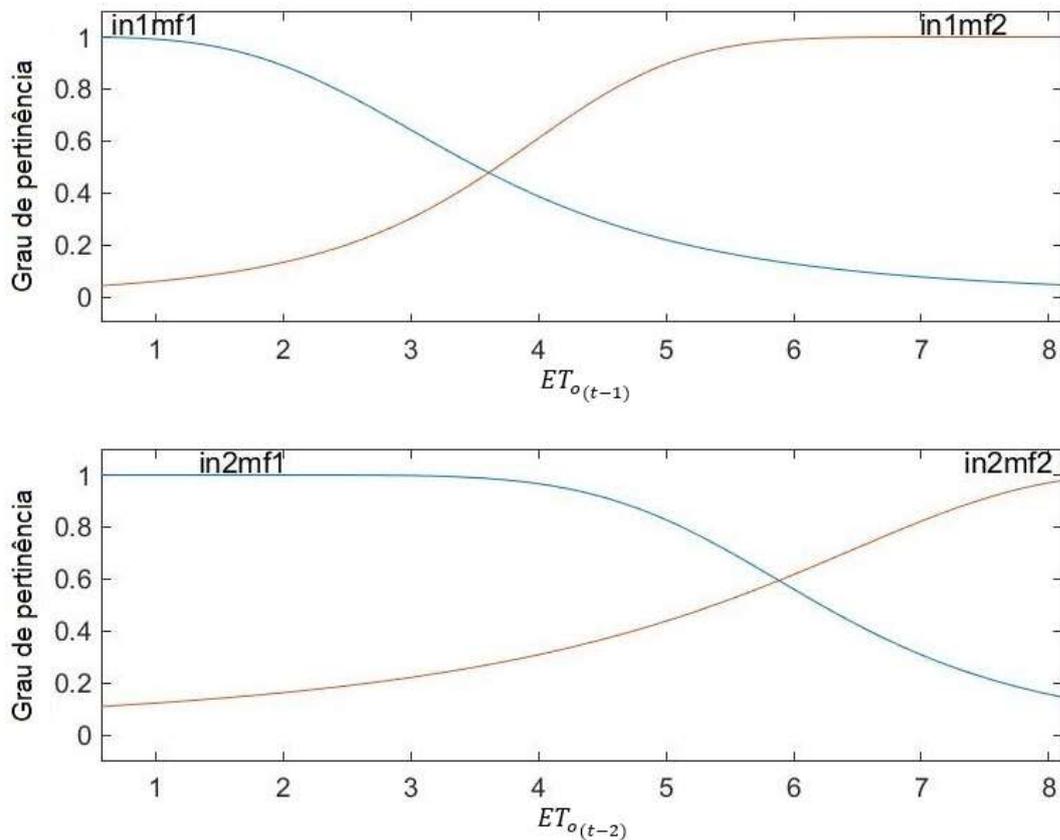
Figura 22 – Arquitetura do modelo ANFIS geral utilizando série temporal de três anos para o município de Tupã – SP.



O modelo neuro-*fuzzy* geral, é composto por duas variáveis de entrada ( $ET_{o(t-1)}$  e  $ET_{o(t-2)}$ ), ou seja, duas informações anteriores à informação a ser prevista na série temporal, sendo que cada uma das variáveis de entrada possui duas funções de pertinência do tipo sino generalizada, definidas pela Equação 21 e representada graficamente na Figura 23. Considerando a largura da função de pertinência, verifica-se que a função de pertinência *in1mf2* da variável  $ET_{o(t-1)}$  e a *in2mf1* da variável  $ET_{o(t-2)}$  são mais amplas, indicando um maior valor no parâmetro  $a$ . Além disso, verifica-se que o grau de pertinência de 0,5 ocorre para valores de  $ET_o$  de,

aproximadamente,  $3,5 \text{ mm/dia}^{-1}$  para a variável  $ET_{o(t-1)}$  e  $6 \text{ mm/dia}^{-1}$  para a variável  $ET_{o(t-2)}$ , demonstrando a série temporal apresenta valores mais altos.

