

**MAURÍCIO BRUNO PRADO DA SILVA**

**ANÁLISE DA EVAPOTRANSPIRAÇÃO DE REFERÊNCIA ( $ET_{oFAO-56}$ ) NO  
PLANALTO OCIDENTAL PAULISTA**

**Botucatu**

**2019**



**MAURÍCIO BRUNO PRADO DA SILVA**

**ANÁLISE DA EVAPOTRANSPIRAÇÃO DE REFERÊNCIA (ETO<sub>FAO-56</sub>) NO  
PLANALTO OCIDENTAL PAULISTA**

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

Orientador: Prof. Dr. João Francisco Escobedo

**Botucatu**

**2019**

S586a Silva, Maurício Bruno Prado da  
Análise da Evapotranspiração de Referência  
(EToFAO-56) no Planalto Ocidental Paulista / Maurício  
Bruno Prado da Silva. -- Botucatu, 2019  
117 p. : il., tabs., mapas

Tese (doutorado) - Universidade Estadual Paulista  
(Unesp), Faculdade de Ciências Agrônômicas, Botucatu  
Orientador: João Francisco Escobedo

1. Evapotranspiração de Referência (EToFAO-56). 2.  
Estimativa de EToFAO-56. 3. Planalto Ocidental Paulista.  
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: ANÁLISE DA EVAPOTRANSPIRAÇÃO DE REFERÊNCIA ( $ET_{o_{FAO-56}}$ ) NO PLANALTO OCIDENTAL PAULISTA**

**AUTOR: MAURÍCIO BRUNO PRADO DA SILVA**

**ORIENTADOR: JOÃO FRANCISCO ESCOBEDO**

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



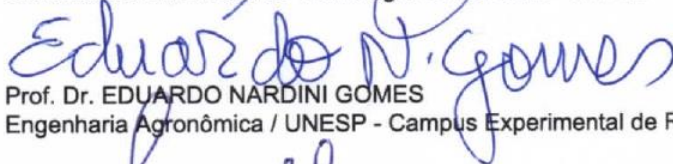
Prof. Dr. JOÃO FRANCISCO ESCOBEDO  
Engenharia Rural / UNESP - Faculdade de Ciências Agrônômicas de Botucatu



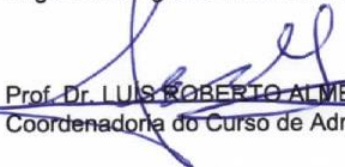
Prof. Dr. DIEGO AUGUSTO DE CAMPOS MORAES  
Análise e Desenvolvimento de Sistemas / Faculdade Eduvale de Avaré



Prof. Dr. CARLOS ROBERTO PEREIRA PADOVANI  
Informática / Faculdade de Tecnologia de Botucatu - FATEC



Prof. Dr. EDUARDO NARDINI GOMES  
Engenharia Agrônômica / UNESP - Campus Experimental de Registro



Prof. Dr. LUIS ROBERTO ALMEIDA GABRIEL FILHO  
Coordenador do Curso de Administração / Faculdade de Ciências e Engenharia - FCE - UNESP - Tupã/SP

Botucatu, 05 de julho de 2019



*Aos meus queridos familiares*

*A minha querida Jéssica*

*Dedico*





## AGRADECIMENTOS

Ao Professor Dr. João Francisco Escobedo pela amizade, competência profissional, excelente orientação e ensinamentos valiosos.

Aos professores que aceitaram o convite para compor a banca avaliadora de minha defesa de tese: Carlos Padovani, Eduardo Nardini, Diego Moraes e Luís Gabriel Filho.

A todos os amigos e colegas de grupo de pesquisa que conviveram comigo ao longo desses anos em Botucatu durante o período de mestrado e doutorado (2014-2019): André Decco, Adriano Marques, Angélica Castilho, Bartolomeu Tangune, Cícero Manoel, Érico Teramoto, Edson Bassetto, Esteban, Jancer Destro, Rocio e Taiza Rossi.

Aos demais amigos e colegas de outros programas de pós-graduação e repúblicas estudantis: Bruno Góes, Bruno Ricardo, Bruno Rodrigues, Cristiane Araújo, Franciana Sousa, Henrique Oldoni, Jannaylton Santos, Nelson Júnior, Lucas Holanda, Ramilos Rodrigues, Mikael Rodrigues, Natália Soares, Maria Clara Gama, Maryjane Gomes, Rafael Lima, Ricardo Vasconcelos, Thalyson Medeiros, Wendell Silva, entre tantos outros.

Aos amigos do Centro Espírita Fraternidade pelos bons momentos compartilhados e pelos intensos trabalhos em prol de ajudar o próximo.

A minha querida namorada Jéssica Maiara Ferrari por todo amor e carinho dedicado ao longo de minha jornada no doutorado.

Aos meus queridos e amados familiares: Darcy Prado (voinha), Jailto Santos (voinho), Maurício Castro (painho), Nancy Prado (mainha), Nelcy Prado (tia), Dedé (tia), Jailto Prado (irmão), que, mesmo à distância, sempre estiveram vibrando e torcendo muito por mim nesta jornada. Também gostaria de agradecer nessas considerações a minha avó materna Maria Castro da Silva (*In memoriam*) por tudo que representou em minha família.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão da bolsa de auxílio financeiro.



## RESUMO

Neste trabalho são desenvolvidos modelos de estimativa e interpolação espacial para a Evapotranspiração de Referência ( $ET_{FAO-56}$ ), modelo de Penman-Monteith, na Região do Planalto Ocidental Paulista (composta pelas regiões intermediárias de São José do Rio Preto e Ribeirão Preto do Estado de São Paulo/Brasil): Análise de Componentes Principais (ACP), Regressões Múltiplas (RM) [Regressões Lineares Múltiplas (RLM) e Regressões Não Lineares Múltiplas (RNLM)] e técnicas de Aprendizado de Máquinas (AM) [*Multilayer Perceptron* (MLP), *Adaptative Neuro Fuzzy Inference System* (ANFIS), *General Regression Neural Network* (GRNN) e *Support Vector Machine* (SVM)], no primeiro capítulo; Krigagem Ordinária (KO) e Mínima Curvatura (MC), no segundo capítulo. O desenvolvimento deste tipo de estudo é de grande importância para otimizar a gestão de recursos hídricos e planejamento da irrigação. A base de dados climáticos utilizada no cálculo de  $ET_{FAO-56}$  é do período de 2013-2017 de 30 localidades da região de estudo. Para a validação dos modelos nos dois estudos foi considerada a  $ET_{FAO-56}$  como referência por meio dos indicadores estatísticos: coeficiente de correlação ( $r$ ), coeficiente de determinação ( $R^2$ ), *Mean Square Error* (MSE), *Root Mean Square Error* (RMSE), *Mean Bias Error* (MBE), e índice de concordância de Willmott ( $d$ ). Os resultados do primeiro estudo mostraram os melhores desempenhos em sequência:  $ET_{OMLP4}$  ( $rRMSE = 0,62\%$ ),  $ET_{OANFIS4}$  ( $rRMSE = 0,75\%$ ),  $ET_{OSVM4}$  ( $rRMSE = 1,19\%$ ),  $ET_{ORLM4}$  ( $rRMSE = 5,23\%$ ),  $ET_{ORNLM4}$  ( $rRMSE = 6,39\%$ ) e  $ET_{OACP}$  ( $rRMSE = 9,32\%$ ). Os resultados obtidos do segundo estudo mostraram que a performance (RMSE) dos modelos anuais e sazonais de interpolação espacial  $ET_{OMC}$  variaram entre 0,11 – 0,14.

Palavras-chave: Evapotranspiração de Referência ( $ET_{FAO-56}$ ); Estimativa de  $ET_{FAO-56}$ ; Planalto Ocidental Paulista



## ABSTRACT

In this work, estimation and spatial interpolation models are developed for the Evapotranspiration Reference ( $ET_{FAO-56}$ ), Penman-Monteith model, in the Western Plateau Paulista Region (composed by the intermediate regions of São José do Rio Preto and Ribeirão Preto of the State of São Paulo/Brazil): Principal Component Analysis (PCA), Multiple Regressions (RM) [Multiple Linear Regressions (RLM) and Multiple Nonlinear Regressions (RNLM)] and Machine Learning (ML) techniques [Multilayer Perceptron (MLP) Adaptive Neuro Fuzzy Inference System (ANFIS), General Regression Neural Network (GRNN) and Support Vector Machine (SVM)] in the first chapter; Ordinary Kriging (OK) and Minimum Curvature (MC) in the second chapter. The development of this type of study is of great importance to optimize water resource management and irrigation planning. The climate database used in the calculation of  $ET_{FAO-56}$  is from the 2013-2017 period from 30 locations in the study region. For the validation of the models in the two studies, the  $ET_{FAO-56}$  was considered as a reference by means of statistical indications: correlation coefficient ( $r$ ), determination coefficient ( $R^2$ ), Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Bias Error (MBE), and Willmott's agreement index ( $d$ ). The results of the first study showed the best performance in sequence:  $ET_{OMLP4}$  (rRMSE = 0.62%),  $ET_{OANFIS4}$  (rRMSE = 0.75%),  $ET_{OSVM4}$  (rRMSE = 1.19%),  $ET_{ORLM4}$  (rRMSE = 5.23%),  $ET_{ORNLM4}$  (rRMSE = 6.39%) and  $ET_{OACP}$  (rRMSE = 9.32%). The results obtained from the second study showed that the performance (RMSE) of the annual and seasonal  $ET_{OMC}$  spatial interpolation models ranged from 0.11 to 0.14.

**Keywords:** Reference Evapotranspiration ( $ET_{FAO-56}$ );  $ET_{FAO-56}$  estimate; Western Plateau Paulista



## SUMÁRIO

INTRODUÇÃO GERAL.....	15
CAPÍTULO 1 .....	17
ESTIMATIVA DA EVAPOTRANSPIRAÇÃO DE REFERÊNCIA ( $ET_{O_{FAO-56}}$ ) NO PLANALTO OCIDENTAL PAULISTA (BRASIL) POR MEIO DE ANÁLISE DE COMPONENTES PRINCIPAIS (ACP), REGRESSÕES MÚLTIPLAS (RM) E TÉCNICAS DE APRENDIZADO DE MÁQUINAS (AM).....	17
1.1 INTRODUÇÃO .....	17
1.1.1 Evapotranspiração de Referência ( $ET_{O_{FAO-56}}$ ) .....	21
1.2 MATERIAL E MÉTODOS .....	23
1.2.1 Região de estudo e informações climáticas .....	23
1.2.2 Economia regional .....	25
1.2.3 Base de dados .....	25
1.2.4 Processamento da base de dados climáticos .....	27
1.2.5 Modelagem da Evapotranspiração de Referência ( $ET_{O_{FAO-56}}$ ) .....	28
1.2.5.1 Técnica Multivariada: Análise de Componentes Principais (ACP).....	28
1.2.5.2 Técnicas de Regressões Múltiplas (RM).....	30
1.2.5.3 Técnicas de Aprendizado de máquinas (AM).....	31
1.2.6 Análise dos dados e validação de modelos.....	36
1.3 RESULTADOS E DISCUSSÕES .....	39
1.3.1 Modelo 1: Análise de Componentes Principais (ACP).....	39
1.3.2 Modelos 2.....	43
1.3.2.1 Análise exploratória dos dados climáticos.....	43
1.3.2.2 Modelo 2: RLM e RNLM.....	46
1.3.3 Modelo 3.....	53
1.3.3.1 Modelo 3: AM: MLP, ANFIS, GRNN e SVM .....	53
1.3.4 Comparação de desempenho estatísticos dos modelos.....	63
1.4 CONCLUSÃO .....	66
REFERÊNCIAS.....	67
CAPÍTULO 2 .....	77
COMPARAÇÃO DE DIFERENTES MÉTODOS DE INTERPOLAÇÃO ESPACIAL DA EVAPOTRANSPIRAÇÃO DE REFERÊNCIA ( $ET_{O_{FAO-56}}$ ) NO PLANALTO OCIDENTAL PAULISTA.....	77
RESUMO.....	77
2.1 INTRODUÇÃO .....	77

<b>2.2 MATERIAL E MÉTODOS</b> .....	<b>79</b>
2.2.1 Região de estudo e informações climáticas .....	79
2.2.2 Economia regional.....	81
2.2.3 Base de dados.....	81
2.2.4 Processamento da base de dados climáticos.....	83
2.2.5 Estatística descritiva .....	84
2.2.6 Métodos de Interpolação Espacial (IE) .....	84
2.2.7 Análise da precisão dos modelos de IE.....	88
2.2.7.1 Validação cruzada dos métodos de IE.....	88
<b>2.3 RESULTADOS E DISCUSSÃO</b> .....	<b>89</b>
2.3.1 Análise da estatística descritiva.....	90
2.3.2 Validação cruzada dos métodos: KO e MC .....	91
<b>2.4 CONCLUSÃO</b> .....	<b>96</b>
<b>REFERÊNCIAS</b> .....	<b>97</b>
<b>CONSIDERAÇÕES FINAIS</b> .....	<b>101</b>
<b>REFERÊNCIAS</b> .....	<b>103</b>
<b>APÊNDICE A – Produção Agrícola do Planalto Ocidental Paulista</b> .....	<b>106</b>
<b>APÊNDICE B – Irradiação solar global no topo da atmosfera (Ho)</b> .....	<b>108</b>
<b>APÊNDICE C - Percentual de dados presentes</b> .....	<b>108</b>
<b>APÊNDICE D - Componentes Principais para cada variável climática</b> .....	<b>109</b>
- Temperatura máxima: .....	109
- Temperatura mínima: .....	110
- Umidade relativa média: .....	111
- Velocidade dos Ventos: .....	112
<b>APÊNDICE E – Distribuições de Frequência de EToFAO-56</b> .....	<b>113</b>
<b>APÊNDICE F – Matriz de pesos de (MLP)</b> .....	<b>114</b>
<b>APÊNDICE G – Modelo de Hargreaves-Samani (H-S)</b> .....	<b>115</b>
<b>APÊNDICE H – Performance do treinamento das técnicas de Aprendizado de Máquinas (AM)</b> .....	<b>115</b>
<b>APÊNDICE I – Variogramas</b> .....	<b>116</b>



## INTRODUÇÃO GERAL

O processo de obtenção de medidas diretas de evapotranspiração geralmente é bastante complexo e difícil, de forma que a obtenção de medidas geralmente está restrita a institutos de pesquisa, tendo em vista os altos custos para implantação de técnicas micrometeorológicas e lisimétricas.

Outra dificuldade encontrada decorrente da obtenção de medidas diretas de evapotranspiração é o manuseio com complexos equipamentos de monitoramento de técnicas micrometeorológicas ou lisimétricas, além do processamento do alto volume de dados (SENTELHAS et al., 2010).

Como consequência das diversas dificuldades existentes para obtenção de medidas diretas de evapotranspiração, métodos indiretos são indicados para estimativa de evapotranspiração por meio de equações matemáticas capazes de ajustarem às condições climáticas locais. Atualmente, o método recomendado pela *Food and Agriculture Organization* (FAO) é o método padrão de Penman-Monteith (P-M) (ALLEN et al., 1998), bastante difundido no mundo. O método de P-M é considerado o mais completo por seu embasamento físico e capacidade de ser adaptado a qualquer localidade, e sua principal desvantagem é a grande quantidade de variáveis climáticas utilizadas (ALLEN et al., 1998; LANDERAS et al., 2008).

Geralmente nações em desenvolvimento possuem poucos recursos para aquisição de sensores automáticos necessários à instalação de Estações Meteorológicas Automáticas (EMA's) (MARTÍ; ZARZO, 2012; SILVA et al., 2017). Situação semelhante ocorre também no Brasil, onde grande parte do território nacional não é contemplada pelas redes de estações meteorológicas. A maior parte das EMA's do país está concentrada na região Sudeste, principalmente no Estado de São Paulo (SANTOS et al., 2014; SILVA et al., 2017). Por isso, em diversas localidades do mundo é comum obter parcialmente algumas variáveis para fins de modelagem, como a temperatura e a umidade relativa do ar. Sendo, então, necessário recorrer a métodos alternativos mais simples e que exijam menos variáveis climáticas de entrada, como a temperatura do ar e a radiação solar, por exemplo. Por isso, o boletim FAO-56 (ALLEN et al., 1998) relata que na impossibilidade de utilização do modelo de P-M numa dada região, devido à insuficiência de variáveis climáticas de entrada, recomenda-se usar o modelo desenvolvido por Hargreaves e Samani (1985), em razão de ser um modelo prático e simples, que necessita apenas do aporte da temperatura do ar como variável climática de entrada.

Outros métodos são relatados na literatura como alternativos na estimativa de  $ET_{\text{FAO-56}}$ , pois podem ajustar-se às condições locais e às variáveis climáticas disponíveis: baseados na temperatura do ar (THORNTHWAITE, 1948; BLANEY e CRIDDLE, 1950; HAMON, 1961; BENEVIDES e LOPEZ, 1970; HARGREAVES e SAMANI, 1985), na radiação solar (MAKKINK, 1957; TURC, 1961; JENSEN e HAISE, 1963; ABTEW, 1966; PRIESTLEY e TAYLOR, 1972; IRMAK et al., 2003), na transferência de massa (DALTON, 1802; TRABERT, 1896; MEYER, 1926; ROHWER, 1931; PENMAN, 1948; ALBRECHT, 1950; ROMANENKO, 1961; BROCKAMP e WENNER, 1963; MAHRINGER, 1970) e no Tanque Classe A (CUENCA, 1989; ALLEN e PRUITT, 1991; SNYDER, 1992; PEREIRA et al., 1995; ALLEN et al., 1998).

Atualmente, outras técnicas computacionais vêm sendo testadas frequentemente em estudos com  $ET_{\text{FAO-56}}$  com sucesso, principalmente em razão da capacidade dessas técnicas em lidar com problemas considerados de natureza não linear, como a  $ET_{\text{FAO-56}}$ , por exemplo. Dentre as técnicas computacionais mais difundidas na modelagem de  $ET_{\text{FAO-56}}$ , destacam-se as Regressões Múltiplas (RM) e o Aprendizado de Máquinas (AM) (ALTHOFF et al., 2018; TANGUNE e ESCOBEDO, 2018), sobretudo em estudos com abrangência local. No entanto, em áreas de grandes dimensões (regiões, perímetro de irrigação ou bacias hidrográficas, por exemplo), outras técnicas como Análise de Componentes Principais (ACP), interpolação espacial [Krigagem Ordinária (KO)] e o Sensoriamento Remoto (SR), vêm ganhando espaço em estudos da estimativa da  $ET_{\text{FAO-56}}$  (VANDERLINDEN et al., 2008; MARTÍ e ZARZO, 2012; FERREIRA SILVA et al., 2018).

O presente trabalho tem como objetivo avaliar o desempenho dessas técnicas na geração de modelos no Planalto Ocidental Paulista (integração das regiões intermediárias de São José do Rio Preto e Ribeirão Preto do Estado de São Paulo), como forma de auxiliar no gerenciamento de recursos hídricos em projetos de irrigação para fornecimento adequado de suprimento hídrico em cultivares predominantes desta região. Com isso, são levantadas duas hipóteses e conclui-se que:

- A técnicas ACP é viável no preenchimento de dados faltantes e estimativa de  $ET_{\text{FAO-56}}$ , assim com as técnicas de AM (MLP, ANFIS e SVM), mediante treinamento;
- A técnica de interpolação espacial Krigagem Ordinária (KO) não é viável para a realização da interpolação espacial de  $ET_{\text{FAO-56}}$ .

# CAPÍTULO 1

## ESTIMATIVA DA EVAPOTRANSPIRAÇÃO DE REFERÊNCIA ( $ET_{O_{FAO-56}}$ ) NO PLANALTO OCIDENTAL PAULISTA (BRASIL) POR MEIO DE ANÁLISE DE COMPONENTES PRINCIPAIS (ACP), REGRESSÕES MÚLTIPLAS (RM) E TÉCNICAS DE APRENDIZADO DE MÁQUINAS (AM)

Autor: Maurício Bruno Prado da Silva

### RESUMO

Neste trabalho é apresentado um estudo sobre modelagem com Evapotranspiração de Referência ( $ET_{O_{FAO-56}}$ ), modelo de Penman-Monteith, usando modelos de Análise de Componentes Principais (ACP), Regressões Múltiplas (RM) [Regressões Lineares Múltiplas (RLM) e Regressões Não Lineares Múltiplas (RNLM)] e técnicas de Aprendizado de Máquinas (AM) [*Multilayer Perceptron* (MLP), *Adaptative Neuro Fuzzy Inference System* (ANFIS), *General Regression Neural Network* (GRNN) e *Support Vector Machine* (SVM)], para a região do Planalto Ocidental Paulista (composta das regiões intermediárias de São José do Rio Preto e Ribeirão Preto). O desenvolvimento do projeto é importante para otimizar a gestão de recursos hídricos e planejamento da irrigação. Para a validação dos modelos nos dois estudos foi considerada a  $ET_{O_{FAO-56}}$  como referência por meio dos indicativos estatísticos: coeficiente de correlação ( $r$ ), coeficiente de determinação ( $R^2$ ), *Mean Square Error* (MSE), *Root Mean Square Error* (RMSE), *Mean Bias Error* (MBE), e índice de concordância de Willmott ( $d$ ). Os resultados do primeiro estudo mostraram os melhores desempenhos em sequência:  $ET_{O_{MLP4}}$  (rRMSE = 0,62%),  $ET_{O_{ANFIS4}}$  (rRMSE = 0,75%),  $ET_{O_{SVM4}}$  (rRMSE = 1,19%),  $ET_{O_{RLM4}}$  (rRMSE = 5,23%),  $ET_{O_{RNLM4}}$  (rRMSE = 6,39%) e  $ACP_{ET_0}$  (rRMSE = 9,32%).

**Palavras – chave:** Evapotranspiração de Referência ( $ET_{O_{FAO-56}}$ ); Estimativa de  $ET_{O_{FAO-56}}$ ; Planalto Ocidental Paulista.

### 1.1 INTRODUÇÃO

A evapotranspiração é um fenômeno físico que promove a complexa transferência de água para a atmosfera por meio da relação entre o balanço hídrico climatológico, evaporação da água da superfície e transpiração das culturas agrícolas (ALLEN et al., 1998; SHIRI et al., 2015). É um fenômeno multidimensional, pois depende da interação entre as variáveis climáticas (radiação solar, temperatura do ar, precipitação, umidade relativa do ar e velocidade do vento) e das condições

fitossanitárias das culturas agrícolas (ALLEN et al., 1998; SENTELHAS et al., 2010; SNYDER, 2017).

O conhecimento da evapotranspiração é de grande importância no planejamento de sistemas de irrigação em regiões de escassez hídrica, na melhoria das estimativas de requerimento de água das culturas agrícolas (ADELOYE et al., 2012; SHIRI et al., 2012) e também em estudos hidrológicos e ambientais (MARTÍ e GASQUE, 2010).

Entretanto, obter medidas confiáveis de evapotranspiração é bastante difícil, devido aos altos custos para implantação de técnicas micrometeorológicas e lisimétricas, além das dificuldades técnicas para operar e fazer manutenção dos equipamentos necessários à utilização dessas técnicas.

As técnicas micrometeorológicas de correlações turbulentas, razão de Bowen e *eddy covariance* geralmente são técnicas complexas que levam em consideração medidas dos fluxos de água, energia e dióxido de carbono (CO<sub>2</sub>) que são de difícil obtenção em razão do complexo manuseio e altos custos de aquisição de equipamentos (BOWEN, 1926; SWINBANK, 1951; SANTOS et al., 2017; SILVA et al., 2017a).

A lisimetria é uma técnica de tentativa de obtenção de medidas diretas de evapotranspiração num dado ambiente. Essa técnica é baseada no balanço hídrico climatológico que possibilita quantificar a água que é perdida para a atmosfera. Entre as técnicas lisimétricas conhecidas estão: lisímetros de pesagem, lisímetros de drenagem e lisímetros de lençol freático constante. Com destaque para a lisimetria de pesagem, que com o advento da eletrônica moderna, possibilitou a obtenção de medidas mais confiáveis em curto espaço de tempo (FARIA et al., 2006). Entretanto, a lisimetria de pesagem apresenta algumas importantes desvantagens, como custo relativamente alto para aquisição, problemas para lidar com a calibração, entre outros (ABDOU e FLURY, 2004).

Como consequência das diversas dificuldades existentes para obtenção de medidas diretas de evapotranspiração, métodos indiretos são indicados por meio de equações matemáticas capazes de se ajustarem às condições climáticas locais. Atualmente, o método recomendada pela *Food and Agriculture Organization* (FAO) é o método padrão de Penman-Monteith (ET<sub>oFAO-56</sub>) (ALLEN et al., 1998), bastante difundido no mundo. O método de P-M é considerado o mais completo por seu embasamento físico e capacidade de ser adaptado a qualquer localidade (ALLEN et

al., 1998; LANDERAS et al., 2008). A única desvantagem do método de P-M é o elevado número de variáveis climáticas utilizadas (radiação solar, temperatura do ar, umidade relativa do ar e velocidade do vento).

Normalmente, estas variáveis não estão à disposição para estimar  $ET_{FAO-56}$ , principalmente em países em desenvolvimento onde faltam recursos para instalação de Estações Meteorológicas Automáticas (EMA's) e aquisição de sensores (MARTÍ e ZARZO, 2012; SILVA et al., 2017b). Situação semelhante ocorre também no Brasil, onde grande parte do território nacional não é contemplada pelas redes de estações meteorológicas. A maior parte das EMA's está concentrada na região Sudeste, principalmente no Estado de São Paulo (SANTOS et al., 2014; SILVA et al., 2017b).

Em diversos países a obtenção de variáveis climáticas é restrita à temperatura e umidade relativa do ar para modelagem da evapotranspiração, e recorrer a métodos alternativos, com exigência de menos variáveis climáticas de entrada é a alternativa. Por isso, o boletim FAO-56 relata que na impossibilidade de utilização do modelo de P-M em uma determinada região, em função da insuficiência de variáveis climáticas de entrada, recomenda-se o uso do modelo desenvolvido por Hargreaves e Samani (1985). Esse modelo é mais prático e simples, necessitando apenas do aporte da temperatura do ar como variável climática de entrada.

Atualmente, existem diversos métodos alternativos para a estimativa de  $ET_{FAO-56}$  que são ajustáveis às condições locais e aos dados climáticas disponíveis. Esses métodos foram divulgados ao longo do século XX e já são bastante difundidos na literatura pesquisada e classificados de acordo com sua principal variável de entrada como os baseados na temperatura do ar (THORNTHWAITE, 1948; BLANEY e CRIDDLE, 1950; HAMON, 1961; BENEVIDES e LOPEZ, 1970; HARGREAVES e SAMANI, 1985), na radiação solar (MAKKINK, 1957; TURC, 1961; JENSEN e HAISE, 1963; ABTEW, 1966; PRIESTLEY e TAYLOR, 1972; IRMAK et al., 2003), na transferência de massa (DALTON, 1802; TRABERT, 1896; MEYER, 1926; ROHWER, 1931; PENMAN, 1948; ALBRECHT, 1950; ROMANENKO, 1961; BROCKAMP e WENNER, 1963; MAHRINGER, 1970) e Tanque Classe A (CUENCA, 1989; ALLEN e PRUITT, 1991; SNYDER, 1992; PEREIRA et al., 1995; ALLEN et al., 1998).

Outras técnicas computacionais atualmente estão bastante difundidas e utilizadas em diversos estudos relacionados com a predição de  $ET_{FAO-56}$  como a Análise de Componentes Principais (ACP) (MARTÍ e ZARZO, 2012), o Aprendizado de Máquinas (AM) (TANGUNE e ESCOBEDO, 2018), a geoestatística

(VANDERLINDEN et al., 2008), as Regressões Múltiplas (RM) (ALTHOFF et al., 2018) e o Sensoriamento Remoto (SR) (FERREIRA SILVA et al., 2018).

Entre as mais difundidas e utilizadas estão às técnicas de Aprendizado de Máquinas (AM) que são baseadas no comportamento de cérebro humano e indicadas para modelar problemas não lineares. Entre as AM mais conhecidas e utilizadas estão a *Artificial Neural Network* (ANN), *Support Vector Machine* (SVM), entre outras. Entre as ANN mais difundidas na literatura estão: *Adaptative Neuro Fuzzy Inference System* (ANFIS), *Generalized Regression Neural Network* (GRNN) e *Multlayer Perceptron* (MLP). Com destaque para MLP, a técnica de ANN mais utilizada entre as demais técnicas.

Kumar et al. (2011) e Tabari et al. (2012) detalham em seus trabalhos um pouco do histórico de sucesso de aplicação dessas técnicas de AM na estimativa de  $ET_{OFAO-56}$  como função das variáveis climáticas. Entre essas técnicas destacam-se a MLP, ANFIS, GRNN e SVM. No caso da técnica GRNN, ainda há poucos trabalhos na literatura, mas que confirma sua viabilidade para modelagem da  $ET_{OFAO-56}$  (KISI, 2006; LADLANI et al., 2012; FENG et al., 2017). Entretanto, a principal dificuldade para trabalhar com essas técnicas é a falta de conhecimentos computacionais (linguagem de programação, usar adequadamente os softwares e seus pacotes, entre outros).

Devido às dificuldades em trabalhar com as técnicas de AM, outras técnicas mais práticas, como as Regressões Múltiplas (RM), podem ser utilizadas. A utilização desta técnica é relatada na literatura por diversos pesquisadores que detalham o bom desempenho na estimativa de  $ET_{OFAO-56}$  (TABARI et al., 2012; SILVA et al., 2016a; ALTHOFF et al., 2018).

A técnica multivariada Análise de Componentes Principais (ACP) vem sendo apontada como alternativa a outros métodos estatísticos em razão de sua versatilidade. É utilizada principalmente para preenchimento de dados faltantes em séries temporais e análise exploratória de variáveis (JOSSE et al., 2009; JOSSE e HUSSON, 2012a). Esta técnica pode ser alternativa a outras técnicas caracterização da variabilidade espacial dos dados, como as de interpolação espacial (GARCÍA-DIEGO e ZARZO, 2010, MARTÍ; ZARZO, 2012).

É importante ressaltar que no Brasil e em muitos países do Hemisfério Sul, existem muito poucos estudos relatando a aplicação de técnicas de AM e ACP para estimativa de importantes variáveis climáticas (SANTOS et al., 2016b; SILVA et al.,

2017b). Em especial, estudos da modelagem da evapotranspiração (ZANETTI et al., 2007).

O objetivo deste trabalho é auxiliar no gerenciamento de recursos hídricos em projetos de irrigação para fornecimento adequado de suprimento hídrico em cultivares predominantes na região do Planalto Ocidental Paulista (integração das regiões intermediárias de São José do Rio Preto e Ribeirão Preto no Estado de São Paulo). Para tanto, é realizado um estudo de modelagem de  $ET_{FAO-56}$  com as técnicas: Análise de Componentes Principais (ACP), Regressões Múltiplas (RM) e Aprendizado de Máquinas (AM).

### 1.1.1 Evapotranspiração de Referência ( $ET_{FAO-56}$ )

O termo Evapotranspiração de Referência ( $ET_{FAO-56}$ ) foi proposto pela primeira vez por Doorembos e Pruitt (1977) como forma de substituir o termo Evapotranspiração Potencial (ETp), até então vigente, e proporcionar algum padrão que só viria a ser obtido em meados de 1990, após a conferência da *Food and Agricultural Organization* (FAO) em Roma, na Itália, que resultou na adoção do método de Penman-Monteith (P-M) como padrão até os dias atuais (ALLEN et al., 1998).

Atualmente, mesmo após mais de duas décadas de divulgação oficial (boletim número 56 da FAO), esse modelo ainda é o mais indicado em razão de seu embasamento físico, embora possua características empíricas, e capacidade de adaptação a qualquer localidade, (ALLEN et al., 1998; LANDERAS et al., 2008).

O termo “evapotranspiração” foi mencionado pela primeira vez em um estudo divulgado em um estudo realizado por Thorntwaite e Wilm (1944), contudo foi a partir de 1948 que outros trabalhos, como os publicados por Thorntwaite (1948) e Penman (1948), detalharam mais aspectos teóricos e físicos sobre a evapotranspiração (CARVALHO et al., 2011). Esses trabalhos possuíam caráter diferente, pois, no caso do estudo de Thorntwaite, relata-se a variação sazonal do balanço hídrico local e a tentativa de construir uma relação empírica com a temperatura média do ar considerando a evapotranspiração como parte do processo do balanço hídrico. Diferentemente do trabalho de Penman, que possuía um caráter de maior abordagem dos processos físicos que regem a evaporação da água na superfície terrestre e sua relação com o efeito conjunto das variáveis climáticas.

Os estudos iniciais sobre evapotranspiração relatados no parágrafo anterior permitiram Monteith na década de 60 realizar modificações no modelo de Penman que resultaram no acréscimo de novos parâmetros a serem considerados como: cobertura vegetal (alfafa ou grama) extensa em pleno crescimento ativo, elevado índice de área foliar e solo em condição de capacidade de campo e a inserção de um número maior de variáveis climáticas (radiação solar, temperatura do ar, umidade relativa do ar e velocidade do vento) (CARVALHO et al., 2011). Como consequência disso, uma nova nomenclatura, denominada Evapotranspiração Potencial (ETp), surge para expressar a demanda atmosférica e seus novos conceitos.

O conceito padronizado e de referência da evapotranspiração é relatado pela primeira vez uma década após a divulgação do trabalho de Monteith no boletim número 24 da FAO de autoria de Dorembos e Pruitt (1977). Nesse boletim esses autores também divulgam os primeiros conceitos de Evapotranspiração de Cultura (ETc) e coeficiente de cultura (Kc), que até os dias atuais são bastante úteis à práticas de irrigação.

Contudo, apenas em 1990, após convenção da FAO, é que denominação Evapotranspiração de Referência ( $ET_{OFAO-56}$ ) está vigente e foi definitivamente divulgada no boletim número 56 em 1998 (ALLEN et al., 1998).

Com o advento definitivo do boletim número 56 da FAO, o atual formato do modelo de estimativa da Evapotranspiração de Referência ( $ET_{OFAO-56}$ ) passou a considerar uma superfície de um cultivo hipotético de alfafa ou grama em estado de desenvolvimento ativo, cobrindo completamente o solo, sem deficiência hídrica e livre de pragas e doenças. A grama (*Paspalum notatum Flüggé*), considerada uma superfície hipoteticamente de referência, assume uma altura de aproximadamente 0,12 m, com resistência da superfície igual a  $70 \text{ s m}^{-1}$  e um albedo de 0,23. A Equação (1) destaca o atual formato do modelo de  $ET_{OFAO-56}$ :

$$ET_{OFAO-56} = \frac{0,408\Delta(R_n - G) + \left(\gamma \frac{900}{T_m + 273}\right)u_2(e_s - e_a)}{\Delta + [\gamma(1 + 0,34U_2)]} \quad (1)$$

em que: saldo de radiação ( $R_n$ ,  $\text{MJ m}^{-2} \text{ dia}^{-1}$ ), fluxo de calor no solo ( $G$ ,  $\text{MJ m}^{-2} \text{ dia}^{-1}$ ), Velocidade do vento ( $U_2$ ,  $\text{m s}^{-1}$ ), Pressão de saturação de vapor d'água do ar ( $e_s$ , kPa), Pressão atual de vapor d'água do ar ( $e_a$ , kPa), Temperatura média do ar ( $T_m$ , °C),



Inclinação da curva de pressão de vapor saturado ( $\Delta$ , kPa °C<sup>-1</sup>), Coeficiente psicrométrico ( $\gamma$ , kPa °C<sup>-1</sup>).

Para obtenção do saldo de radiação ( $R_n$ ), foi considerada a metodologia do balanço de radiação sugerida no Boletim FAO-56 (ALLEN et al., 1998). O fluxo de calor do solo, quando considerado em partição diária, é praticamente igual a zero. Os valores medidos da velocidade do vento ( $U_2$ ), considerada no modelo  $ET_{OFAO-56}$  é obtida de anemômetros instalados a 2 m de altura da superfície do solo. Do contrário, os valores medidos de  $U_2$  em alturas superiores são convertidos para 2 m de altura da superfície. Mais detalhes sobre o processo de conversão podem ser encontrados em Allen et al. (1998).

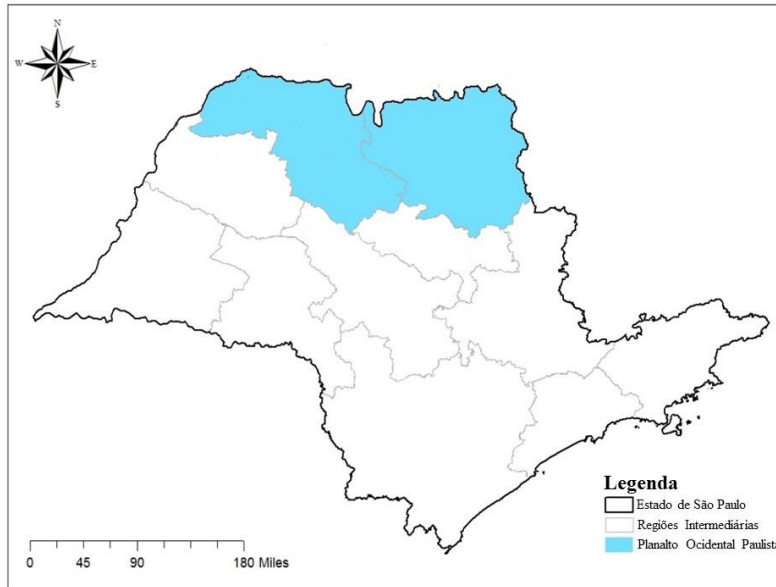
A viabilidade deste método é comprovada mediante inúmeros testes em diversas localidades do mundo, onde o modelo de referência de Penman-Monteith foi validado satisfatoriamente por meio de lisimetria, (ALLEN et al., 1998; SENTELHAS, et al., 2010). Pereira et al. (2015) relata, em suas considerações sobre futuras projeções do modelo de referência de Evapotranspiração de Referência ( $ET_{OFAO-56}$ ) e Evapotranspiração de Cultura ( $ET_c$ ), que é de extrema importância a continuação de estudos que busquem melhorar a eficiência dos modelos.

## 1.2 MATERIAL E MÉTODOS

### 1.2.1 Região de estudo e informações climáticas

Na Figura 1 é apresentado o Planalto Ocidental Paulista (destacado em azul), localizado a noroeste do Estado de São Paulo. Essa região é composta pela integração de duas regiões intermediárias do Estado de São Paulo: São José do Rio Preto e Ribeirão Preto (latitude 19°32' – 21°57'S, longitude 46°38' – 51°28'O e altitude 335,00 – 1026,00 m).

**Figura 1.** Localização do Planalto Ocidental Paulista no Estado de São Paulo.



O Instituto Brasileiro de Geografia e Estatística desde o ano de 2017 vem adotando o termo “Região Intermediária” como forma de delimitar as diferentes regiões do Estado de São Paulo em razão de suas características locais (IBGE, 2018). A área total destas regiões integradas neste estudo é de aproximadamente 53.492,00 km<sup>2</sup>, abrangendo 164 municípios, ou seja, 25,43% dos municípios do Estado de São Paulo, com população de aproximadamente 4.188,425 de habitantes (9,29% da população total do estado) (IBGE, 2018).

O clima predominante na região de estudo é do tipo Aw (CEPAGRI, 2018; ROLIM, 2007; TANGUNE e ESCOBEDO, 2018), tropical (megatérmico), com verão chuvoso e inverno seco (KÖPPEN, 1928). Com temperatura média do ar variando de 19,40 a 24,20°C e intervalo de variação dos meses chuvosos de 9,0 – 300,0 mm de agosto a janeiro.

O regime de chuvas (ou período úmido) da Região Sudeste do Brasil (em especial, no Estado de São Paulo) é normalmente influenciado pelo transporte de umidade em níveis mais baixos da atmosfera (850 hPa) por meio de Jatos de Baixos Níveis (JBN) que transportam umidade do Sul da Região Amazônica até o Oceano Atlântico (NOBRE et al., 2012). Entretanto, em anos atípicos, o regime de chuvas é drasticamente interrompido, como entre os anos 2013/2014 e 2014/2015, que foram classificados como anos: excepcionalmente seco e muito seco, respectivamente (COELHO et al, 2016a). Nesses anos o padrão de chuvas na região Sudeste foi interrompido por uma anomalia da circulação anti-ciclônica, no Oceano Atlântico

próximo à costa da região Sudeste do Brasil. Este tipo de circulação inibiu o transporte para esta região, refletindo baixa quantidade de chuvas precipitadas. Simultaneamente foi verificado um aumento da temperatura do Oceano Atlântico que contribuiu para a redução da precipitação na região (COELHO et al., 2016b).

No caso do período seco, é conhecido que o regime de chuvas é do tipo frontal, originada do encontro das massas frias e secas vindas da Região Sul com as massas quentes e úmidas, da Região Sudeste. Sua distribuição gera aumento na nebulosidade, com ocorrência de chuvas generalizadas na região, com intensidade de precipitação variando de fraca a moderada (SATYAMURTY et al., 1998).

### **1.2.2 Economia regional**

Com as condições climáticas favoráveis às atividades da agricultura, o Planalto Ocidental Paulista tem sua economia predominantemente destinada ao agronegócio. Sendo esse setor responsável pela geração de empregos e movimentação da economia local.

O relatório anual do Instituto de Economia Agrícola (IEA), vinculado à Secretaria de Agricultura e Abastecimento do Estado de São Paulo e elaborado por Silva et al. (2019), mostra as principais *commodities* agrícolas produzidas pelo Estado de São Paulo e suas regiões: amendoim (água/seca), café, cana-de-açúcar, eucalipto, feijão (água/seca/com irrigação/sem irrigação), laranja, mandioca (indústria/mesa), milho (com irrigação/sem irrigação), seringueira e soja. Nesse relatório é possível verificar a posição de destaque das regiões intermediárias que compõem o Planalto Ocidental Paulista.

No Apêndice A, por meio da tabela em destaque, é possível verificar do percentual da área de produção agrícola, produção relativa e a posição no ranking de maior produção em 2017 de amendoim, cana-de-açúcar, café, feijão, laranja, mandioca, milho, seringueiras e soja.

### **1.2.3 Base de dados**

A base de dados climáticos monitorada na região de estudo é composta das variáveis climáticas: irradiação solar global ( $H$ , MJ m<sup>-2</sup> hora<sup>-1</sup>); temperaturas máxima e mínima do ar ( $T_{max}$  e  $T_{min}$ , °C), umidade relativa do ar (UR, %) e velocidade do vento ( $U_2$ , m s<sup>-1</sup>) medida a 10 metros de altura da superfície em partição horária –

convertida posteriormente para 2 metros de altura conforme metodologia descrita por Allen et al. (1998) no boletim FAO-56.

A variável astronômica Irradiação solar global no topo da atmosfera ( $H_0$ , MJ m<sup>-2</sup> hora<sup>-1</sup>) também compôs a base de dados climáticos deste estudo. Para obtenção desta variável foi utilizada a metodologia descrita por Iqbal (1983) e detalhada no Apêndice B.

Na Tabela 1 são mostradas as informações geográficas (localidade, latitude, longitude, altitude e distância do mar), classificação climática e fonte de aquisição da base climática de dados das 30 localidades situadas na região de estudo do Planalto Ocidental Paulista.

**Tabela 1.** Informações geográficas das EMA's [localidades, latitudes (°), longitude (°), altitude (m)], classificação climática (clss. climática), fontes dos dados.