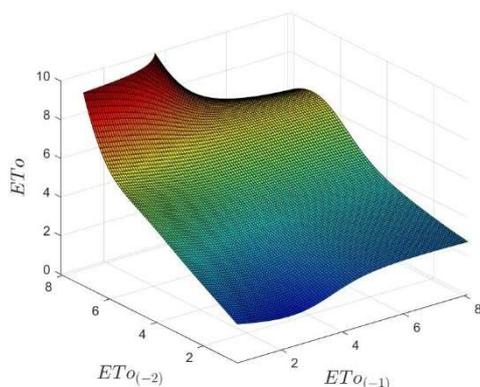
Figura 23 – Funções de pertinência das variáveis de entrada  $ET_{o(t-1)}$  e  $ET_{o(t-2)}$  do modelo ANFIS de Tupã – SP.



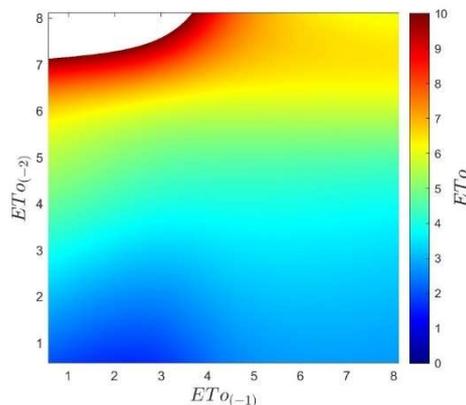
Considerando que as duas funções de pertinência das duas variáveis de entrada são combinadas para a obtenção das regras do sistema neuro-*fuzzy*, tem-se um sistema com 4 regras ( $2^2$ ). A partir dessas de regras a rede é treinada e são obtidos os resultados de previsão do valor de  $ET_o$ , com o menor erro associado, cujos resultados podem ser vistos no gráfico de superfície (Figura 24A) e contorno (Figura 24B).

Figura 24 – Gráficos de previsão de  $ET_o$  a partir da variação da  $ET_{o(t-1)}$  e  $ET_{o(t-2)}$ , para o município de Tupã – SP.

A.



B.



Assim, a combinação das variáveis  $ET_{o(t-1)}$  e  $ET_{o(t-2)}$  para o município de Tupã – SP não apresentam resultados quando a primeira (*lag* 1) apresenta valor acima de  $7 \text{ mm/dia}^{-1}$  e a segunda (*lag* 2) abaixo de  $4 \text{ mm/dia}^{-1}$ . Isso ocorre pois, apesar de graficamente a combinação ser possível, a série temporal não apresentou tal combinação, já que na prática tais valores com ocorrência consecutiva serão improváveis de ocorrer. Comumente, a  $ET_o$  apresenta valores de, aproximadamente,  $4 \text{ mm/dia}^{-1}$  (região azul claro) combinando valores de  $ET_o$  do dia anterior (*lag* 1) entre 0 e  $4 \text{ mm/dia}^{-1}$  com *lag* 2 de 2 a  $4 \text{ mm/dia}^{-1}$  ou *lag* 1 entre 4 e  $8 \text{ mm/dia}^{-1}$  e *lag* 2 entre 3 e  $5 \text{ mm/dia}^{-1}$ .

## 2.4 Conclusões

A partir da composição das séries temporais de evapotranspiração de referência utilizando o método de Penman-Monteith/FAO 56 foi possível realizar as simulações do modelo neuro-fuzzy ANFIS que possuíssem arquitetura com o menor erro de treinamento associado. Assim, os modelos que utilizaram dados relativos às estações do ano (outono, inverno, primavera e verão) apresentaram resultados das medidas estatísticas de erro semelhantes ao modelo geral. Portanto, considerou-se o modelo geral, treinado a partir de série temporal com três anos de dados, mais vantajoso de ser utilizado, já que por contemplar maior série temporal para o treinamento da rede e incluir a sazonalidade existente no período, apresentou boa acurácia, além de ser um único instrumento de previsão aplicável a todo o período.

O modelo de Botucatu – SP caracterizou-se pela utilização de três observações anteriores à observação prevista enquanto o modelo de Tupã – SP utilizou duas, sendo que ambos os modelos empregaram duas funções de pertinência para a obtenção dos parâmetros das regras de combinação que produziram o menor erro associado, sendo que o modelo de Botucatu – SP estabeleceu erro de 0,8194 na 47° época e o de Tupã – SP apresentou erro de 0,990 na 81° época.

Desta forma, a utilização dos gráficos de superfície e contorno permite perceber com facilidade o provável valor de  $ET_o$  quando se combina valores obtidos nos dias anteriores ao dia que se quer prever. Assim, torna-se uma ferramenta importante para o planejamento de safra, uma vez que o conhecimento prévio dos valores de evapotranspiração de referência de uma determinada região permitem maior eficiência no dimensionamento de sistemas de irrigação e programação de seu acionamento, aumentando potencialmente a margem de lucro do produtor e uso racional de recursos naturais para a produção de alimentos.

## Referências

- AL-DUNAINAWI, Y.; ABBOD, M. F.; JIZANY, A. A New MIMO ANFIS-PSO Based NARMA-L2 Controller for Nonlinear Dynamic Systems. **Engineering Applications of Artificial Intelligence**, v. 62, p. 265–275, 2017.
- ALBUQUERQUE, R. C. DE. **MODELAGEM EM SÉRIES TEMPORAIS: APLICAÇÃO EM DADOS DE PRECIPITAÇÃO NA REGIÃO DO SERTÃO DE PERNAMBUCO-BRASIL**. Dissertação (Mestrado em Biometria e Estatística Aplicada). Universidade Federal Rural de Pernambuco, 2019.
- ALLEN, R. et al. **Crop evapotranspiration** – Guidelines for computing crop water requirements. Rome: FAO, 1998.
- ATHANASSOPOULOS, E.; VOSKOGLU, M. G. A Philosophical Treatise on the Connection of Scientific Reasoning with Fuzzy Logic. **Mathematics**, v. 8, n. 6, p. 875–890, 2020.
- BAYER, S. L.; BELLI, F. Proposta de melhoria no processo de previsão de vendas, com foco no indicador MAPE – um estudo de caso. **FFBusiness**, v. 14, n. 19, 2016.
- BOX, G. E. P.; JENKINS, G. M. **Time series analysis: forecasting and control**. San Francisco, Holden-Day, 1976.
- CAMELO, H. DO N. et al. Métodos de Previsão de Séries Temporais e Modelagem Híbrida ambos Aplicados em Médias Mensais de Velocidade do Vento para Regiões do Nordeste do Brasil. **Revista Brasileira de Meteorologia**, v. 32, n. 4, p. 565–574, 2017.
- CARVALHO, L. G. DE et al. EVAPOTRANSPIRAÇÃO DE REFERÊNCIA: UMA ABORDAGEM ATUAL DE DIFERENTES MÉTODOS DE ESTIMATIVA. **Pesquisa Agropecuária Tropical**, v. 41, n. 3, p. 456–465, 2011.
- CRONE, S. F.; HIBON, M.; NIKOLOPOULOS, K. Advances in forecasting with neural networks? Empirical evidence from the NN3 competition on time series prediction. **International Journal of Forecasting**, v. 27, n. 3, p. 635–660, 2011.
- DEB, C. et al. A review on time series forecasting techniques for building energy consumption. **Renewable and Sustainable Energy Reviews**, v. 74, p. 902–924, 2017.
- DUKA, A. V. ANFIS Based Solution to the Inverse Kinematics of a 3 DOF Planar Manipulator. **Procedia Technology**, v. 19, p. 526–533, 2015.
- EMBRAPA. **Circular Técnica 65: Roteiro de cálculo da evapotranspiração de referência pelo método de Penman- Monteith-FAO**. Bento Gonçalves: EMBRAPA, 2006.
- FERREIRA, R. C. Tendência dos índices de instabilidade na América do Sul. **Ciência e Natura**, v. 38, n.1, p. 164–169, 2016.
- FERRONATO, A. et al. Métodos de estimativa da evapotranspiração de referência para Santo Antônio do Leverger-MT. **Revista de Ciências Agroambientais**, v. 14, n. 1, p. 110–118, 29 ago. 2016.
- GIUSTI, E.; MARSILI-LIBELLI, S. A Fuzzy Decision Support System for irrigation and water conservation in agriculture. **Environmental Modelling and Software**, v. 63, p.