Localidades	Latitude	Longitude	Altitude	Dist. do mar	Clss. Climática	Fontes
	(°)	(°)	(m)	(km)		
Campina Verde*	19°32'S	49°32'W	555,2	587,0	Aw	INMET
Paranaíba**	19°41'S	51°10'W	424,0	668,0	Aw	INMET
Populina	19°52'46"S	50°28'13"W	394,0	610,0	Aw	UNESP
Sacramento*	19°53'S	47°26'W	912,0	465,0	Csa	INMET
Conceição das Alagoas*	19°59'S	48°09'W	568,0	494,0	Aw	INMET
Paranapuã	20°01'29"S	50°33'57"W	436,1	604,0	Aw	UNESP
Jales	20°09'S	50°35'W	457,0	581,0	Aw	INMET
Ituverava	20°21'S	47°46'W	600,0	440,0	Aw	INMET
Ilha Solteira	20°25'24"S	51°21'13.1"W	337,0	622,0	Aw	UNESP
Marinópolis	20°26'48"S	50°48'26.1"W	370,0	585,0	Aw	UNESP
Pereira Barreto I	20°31'42"S	51°14'58"W	426,0	589,0	Aw	UNESP
Barretos	20°33'S	48°32'W	533,0	444,0	Aw	INMET
Franca	20°34'S	47°22'W	1026,0	387,0	Aw	INMET
Itapura	20°38'28"S	51°28'29.9"W	335,0	608,0	Aw	UNESP
Pereira Barreto II	20°40'23"S	51°2'2.1"W	357,0	579,0	Aw	UNESP
Sud Mennucci	20°43'42"S	50°57'30"W	350,0	565,0	Aw	UNESP
Passos*	20°44'S	46°38'W	784,0	328,0	Cwa	INMET
Monte Azul Paulista	20°54'26"S	48°38'29"W	611,0	410,0	Aw	IAC
Brodowsky	20°59'03"S	47°39'33"W	863,0	359,0	Cwa	IAC
José Bonifácio	21°02'S	49°41'W	405,0	465,0	Aw	INMET
Airanhã	21°07'S	48°50'W	525,0	398,0	Aw	INMET
Sertãozinho	21°08'16"S	47° 59' 24" W	579,0	362,0	Aw	IAC
Cássia dos Coqueiros	21°16'58"S	47°10'11"W	890,0	303,0	Cwb	IAC
Valparaíso	21°19'S	50°55'W	374,0	526,0	Aw	INMET
Pradópolis	21°20'S	48°06'W	544,0	339,0	Aw	INMET
São Simão	21°28'44"S	47° 33' 3" W	620,0	304,0	Aw	IAC
Lins	21°39'S	49°44'W	459,0	415,0	Aw	INMET
Tambaú	21°42'18"S	47° 16' 26" W	698,0	268,0	Aw	IAC
Ibitinga	21°51'S	48°48'W	492,0	344,0	Aw	INMET
São Carlos	21°57'S	47°52'W	863,0	267,0	Cwa	INMET

\*Estado de Minas Gerais, \*\*Estado de Mato Grosso do Sul.

Entre as 30 (trinta) EMA's, 5 (cinco) estão situadas fora dos limites territoriais do Estado de São Paulo e da região de estudo: Campina Verde, Conceição das Alagoas, Sacramento e Passos (Estado de Minas Gerais); e Paranaíba (Estado de Mato Grosso do Sul). Essas localidades foram selecionadas em razão de sua proximidade com os limites extremos do Estado de São Paulo. Outras localidades de

regiões intermediárias paulistas foram inseridas neste estudo (Ibitinga, Ilha Solteira, Itapura, Lins, Pereira Barreto I e II, São Carlos, Sud Mennucci, Valparaíso) devido à proximidade com a região de estudo.

Contudo, é importante destacar que o número de EMA's operando no Estado de São Paulo pode ser ainda maior, pois outras instituições como a Universidade de São Paulo (USP), usinas de setor sucroalcooleiro, associações de produtores de café, também operam com esse tipo de estação no Estado (AGRITEMPO, 2018).

#### **1.2.4 Processamento da base de dados climáticos**

Em razão do grande volume de informações, a necessidade e a demanda específica de alto desempenho de processamento, foi utilizada a tecnologia de gerenciamento de banco de dados MySQL 5.7 (*Relational Database Management System*) da *Oracle* em sua versão de código aberto (licença GNU), atuando como um servidor local (ORACLE CORPORATION, 2019a). Todas as operações computacionais em formato de algoritmos e funções foram desenvolvidas através da linguagem de alto nível de manipulação de dados SQL (*Structured Query Language*).

O MySQL 5.7 é um servidor de banco de dados *multithread* (encadeamento de execução de processos) e multiusuário (acesso simultâneo de diversos usuários). Seus recursos são planejados para aplicações em grandes volumes de dados, softwares de missão crítica e sistemas computacionais implementados em massa (ORACLE CORPORATION, 2019a). Sua principal função neste estudo foi armazenar, processar e padronizar o extenso volume de parâmetros meteorológicos de 30 Estações Meteorológicas Automáticas (EMA's) espalhadas pelo Planalto Ocidental Paulista.

A função *average* (AVG) da linguagem SQL foi utilizada para se chegar aos valores médios da radiação solar global, temperatura do ar, umidade relativa do ar e velocidade do vento a 10 metros de altura. As funções de máximo (MAX) e mínimo (MIN) foram utilizadas para se obter os valores máximos e mínimos referentes a temperatura e umidade relativa do ar. A função de soma (SUM) foi utilizada para se obter os valores de precipitação pluviométrica acumulados para o dia (ORACLE CORPORATION, 2019b).

### 1.2.5 Modelagem da Evapotranspiração de Referência ( $ET_{FAO-56}$ )

Nesta seção é descrita detalhadamente as técnicas utilizadas neste estudo para modelagem da  $ET_{FAO-56}$  que estão divididas em Análise de Componentes Principais (ACP), Regressões Múltiplas (RM) e Aprendizado de Máquinas (AM).

Para o processamento da base de dados climáticos e obtenção de estimativas de  $ET_{FAO-56}$  para as 30 EMA's foram utilizados os *softwares*: *MATrix LABoratory* (Matlab), com as técnicas de Aprendizado de Máquinas (AM): *Multilayer Perceptron* (MLP), *Adaptive Neuro-Fuzzy Inference System* (ANFIS) e *Generalized Regression Neural Network* (GRNN); *Waikato Environment for Knowledge Analysis* (WEKA), com a técnica de AM: *Support Vector Machine* (SVM); *Microcal™ Origin 6.0®*, com as técnicas de Regressões Múltiplas; e, por fim, *R-GUI*, com a técnica multivariada de Análise de Componentes Principais (ACP) (MICROCAL SOFTWARE, 1999; MATHWORKS, 2017; R Core Team, 2018; WEKA, 2018). Em decorrência de ajustes dos parâmetros de cada técnica implementada nesses *softwares*, a seguir são detalhados brevemente cada software e técnica utilizada.

#### 1.2.5.1 Técnica Multivariada: Análise de Componentes Principais (ACP)

A técnica ACP é uma ferramenta exploratória que permite revelar a existência de anomalias na relação entre as variáveis medidas e agrupamentos entre as amostras ou indivíduos do ponto de vista multidimensional (JOSSE et al., 2009; LYRA et al., 2010; JOSSE e HUSSON, 2012a). Para Mingoti (2017) o objetivo principal da técnica ACP é explicar a estrutura de variância e covariância de um vetor aleatório, composto de  $p$ -variáveis aleatórias, por meio da construção de combinações lineares das variáveis originais. Dentre as aplicações dessas técnicas pode se destacar a análise exploratória de variáveis por meio da redução do espaço paramétrico e preenchimento de dados perdidos ou faltantes.

Para a realização da Análise de Componentes Principais (ACP) observa-se a premissa de o número de elementos ( $n$ ) ser maior que o número de variáveis ( $p$ ), ou seja, ( $n > p$ ). Do contrário, o número de elementos ( $n$ ) será menor que o número de variáveis ( $p$ ): ( $n \leq p$ ) e ocorrerá uma situação de alta dimensionalidade das variáveis que para resolução, o mais indicado é optar por métodos iterativos com a utilização de algoritmos computacionais.

A seguir são detalhadas as duas formas de utilização da técnica Análise de Componentes Principais (ACP):

- Análise de Componente Principal para casos em que o número de elementos é maior do que o número de variáveis ( $n \geq p$ )

Algebricamente a Componente Principal é um método capaz de transformar um conjunto de  $p$  variáveis originais  $X_1, X_2, \dots, X_p$  mensuradas nos  $n$  elementos em um novo conjunto de variáveis,  $Y_1, Y_2, \dots, Y_k$ , com  $k$  menor ou igual  $p$ , em que cada CP será uma combinação linear das variáveis originais (Equação 2), não correlacionadas entre si, explicando parte da variabilidade máxima dessas variáveis (STRAPASSON et al., 2000; MELO PRADO et al., 2016). Morrison (1976) relata que o primeiro CP (ou  $Y_1$ ) de um conjunto de  $p$  variáveis de um vetor aleatório  $\mathbf{X}' = (X_1, X_2, \dots, X_p)$  é definido como combinação linear das variáveis:

$$Y_1 = b_{11}X_1 + b_{21}X_2 + \dots + b_{p1}X_p = \mathbf{b}'_1\mathbf{X} \quad (2)$$

em que os coeficientes  $b_{i1}$  são elementos do autovetor  $\mathbf{b}_1$  associado ao maior autovalor ( $\lambda_1$ ) obtidos a partir da decomposição espectral da matriz de correlação (R) ou de covariância (S) do vetor aleatório  $\mathbf{X}$ . Assim, primeira componente ( $Y_1$ ) explica a maior proporção da variabilidade total dos dados, e assim sucessivamente. A matriz de correlação (R) é mais apropriada para o caso de variáveis com unidades diferentes. O segundo Componente Principal estará associado ao segundo maior autovalor, bem como seus autovetores, e assim sucessivamente.

- Análise de Componente Principal quando o número de elementos é menor que o número de variáveis ( $n < p$ )

No caso de dados com alta dimensão, é comum observar um número de elementos amostrais menores ao número de variáveis. Neste caso, a estimativa da matriz R ou S apresentada anteriormente não é eficiente, sendo mais apropriado utilizar algoritmos iterativos para obtenção das combinações lineares (CP), possibilitando estimar a matriz de dados  $X$ , de modo que  $X = \mathbf{T}\mathbf{L}' + \mathbf{E}$ , em que  $L$  é a matriz de cargas,  $T$  a matriz de escores e  $E$  é a matriz de resíduos. Entre os algoritmos numéricos para a decomposição da matriz de dados  $X = \mathbf{T}\mathbf{L}' + \mathbf{E}$ , destacam-se o NIPALS ou EM. No caso do algoritmo EM, Josse e Husson (2012a) destaca que esse algoritmo possui melhor desempenho estatístico do que NIPALS.

O algoritmo EM é relatado pela primeira vez por Dempster et al. (1977) como uma ferramenta de cálculos iterativos de estimativas de máxima verossimilhança quando as observações podem ser vistas como dados incompletos. Nesse algoritmo por meio do método de imputações múltiplas é possível realizar a reconstrução da matriz de dados das variáveis aleatórias (processo inverso da decomposição espectral da matriz) associado ao número de CP determinados (KIERS, 1997; JOSSE et al., 2009; JOSSE e HUSSON, 2012b). Esse algoritmo está disponibilizado no pacote *missMDA* do ambiente computacional R-Gui (JOSSE e HUSSON, 2016).

O pacote *missMDA* do R foi utilizada neste estudo com a finalidade de criar uma nova base de dados climáticos para preencher os valores faltantes e também para criar uma nova matriz de  $ET_{FAO-56}$ , para a finalidade de validar a estimativa obtida da reconstrução da matriz com a do modelo padrão da FAO. Detalhes sobre o percentual de dados faltantes e informações sobre a recomposição de valores estimados recompondo a base de dados estão disponíveis nos Apêndices C e D, respectivamente.

#### 1.2.5.2 Técnicas de Regressões Múltiplas (RM)

As Regressões Múltiplas (RM) são técnicas estatísticas que comumente são utilizadas para descrever a relação entre variáveis dependentes e mais de duas variáveis independentes (JATO-ESPINO et al., 2017). As RM são expressas na maneira que mais se adequem à natureza dos dados. Assim, são conhecidas as suas formas lineares e não lineares.

##### - Regressões Lineares Múltiplas (RLM):

As RLM são equações lineares que relacionam um conjunto de variáveis independentes a uma variável resposta dependente expressa por (Equação 3):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \quad (3)$$

sendo Y a variável dependente;  $\beta_0$ - $\beta_n$  são os parâmetros da equação da relação linear;  $X_1$  a  $X_n$ , são as variáveis independentes (OZBAYOGLU e OZBAYOGLU, 2006; TABARI et al., 2012).

##### - Regressões Não Lineares Múltiplas (RNLM):

As RNLM são equações não lineares que relacionam um conjunto de variáveis independentes a uma variável resposta dependente que podem ser expressa de forma logarítmica, exponencial ou trigonométrica (MAROFI et al., 2011; TABARI et al., 2012). A Equação (4) mostra uma representação genérica de uma possível RNLM:



$$Y = \alpha_0 (X_1^{\alpha_1}) (X_2^{\alpha_2}) \dots (X_n^{\alpha_n}) \quad (4)$$

onde Y representa a variável dependente;  $\alpha_0$  a  $\alpha_n$ , são os parâmetros da equação para a relação não linear;  $X_1$  a  $X_n$  são as variáveis independentes (BILGILI, 2010; TABARI et al., 2012).

Os modelos não lineares podem ser transformados para um domínio linear adequadamente por meio de alguma transformação nas variáveis, tal como a transformação logaritmo, como mostra a Equação (5):

$$\log(Y) = \log(\alpha_0) + \alpha_1 \log(X_1) + \alpha_2 \log(X_2) + \dots + \alpha_n \log(X_n) \quad (5)$$

uma regressão de  $\log(Y)$  sobre  $\log(X_1)$ ,  $\log(X_2)$ , ...,  $\log(X_n)$  é utilizada para estimar parâmetros  $\alpha_0$ ,  $\alpha_1, \dots, \alpha_n$  (BILGILI, 2010; TABARI et al., 2012).

A geração de modelos de RM para estimativa de  $ET_{FAO-56}$  foi realizada utilizando o software Microcal™Origin 6.0® por meio da janela ajuste de funções não lineares. Esta janela utiliza o algoritmo *Levenberg-Marquardt* para minimização da soma dos desvios quadráticos relativos (Qui-quadrado –  $Q^2$ ) nos parâmetros das funções ajustadas. Este procedimento ocorre iterativamente e parte de estimativas iniciais dos parâmetros que a cada iteração retorna com uma nova estimativa. Quando a variação de  $Q^2$  diminui de uma iteração para outra, a estimativa pode convergir ou divergir para um mínimo local de  $Q^2$ .

#### 1.2.5.3 Técnicas de Aprendizado de máquinas (AM)

O Aprendizado de Máquina (AM) é uma técnica computacional que realiza extração de informações de um conjunto de dados. Com o emprego de um princípio de inferência denominada de indução que resultará em aprendizagem supervisionada (LORENA et al., 2011). Para tanto, o algoritmo irá reconhecer padrões dos dados de entrada, para que produzam resultados coerentes para os de saída. Assim, o problema será modelado e apresentado por conjunto de dados composto por pares de dados de entrada e saída (MITCHELL, 1997). Essas técnicas resultam da interação de dois campos científicos como a Inteligência Artificial (IA) e a estatística.

Por meio da interação da inteligência artificial e da estatística é possível realizar extrações de informações de conjuntos de dados. Essa técnica é conhecida como Aprendizado de Máquinas (AM), que por meio de seu algoritmo utiliza de modelos que permitam previsões de novos dados (MITCHELL, 1997; LORENA et al., 2011).

Entre as Técnicas de Aprendizado de Máquinas (AM) utilizadas para estimar a Evapotranspiração de Referência ( $ET_{FAO-56}$ ) destacam-se as *Artificial Neural Network*

(ANN), *Adaptive Neuro-fuzzy Inference System* (ANFIS), *Co – Active Neuro Fuzzy Inference System* (CANFIS), *Gene Expression Programming* (GEP), *Generalized Regression Neural Networks* (GRNN) e *Support Vector Machine* (SVM). Com diversas aplicações bem sucedidas em diversas localidades do mundo (KISI, 2006; KISI e OSTURK, 2007; AYTEK, 2009; COBANER, 2011; SHIRI et al., 2012; TABARI et al., 2012). Para a realização de estimativas de  $ET_{\text{FAO-56}}$  neste trabalho, por meio das técnicas de AM nas fases de treinamento e validação dos modelos, foram utilizados os softwares *Matlab* e *WEKA*.

### *Artificial Neural Network (ANN)*

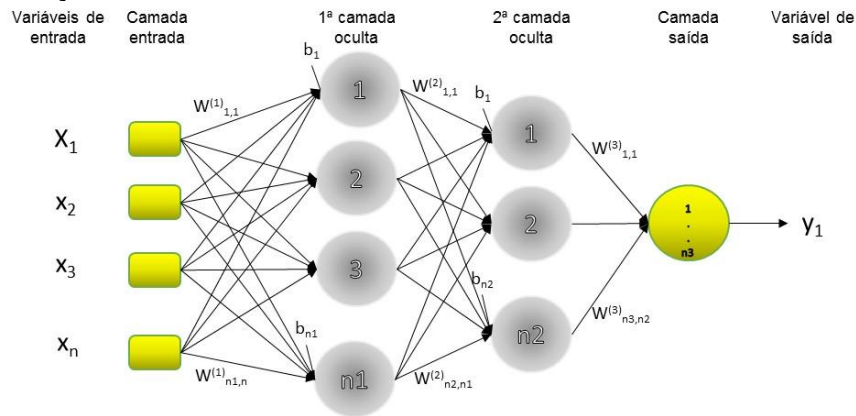
Formada por neurônios e distribuídos de maneira paralela, as *Artificial Neural Network* (ANN) são técnicas compostas por camadas que processam informações entre si. Essas técnicas são capazes de lidar com funções não lineares consideradas complexas, realizar reconhecimento e classificação de padrões, agrupamentos, entre outros (LANDERAS et al., 2008; LORENA et al., 2011).

São relatados na literatura diversos tipos de ANN como as Redes recorrentes de Hopfield (HOPFIELD, 1982), as Redes auto-organizáveis de Kohonen (KOHONEN, 1982), a *Learning Vector Quantization* (LVQ) (KOHONEN, 1990), *Counter Propagation* (CP) (CHANG e SU, 2005) e as Redes *Adaptative Resonance Theory* (ART) (GROSSBERG, 1976a; GROSSBERG, 1976b). Que são utilizadas para diversos tipos de aplicações, como relata Silva et al. (2016a).

### *Multilayer Perceptron (MLP)*

Geralmente a técnica MLP é utilizada com uma ou duas camadas intermediárias que estão localizadas entre a entrada e saída de variáveis. Essa técnica é classificada como rede do tipo *feedforward*, conhecida por possuir alimentação a frente ou única direção de camadas múltiplas (Figura 2) (HAYKIN, 1998).

**Figura 2.** Ilustração de uma rede MLP.



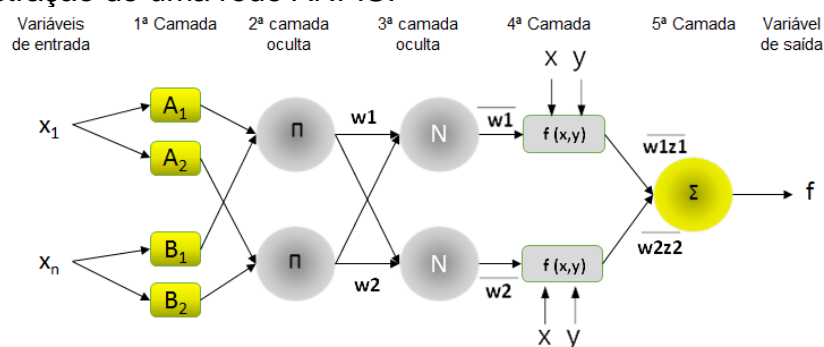
Fonte: Adaptado de Kumar et al. (2011).

O algoritmo realiza o treinamento da técnica MLP de maneira supervisionada e em duas fases. Na primeira fase as informações das variáveis de entrada e os pesos iniciais se propagam e ao final desse processo há comparação das informações iniciais e de saída; na segunda fase, ocorre o processo inverso e as informações de saída são comparadas com as iniciais (LAM et al., 2008; SILVA et al., 2017b).

#### *Adaptive Neuro-Fuzzy Inference System (ANFIS)*

A lógica de processamento desta técnica consiste em utilizar a forma de interpretação da Lógica *Fuzzy* e as características de aprendizado da ANN compondo um sistema híbrido. A Figura 3 apresenta a estrutura da ANFIS para duas variáveis de entrada, uma saída e duas regras

**Figura 3.** Ilustração de uma rede ANFIS.



Fonte: adaptado de Jang (1993).

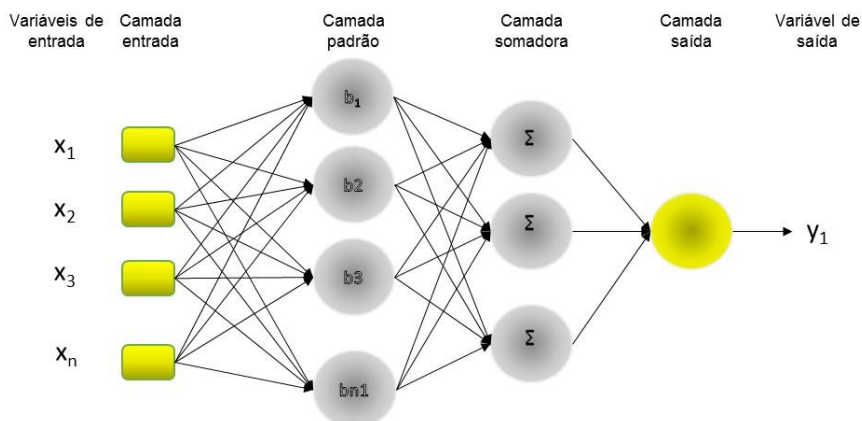
Os nós adaptativos representados por quadrados e os fixos por círculos e divididos em interações entre cinco camadas que acontecem em dois processos: fuzzificação e defuzzificação (ZADEH, 1968; TAKAGY & SUGENO, 1985).

Para a utilização da técnica ANFIS, no script para treinamento, é utilizada a função *genfis2* com um algoritmo de clusterização subtrativa pela qual gera uma estrutura FIS (*Fuzzy Inference System*) do tipo *Takagi-Sugeno* (TAKAGI & SUGENO, 1985). Que viabiliza um treinamento com baixo esforço computacional, mesmo com aumento do número de variáveis, e menor valor de MSE. Para tanto, é necessário um ajuste adequado do (*radii*). Este parâmetro define o vetor que especifica o alcance de influência do centro do *cluster* em cada dimensão de dados e o número de épocas de interação.

#### Generalized Regression Neural Network (GRNN)

A rede GRNN é uma versão recente e otimizada da rede *Radial Basis Function* (RBF); e não requer um treinamento iterativo. Consiste de duas camadas (padrão e somadora) além da camada de entrada e saída (Figura 4).

**Figura 4.** Ilustração de uma rede GRNN.



Fonte: Adaptado de Khatib e Elmenreich (2015).

Nesta estrutura o número de neurônio da camada padrão é igual ao número de amostras da camada de entrada. A camada padrão é conectada a camada somadora que funcionam juntas para normalizar os valores de saída.

Essa técnica difere de outras de ANN por possuir um processo de aprendizado que não faz uso dos ajustes simultâneos para adquirir conhecimento no treinamento, porém, possui capacidade de aprender quase que instantaneamente com os dados apresentados na base de treinamento (HANSEN & MESERVY, 1996).

Na técnica de ANN do tipo GRNN, é utilizada a função *newgrnn*. Sua parametrização ocorre com a definição do ajustamento para definir a abertura da

função de ativação de arquitetura gaussiana, que define quantos neurônios serão utilizados na modelagem.

### *Support Vector Machine (SVM)*

A Support Vector Machine (SVM) é uma técnica derivada da teoria de aprendizagem estatística e foi primeiramente introduzida por Vapnik em 1995 (TABARI et al., 2012). Essa técnica tenta, por meio do princípio da minimização do risco estrutural, minimizar um limite superior de erro de generalização em vez de minimizar o erro de treinamento local (VAPNIK, 1995; VAPNIK, 1998; CHEN et al., 2015). A solução de problemas de regressão usando a SVM pode ser dada por meio de um algoritmo iterativo, chamado de *Sequential Minimal Optimization* (SMO) (SMOLA & SCHÖLKOPF, 1998).

A regressão com SVM consiste em estimar uma função de acordo com um determinado conjunto de dados definidos:  $\{(x_i, y_i)\}_{i=1}^n$ , onde  $x_i \in \mathfrak{R}^n$  é o vetor de caracterização que representa as variáveis de entrada;  $y_i \in \{-1, +1\}$  é a saída e  $n$  é o número total de conjuntos de dados (TABARI et al., 2012).

Para a técnica SVM foi utilizado o algoritmo SMO (Otimização Mínima Sequencial) com função de transformação de base radial (RBF) de *Kernel* com método de validação cruzada (PLATT, 1998; SMOLA & SCHÖLKOPF, 2004). Na utilização da RBF é necessário ajustar adequadamente os parâmetros da configuração básica do método SVM/SMO: Custo ( $C$ ), determina o ponto de equilíbrio entre a maximização da margem e minimização do erro e que controla a influência de cada vetor; Gamma ( $\gamma$ ), controla a flexibilidade da função de *Kernel* e verifica seu ajustamento (*overfitting* e *underfitting*); e épsilon ( $\epsilon$ ), erro de tolerância. A principal desvantagem da técnica de SVM é que ela não tem base física e pertence a uma classe de abordagens de caixa preta orientadas por dados (TABARI et al., 2012).

A técnica de AM do tipo Support Vector Machine (SVM) foi treinada e avaliada por meio do software *Waikato Environment for Knowledge Analysis* (WEKA). Esse software consiste num conjunto de algoritmos, disponível em <http://www.cs.waikato.ac.nz/~ml/weka/>, que contém ferramentas para pré-processamento de dados, classificação, regressão, clustering, regras de associação e visualização (WITTEN et al., 2011).

### 1.2.6 Análise dos dados e validação de modelos

A base de dados totalizou 49.171 pontos e foi separada em 70% (34.419 pontos) para treinamento e 30% para validação (14.752 pontos).

Para a realização da estatística descritiva da base de dados foram utilizadas as seguintes técnicas:

- Média aritmética:

$$\bar{x} = \frac{\sum x_i}{n} \quad (6)$$

- Desvio Padrão:

$$SD = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} \quad (7)$$

- Erro Padrão (SE):

$$SE = \frac{SD}{\sqrt{n}} \quad (8)$$

- Coeficiente de Variação (CV, %):

$$CV = \frac{SD}{\bar{x}} \quad (9)$$

“n” representa o número de indivíduos e  $x_i$  os valores medidos das variáveis.

Para auxiliar na interpretação dos resultados obtidos em estatística descritiva com o Coeficiente de Variação (CV, %), é utilizado neste estudo a escala classificatória elaborada por Pimentel Gomes (1984): Baixo: < 10%; Médio: 10 – 20%; Alto: 20 – 30%; Muito Alto: > 30, para auxiliar a analisar o nível de dispersão dos dados em relação ao seu valor médio.

Para a avaliação do desempenho estatístico dos modelos de estimativa de  $ET_{\text{FAO-56}}$  foram utilizados em conjunto os seguintes indicativos estatísticos: Weibull, coeficiente de determinação ( $R^2$ ), coeficiente de correlação ( $r$ ), *Mean Bias Error* (MBE), *Root Mean Square Error* (RMSE) o teste  $t$  de Student e o índice de concordância de Willmott ( $d$ ). A utilização conjunta desses indicativos auxilia na análise dos modelos de melhor desempenho na estimativa de  $ET_{\text{FAO-56}}$ .

O método de Weibull considera a probabilidade de densidade referente a uma função de distribuição de frequência de um conjunto de dados (WEIBULL, 1951). Utiliza-se nesse conjunto de dados técnicas analíticas e gráficas para determinação de um modelo estatístico.

A função de distribuição acumulada é representada por meio da Equação (10):

$$f(u) = \frac{k}{c} \left(\frac{u}{c}\right)^{k-1} \exp\left\{-\left[\left(\frac{u}{c}\right)^k\right]\right\} \quad (10)$$

quando  $u \geq 0$ , e,  $f(u) = 0$  quando  $< 0$ .

em que:  $u$  (variável);  $c$  (parâmetro da escala com dimensões) e  $k$  (parâmetro de forma sem dimensões).

Usualmente, este método de distribuição de probabilidade Weibull, é utilizado no âmbito de pesquisa do campo florestal e variáveis climáticas (ASSAD e CASTRO, 1991; BURIOL et al., 2000; BURIOL et al., 2001; CATALUNHA et al., 2002; CARGNELUTTI FILHO et al., 2004; SANSIGOLO, 2005; BINOTI et al., 2010; GABRIEL FILHO et al., 2011; EBLING et al., 2014; ULIANA et al., 2017; ROSA FILHO et al., 2018).

Em que o coeficiente de determinação ( $R^2$ ) indica o nível de variação da variável dependente  $ET_{\text{FAO-56}}$  atribuída às variáveis climáticas (variáveis independentes) (Equação (11)):

$$R^2 = \frac{[\sum_{i=1}^n (Y_i - \bar{Y}_i)(X_i - \bar{X}_i)]^2}{\sum_{i=1}^n (Y_i - \bar{Y}_i)^2 \sum_{i=1}^n (X_i - \bar{X}_i)^2} \quad (11)$$

O coeficiente de correlação ( $r$ ) será utilizado para revelar o grau de relação entre  $ET_{\text{FAO-56}}$  e estimativas de  $ET_{\text{FAO-56}}$  por meio de outras técnicas estatísticas (Equação 12):

$$r = \frac{\sum_{i=1}^n (Y_i - \bar{Y}_i)(X_i - \bar{X}_i)}{\sqrt{\sum_{i=1}^n (Y_i - \bar{Y}_i)^2 \sum_{i=1}^n (X_i - \bar{X}_i)^2}} \quad (12)$$

O *Mean Bias Error* (MBE) (Equação 13) indica por meio da diferença de valores estimados e medidos, o desvio da média para indicar se existe superestimação ou subestimação na estimativa de  $ET_{\text{FAO-56}}$  (JACOVIDES et al., 1995). E quanto mais próximo de zero, melhor o ajuste do modelo.

$$MBE = \frac{\sum_{i=1}^n (Y_i - X_i)}{n} \quad (13)$$

O *Root Mean Square Error* (RMSE) (Equação 14) fornece informações sobre o desempenho de modelos por meio da raiz quadrada do erro médio. Seu valor é sempre positivo e quanto menor os valores obtidos de *RMSE* (espalhamento), melhor desempenho dos modelos (STONE, 1993; JACOVIDES et al., 1995; SILVA et al., 2017b). Ou seja,

$$RMSE = \left[ \frac{\sum_{i=1}^n (Y_i - X_i)^2}{n} \right]^{\frac{1}{2}} \quad (14)$$

$Y_i$  representa os valores estimados,  $\bar{Y}_i$  são os valores médios estimados,  $X_i$  são os valores medidos,  $\bar{X}$  são os valores médios medidos e  $n$  é o número de observações.

Como forma de auxiliar na análise do desempenho dos modelos validados, também são utilizadas neste estudo as versões dos indicativos MBE e RMSE adaptadas em porcentagem [Equações (15) e (16)]:

$$rMBE (\%) = 100 \times \frac{\sum_{i=1}^n (y_i - x_i)}{\bar{X}} \quad (15)$$

$$rRMSE (\%) = 100 \times \left[ \frac{\sum_{i=1}^n (Y_i - X_i)^2}{\bar{X}} \right]^{\frac{1}{2}} \quad (16)$$

Para análise do desempenho dos modelos de estimativa de  $ET_{OFAO-56}$ , é utilizada uma escala classificatória para os diferentes intervalos de  $rRMSE$  (JAMIESON et al., 1991; SILVA et al., 2017b): Excelente se  $rRMSE < 10\%$ ; Bom se  $10\% \leq rRMSE < 20\%$ ; Aceitável se  $20\% \leq rRMSE < 30\%$ ; Pobre se  $rRMSE \geq 30\%$ .

Para complementação da validação dos modelos, o teste  $t$  de Student (Equação 17) é utilizado para avaliar a significância dos modelos de estimativa de  $ET_{OFAO-56}$ .

$$t = \frac{\frac{1}{n} \sum_{i=1}^n y_i - x_i}{SD / \sqrt{n}} \quad (17)$$

Stone (1993) adaptou o teste  $t$  de Student (17) por meio de interpolação com os indicadores estatísticos MBE e RMSE (Equação 18):

$$t = \left( \frac{(N-1) \times MBE^2}{RMSE^2 - MBE^2} \right)^{1/2} \quad (18)$$

para verificar a significância de modelos de estimativa de radiação solar. Posteriormente Jacovides e Kontoyiannis (1995) adaptaram para modelagem da evapotranspiração. Neste estudo, a hipótese de nulidade é rejeitada considerando uma diferença significativa a 5% de probabilidade erro. No caso, entre os valores calculados de  $ET_{OFAO-56}$  e as estimativas por meio das técnicas estatísticas utilizadas.

O índice de concordância de Willmott "d" (Equação 19) também é utilizado no processo de validação dos modelos de estimativa de  $ET_{OFAO-56}$  para indicar o quão perto as estimativas estão das medidas correspondentes a linha de comparação 1:1. O valor de "d" igual a 1 indica que o modelo pode estimar exatamente igual aos valores de referência ou observados.

$$d = 1 - \frac{\sum_{i=1}^M (Y_i - X_i)^2}{\sum_{i=1}^M (|Y_i - \bar{X}| + |X_i - \bar{X}_1|)} \quad (19)$$

Este índice foi divulgado pela primeira vez por Willmott (1981) e este autor considerava que somente o coeficiente de correlação ( $r$ ) seria insuficiente para mostrar o ajuste de modelos. Neste trabalho, Willmott (1981) mostra que esse indicativo pode perfeitamente contribuir na análise do desempenho de modelos ao lado de outros indicativos, como o RMSE. Recentemente Willmott et al. (2012) vem sugerindo alguns aprimoramentos neste indicativo.



### 1.3 RESULTADOS E DISCUSSÕES

Com base nos dados climáticos obtidos das 30 EMA's da região de estudo (integração das regiões intermediárias de São José do Rio Preto e Ribeirão Preto), foram gerados modelos para estimativa de  $ET_{FAO-56}$  por meio de três diferentes técnicas: (1) Análise de Componentes Principais – ACP, técnica multivariada utilizada para predição de  $ET_{FAO-56}$  (e também como técnica para preenchimento de dados faltantes da variáveis climáticas) em escala regional e, posteriormente, para determinação das variáveis climáticas que mais tem influência sobre o processo de evapotranspiração; (2) Regressões Múltiplas - RM, para estimar  $ET_{FAO-56}$  por meio de Regressões Lineares Múltiplas (RLM) e Regressões Não Lineares Múltiplas (RNLM); e (3) Aprendizado de Máquinas (AM), que utilizou as técnicas: ANN (MLP, ANFIS e GRNN) e SVM. Para as técnicas RM e AM foram utilizando quatro diferentes combinações de entradas de variáveis climáticas.

Por fim, os modelos de melhor desempenho estatístico de cada técnica foram comparados entre si.

#### 1.3.1 Modelo 1: Análise de Componentes Principais (ACP)

Martí e Zarzo (2012) demonstraram em seu trabalho que a técnica multivariada Análise de Componente Principais (ACP) pode ser utilizada para estimativa de  $ET_{FAO-56}$  por meio da reconstrução da matriz de dados das variáveis aleatórias com a utilização de algoritmos iterativos.

Neste estudo, o algoritmo EM foi utilizado para reconstrução da base de dados climáticos (H, Tmax, Tmim, UR,  $U_2$ ) de 30 EMA's do Planalto Ocidental Paulista medidas no período 2013-2017. Utilizando metodologia proposta por Josse e Husson (2012a) para preenchimento de valores faltantes na base de dados. Posteriormente a recomposição de valores faltantes das variáveis climáticas, esta é utilizada para estimativa da Evapotranspiração de Referência ( $ET_{FAO56}$ ), método proposto por Allen et al. (1998) no boletim FAO-56.

Neste estudo, a técnica ACP também é utilizada para predição de  $ET_{FAO-56}$ . Similarmente aos estudos relatados por Martí e Zarzo (2012) e os resultados são apresentados na Tabela 3. Por meio desta tabela é possível analisar que a primeira Componente Principal (CP1) consegue explicar 30,59% da variação total dos dados de  $ET_{FAO-56}$  e possui autovalor ( $\lambda$ ) igual a 501,39; enquanto que a segunda Componente Principal (CP2) pode explicar 15,30% da variabilidade total residual dos

dados, 45,89% da variabilidade total acumulada e autovalor ( $\lambda$ ) igual a 250,72; em seguida, a terceira Componente Principal (CP3) é capaz de explicar 9,11% da variabilidade total residual dos dados, 55,0% da variabilidade total acumulada e autovalor ( $\lambda$ ) igual a 149,35; e, por fim, a Componente Principal 4 (CP4) que pode explicar 8,40 da variabilidade residual total dos dados, 63,40% da variabilidade total acumulada e autovalor ( $\lambda$ ) igual a 137,67.

**Tabela 2.** Componentes Principais (CP), variabilidade ( $\phi$ , %), variabilidade acumulada ( $\phi_{acm}$ , %) e autovalores ( $\lambda$ ).

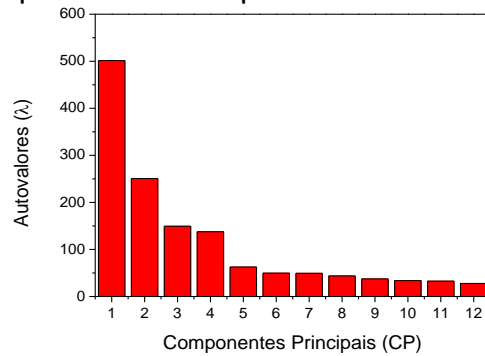
CP	$\phi$	$\phi_{acm}$	$\lambda$
	(%)	(%)	(-)
1	30,59	30,59	501,39
2	15,30	45,89	250,72
3	9,11	55,00	149,35
4	8,40	63,40	137,67

Outros autores consideram valores similares aos obtidos neste trabalho em seus estudos com evapotranspiração e variáveis climáticas (ANDRADE et al., 2003; MARTÍ e ZARZO, 2012; MELO PRADO et al., 2016).

Para a determinação do número mínimo de Componentes Principais, associada à máxima variabilidade acumulada dos dados, é utilizado o método heurístico gráfico do *scree plot*, embora na literatura existam alguns métodos computacionais para essa finalidade (JOSSE e HUSSON, 2012b). Esse método é considerado empírico e observacional, está baseado na escolha do número mínimo de CP associada a uma máxima variância acumulada.

A Figura 5 mostra a relação entre as barras vermelhas e os autovalores ( $\lambda$ ) em cada CP. Os autovalores ( $\lambda$ ) representam a variância acumulada. Nessa figura se verifica que entre as CP4 e CP5 existe uma mudança abrupta na curva de inclinação de ( $\lambda$ ) e que até a quarta CP pode ser escolhida para este estudo. Rencher (2002) e Melo Prado (2016) destacam que o valor do nível de explicação da variação global dos dados deve estar pelo menos entre 60-70% e que sempre é importante escolher um número mínimo de CP.

**Figura 5.** Gráfico de *scree plot* indicando a variabilidade dos dados pela relação: Componentes Principais x Autovalores ( $\lambda$ ).



A estatística descritiva apresentada na Tabela 3 mostra: a média ( $\bar{X}$ ), Desvio Padrão (SD), Erro Padrão (SE), Mínimo (min), Máximo (Max) e Coeficiente de Variação (CV) dos valores estimados de  $ET_{\text{FAO-56}}$  pela técnica multivariada ACP no período avaliado (2013-2017).

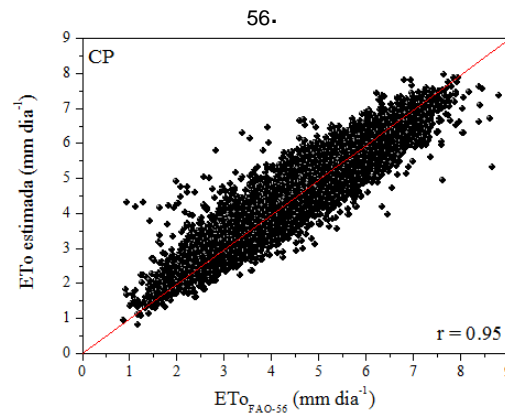
**Tabela 3.** Estatística descritiva do modelo de Análise de Componentes Principais (CP).

Modelos	$\bar{X}$	SD	SE	Min	Max	CV
	(mm dia <sup>-1</sup> )	(mm dia <sup>-1</sup> )	(mm dia <sup>-1</sup> )	(mm dia <sup>-1</sup> )	(mm dia <sup>-1</sup> )	(%)
$ET_{\text{FAO-56}}$	4,41	1,25	0,01	0,87	8,82	28,45
$ET_{\text{ACP}}$	4,41	1,19	0,01	0,82	7,97	26,93

Nível de confiança: 95%.

O modelo  $ET_{\text{ACP}}$  apresentou valor médio similar ao de  $ET_{\text{FAO-56}}$ , com valor máximo estimado de evapotranspiração 9,64% abaixo do obtido por  $ET_{\text{FAO-56}}$ . A distribuição de frequência de  $ET_{\text{ACP}}$  encontra-se disponível no Apêndice D.

A Figura 6, mostra a correlação entre as estimativas de Evapotranspiração de Referência pela técnica Componentes Principais ( $ET_{\text{CP}}$ ) com 4 componentes e os valores calculados por  $ET_{\text{FAO-56}}$ . A distribuição dos valores estimados por  $ET_{\text{CP}}$  e  $ET_{\text{FAO-56}}$  estão em concordância linear com a linha ideal de comparação (1:1). O valor do coeficiente de correlação linear ( $r$ ) obtido é ( $r = 0,95$ ).

**Figura 6.** Comparação entre os valores estimados por  $ET_{O_{CP}}$  e calculado de  $ET_{O_{FAO-56}}$ 

A Tabela 4 mostra os resultados obtidos após a validação da técnica multivariada Análise de Componentes Principais (ACP) com os diferentes indicativos estatísticos utilizados para avaliar seu desempenho. Nesta tabela, verifica-se que os seguintes valores obtidos pelos indicativos de validação:  $rRMSE_{ACP} = 9,33\%$ ;  $rMBE_{ACP} = 0,03$ ;  $d_{ACP} = 0,97$ ;  $t_{cac} = 0,42$ . Indicando que o modelo de  $ET_{O_{ACP}}$  tem espalhamento em torno de 9,33%, valor considerado por Jamieson (1991) como excelente para avaliação de desempenho estatístico de modelos; baixo valor de  $rMBE$  mostrando que o modelo de  $ET_{O_{ACP}}$  não tem tendência de subestimar ou superestimar o modelo  $ET_{O_{FAO-56}}$ ; valor de  $d$  próximo de 1; Pelo teste  $t$ , verifica-se que a média dos valores previstos de  $ET_{O_{FAO-56}}$  pelo modelo ( $ET_{O_{ACP}}$ ) não diferem estatisticamente do valor médio observado em  $ET_{O_{FAO-56}}$ .

Ou seja, os indicativos estatísticos, utilizados para avaliar o desempenho do modelo  $ET_{O_{ACP}}$ , revelam que esse modelo estima  $ET_{O_{FAO-56}}$  com excelente precisão e exatidão de maneira significativa.

**Tabela 4.** Indicativos estatísticos e comparação com a literatura.

Autores	Localidades	Latitudes	$rRMSE$	RMSE	$rMBE$	MBE	$d$	$t_{calc}$	$t_{crit}$
	(n°)	(°)							
Martí e Zarzo (2012)	30 (Espanha)	37°52'12" - 40°60' N	-	0,32 - 0,33	-	-	-	-	-
<b>Presente estudo</b>	<b>Região de estudo</b>	<b>19°32' - 21°57' S</b>	<b>9,33</b>	<b>0,41</b>	<b>0,03</b>	<b>0,00</b>	<b>0,97</b>	<b>0,42</b>	<b>1,96</b>

Os valores obtidos pelo índice estatístico RMSE neste estudo ( $RMSE_{ACP} = 0,411 \text{ mm dia}^{-1}$ ), foi superior aos encontrados por Martí e Zarzo (2012) na Espanha, como mostrado na Tabela 5. Esses autores também trabalharam dados climáticos de 30 localidades da região de Valência.