73–86, 1 jan. 2015.

GOCIC, M.; TRAJKOVIC, S. Analysis of trends in reference evapotranspiration data in a humid climate. **Hydrological Sciences Journal**, v. 59, n. 1, p. 165–180, 2013.

HALLAK, R.; PEREIRA FILHO, A. J. Metodologia para análise de desempenho de simulações de sistemas convectivos na região metropolitana de São Paulo com o modelo ARPS: sensibilidade a variações com os esquemas de advecção e assimilação de dados. **Revista Brasileira de Meteorologia**, v. 26, n. 4, p. 591–608, dez. 2011.

IAG USP - INSTITUTO DE ASTRONOMIA GEOFÍSICA E CIÊNCIAS ATMOSFÉRICAS DA UNIVERSIDADE DE SÃO PAULO. **Início das estações do ano (2005–2020)**. Disponível em: <<https://www.iag.usp.br/astronomia/inicio-das-estacoes-do-ano>>. Acesso em: 28 jul. 2021.

KESHTEGAR, B. et al. Training ANFIS structure using simulated annealing algorithm for dynamic systems identification. **Neurocomputing**, v. 302, p. 66–74, 2018a.

KESHTEGAR, B. et al. Subset Modeling Basis ANFIS for Prediction of the Reference Evapotranspiration. **Water Resources Management**, v. 32, p. 1101–1116, 2018b.

KHAN, M. Z.; KHAN, M. F. Application of ANFIS, ANN and fuzzy time series models to CO<sub>2</sub> emission from the energy sector and global temperature increase. **International Journal of Climate Change Strategies and Management**, v. 11, n. 5, p. 622–642, 2019.

LIMA, J. E. C.; CASTRO, L. F. DE; CARTAXO, G. A. A. Aplicação do Modelo SARIMA na Previsão de Demanda no Setor Calçadista. **Revista de psicologia**, v. 13, n. 46, p. 892–913, 2019.

LOPES, D. M. et al. Estimativa da evapotranspiração de referência por diferentes métodos para o município de Areia-PB. *In: Congresso Técnico Científico da Engenharia e da Agronomia, 2018, Maceió. Anais...Maceió: CONTECC, 2018*

LYNCH, C. J. .; GORE, R. Application of one-, three-, and seven-day forecasts during early onset on the COVID-19 epidemic dataset using moving average, autoregressive, autoregressive moving average, autoregressive integrated moving average, and naïve forecasting methods. **Data in brief**, v. 35, 2021.

MAMAK, M. et al. Evapotranspiration prediction using adaptive neuro-fuzzy inference system and Penman FAO 56 equation for St. Johns, FL, USA. *In: Environmental Engineering International Conference, 10., 2017, Lithuania. Anais...Lithuania: Vilnius Gediminas Technical University, 2017*

MANIKUMARI, N.; MURUGAPPAN, A.; VINODHINI, G. Time Series Forecasting of Daily Reference Evapotranspiration by Neural Network Ensemble Learning for Irrigation System. **IOP Conference Series: Earth and Environmental Science**, v. 80, 2017.

MATHUR, N.; GLESK, I.; BUIS, A. Comparison of adaptive neuro-fuzzy inference system (ANFIS) and Gaussian processes for machine learning (GPML) algorithms for the prediction of skin temperature in lower limb prostheses. **Medical Engineering and Physics**, v. 38, n. 10, p. 1083–1089, 2016.