### 1.3.2 Modelos 2

Nesta seção é realizada inicialmente uma análise exploratória das variáveis climáticas (H, Ho, Tmax, Tmin, UR e U<sub>2</sub>) por meio da técnica ACP com o objetivo de verificar o agrupamento de variáveis climáticas mais relevantes nos processos físicos da evapotranspiração. Esses agrupamentos servirão de base para a construção de diferentes modelos de estimativa de ET<sub>FAO-56</sub> com a utilização de Regressões Múltiplas (RM) neste item 2 e com o auxílio de técnicas de Aprendizado de Máquinas (AM) no item 3.

#### 1.3.2.1 Análise exploratória dos dados climáticos

A estatística descritiva apresentada na Tabela 5 mostra: a média ( $\bar{X}$ ), Desvio Padrão (SD), Erro Padrão (SE), Mínimo (min), Máximo (Max) e Coeficiente de Variação (CV) da base de dados climáticos [Temperatura máxima (Tmax), Temperatura mínima (Tmin), Irradiação solar global no topo da atmosfera (Ho), Irradiação solar global (H), Umidade relativa (UR) e Velocidade do vento (U<sub>2</sub>)] no período avaliado (2013-2017). Totalizando 1639 pontos para cada variável climática.

**Tabela 5.** Estatística descritiva da base dados climáticos.

Variáveis climáticas	X (mm dia <sup>-1</sup> )	SD (mm dia <sup>-1</sup> )	SE (mm dia <sup>-1</sup> )	Mínimo (mm dia <sup>-1</sup> )	Máximo (mm dia <sup>-1</sup> )	CV (%)
Tmax	30,42	3,82	0,02	10,40	43,90	12,56
Tmin	17,60	3,82	0,02	-1,00	30,20	21,69
Ho	33,82	6,01	0,03	23,69	41,26	17,76
H	17,42	5,62	0,03	0,00	35,51	32,29
UR	67,36	11,17	0,05	19,00	100,00	16,59
U <sub>2</sub>	1,42	0,64	0,00	0,00	9,07	45,37

Nível de confiança: 95%.

Analisando-se os resultados obtidos nesta Tabela, nota-se que as variáveis climáticas Ho, H e UR apresentam os maiores valores de Desvio Padrão (SD). Maiores valores de SD podem significar que essas variáveis podem ser importantes na Análise de Componentes Principais (ACP). Melo Prado et al. (2016) analisando o clima de Uberlândia (Minas Gerais - Brasil) verificou que as variáveis com maiores valores de SD tiveram alguma influência sobre ACP.

Posteriormente aos processos de preenchimento de dados faltantes e padronização da base climática, a matriz de dados composta pelas variáveis climáticas, é transformada em matriz de correlação. Essa transformação é realizada em razão das variáveis climáticas possuírem diferentes variâncias e unidades de

medida e, após a padronização dessas variáveis, um novo conjunto de dados com média ( $\bar{x}$ ) igual a zero e desvio padrão (SD) igual a um (1) é obtido.

Após a padronização das variáveis climáticas é realizada ACP e seus resultados são mostrados na Tabela 6. Onde é possível analisar que a primeira Componente Principal (CP1) pode explicar 41,0% da variação total dos dados e possui autovalor ( $\lambda$ ) igual 2,46; enquanto que a segunda Componente Principal (CP2) pode explicar 26,0% da variação residual dos dados com autovalor ( $\lambda$ ) igual 1,57; e a terceira Componente Principal (CP3) pode explicar 17,0% da variação residual dos dados com autovalor ( $\lambda$ ) igual 1,05. Essas três Componentes Principais (CP) totalizam variabilidade acumulada de 85,4%.

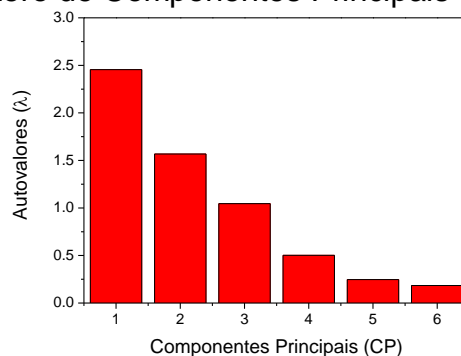
**Tabela 6.** Parâmetros: Autovalores ( $\lambda$ ), variabilidade (Variab., %) e variabilidade acumulada (Variab. Acm., %).

Parâmetros	CP1	CP2	CP3	CP4	CP5	CP6
$\lambda$	2,46	1,57	1,05	0,50	0,25	0,18
Variab. (%)	41,00	26,0	17,0	0,08	0,04	0,03
Variab. Acm. (%)	41,00	67,0	85,0	0,93	0,97	1,00

Outros autores relatam o sucesso da utilização da técnica CP para análise exploratória de dados climáticos, com a associação da máxima variabilidade de dados acumulados associada a um número mínimo de Componentes Principais (ANDRADE et al., 2003; IKUDAYISI e ADEYERNO, 2016; MELO PRADO et al., 2016).

A Figura 7 mostra a relação entre as barras vermelhas e os autovalores ( $\lambda$ ) em cada CP. Os autovalores ( $\lambda$ ) representam a variância acumulada. Nessa figura, verifica-se que entre as CP3 e CP4 existe uma mudança abrupta na curva de inclinação de ( $\lambda$ ) e que até terceira CP podem ser escolhida para este estudo. A variância acumulada dos três primeiros CP totalizam 84,50%.

**Figura 7.** Scree Plot mostrando a variabilidade dos dados por meio da relação: autovalores ( $\lambda$ ) versus o número de Componentes Principais (CP).



Na tabela 7 são apresentados os valores obtidos para as cargas dos Componentes Principais (CP) associado a cada variável climática. Estas cargas possibilitam avaliar ao grau de relação das variáveis climáticas com a respectiva CP.

**Tabela 7.** Cargas das Componentes Principais (CP) associadas às variáveis climáticas.

Variáveis	CP1	CP2	CP3
$T_{\text{máx}}$	0,55	0,14	0,23
$T_{\text{mín}}$	0,42	-0,51	-0,02
$H_o$	0,50	-0,34	-0,21
$H$	0,49	0,32	0,02
$UR$	-0,20	-0,70	-0,01
$U_2$	0,03	0,13	-0,95

Nesta tabela, verificam-se associações mais altas e positivas da CP1 com as variáveis climáticas:  $T_{\text{máx}}$  (0,55),  $T_{\text{mín}}$  (0,42),  $H_o$  (0,50) e  $H$  (0,49); contrastando com  $UR$  (-0,20), que detém associação negativa com CP1; e  $U_2$  (0,03) que detém baixa associação com esta CP. Essas informações indicam que há um efeito conjunto das variáveis de maior associação na CP1 com o processo de evapotranspiração, desde que seja mantida uma relação inversamente proporcional com a  $UR$  para que CP1 possa explicar 41,0% de toda variabilidade dos dados.

Na CP2, após análise da variabilidade residual do conjunto de dados das variáveis climáticas, verifica-se associações mais altas e negativas com as variáveis climáticas:  $T_{\text{mín}}$  (-0,51),  $H_o$  (-0,34) e  $UR$  (-0,70); em contraste com as relações positivas obtidas com as variáveis:  $T_{\text{máx}}$  (0,14),  $H$  (0,32) e  $U_2$  (0,13). Indicando que nesta CP a variável climática de maior importância é  $UR$  (-0,70) associada às demais variáveis, que podem explicar 26,0% da variabilidade total dos dados e 67,0% da variabilidade total acumulada dos dados.

E por fim, na CP3, verifica-se que a variável  $U_2$  (-0,95) detém a associação mais alta e negativa com esta CP, sendo a variável mais influente no processo de evapotranspiração encontrada na análise desta variabilidade de dados residual; seguido posteriormente pelas variáveis climáticas de associação negativa com CP3:  $H_o$  (-0,21),  $T_{\text{mín}}$  (-0,02) e  $UR$  (-0,01); contrastando com os valores de associações positivas obtidas por:  $T_{\text{máx}}$  (0,23) e  $H$  (0,02). Essas informações indicam que a variável climática  $U_2$  é a mais influente no processo de evapotranspiração quando analisado a terceira componente, considerando um efeito conjunto com outras variáveis, pois esta componente é capaz de explicar 17,0% da variabilidade residual dos dados e 85,0% da variabilidade residual acumulada dos dados.

Os resultados obtidos na análise exploratória de dados indicaram que as variáveis climáticas Tmax, Tmin, Ho, UR e U<sub>2</sub> tem mais influência e efeito direto sobre os processos de evapotranspiração, quando combinadas. Esses resultados foram também constatados por outros pesquisadores em seus trabalhos (MOHAN e ARUMUGAM, 1996; IKUDAYISI e ADEYERNO, 2016; MELO PRADO, 2016).

Após essas constatações, foi possível construir quatro diferentes combinações de variáveis de entrada para modelos de estimativa de ETo<sub>FAO-56</sub> por meio de técnicas de Regressões Múltiplas (RM) e Aprendizado de Máquinas (AM). Para em seguida, avaliar o desempenho estatístico do efeito conjunto dessas variáveis climáticas na estimativa de ETo<sub>FAO-56</sub>.

### 1.3.2.2 Modelo 2: RLM e RNLM

Com base na determinação das variáveis climáticas de maior relevância para a geração de modelos de estimativas de ETo<sub>FAO-56</sub>, foram gerados quatro modelos de Regressões Lineares Múltiplas (RLM) e Regressões Não Lineares Múltiplas, apresentados em sua forma linear pela transformação logarítmica (RNLM).

Na Tabela 8 são apresentadas as equações obtidas para as quatro diferentes combinações de RLM e RNLM com seus respectivos coeficientes de determinação (R<sup>2</sup>).

**Tabela 8.** Equações geradas por meio das técnicas RLM e RNLM para cada combinação adotada e seus respectivos R<sup>2</sup>.

Modelos	Equações	R <sup>2</sup>
RLM1	$ET_o = -3,16 + 0,17(T_{max}) - 0,09(T_{min}) + 0,12(H_o)$	0,59
RLM2	$ET_o = -1,47 + 0,053(T_{max}) - 0,01(T_{min}) + 0,065(H_o) + 0,128(Rn)$	0,75
RLM3	$ET_o = 1,07 + 0,01(T_{max}) + 0,02(T_{min}) + 0,08(H_o) + 0,11(Rn) - 0,03(UR)$	0,78
RLM4	$ET_o = -2,38 + 0,07(T_{max}) + 0,02(T_{min}) + 0,05(H_o) + 0,12(Rn) - 0,01(UR) + 0,93(U_2)$	0,98
RNLM1	$\log(ET_o) = \log(0,02) + 0,92 \log(T_{max}) - 0,20 \log(T_{min}) + 0,80 \log(H_o)$	0,76
RNLM2	$\log(ET_o) = \log(0,05) + 0,4 \log(T_{max}) - 0,04 \log(T_{min}) + 0,62 \log(H_o) + 0,38 \log(Rn)$	0,85
RNLM3	$\log(ET_o) = \log(0,1) + 0,44 \log(T_{max}) - 0,03 \log(T_{min}) + 0,71 \log(H_o) + 0,320 \log(Rn) - 0,25 \log(UR)$	0,86
RNLM4	$\log(ET_o) = \log(0,01) + 0,65 \log(T_{max}) - 0,01 \log(T_{min}) + 0,43 \log(H_o) + 0,39 \log(Rn) + 0,21 \log(UR) + 0,30 \log(U_2)$	0,96

Verifica-se na Tabela 8 que os valores de R<sup>2</sup> crescem no sentido do aumento das variáveis climáticas melhorando o ajuste das equações. O maior ganho no desempenho ocorre da terceira para a quarta combinação com a entrada da variável



$U_2$ . Os efeitos da inserção dessa variável climática são esperados em razão de sua importância física para a ocorrência da evapotranspiração no ambiente. A variável  $U_2$  promove a substituição do ar saturado por água (úmido) evapotranspirada por ar mais seco (ALLEN et al., 1998). Com isso, sempre haverá a manutenção do déficit de saturação de vapor d' água na atmosfera ( $e_s - e_a$ ) para regular a taxa de evapotranspiração a todo o momento.

As equações que são propostas e mostradas na Tabela 8, para os modelos RLM e RNLM, são similares às equações de trabalhos pesquisados na literatura. O coeficiente de determinação para RLM4 ( $R^2 = 0,98$ ) é similar aos valores obtidos, o que mostra o quanto este modelo consegue explicar os valores observados. Inferior aos valores obtidos por Sriram e Rashmi (2014) em Chennai (Índia) com  $R^2 = 0,99$ ; superior aos valores obtidos por Cristea et al. (2013) nos Estados Unidos, com intervalo de variação de  $R^2 = 0,79$  a  $0,98$ ; aos valores obtidos por Tabari e Grismer (2013) em Rash (Irã), com intervalo de variação de  $R^2 = 0,94$  a  $0,98$ ; aos valores obtidos por Khanmohammadi et al. (2018) no Irã, com intervalo de  $R^2 = 0,82$  a  $0,97$ ; aos valores obtidos por Ye et al. (2014) na China com  $R^2 = 0,94$ ; aos valores obtidos por Bogawski e Bednorz (2014) na Polônia, com intervalo de variação de  $R^2 = 0,95$  a  $0,96$ ; aos valores obtidos por Mallikarjuna et al. (2013) na Índia, com intervalo de variação de  $R^2 = 0,89$  a  $0,97$ .

*Estatística descritiva e validação dos modelos de Regressões Múltiplas (RM): Regressões Lineares Múltiplas (RLM) e Regressões Não Lineares Múltiplas (RNLM)*

A estatística descritiva apresentada na Tabela 9 mostra: a média ( $\bar{X}$ ), Desvio Padrão (SD), Mínimo (min) e Máximo (Max) dos valores estimados de  $ET_{O_{FAO-56}}$  pelas técnicas de Regressões Múltiplas (RM) no período avaliado (2013-2017) com as quatro combinações de entrada desses modelos.

**Tabela 9.** Estatística descritiva dos modelos de Regressões Múltiplas (RM): Regressões Lineares Múltiplas (RLM) e Regressões Não Lineares Múltiplas (RNLM).

Modelos	$\bar{X}$	SD	SE	Min	Max	CV
	(mm dia <sup>-1</sup> )	(mm dia <sup>-1</sup> )	(mm dia <sup>-1</sup> )	(mm dia <sup>-1</sup> )	(mm dia <sup>-1</sup> )	(%)
ET <sub>FAO-56</sub>	4,41	1,25	0,01	0,87	8,82	28,45
Log(ET <sub>FAO-56</sub> )	0,62	0,13	0,00	0,01	0,95	21,44
RLM1	4,49	0,97	0,01	1,08	7,22	21,54
RLM2	4,40	1,09	0,01	0,90	7,23	24,68
RLM3	4,34	1,12	0,01	0,67	7,49	25,88
RLM4	4,53	1,29	0,01	0,13	9,16	28,42
RNLM1	0,62	0,09	0,00	0,19	0,80	14,05
RNLM2	0,63	0,11	0,00	0,03	0,83	18,14
RNLM3	0,63	0,12	0,00	0,01	0,89	18,91
RNLM4	0,62	0,13	0,00	0,01	0,96	20,86

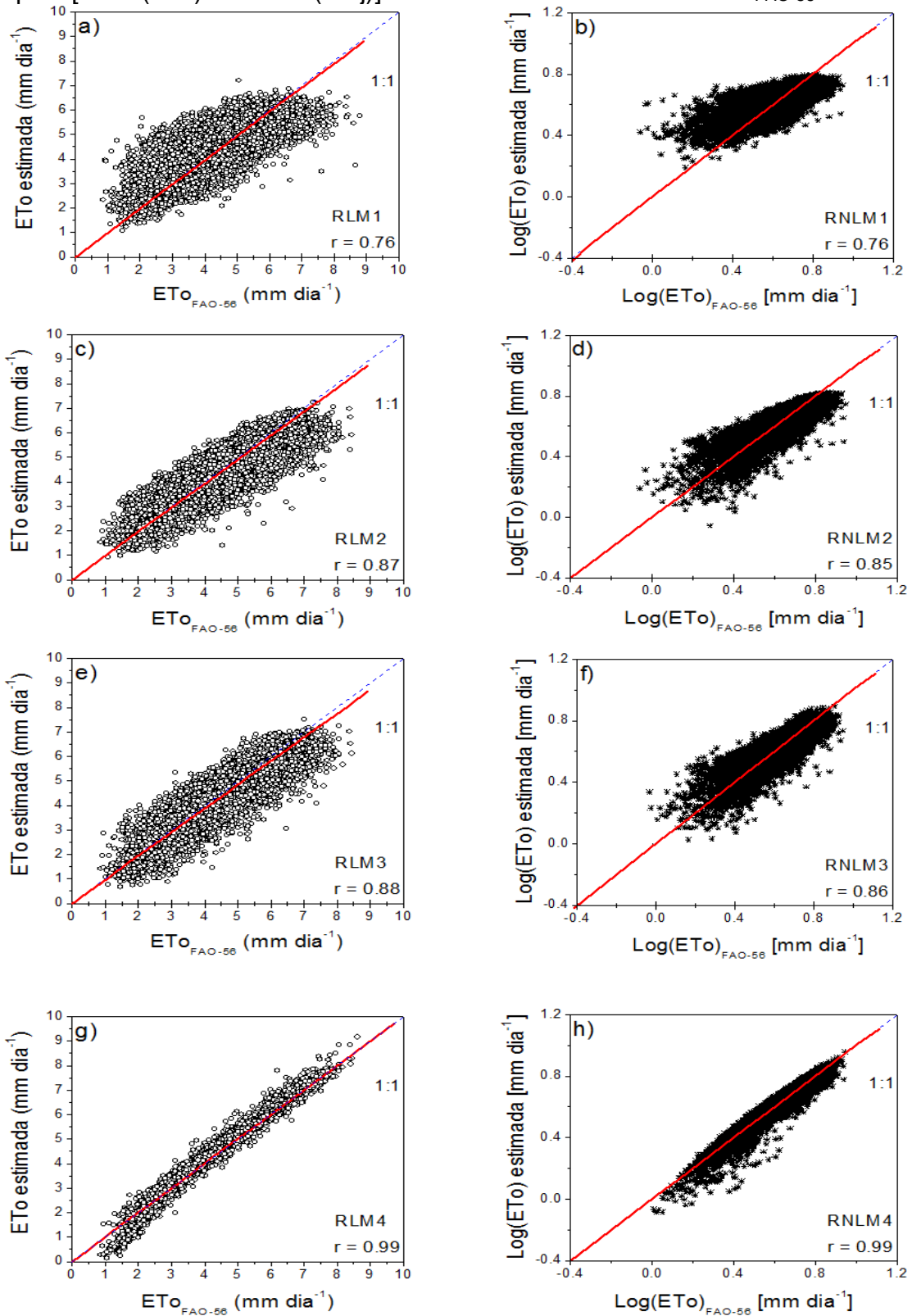
Nível de confiança: 95%.

Os resultados mostrados na Tabela 9 mostram que os resultados dos modelos RLM4 e RNLM4 apresentam os maiores SD e valores máximos de estimativa de ET<sub>FAO-56</sub>. Os resultados obtidos pelos modelos da quarta combinação indicam a existência de maior variabilidade de seus valores em relação as suas médias. A distribuição de frequência dos modelos de RM estimando ET<sub>FAO-56</sub> estão inseridas no Apêndice E.

Para avaliar o desempenho dos modelos RLM e RNLM foram utilizados os indicadores estatísticos: coeficiente de correlação ( $r$ ), *Root Mean Squared Error* (RMSE), *Mean Bias Error* (MBE), Índice de Concordância de Willmott ( $d$ ) e teste  $t$ . No caso dos indicadores estatísticos: RMSE e MBE. Eles são utilizadas adicionalmente em versão de porcentagem: *Relative Root Mean Square Error* (rRMSE) e *Relative Mean Bias Error* (rMBE). Para auxiliar na interpretação do desempenho destes modelos.

A Figura 8 (a - j) mostra a correlação entre os valores estimados de ET<sub>FAO-56</sub> pelos modelos de Regressões Múltiplas (RLM e RNLM) e os calculados pelo método padrão do boletim FAO-56. E suas retas obtidas por regressão linear com os coeficientes de correlação ( $r$ ) para as quatro combinações. A cada combinação houve a inserção de uma variável climática que alteraram as estimativas de ET<sub>FAO-56</sub> nessas combinações. Como consequência, os valores das correlações e das dispersões foram sendo alteradas gradativamente.

**Figura 8 (a - j).** Correlações entre os valores estimados pelos modelos de Regressões Múltiplas [RLM (a - i) e RNLM (b - j)] e os valores calculados de  $ET_{FAO-56}$ .



Da primeira a quarta combinação há um gradual aumento do valor de  $r$  à medida que são acrescentadas às variáveis climáticas de entrada dos modelos RM.

Essa introdução de variáveis melhora o desempenho com ganho nos valores de  $r$  e aumento no intervalo de variação dos modelos da quarta combinação em relação às combinações anteriores e as correlações apresentam retas cada vez mais próximas da reta ideal (1:1). No apêndice G está inserido o gráfico de correlação do modelo de Hargreaves-Samani (H-S). Observando-o, é possível identificar que o valor de  $r$  deste modelo é similar aos valores da combinação 1.

O coeficiente de correlação para RLM4 ( $r = 0,99$ ) é similar e está no mesmo intervalo de variação dos valores obtidos por Kisi (2011) nos Estados Unidos, com  $r = 0,86$  a  $0,99$ , por Martí e Gonzáles-Altozano (2011) na Espanha - com  $r = 0,95$  a  $0,98$ , para Tabari et al. (2012) em Hamadã (Irã) - com  $r = 0,96$  a  $0,98$ , Manikumari e Vinodhini (2016) em Annamalai Nagar (Índia) - com  $r = 0,98$ , e superior aos valores obtidos por Mahida e Patel (2015) em Bhaniyara (Índia), com intervalo de variação de  $r = 0,75$  a  $0,95$ .

O coeficiente de correlação para RNLM4 ( $r = 0,99$ ) é similar e está no mesmo intervalo de variação dos valores obtidos por Tabari et al. (2012) em Hamadã (Irã) com  $r = 0,89$  a  $0,99$ .

Na Tabela 10 são apresentados os valores obtidos pelos indicativos estatísticos utilizados para avaliar o desempenho dos modelos de RM: *Root Mean Squared Error* (RMSE), *Mean Bias Error* (MBE), Índice de Concordância de Willmott ( $d$ ) e teste  $t$ . No caso dos indicativos estatísticos RMSE e MBE, são utilizadas adicionalmente suas versões em porcentagem: *Relative Root Mean Square Error* (rRMSE) e *Relative Mean Bias Error* (rMBE). Para auxiliar na interpretação do desempenho deste modelo.

**Tabela 10.** Arquitetura e desempenho estatístico dos modelos de RLM e RNLM.

Modelos	Arquitetura	rRMSE	RMSE	rMBE	MBE	$d$	$t$	$t_{crit}$
		(%)	(mm dia <sup>-1</sup> )	(%)	(mm dia <sup>-1</sup> )	(-)	(-)	(-)
RLM1	Tmax, Tmin, Ho	18,54	0,82	2,03	0,09	0,85	13,39*	1,96
H-S	// // //	19,42	0,86	6,56	0,29	0,87	43,62*	1,96
RLM2	Tmax, Tmin, Ho, H	14,25	0,63	-0,10	0,00	0,92	0,87	1,96
RLM3	Tmax, Tmin, Ho, H, UR	13,52	0,60	-1,58	-0,07	0,93	14,33*	1,96
RLM4	Tmax, Tmin, Ho, H, UR, U <sub>2</sub>	5,23	0,23	2,92	0,13	0,99	81,81*	1,96
RNLM1	Tmax, Tmin, Ho	14,14	0,09	-0,07	0,00	0,83	0,64	1,96
H-S	// // //	19,42	0,86	6,56	0,29	0,87	43,62*	1,96
RNLM2	Tmax, Tmin, Ho, H	11,44	0,07	0,06	0,00	0,91	0,59	1,96
RNLM3	Tmax, Tmin, Ho, H, UR	10,90	0,07	0,02	0,00	0,92	0,20	1,96
RNLM4	Tmax, Tmin, Ho, H, UR, U <sub>2</sub>	6,39	0,04	-0,26	0,00	0,98	4,87*	1,96

\*Significativo a  $p < 0,05$  pelo teste  $t$ .

A seguir são descritos detalhadamente os valores obtidos pelos indicativos estatísticos rRMSE, rMBE e  $d$  em cada uma das quatro combinações utilizadas neste

estudo (Tabela 10). Os indicativos RMSE e MBE, em sua versão percentual (rRMSE e rMBE), são utilizados para facilitar o entendimento do desempenho dos modelos.

Na comparação entre os modelos da primeira combinação de RM (RLM1 e RNLM1), verifica-se que a média dos valores previstos de  $ET_{\text{FAO-56}}$  por RLM1 ( $ET_{\text{ORLM1}}$ ) diferem estatisticamente do valor médio observado em  $ET_{\text{FAO-56}}$ , diferentemente de RNLM1 ( $ET_{\text{ORNLM1}}$ ). Ambos os modelos de RM da primeira combinação obtiveram desempenho estatístico superior ao modelo de H-S. Esse modelo possui as mesmas variáveis de entrada dos modelos de RM:  $T_{\text{máx}}$ ,  $T_{\text{mín}}$  e  $H_0$ .

Os resultados mostrados na Tabela 10 revelam que a inserção de novas variáveis a cada combinação resulta em modificações do valor estimado de  $ET_{\text{FAO-56}}$  e alteração da dispersão de cada modelo. Sobretudo, da terceira para a quarta combinação, com a introdução da variável climática vento ( $U_2$ ). Que fisicamente é bastante importante, em efeito conjunto com as demais variáveis climáticas, para os processos de evapotranspiração, com a regulação das taxas locais de evapotranspiração mediante a substituição de ar saturado de vapor de água por ar seco.

Verifica-se ainda na Tabela 10 que, nas quatro primeiras combinações, os modelos de RNLM possuem melhor desempenho do que RLM. E isso se mantém a cada inserção de uma nova variável. Entretanto, na quarta combinação, após a inserção da variável climática vento ( $U_2$ ), e o modelo RLM4 (rRMSE = 5,23%) obtém desempenho estatístico um pouco superior a RNLM4 (rRMSE = 6,39%). Em que não é conclusivo afirmar que RLM4 possua desempenho estatístico superior a RNLM4, pois o índice rRMSE considera o valor médio da base da validação, em que valores pequenos da média da base de validação podem levar a um aumento desproporcional de rRMSE. Ao contrário do índice estatístico RMSE, que considera o número de indivíduos na base da validação, mais detalhes sobre os índices de validação podem ser vistos entre as Equações 14 - 16.

As comparações entre os modelos RLM4 e RNLM4, por meio do indicativo estatístico RMSE, revelam desempenho de RNLM4 um pouco superior que o de RLM4.

A Tabela 11 mostra os valores obtidos pelos indicativos estatísticos obtidos com as técnicas RLM4 e RNLM4 neste estudo e os obtidos por outros pesquisadores em outras localidades do mundo.

**Tabela 11.** Indicativos estatísticos obtidos para RLM e RNLM no Planalto Ocidental Paulista e em outras localidades do mundo.

Autores	Modelos	Localidades	Latitudes	rRMSE	RMSE	rMBE	MBE
		(n°)	(°)	(%)	(mm dia <sup>-1</sup> )	(%)	(mm dia <sup>-1</sup> )
Kisi (2011)	RLM	3 (Estados Unidos)	38°24'04" - 38°31'35" N	-	0,31 - 1,1	-	-
Martí e Gonzáles-Altozano (2011)	RLM	30 (Espanha)	37°52'12" - 40°60' N	-	0,33 - 0,65	-	0,001 - 0,004
Tabari et al. (2012)	RLM	Hamadã (Irã)	35°12' N	-	0,55 - 0,76	-	-
Almedeij (2012)	RLM	Cidade do Kuwait (Kuwait)	29°22'10" N	-	1,68 - 5,30	-	-
Laaboudi et al. (2012)	RLM	Adrar (Argélia)	27°54' N	-	0,45	-	-
Mallikarjuna et al. (2013)	RLM	5 (Índia)	13°05' - 17°19' N	-	0,26 - 0,49	-	-
Xu et al. (2013)	RLM	Nanjing (China)	31°15'15" N	-	0,52	-	-
Tabari e Grismer (2013)	RLM	Rash (Irã)	37°15' N	-	0,18 - 0,26	-	(-0,02) - 0,01
Ladlani et al. (2014)	RLM	Dar el Beida (Argélia)	36°43' N	-	0,59 - 1,26	-	-
Mahida e Patel (2015)	RLM	Bhaniyara (Índia)	22°23' N	-	0,32 - 0,54	-	-
Kisi e Demir (2016)	RLM	Antália (Turquia)	36°42' N	-	0,50	-	-
Silva et al. (2016b)*	RLM	Rio Branco (Brasil)	9°58'29" S	-	0,04 - 0,13	-	-
Manikumari e Vinodhini (2016)	RLM	Annamalai Nagar (Índia)	11°25' N	-	0,26	-	-
Althoff et al. (2018)*	RLM	5 (Brasil)	14°55'48" - 19°36'0" S	-	0,29 - 0,92	-	-
<b>Presente estudo</b>	<b>RLM4</b>	<b>Região de estudo</b>	<b>19°32' - 21°57' S</b>	<b>5,23</b>	<b>0,23</b>	<b>2,92</b>	<b>0,13</b>
Tabari et al. (2012)	RMNL	Hamadã (Irã)	35°12' N	-	0,45 - 1,07	-	-
<b>Presente estudo</b>	<b>RNLM4</b>	<b>Região de estudo</b>	<b>19°32' - 21°57' S</b>	<b>6,39</b>	<b>0,04</b>	<b>-0,26</b>	<b>0,00</b>

Estados brasileiros: <sup>1</sup>Acre e <sup>2</sup>Minas Gerais.

A comparação realizada entre os valores obtidos pelos indicativos estatísticos deste estudo e os divulgados por outros pesquisadores na literatura indicam que o valor de RMSE para RLM4 obtido neste trabalho (Tabela 11) está no mesmo intervalo de variação dos obtidos em Rash (Irã), localizado na latitude 37°15' N e inferior aos resultados obtidos nos Estados Unidos (região localizada na latitude: 38°24'04" - 38°31'35" N), na Espanha (localizada na latitude: 37°52'12" - 40°60' N), em Hamadã no Irã (localizada na latitude: 35°12' N), na Cidade do Kuwait no Kuwait (localizada na latitude: 29°22'10" N), em Adrar na Argélia (localizada na latitude: 27°54' N), na Índia (localizada na latitude: 13°05' - 17°19' N), em Nanjing na China (localizada na latitude: 31°15'15" N), em Dar el Beida na Argélia (localizada na latitude: 36°43' N), em Bhaniyara na Índia (localizada na latitude: 22°23' N), em Antália na Turquia (localizada na latitude: 36°42' N), em Annamalai Nagar na Índia (localizada na latitude: 11°25' N). No Brasil no estado de Minas Gerais (localizada na latitude: 14°55'48" - 19°36'0" S) e superior aos resultados obtidos em Rio Branco no Acre/Brasil (localizada na latitude: 14°55'48" - 19°36'0" S).

Para RNL4, o valor de RMSE obtido neste trabalho é inferior ao obtido em Hamadã no Irã (localizada na latitude: 35°12' N).

### 1.3.3 Modelo 3

A arquitetura utilizada para a construção das combinações dos modelos de técnicas de AM foi à mesma das RM após análise exploratória das variáveis climáticas realizada na seção 1.3.2.1.

#### 1.3.3.1 Modelo 3: AM: MLP, ANFIS, GRNN e SVM

Neste item é inicialmente realizada uma estatística descritiva com as estimativas de  $ET_{FAO-56}$  por meio das técnicas de AM. Posteriormente é realizada a validação dos modelos de AM.

A estatística descritiva apresentada na Tabela 12 mostra: a média ( $\bar{X}$ ), Desvio Padrão (SD), Mínimo (min) e Máximo (Max) dos valores estimados de  $ET_{FAO-56}$  pelas técnicas de Aprendizado de Máquinas (AM) no período avaliado (2013-2017) com as quatro combinações de entrada desses modelos.

**Tabela 12.** Estatística descritiva dos modelos de Aprendizado de Máquinas (AM).

Modelos	$\bar{X}$	SD	SE	Min	Max	CV
	(mm dia <sup>-1</sup> )	(mm dia <sup>-1</sup> )	(mm dia <sup>-1</sup> )	(mm dia <sup>-1</sup> )	(mm dia <sup>-1</sup> )	(%)
ET <sub>FAO-56</sub>	4,41	1,25	0,01	0,87	8,82	28,45
MLP1	4,41	1,02	0,01	1,08	6,66	23,12
MLP2	4,41	1,11	0,01	1,47	7,98	25,19
MLP3	4,41	1,13	0,01	1,30	7,63	25,66
MLP4	4,40	1,25	0,01	0,88	8,85	28,44
ANFIS1	4,41	1,02	0,01	1,08	6,66	23,12
ANFIS2	4,41	1,11	0,01	1,36	7,46	25,07
ANFIS3	4,41	1,13	0,01	1,06	7,55	25,58
ANFIS4	4,40	1,25	0,01	0,81	8,65	28,43
GRNN1	4,40	1,23	0,01	0,98	8,39	27,92
GRNN2	4,40	1,26	0,01	0,84	8,70	28,62
GRNN3	4,40	1,25	0,01	0,82	8,73	28,43
GRNN4	4,39	1,23	0,01	0,95	8,52	28,04
SVM1	4,39	0,99	0,01	1,56	7,31	22,51
SVM2	4,36	1,09	0,01	1,46	7,56	25,08
SVM3	4,36	1,12	0,01	1,48	7,66	25,63
SVM4	4,41	1,25	0,01	0,82	8,71	28,35

Os resultados mostrados na Tabela 12 revelam que os resultados dos modelos MLP4, ANFIS4, e SVM4 apresentam os maiores SD e valores máximos de estimativa de ET<sub>FAO-56</sub>. Assim como o menor valor de estimativa de ET<sub>FAO-56</sub>. Apenas os resultados obtidos na combinação quatro do modelo GRNN contrasta com a dos demais modelos. Em geral, em todos os casos a distribuição normal mostrou-se adequada para a descrição dos modelos propostos (Apêndice E).

Os resultados obtidos pelos modelos da quarta combinação indicam a existência de maior variabilidade de seus valores em relação as suas médias.

As combinações de entrada das variáveis para os modelos de AM foram às mesmas utilizadas pelos modelos de RLM e RNLM. Estas combinações foram determinadas por meio da utilização da técnica multivariada ACP para exploração das variáveis climáticas. Para identificação das variáveis de maior relevância para a geração dos modelos de estimativa de ET<sub>FAO-56</sub>.

Antes da validação dos modelos de AM, foram realizados diversos ensaios para determinação ajuste dos parâmetros de cada técnica que se adequasse a um melhor desempenho estatístico. No caso da técnica MLP, foi verificado que após diversos ensaios realizados, a minimização de erros mostrou-se mais eficiente com 25 neurônios (25 N) (a matriz de peso desta técnica é mostrada no Apêndice F); na técnica ANFIS, o *radii* igual a 0,5 mostrou-se mais adequado à geração de modelos com a utilização de função do tipo gaussiana e número de épocas fixado em 200 interações; na técnica GRNN o valor do parâmetro *spread* mais adequado obtido foi

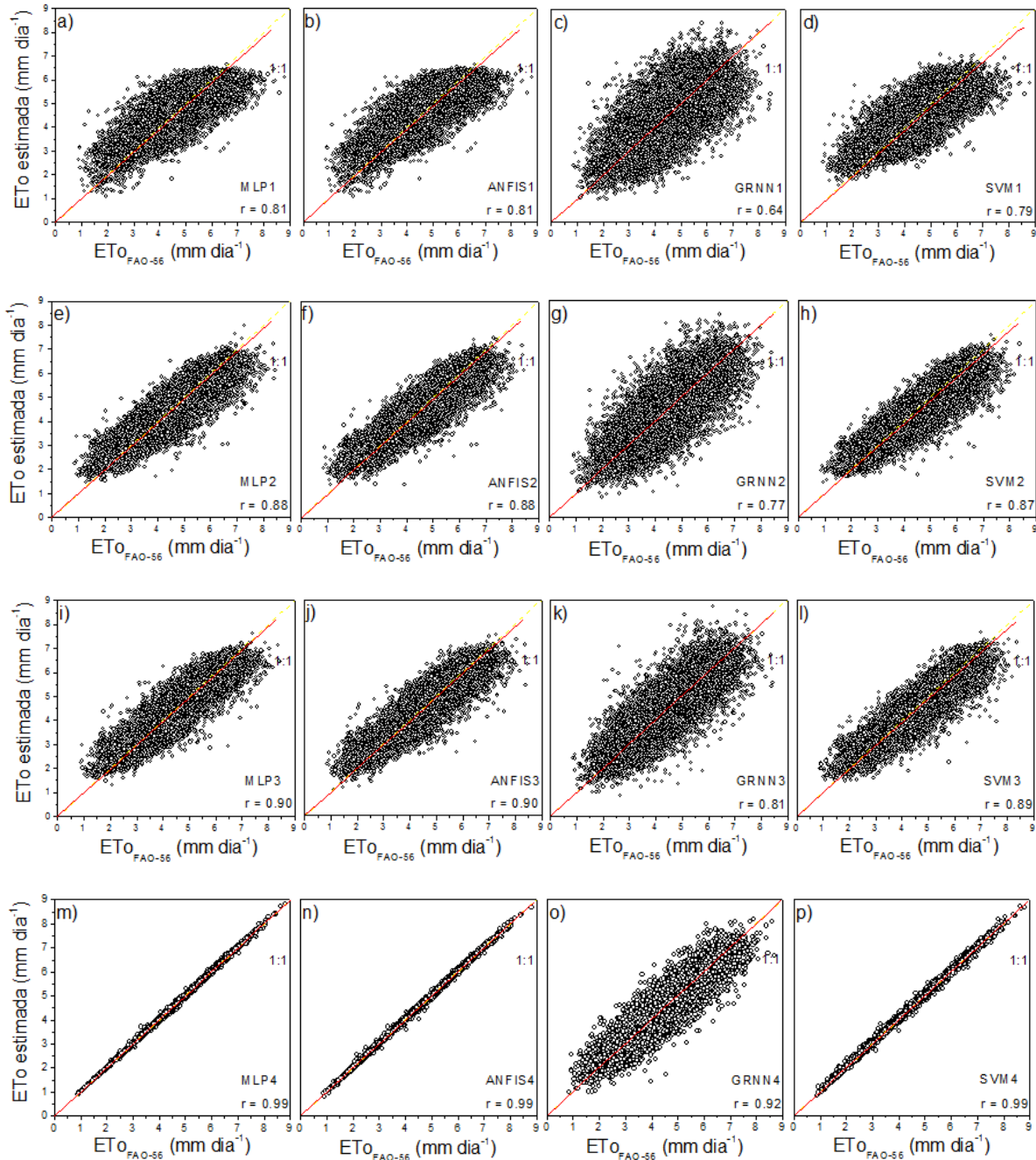


de 0,05; e na técnica SVM os parâmetros de ajuste ( $C$ ,  $\gamma$  e  $\epsilon$ ) da configuração básica do algoritmo SMO resultaram nos seguintes resultados:  $C = \{500\}$ ,  $\gamma = \{0,02\}$  e  $\epsilon = \{0,001\}$ .

#### *Validação dos modelos de AM*

A Figura 9 (a - t) mostra a correlação entre os valores estimados de  $ET_{FAO-56}$  pelos modelos de AM (MLP, ANFIS, GRNN e SVM) e os medidos. As retas obtidas por regressão linear e os coeficientes de correlação ( $r$ ) para as quatro combinações. A inserção progressiva de uma nova variável climática a cada combinação modifica a estimativa de  $ET_{FAO-56}$ , alterando as correlações e a dispersão dos valores da base de validação. As diferenças da dispersão entre as combinações podem ser observadas pela Figura 11 (a - t), comparando-se as dispersões das quatro combinações.

**Figura 9 (a - p).** Correlações entre os valores estimados pelos modelos de AM (MLP, ANFIS, GRNN e SVM) e os valores medidos de  $ET_{FAO-56}$ .



Da primeira a quarta combinação há um gradual aumento do valor de  $r$  à medida que são acrescentadas às variáveis climáticas de entrada dos modelos RM. Essa introdução de variáveis melhora o desempenho com ganho nos valores de  $r$  e aumento no intervalo de variação dos modelos da quarta combinação em relação às combinações anteriores e as correlações apresentam retas cada vez mais próximas da reta ideal (1:1). No apêndice G está inserido o gráfico de correlação do modelo de

Hargreaves-Samani (H-S). Observando-o, é possível identificar que o valor de  $r$  deste modelo é similar aos valores da combinação 1.

O coeficiente de correlação para ANFIS4 ( $r = 0,99$ ) é similar aos valores obtidos por Tabari et al. (2012), com intervalo de variação de  $r = 0,98$  a  $0,99$  e superior aos valores obtidos por Shiri et al. (2012), com intervalo de variação de  $r = 0,92$  a  $0,98$ . O coeficiente de correlação para SVM4 ( $r = 0,99$ ) é igual ao obtido por Manikumari e Vinodhini (2016), com  $r = 0,99$  e similar aos valores obtidos por Kisi e Çimen (2009) com intervalo de variação de  $r = 0,97$  a  $0,99$ , Tabari et al. (2012) com intervalo de variação de  $r = 0,97$  a  $0,99$ , Gocic et al. (2015) com intervalo de variação de  $r = 0,98$  a  $0,99$ , Tangune e Escobedo (2018) com intervalo de variação de  $r = 0,73$  a  $0,99$  e superior aos obtidos por Wen et al. (2015) com intervalo de variação de  $r = 0,77$  a  $0,99$ .

A seguir são descritos detalhadamente os valores obtidos pelos indicativos estatísticos  $rRMSE$ ,  $rMBE$  e  $d$  em cada uma das quatro combinações utilizadas neste estudo (Tabela 13). Os indicativos  $RMSE$  e  $MBE$ , em sua versão percentual ( $rRMSE$  e  $rMBE$ ), são utilizados para facilitar o entendimento do desempenho dos modelos.

**Tabela 13.** Arquitetura e desempenho estatístico dos modelos de AM: MLP, ANFIS, SVM, GRNN e SVM.

Modelos	Combinações	Validação (30% da base total de dados)						
		rRMSE	RMSE	rMBE	MBE	d	t	t <sub>crit</sub>
		(%)	(mm dia <sup>-1</sup> )	(%)	(mm dia <sup>-1</sup> )	(-)	-	-
MLP1 H-S	Tmax, Tmin, Ho	16,64	0,73	0,19	0,01	0,89	0,84	1,96
	// // //	19,42	0,86	6,56	0,29	0,87	43,62*	1,96
MLP2	Tmax, Tmin, Ho, H	13,49	0,59	0,20	0,01	0,93	1,81	1,96
MLP3	Tmax, Tmin, Ho, Rn, UR	12,43	0,55	0,17	0,01	0,95	1,65	1,96
MLP4	Tmax, Tmin, Ho, Rn, UR, U <sub>2</sub>	0,62	0,03	0,00	0,00	0,99	0,09	1,96
ANFIS1 H-S	Tmax, Tmin, Ho	16,81	0,74	0,12	0,01	0,89	0,84	1,96
	// // //	19,42	0,86	6,56	0,29	0,87	43,62*	1,96
ANFIS2	Tmax, Tmin, Ho, H	13,52	0,60	0,14	0,01	0,93	1,25	1,96
ANFIS3	Tmax, Tmin, Ho, Rn, UR	12,60	0,56	0,15	0,01	0,94	1,45	1,96
ANFIS4	Tmax, Tmin, Ho, H, UR, U <sub>2</sub>	0,75	0,03	0,00	0,00	0,99	0,38	1,96
GRNN1 H-S	Tmax, Tmin, Ho	21,82	0,96	-0,13	-0,01	0,83	0,72	1,96
	// // //	19,42	0,86	6,56	0,29	0,87	43,62*	1,96
GRNN2	Tmax, Tmin, Ho, H	18,30	0,81	-0,09	0,00	0,89	0,56	1,96
GRNN3	Tmax, Tmin, Ho, Rn, UR	16,69	0,74	-0,12	-0,01	0,91	0,79	1,96
GRNN4	Tmax, Tmin, Ho, H, UR, U <sub>2</sub>	11,05	0,49	-0,46	-0,02	0,96	5,02*	1,96
SVM1 H-S	Tmax, Tmin, Ho	17,46	0,77	-0,42	-0,02	0,87	2,89*	1,96
	// // //	19,42	0,86	6,56	0,29	0,87	43,62*	1,96
SVM2	Tmax, Tmin, Ho, H	13,83	0,61	-1,12	-0,05	0,93	9,88*	1,96
SVM3	Tmax, Tmin, Ho, Rn, UR	12,89	0,57	-0,99	-0,04	0,94	9,33*	1,96
SVM4	Tmax, Tmin, Ho, Rn, UR, U <sub>2</sub>	1,19	0,05	0,07	0,00	0,99	7,56*	1,96

\*Significativo a  $p < 0,05$  pelo teste  $t$ .