MIGLIATO, A. L. T.; PONTI, M. A. Estudo de Predição em Séries Temporais Sazonais Comparando SARIMA e Redes Neurais LSTM e GRU. *In: Workshop de Matemática, Estatística e Computação Aplicadas à Indústria, 2021, São Carlos.*

**Anais...**São Carlos: USP, 2021

NAVARRO-HELLÍN, H. et al. A decision support system for managing irrigation in agriculture. **Computers and Electronics in Agriculture**, v. 124, p. 121–131, 2016.

NAZARÉ, G.; CASTRO, R.; GABRIEL FILHO, L. R. A. Wind power forecast using neural networks: Tuning with optimization techniques and error analysis. **Wind Energy**, v. 23, n. 3, p. 810–824, 2020.

NOURANI, V.; KOMASI, M. A geomorphology-based ANFIS model for multi-station modeling of rainfall-runoff process. **Journal of Hydrology**, v. 490, p. 41–55, 2013.

PENG, H. W. et al. Time series forecasting with a neuro-fuzzy modeling scheme. **Applied Soft Computing**, v. 32, p. 481–493, 1 jul. 2015.

RIBEIRO, A. A.; SIMEÃO, M.; SANTOS, A. R. B. Comparação de métodos de estimativa da evapotranspiração de referência no período chuvoso e seco em Piripiri (PI). **Revista Agrogeoambiental**, v. 8, n. 3, p. 89–100, 2016.

ROSA, J. A.; FIALHO, T. J.; DOETZER, B. H. W. PROGRAMA DE IRRIGAÇÃO NOTURNA NO ESTADO DO PARANÁ. *In: III Workshop Internacional de Inovações tecnológicas na Irrigação & II Conferência sobre Recursos Hídricos do Semiárido Brasileiro.* **Anais...** Fortaleza: WINOTEC, 2010.

ROY, D. K. et al. Using ensembles of adaptive neuro-fuzzy inference system and optimization algorithms to predict reference evapotranspiration in subtropical climatic zones. **Journal of Hydrology**, v. 591, p. 125509, 2020.

SANTOS, A. A. DOS. **Evapotranspiração de referência horária e diária para o estado do Paraná.** Dissertação (Mestrado em Ciência do Solo) - Universidade Federal do Paraná, 2020.

SILVA, C. F. D. et al. A fuzzy approach to the strategic development of human capital in the electric sector. **Computers and Industrial Engineering**, v. 149, 2020.

SPACCA, J. L. C. **Usando o sistema de inferência neuro fuzzy - ANFIS para o cálculo da cinemática inversa de um manipulador de 5 DOF.** Dissertação (Mestrado em Engenharia Elétrica). Faculdade de Engenharia de Ilha Solteira, Universidade Estadual Paulista, 2019.

VIE, A.; KLEINNIJENHUIS, A. M.; FARMER, D. J. **Qualities, challenges and future of genetic algorithms:** a literature review. In press, 2020.

VLASENKO, A. et al. A Novel Neuro-Fuzzy Model for Multivariate Time-Series Prediction. **Data**, v. 62, n. 2, p. 352–356, 2018.

WERNER, L.; BISOGNIN, C.; MENEGOTTO, L. PREVISÃO DA UMIDADE RELATIVA DO AR DIÁRIA DE PORTO ALEGRE. **Cadernos do IME - Série Estatística**, v. 43, p. 39, 2017.

ZADEH, L. A. Fuzzy sets. **Information and Control**, v. 8, n. 3, p. 338–353, 1965.

## CONSIDERAÇÕES FINAIS

Verifica-se que o modelo neuro-*fuzzy* (ANFIS) geral, com rede treinada com série temporal de três anos, mostra-se eficiente na função de prever o valor de  $ET_o$  com base em seus valores anteriores, tendo erro de ajustes (MAPE) de 0,25% para Botucatu – SP e 0,26% para Tupã – SP. Apesar da qualidade de previsão ser semelhante aos dois municípios, seus erros quadráticos evidenciam a ocorrência de maior variabilidade nos valores da série temporal de Tupã – SP, fato ocasionado pelas características climáticas do município.

Vale ressaltar, que o procedimento de imputação para a reconstrução da série temporal de dados meteorológicos foi eficiente e fundamental para a qualidade do modelo ANFIS obtido, já que possibilitou a reconstrução do banco de dados considerando a variabilidade e incerteza do valor perdido, o que permitiu aumentar a qualidade do treinamento da rede ANFIS, já que as séries temporais estavam completas.

Desta maneira, a utilização dos modelos matemáticos para a previsão do valor de  $ET_o$  apresentam-se como uma ferramenta confiável para serem incorporadas por produtores e pesquisadores para a gestão e dimensionamento de sistemas de irrigação, já que apresentam baixo desvio do valor real. Assim, ao associar valor de  $ET_o$ , ao coeficiente de cultura ( $K_c$ ), é possível ganhos de produtividade com baixo custo, já que se pode suprir a demanda hídrica da cultura evitando desperdícios de recursos naturais. Desta forma, há produção de alimentos de maneira sustentável, sendo esta uma questão essencial para todo o mundo, principalmente, para regiões dependentes da irrigação para a produção de alimentos.

Considerando, ainda, tornar os modelos elaborados uma ferramenta de fácil utilização, é possível o desenvolvimento de um aplicativo para celulares, em linguagem interativa e amigável, que automatize a busca pelo resultado, permitindo a transferência do conhecimento para o campo. Além desse avanço, é recomendável ainda a realização de outros estudos que verifiquem a qualidade das previsões para outras localidades, testando ou associando outras metodologias que permitam reduzir o erro ainda existente.

## REFERÊNCIAS

- ADNAN, S.; ULLAH, K.; AHMED, R. Variability in meteorological parameters and their impact on evapotranspiration in a humid zone of Pakistan. **Meteorological Applications**, v. 27, n. 1, 2020.
- ALLEN, R. et al. **Crop evapotranspiration** – Guidelines for computing crop water requerimentos. Rome: FAO, 1998.
- AYDIN, M. C.; KAYISLI, K. Prediction of Discharge Capacity over Two-Cycle Labyrinth Side Weir Using ANFIS. **Journal of Irrigation and Drainage Engineering**, v. 142, n. 5, 2016.
- BABA, R. K.; VAZ, M. S. M. G.; COSTA, J. DA. Correção de dados agrometeorológicos utilizando métodos estatísticos. **Revista Brasileira de Meteorologia**, v. 29, n. 4, p. 515–526, 2014.
- BASSANEZI, R. C. **Ensino-aprendizagem com modelagem matemática: uma nova estratégia**. 3 ed. São Paulo: Contexto, 2002.
- CARDOSO, G. M.; JUSTINO, F. Simulação dos componentes da evapotranspiração sob condições climáticas atuais e de cenários climáticos futuros de aquecimento global com o uso de modelos de clima-vegetação. **Revista Brasileira de Meteorologia**, v. 29, n. 1, p. 85–95, 2014.
- CARVALHO, L. G. et al. EVAPOTRANSPIRAÇÃO DE REFERÊNCIA: UMA ABORDAGEM ATUAL DE DIFERENTES MÉTODOS DE ESTIMATIVA. **Pesquisa Agropecuária Tropical**, v. 41, n. 3, p. 456–465, 2011.
- GABRIEL FILHO, L. R. A. et al. Aplicação da lógica fuzzy para avaliação da eficiência e racionalidade de usinas sucroalcooleiras. *In*: BEDREGAL, B. et al. (eds.). **Recentes Avanços em Sistemas Fuzzy**. 1 ed. São Carlos: SBMAC - Sociedade Brasileira de Matemática Aplicada e Computacional, 2012, p. 405–423.
- GOYAL, M. K. et al. Modeling of daily pan evaporation in sub tropical climates using ANN, LS-SVR, Fuzzy Logic, and ANFIS. **Expert Systems With Applications**, v. 41, n. 11, p. 5267–5276, 2014.
- JIANG, Y. et al. Optimizing regional irrigation water use by integrating a two-level optimization model and an agro-hydrological model. **Agricultural Water Management**, v. 178, p. 76–88, 2016.
- KLAR, A. E. et al. DISTRIBUIÇÃO DA ENERGIA NO INTERIOR DE UMA ESTUFA PLÁSTICA NO PERÍODO DO INVERNO. **IRRIGA**, v. 11, n. 2, p. 257–265, 2006.
- LABABIDI, H. M. S.; BAKER, C. G. J. Fuzzy Modeling. *In*: SABLANI, S. et al. (eds.). **Handbook of Food and Bioprocess Modeling Techniques**. Boca Rator: CRC Press, 2006, p. 451–498.
- MACARTHUR, B. D.; STUMPF, P. S.; OREFFO, R. O. C. From mathematical modeling and machine learning to clinical reality. *In*: LANZA, R. et al. (eds.). **Principles of Tissue Engineering**. 5 ed. Cambridge: Academic Press, 2018, p. 37–51.
- NAGHDI, R.; GHAJAR, I.; TSIORAS, P. A. Time prediction models of grapple skidder HSM 904 using multiple linear regressions (MLR) and adaptive neuro-fuzzy inference