A seguir são descritos detalhadamente os valores obtidos pelos indicativos estatísticos rRMSE, rMBE e d em cada uma das quatro combinações utilizadas neste estudo (Tabela 13). Os indicativos RMSE e MBE, em sua versão percentual (rRMSE e rMBE), são utilizados para facilitar o entendimento do desempenho dos modelos.

Na comparação entre os modelos da primeira combinação de AM (MLP1, ANFIS1, GRNN1 e SVM1), verifica-se que a média dos valores previstos de  $ET_{\text{FAO-56}}$  por SVM1 ( $ET_{\text{SVM1}}$ ) diferem estatisticamente do valor médio observado em  $ET_{\text{FAO-56}}$ , diferentemente dos demais modelos dessa combinação. Todos os modelos dessa combinação, exceto GRNN1 ( $ET_{\text{GRNN1}}$ ), obtiveram desempenho estatístico superior ao modelo de H-S. Esses modelos possuem as mesmas variáveis de entrada dos modelos de AM:  $T_{\text{máx}}$ ,  $T_{\text{mín}}$  e  $H_0$ .

Os resultados mostrados na Tabela 13 revelam que a inserção de novas variáveis a cada combinação resulta em modificações do valor estimado de  $ET_{\text{FAO-56}}$  e alteração da dispersão de cada modelo. Sobretudo, da terceira para a quarta combinação, com a introdução da variável climática vento ( $U_2$ ). Que fisicamente é bastante importante, em efeito conjunto com as demais variáveis climáticas, para os processos de evapotranspiração, com a regulação das taxas locais de evapotranspiração mediante a substituição de ar saturado de vapor de água por ar seco.

Verifica-se ainda na Tabela 13 que, nas quatro primeiras combinações, os modelos de MLP possuem melhor desempenho do que os demais modelos de AM (ANFIS, GRNN e SVM). E isso se mantém a cada inserção de uma nova variável. Entretanto, na quarta combinação, após a inserção da variável climática vento ( $U_2$ ), e o modelo MLP4 (rRMSE = 0,62%) obtém desempenho estatístico um pouco superior a ANFIS4 (rRMSE = 0,75%).

Observa-se através dos indicativos estatísticos apresentados na Tabela 14 que as técnicas MLP, ANFIS e SVM proporcionaram resultados próximos entre as combinações conforme a inserção progressiva das variáveis. Exceto no caso da técnica GRNN.

Deve-se ressaltar que em função do tipo de treinamento e parametrização, a técnica MLP mostra-se a mais simplista, tendo em vista a sua facilidade de implementação, que são dependentes do número de camadas, do número de neurônios, da função de ativação e do tipo de algoritmo de treinamento, que podem ANFIS, existe uma dependência muito grande no número e do tipo de funções de

proporcionar uma convergência e uma aprendizagem melhor. No caso da técnica ativação a ser utilizada, que demanda de esforço computacional grande que nem sempre é satisfatório. Para a técnica SVM, assim como na técnica anterior, a definição do algoritmo de treinamento e a definição da função de transformação são preponderantes para eficiência do treinamento, na qual influenciam o ajustamento e os parâmetros de erro. Já as diferenças apresentadas pela técnica GRNN em relação às outras técnicas, podem estar relacionadas à forma de treinamento da técnica ou pelo número de dados utilizados, que dificultam o treinamento dessa técnica por não ser supervisionado.

A Tabela 14 mostra uma comparação dos indicativos estatísticos [rRMSE (%), RMSE ( $\text{mm dia}^{-1}$ )] obtidos na validação entre os valores medidos e os estimados deste trabalho, específicos para a localidade estuda com as técnicas de AM e suas parametrizações, com os de outros pesquisadores que divulgaram trabalhos semelhantes em outras localidades do mundo.

**Tabela 14.** Indicativos estatísticos para técnicas de AM na região de estudo e em outras localidades do mundo.

Autores	Modelos	Localidades		rRMSE	RMSE
		(n°)	(°)	(%)	(mm dia <sup>-1</sup> )
Kisi (2007)	MLP	3 (Estados Unidos)	34°30' - 34°70' N	-	0,17 - 0,57
Kisi e Ozturk (2007)	MLP	3 (Estados Unidos)	34°02' - 34°03' N	-	0,31 - 0,35
Zanetti et al. (2007) <sup>1,3</sup>	MLP	2 (Brasil)	20°45' - 21°45' S	-	0,59 - 0,79
Landeras et al. (2008)	MLP	4 (Espanha)	42°30' - 43°N	-	0,24 - 0,53
Martí e Gasque (2010)	MLP	30 (Espanha)	37°52'12" - 40°60' N	-	0,32 - 0,66
Traore et al. (2010)	MLP	Bobo Diulasso (Burkina Faso)	11°17' N	-	0,05 - 0,44
Cobaner (2011)	MLP	Santa Mônica (Estados Unidos)	34°02' N	-	0,45 - 1,36
Martí e Gonzáles-Altozano (2011)	MLP	30 (Espanha)	37°52'12" - 40°60' N	-	0,34 - 0,68
Adeloye et al. (2012)	MLP	Edimburgo (Escócia)	55°57' N	-	0,00 - 0,03
Huo et al. (2012)	MLP	3 (China)	37°02' - 38°63' N	-	0,07 - 0,65
Laaboudi et al. (2012)	MLP	Adrar (Argélia)	27°54' N	-	0,27
Mallikarjuna et al. (2013)	MLP	5 (Índia)	13°05' - 17°19' N	-	0,18 - 0,27
Adamala et al. (2014)	MLP	17 (Índia)	9°10' - 32°06' N	-	0,07 - 0,30
Deo e Sahin (2015)	MLP	8 (Austrália)	29°25'48" - 39°07'48" S	-	0,07
Kisi e Demir (2016)	MLP	Antália (Turquia)	36°42' N	-	0,07 - 0,51
Yassin et al. (2016)	MLP	19 (Arábia Saudita)	16°22'46" - 32°14' N	-	0,21 - 3,19
Landeras et al. (2017)	MLP	4 (Gana)	5°36' - 9°36' N	-	0,49 - 0,84
Tangune e Escobedo (2018) <sup>3</sup>	MLP	22 (Brasil)	20°0'36" - 23°58'12" S	-	0,07 - 0,77
<b>Presente estudo</b>	<b>MLP4</b>	<b>Região de estudo</b>	<b>19°32' - 21°57' S</b>	<b>0,62</b>	<b>0,03</b>
Kisi e Ozturk (2007)	ANFIS	2 (Estados Unidos)	34°02' - 34°03' N	-	0,20 - 0,26
Cobaner (2011)	ANFIS	Santa Mônica (Estados Unidos)	34°02' N	-	0,16 - 1,33
Shiri et al. (2012)	ANFIS	4 (Espanha)	42°30' - 43°0' N	-	0,32 - 1,07
Tabari et al. (2012)	ANFIS	Hamedan (Irã)	35°12' N	-	0,06 - 0,08
Ladlani et al. (2014)	ANFIS	Dar el Beida (Argélia)	36°43' N	-	0,45 - 0,85
Shiri et al. (2014)	ANFIS	29 (Irã)	26°32' - 38°45' N	-	0,56 - 0,59
Petkovic et al. (2015)	ANFIS	12 (Sérvia)	42°33' - 45°47' N	-	0,26 - 0,44
<b>Presente estudo</b>	<b>ANFIS4</b>	<b>Região de estudo</b>	<b>19°32' - 21°57' S</b>	<b>0,75</b>	<b>0,03</b>
Kisi (2006)	GRNN	2 (Estados Unidos)	34°02' - 34°03' N	-	0,24 - 0,41
Ladlani et al. (2012)	GRNN	Dar el Beida (Argélia)	36°43' N	-	0,52 - 0,90
Feng et al. (2017)	GRNN	2 (China)	30°48' - 30°40'12" N	-	0,08 - 0,38
<b>Presente estudo</b>	<b>GRNN4</b>	<b>Região de estudo</b>	<b>19°32' - 21°57' S</b>	<b>11,05</b>	<b>0,49</b>
Kisi e Çimen (2009)	SVM	3 (Estados Unidos)	38°24'04" - 38°31'35" N	-	0,11 - 0,54
Tabari et al. (2012)	SVM	Hamedan (Irã)	35°12' N	-	0,02 - 0,07
Shiri et al. (2014)	SVM	29 (Irã)	26°32' - 38°45' N	-	0,53 - 0,61
Gocic et al. (2015)	SVM	12 (Sérvia)	42°33' - 45°47' N	-	0,23 - 0,30
Manikumari e Vinodhini (2016)	SVM	Annamalai Nagar (Índia)	11°25' N	-	0,19
Tangune e Escobedo (2018) <sup>3</sup>	SVM	22 (Brasil)	20°0'36" - 23°58'12" S	-	0,05 - 0,85
Wen et al. (2015)	SVM	Ejina (China)	41°58'54" N	-	0,26 - 0,54
Karimi et al. (2017)	SVM	8 (Coréia do Sul)	35°60'18" - 37°45'4" N	-	0,29 - 0,50
Fan et al. (2018)	SVM	8 (China)	23°11'60" - 45°47'60" N	-	0,11 - 0,87
Althoff et al. (2018) <sup>1</sup>	SVM	5 (Brasil)	14°55'48" - 19°36'0" S	-	0,29 - 0,92
<b>Presente estudo</b>	<b>SVM4</b>	<b>Região de estudo</b>	<b>19°32' - 21°57' S</b>	<b>1,19</b>	<b>0,05</b>

Estados brasileiros: <sup>1</sup>Minas Gerais, <sup>2</sup>Rio de Janeiro e <sup>3</sup>São Paulo no Brasil.

O valor de RMSE para MLP4 obtido neste trabalho (Tabela 14) é similar e está no mesmo intervalo de variação dos valores obtidos em Edimburgo (Escócia) - localizado na latitude 55°57' N, na Índia - entre as latitudes 9°10' - 32°06' N, no Estado de São Paulo (Brasil) - entre as latitudes 20°0'36" - 23°58'12" S; e inferior aos resultados obtidos nos Estados Unidos - entre as latitudes 38°24'04" - 38°31'35" N, nos Estados de Minas Gerais e Rio de Janeiro (Brasil) - entre as latitudes 20°45' - 21°45' S, na comunidade autônoma do País Basco (Espanha) - entre as latitudes 42°30' - 43°N, na comunidade autônoma de Valência (Espanha) - entre as latitudes 37°52'12" - 40°60' N, em Bobo Diulasso (Burkina Faso) - na latitude 11°17' N, em Santa Mônica (Estados Unidos) - na latitude 34°02' N, na China - entre as latitudes 37°02' - 38°63' N, em Adrar (Argélia) - na latitude 27°54' N, na Índia - entre as latitudes 13°05' - 17°19' N, na Austrália - entre as latitudes 29°25'48" - 39°07'48" S, em Antália (Turquia) - na latitude 36°42' N, na Arábia Saudita - entre as latitudes 16°22'46" - 32°14' N e em Gana - entre as latitudes 5°36' - 9°36' N.

O valor de RMSE para ANFIS4 é inferior aos resultados obtidos nos Estados Unidos - entre as latitudes 34°02' - 34°03' N, em Santa Mônica (Estados Unidos) - na latitude 34°02' N, na Espanha - entre as latitudes 42°30' - 43°0' N, em Hamedã (Irã) - na latitude 35°12' N, Dar el Beida (Argélia) - na latitude 36°43' N, no Irã - entre as latitudes 26°32' - 38°45' N, na Sérvia - entre as latitudes 42°33' - 45°47' N.

O valor de RMSE para GRNN4 é inferior aos resultados obtidos em Dar el Beida (Argélia) - na latitude 36°43' N; e superior aos resultados obtidos nos Estados Unidos - entre as latitudes 34°02' - 34°03' S, e na China - entre as latitudes 30°48' - 30°40'12" N.

O valor de RMSE para SVM4 é similar e está no mesmo intervalo de variação dos valores obtidos em Hamadã (Irã) - na latitude 35°12' N, no Estado de São Paulo (Brasil) - entre as latitudes 20°0'36" - 23°58'12" S; e inferior aos valores obtidos nos Estados Unidos - entre as latitudes 38°24'04" - 38°31'35" N, no Irã - entre as latitudes 26°32' - 38°45' N, na Sérvia - entre as latitudes 42°33' - 45°47' N, em Annamalai Nagar (Índia) - na latitude 11°25' N, em Ejina (China) - na latitude 41°58'54" N, na Coreia do Sul - entre as latitudes 35°60'18" - 37°45'4" N, na China - entre as latitudes 23°11'60" - 45°47'60" N e no Estado de Minas Gerais (Brasil) - entre as latitudes 14°55'48" - 19°36'0" S.

Deve-se ressaltar que os valores dos indicativos encontrados são para os arranjos propostos, mas outros devem ser testados. Assim como suas



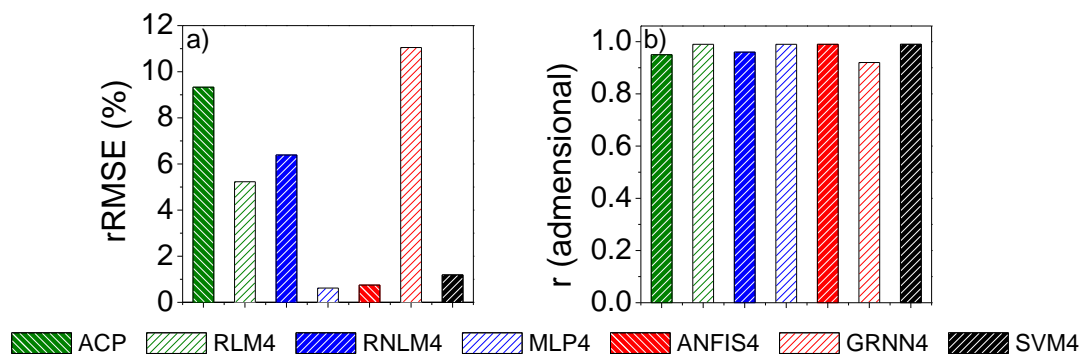
parametrizações a fim de verificar a eficiência dos modelos e da importância das variáveis inseridas nas combinações. E avaliando se estas realmente são impactantes para melhoria dos indicadores estatísticos e se são mais significativas. Por fim, dos resultados apresentados a técnica do tipo MLP é a mais indicada para a estimativa de  $ET_{FAO-56}$  com as combinações e parametrizações propostas.

Adamala (2014) e Pandorfi et al. (2016) relatam o bom desempenho da técnica de AM *Artificial Neural Network* (ANN) na modelagem de evapotranspiração oriundo de medidas lisimétricas. E essa informação indica que possivelmente esta técnica poderia modelar a evapotranspiração melhor que modelo de  $ET_{FAO-56}$ . Desde que essas medidas lisimétricas fossem consideradas confiáveis.

### 1.3.4 Comparação de desempenho estatísticos dos modelos

A Figura 10 (a - b) apresenta os valores obtidos pelos indicadores estatísticos (rRMSE) e (r), por meio de gráficos de barra, para os modelos: ACP ( $ET_{OACP}$ ), RLM4 ( $ET_{ORLM4}$ ), RNL4 ( $ET_{ORNLM4}$ ), MLP4 ( $ET_{OMLP4}$ ), ANFIS4 ( $ET_{OANFIS4}$ ), GRNN4 ( $ET_{OGRNN4}$ ) e SVM4 ( $ET_{OSVM4}$ ). Onde é possível analisar e comparar os valores obtidos pelos indicadores estatísticos desses modelos.

**Figura 10 (a - b).** Comparação do desempenho de modelos por meio dos indicadores estatísticos: *Relative Root Mean Squared Error* (rRMSE) e coeficiente de correlação (r).



Nos gráficos de barras da Figura 10 (a), representando o percentual relativo rRMSE (%), é possível verificar superioridade numérica no desempenho estatístico do modelo MLP4 ( $rRMSE_{MLP4} = 0,62\%$ ) em relação aos demais modelos na seguinte ordem: ANFIS4 ( $rRMSE_{ANFIS4} = 0,75\%$ ), SVM4 ( $rRMSE_{SVM4} = 1,19\%$ ), RLM4

( $rRMSE_{RLM4} = 5,23\%$ ), RNL4 ( $rRMSE_{RNL4} = 6,39\%$ ), ACP ( $rRMSE_{ACP} = 9,33\%$ ) e GRNN4 ( $rRMSE_{GRNN4} = 11,05\%$ )

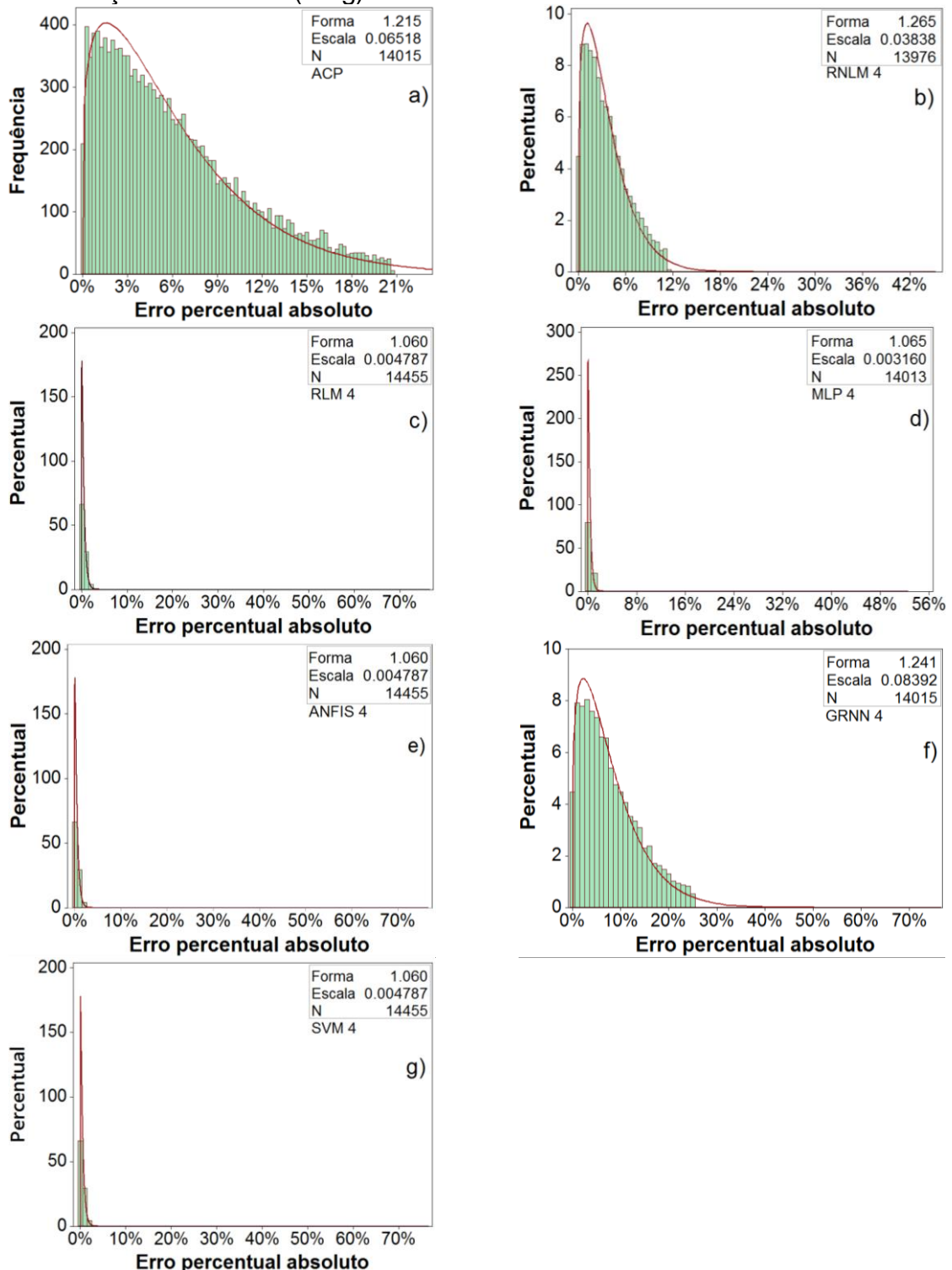
Outros autores como Tabari et al. (2012), Shiri et al. (2014), Tangune e Escobedo (2018) também compararam técnicas entre si e obtiveram intervalo de variação de seus resultados de RMSE semelhantes aos obtidos neste estudo.

Nos gráficos de barras da Figura 10 (b), é possível verificar que o desempenho estatístico do modelo MLP4 ( $r_{MLP4} = 0,99$ ) é similar ao desempenho dos demais modelos: ANFIS4 ( $r_{ANFIS4} = 0,99$ ) e SVM4 ( $r_{SVM4} = 0,99$ ); e numericamente superior aos seguintes modelos na seguinte ordem: GRNN4 ( $r_{GRNN4} = 0,92$ ), ACP ( $r_{ACP} = 0,95$ ), RLM4 ( $r_{RLM4} = 0,96$ ) e RNL4 ( $r_{RNL4} = 0,96$ ).