- system (ANFIS). **Operational Research**, v. 16, n. 3, p. 501–512, 2016.
- PRANDO, E. D. P. et al. SISTEMA WEB DE MANEJO DA IRRIGAÇÃO – SISMI. **IRRIGA**, v. 1, n. 2, p. 121–136, 2015.
- ROSSATO, L. et al. Impact of Soil Moisture on Crop Yields over Brazilian Semiarid. **Frontiers in Environmental Science**, v. 5, 2017.
- SILVA, A. O. EVAPOTRANSPIRAÇÃO DE REFERÊNCIA PARA DIMENSIONAMENTO DE SISTEMAS DE IRRIGAÇÃO EM PETROLINA, PE. **Brazilian Journal of Biosystems Engineering**, v. 9, n. 1, p. 30-38, 2015.
- STAN, F. I. et al. Study on the Evaporation and Evapotranspiration Measured on the Caldarusani Lake (Romania). **Procedia Environmental Sciences**, v. 32, p. 281–289, 2016.
- TEJERO, I. F. G. et al. Sustainable Deficit-Irrigation Management in Almonds (*Prunus dulcis* L.): Different Strategies to Assess the Crop Water Status. *In*: TEJERO, I. F. G.; ZUAZO, V. H. D. (eds.). **Water Scarcity and Sustainable Agriculture in Semiarid Environment**. Cambridge: Academic Press, 2018, p. 271–298.
- TOMAS-BURGUERA, M. et al. Accuracy of reference evapotranspiration (ET<sub>o</sub>) estimates under data scarcity scenarios in the Iberian Peninsula. **Agricultural Water Management**, v. 182, n. 1, p. 103–116, 2017.
- TOUREIRO, C. et al. Irrigation management with remote sensing: Evaluating irrigation requirement for maize under Mediterranean climate condition. **Agricultural Water Management**, v. 184, p. 211–220, 2017.
- VENTURA, T. M. et al. ANÁLISE DA APLICABILIDADE DE MÉTODOS ESTATÍSTICOS PARA PREENCHIMENTO DE FALHAS EM DADOS METEOROLÓGICOS. **Revista Brasileira de Climatologia**, v. 19, n. 1, p. 168–177, 2016.
- ZADEH, L. A. Quantitative fuzzy semantics. **Information Sciences**, v. 3, p. 159–176, 1971.
- ZADEH, L. A. Fuzzy Logic. **Computer**, v. 21, n. 4, p. 83–93, 1988.