Os valores de  $r$  obtidos neste estudo são similares aos divulgados por Tabari et al. (2012), Shiri et al. (2014) e Tangune e Escobedo (2018).

Na Figura 11 a distribuição probabilística de Weibull expressa a frequência dos valores do erro percentual absoluto. Nesta distribuição, verifica-se que os valores do erro percentual absoluto (mostrada na Equação 10), para ACP ( $k = 1,22$  e  $c = 0,07$ ), RNL4 ( $k = 1,27$  e  $c = 0,04$ ), RLM4 ( $k = 1,06$  e  $c = 0,01$ ), MLP4 ( $k = 1,07$  e  $c = 0,003$ ), ANFIS4 ( $k = 1,06$  e  $c = 0,005$ ), GRNN4 ( $k = 1,20$  e  $c = 0,08$ ) e SVM4 ( $k = 1,06$  e  $c = 0,005$ ).

**Figura 11.** Distribuição do erro percentual absoluto da estimativa de  $ET_{FAO-56}$  por meio da função de Weibull (a - g).



Os valores das variações do parâmetro  $c$ , quando se mantém o parâmetro  $k$ , acarretam compressão ou expansão da curva. Uma vez que, aproximadamente, o parâmetro  $k$  não teve variações abruptas, pode-se afirmar que valores menores do parâmetro estão associados a uma compressão das curvas.

É possível verificar o ganho no desempenho dos modelos à medida que ocorre a inserção de novas variáveis. Além disto, comparando-se gradativamente os modelos 1 a 4, percebe-se uma redução sensível no parâmetro de escala, o que ocorre devido à concentração dos valores do erro percentual absoluto próximo a 0%. Isto revela nitidamente que, quanto há uma evolução na acurácia do modelo 1 ao 4, sendo obviamente este último (modelo 4) o que apresenta menores erros possíveis.

Matematicamente as variações dos valores do parâmetro  $c$ , quando se mantém o parâmetro  $k$ , acarretam em compressão ou expansão da curva. Quando não há variações abruptas do parâmetro  $k$ , pode-se concluir que os menores valores deste parâmetro estão associados a uma compressão das curvas, o que é observado gradativamente quando é feita a comparação do modelo 1 até o modelo 4.

Contudo, ressalta-se que há situações nas quais não existem a quantidade de informações meteorológicas disponíveis que tal modelo 4 exige, o que naturalmente faz-se necessário utilizar os demais modelos, de acordo então com o conjunto de informações disponíveis.

Os valores das variações no parâmetro  $c$ , quando se mantém o parâmetro  $k$ , acarretam compressão ou expansão da curva. Uma vez que, aproximadamente, o parâmetro  $k$  não teve variações abruptas, pode-se afirmar que valores menores do parâmetro estão associados a uma compressão das curvas, o que é observado gradativamente quando é feita a comparação do modelo 1 até o modelo 4.

Entretanto, ressalta-se que há situações nas quais não existem a quantidade de informações meteorológicas disponíveis que tal modelo 4 exige, o que naturalmente faz-se necessário utilizar os demais modelos, de acordo então com o conjunto de informações disponíveis.

## 1.4 CONCLUSÃO

Dos três métodos apresentados para modelagem de  $ET_{FAO-56}$  neste estudo, conclui-se que:

- A utilização da técnica multivariada (ACP) é uma ferramenta útil para análise exploratória de variáveis climáticas. Essa técnica analisa o efeito conjunto das variáveis e determina gradualmente a variável mais importante para a construção de modelos de estimativa de  $ET_{FAO-56}$ . E técnica é capaz de realizar estimativa de  $ET_{FAO-56}$  com base na proximidade entre as EMA's com boa precisão e exatidão.

- Os valores dos indicadores estatísticos da base de validação de RLM4 e RNLM4 indicam que ambos os modelos de regressões múltiplas podem ser utilizados na estimativa de  $ET_{OFAO-56}$  com boa precisão e exatidão.
- Os modelos de AM estudados (MLP, ANFIS, GRNN e SVM) corroboraram resultados já observados sobre a eficiência dessas técnicas em estimar  $ET_{OFAO-56}$ . Para estudos futuros, testar a eficiência desta técnica em estimar medidas lisimétricas de evapotranspiração seria de grande relevância para mostrar que o esforço computacional das AM são capazes de estimar a evapotranspiração com desempenho estatístico superior ao do método padrão da FAO ( $ET_{OFAO-56}$ ).
- Na comparação do desempenho estatístico entre as técnicas utilizadas neste estudo, verificou-se melhor desempenho estatístico na seguinte ordem: AM (MLP4, ANFIS4 e SVM4), RM (RLM4 e RNLM4) e ACP.

## REFERÊNCIAS

- ABDOU, H. M.; FLURY, M. Simulation of water flow and solute transport in free-drainage lysimeters and field soils with heterogeneous structures. **European Journal of Soil Science**, v. 55, pp. 229-241, 2004.
- ABTEW, W. Evapotranspiration measurements and modeling for three wetland systems in South Florida. **Journal of the American Water Resources Association**, v. 127, n. 3, pp. 140-147, 1966.
- AGRITEMPO – Sistema de Monitoramento Agrometeorológico. Estações meteorológicas para o Estado de São Paulo. Disponível em: <<http://www.agritempo.gov.br/agritempo/jsp/Estacao/index.jsp?siglaUF=SP>>. Acesso em 28 de dezembro de 2018.
- ADAMALA, S.; RAGHUWANSHI, N. S.; MISHRA, A. et al. Neural Computing Modeling of the reference crop evapotranspiration. **Journal of Hydrologic Engineering**, v. 19, pp. 1131-1140, 2014.
- ADELOYE, A. J.; RUSTUM, R.; KARIYAMA, I. D. Neural Computing Modeling of the reference crop evapotranspiration. **Environmental Modelling e Software**, v. 29, pp. 61-73, 2012.
- ALBRECHT, F. Die Methoden zur Bestimmung Verdunstung der natürlichen Erdoberfläche. **Arch Meteor Geoph Biokl**, Serie, B2, v. 1–38, 1950.
- ALLEN, R. G.; PRUITT, W. O. FAO-24 reference evapotranspiration factors. **Journal of Irrigation and Drainage Engineering**, v. 175, n. 5, pp. 758 - 773, 1991.
- ALLEN, R. G.; PEREIRA, L. S.; RAES, D. et al. Crop evapotranspiration - guidelines for computing crop water requirements. pp. 50. **Food and Agriculture Organization of the United Nations (FAO) - Irrigation and drainage paper 56**, Rome, 1998.
- ALMEDEIJ, J. Modeling Pan Evaporation for Kuwait by Multiple Linear Regression. **The Scientific World Journal**, Article ID 574742, 9 pages, doi: 10.1100/2012/574742, 2012.
- ALTHOFF, D.; BAZAME, H. C.; FILGUEIRAS, R. et al. Heuristic methods applied in reference evapotranspiration modeling. **Ciência e Agrotecnologia**, v. 42, n. 3, pp. 314-324, 2018.

- ANDRADE, E. M.; SILVEIRA, S. S.; AZEVEDO, B. M. Investigações da Estrutura Multivariada da Evapotranspiração na Região Centro Sul do Ceará pela Análise de Componentes Principais. **Revista Brasileira de Recursos Hídricos**, v. 8, pp. 39-44, 2003.
- ASSAD, E.D.; CASTRO, L.H.R. Análise freqüencial da pluviometria para a estação de Sete Lagoas, MG. **Pesquisa Agropecuária Brasileira**, v.26, p.397-402, 1991.
- AYTEK, A. Co-active Neurofuzzy Inference System for Evapotranspiration modeling. **Soft Computing**, v. 13, pp. 691-700, 2009.
- BENEVIDES, J. G; LOPEZ, D. Formula para el caculo de la evapotranspiracion potencial adaptada al tropico (15° N - 15° S). **Agronomia Tropical**, v. 20, n. 5, pp. 335-345, 1970.
- BILGILI, M. Prediction of soil temperature using regression and artificial neural network models. **Meteorology and Atmospheric Physic**, v. 110, pp. 59-70, 2010.
- BINOTI, D. H. B.; LEITE, H. G; NOGUEIRA, G. S. et al. Uso da função Weibull de três parâmetros em um modelo de distribuição diamétrica para plantios de eucalipto submetidos a desbaste. *Revista Árvore*, v.34, p.147-156, 2010.
- BOGAWSKI, P.; Bednorz, E. Comparison and Validation of Selected Evapotranspiration Models for Conditions in Poland (Central Europe). **Water Resources Management**, v. 28, pp. 5021-5038, 2014.
- BOWEN, I. S. The ratio of heat losses by conduction and by evaporation from any water surface. **Physical Review**, v. 27, pp. 779-787, 1926.
- BURIOL, G.A.; ESTEFANEL, V.; ANDRIOLO, J.L.; MATZENAUER, R.; TAZZO, I.F. Disponibilidade de radiação solar para o cultivo do tomateiro durante o inverno no Estado do Rio Grande do Sul. **Pesquisa Agropecuária Gaúcha**, v.6, p113-120, 2000.
- BURIOL, G.A.; HELDWEIN, A.B.; ESTEFANEL, V.; MATZENAUER, R.; MARCON, I.A. Probabilidade de ocorrência de valores de radiação solar prejudiciais ao cultivo do pepineiro na região do baixo vale do Taquari, RS. **Pesquisa Agropecuária Gaúcha**, v.7, p.97-104, 2001.
- BLANEY, H. F.; CRIDDLE, W. D. **Determining Water Requirements in Irrigated Areas from Climatological and Irrigation Data**. USDA. SCSTP- 96, p. 48, 1950.
- BRAY, M.; HAN, D. Identification of support vector machines for runoff modeling. **Journal of Hydroinformatics**, v. 6, n. 4, pp. 265–280, 2004.
- BROCKAMP, B.; WENNER, H. Verdunstungsmessungen auf den Steiner See bei Münster. **Dt Gewsserkuändl**, Mitt. 7, pp. 149–154, 1963.
- CARGNELUTTI FILHO, A.; MATZENAUER, R.; TRINDADE, J. K. Ajustes de funções de distribuição de probabilidade à radiação solar global no Estado do Rio Grande do Sul. *Pesq. agropec. bras.*, Brasília, v.39, n.12, p.1157-1166, 2004.
- CARVALHO, L. G.; RIOS, G. F. A.; MIRANDA, W. L. et al. Evapotranspiração de Referência: uma abordagem atual de diferentes métodos de estimativa. **Pesq. Agropec. Trop.**, Goiânia, v. 41, n. 3, p. 456-465, jul./set. 2011.
- CATALUNHA, M.J.; SEDIYAMA, G.C.; LEAL, B.G.; SOARES, C.P.B.; RIBEIRO, A. Aplicação de cinco funções densidade de probabilidade a séries de precipitação pluvial no Estado de Minas Gerais. **Revista Brasileira de Agrometeorologia**, v.10, p.153-162, 2002.
- CENTRO DE PESQUISAS METEOROLÓGICAS E CLIMÁTICAS APLICADAS A AGRICULTURA - CEPAGRI. **Clima dos Municípios Paulistas**. Disponível em: <<http://www.cpa.unicamp.br/outras-informacoes/clima-dos-municipios-paulistas.html>>. Acesso em: 10 de fev. 2018.
- COBANER, M. Evapotranspiration estimation by two different neuro-fuzzy inference systems. **Journal of Hydrology**, v. 398, pp. 292-302, 2011.
- COELHO, C. A. S.; CARDOSO, D. H. F.; FIRPO, M. A. Precipitation diagnostics of an exceptionally dry event in São Paulo, Brazil. **Theoretical and Applied Climatology**, DOI 10.1007/s00704-015-1540-9, 2016a.

- COELHO, C. A. S.; OLIVEIRA, C. P.; AMBRIZZI, T. et al. The 2014 southeast Brazil austral summer drought: regional scale mechanisms and teleconnections. **Climate Dynamics**, DOI 10.1007/s00704-015-1540-9, 2016b.
- CUENCA, R. H. Irrigation system design: an engineering approach. **Prentice-Hall, Englewood Cliffs**, NJ, p. 133, 1989.
- CHANG, C.; SU, S. J. The application of a full counter-propagation neural network to image water-marking. *Proceeding of the IEEE International Conference Networking, Sensing and Control*, Tucson, Arizona, pp. 993-998, 2005.
- CHEN, J-L.; LI, G-S.; WU, S-J. Assessing the potential of support vector machine for estimating daily solar radiation using sunshine duration. **Energy Conversion and Management**, v.75, 311–318, 2013.
- CHEN, J-L.; LI, G-S.; XIAO, B. B. Assessing the transferability of support vector machine model for estimation of global solar radiation from air temperature. **Energy Conversion and Management**, v. 89, pp. 318–329, 2015.
- CRISTEA, N. C.; KAMPF, S. K.; BURGESS, S. J. Linear models for estimating annual and growing season reference evapotranspiration using averages of weather variables. **International Journal of Climatology**, v. 33, pp. 376–387, 2013.
- DALTON, J. M. Experimental essays on the constitution of mixed gases; on the force of steam of vapour from waters and other liquids in different temperatures, both in a torricellian vacuum and in air on evaporation and on the expansion of gases by heat. **Memoirs and proceedings of the Manchester Literary e Philosophical Society**, v. 5, pp. 535-602, 1802.
- DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum Likelihood from incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, v. 39, n. 1, pp. 1-38, 1977.
- DEO, R. C.; SAHIN, M. Application of the Artificial Neural Network model for prediction of monthly Standardized Precipitation and Evapotranspiration Index using hydrometeorological parameters and climate indices in eastern Australia. **Atmospheric Research**, v. 161-162, pp. 65-81, 2015.
- DOOREMBOS, J.; PRUITT, W. Crop water requirements. Rome: FAO (Irrigation and drainage paper, 24), 1977.
- EBLING, Â. A.; PELISSARI, A. L.; ABRÃO, S. F.; BAMBERG, R. Prognose da estrutura diamétrica de remanescente de floresta com Araucária utilizando a função densidade de probabilidade Weibull. **Revista Agroambiente**, v. 8, p. 112-118, 2014.
- FAN, J.; YUE, W.; WU, L. et al. Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China. **Agricultural and Forest Meteorology**, v. 263, pp. 225-241, 2018.
- FARIA, R. T.; CAMPECHE, F. S.; CHIBANA, E. Y. Construção e calibração de lisímetros de alta precisão. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 10, pp. 237-242, 2006.
- FENG, Y.; CUI, N.; GONG, D. ET al. Evaluation of random forests and generalized regression neural networks for daily reference evapotranspiration modelling. **Agricultural Water Management**, v. 193, pp. 163-173, 2017.
- FERREIRA SILVA, C. O.; MANZIONE, R. L.; ALBUQUERQUE FILHO, J. L. Large-Scale Spatial Modeling of Crop Coefficient and Biomass Production in Agroecosystems in Southeast Brazil. **Horticulturae**, doi: 10.3390/horticulturae 4040044, 2018.
- GABRIEL FILHO, L.R. A.; CREMASCO, C. P.; SERAPHIM, O. J.; CANEPPELE, F.L. Caracterização analítica e geométrica da metodologia geral de determinação de distribuições de Weibull para o regime eólico e suas aplicações. **Engenharia Agrícola**, Jaboticabal, v.31, n.1, p.56-66, 2011.

GARCÍA-DIEGO, F.; F. J.; ZARZO, M. Microclimate monitoring by multivariate statistical control: The renaissance frescoes of the Cathedral of Valencia (Spain). **Journal of Cultural Heritage**, v. 11, pp. 339-344, 2010.

GOCIC, M.; MOTAMEDI, S. SHAMSHIRBAND, S. et al. Soft computing approaches for forecasting reference evapotranspiration. **Computers and Electronics in Agriculture**, v. 113, pp. 164-173, 2015.

GROSBERG, S. Adaptive pattern classification and universal recording: part I (parallel development and coding of neural feature detectors). **Biological Cybernetics**, v. 23, n. 3, pp. 121-134, 1976a.

GROSBERG, S. Adaptive pattern classification and universal recording: part II (Feedback, expectation, olfaction, and illusions). **Biological Cybernetics**, v. 23, n. 3, pp. 187-202, 1976b.

HAMON, W. R. Estimating potential evapotranspiration. **Journal of the Hydraulics Division Proceedings of the American Society of Civil Engineers**, v. 87, pp. 107-120, 1961.

HANSEN, J. V.; Meservy, R. D. Learning experiments with genetic optimization of a generalized regression neural network. **Decision Support System**, v. 18, p. 317 - 325, 1996.

HARGREAVES, G. H, Samani, Z. A. Reference crop evapotranspiration from temperature. **Applied Engineering in Agriculture**, v. 1, n. 2, pp. 96-99, 1985.

HAYKIN, S. **Neural networks: A comprehensive foundation**. 2nd ed. Hamilton: Prentice Hall, 897 p., 1998.

HOPFIELD, J. J. **Neural networks and physical systems with emergence collective computational abilities**. **Proceeding of the National Academy of Science of the USA**. v. 79, n. 8, pp. 2554-2558, 1998.

HUO, Z.; FENG, S.; KANG, S. et al. Artificial neural network models for reference evapotranspiration in an arid area of northwest China. **Journal of Arid Environments**, v. 82, pp. 81-90, 2012.

**Instituto de Economia Agrícola do Estado de São Paulo (IEA)**. Disponível em: <<http://www.iea.agricultura.sp.gov.br/out/Bancodedaos2.html>>. Acesso em: 21 de dezembro de 2017.

**Instituto Brasileiro de Geografia e Estatística (IBGE)**. Acessado em 2018.

IQBAL, M. **An introduction to solar radiation**. New York: Academic Press, 1983. 390 p.  
IRMAK, S.; IRMAK, A.; ALLEN, R. G.; JONES, J. W. Solar and net radiation based equations to estimate reference evapotranspiration in humid climates. **Journal of Irrigation and Drainage Engineering**, v. 129, n. 5, pp. 336-347, 2003.

IKUDAYISI, A.; ADEYEMO, J. Effects of Different Meteorological Variables on Reference Evapotranspiration Modeling: Application of Principal Component Analysis. **World Academy of Science, Engineering and Technology International Journal of Geological and Environmental Engineering**, v. 10, n. 6, pp. 1-5, 2016.

JAMIESON, P. D.; PORTER, J. R.; WILSON, D. R. A test of the computer simulation model ARC - WHEAT1 on wheat crops grown in New Zealand. **Field Crops Research**, v. 27, pp. 337-350, 1991.

JACOVIDES, C. P; KONTOYIANNIS, H. Statistical procedures for the evaluation of evapotranspiration computing models. **Agricultural Water Management**, v. 27, pp. 365-371, 1995.

JANG, J. S. ANFIS : Adaptive Network Based Fuzzy Inference System. **IEEE Transactions on systems, man, and cybernetics**, v. 23, n. 3, 1993.

JATO-ESPINO, D.; CHARLESWORTH, S.; PERALES-MOMPARLER, S. Prediction of Evapotranspiration in a Mediterranean Region Using Basic Meteorological Variables. **Journal of Hydrologic Engineering**, v. 22, n. 4, pp. 04016064, 2017.



- JENSEN, M. E.; HAISE, H. R. Estimating evapotranspiration from solar radiation. **Proceeding of the Journal of Irrigation and Drainage Division: American Society of Civil Engineers**, v. 89, pp. 15-41, 1963.
- JOSSE, J.; HUSSON, F.; PAGÈS, J. Gestion des données manquantes en Analyse en Composantes Principales. **Journal de la Société Française de Statistique**, v. 150, n. 2, 2009.
- JOSSE, J.; HUSSON, F. Handling missing values in exploratory multivariate data analysis methods. **Journal de la Société Française de Statistique**, compilada com jsfds, versão : 2009/12/09, 2012a.
- JOSSE, J.; HUSSON, F. Selecting the number of components in principal component analysis using cross-validation approximations. **Computational Statistics and Data Analysis**, v. 56, pp. 1869-1879, 2012b.
- JOSSE, J.; HUSSON, F. missMDA: A Package for Handling Missing Values in Multivariate Data Analysis. **Journal of Statistical Software**, v. 70, pp. 1-31, 2016.
- KARIMI, S.; KISI, O.; KIM, S. et al. Modelling daily reference evapotranspiration in humid locations of South Korea using local and cross-station data management scenarios. **International Journal of Climatology**, v. 37, pp. 3228-3246, 2017.
- KIERS, H. A. Weighted least squares fitting using ordinary least squares algorithms. **Psychometrika** v. 62, n. 2, pp. 251-266, 1997.
- KISI, O. Generalized regression neural networks for evapotranspiration modelling. **Hydrological Sciences Journal**, v. 51, pp. 1092-1105, 2006.
- KISI, O. Evapotranspiration modelling from climatic data using a neural computing technique. **Hydrological Processes**, v. 21, pp. 1925-1934, 2007.
- KISI, O.; OZTURK, O. Adaptive Neurofuzzy computing technique for evapotranspiration estimation. **Journal of Irrigation and Drainage Engineering**, v. 133, n. 4, pp. 368-379, 2007.
- KISI, O.; CIMEN, M. Evapotranspiration modelling using support vector machines. **Hydrological Sciences–Journal–des Sciences Hydrologiques**, v. 54, n. 5, pp. 918-928, 2009.
- KISI, O.; CIMEN, M. Evapotranspiration modeling using a wavelet regression model. **Irrigation Science**, v. 29, pp. 241-252, 2011.
- KISI, O.; DEMIR, V. Evapotranspiration Estimation using Six Different Multi-layer Perceptron Algorithms. **Irrigation e Drainage Systems Engineering**, v. 5, n. 2, pp. 1-6, 2016.
- KÖPPEN, W.; GEIGER, R. *Klimate der Erde*. Gotha: **Verlag Justus Perthes**. Wall-map 150cmx200cm, 1928.
- KOHONEN, T. Self-organized of topologically correct feature maps. **Biological Cybernetics**, v. 43, pp. 59-69, 1982.
- KOHONEN, T. Improved versions of learning vector quantization. Proceedings of the International Joint Conference on Neural Networks, San Diego, California, pp. 545-550, 1990.
- KUMAR, M.; RAGHUWANSHI, N. S.; SINGH, R. Artificial neural networks approach in evapotranspiration modeling: a review. **Irrigation Science**, v. 29, n. 4, pp. 11-25, 2011.
- KHATIB, T., ELMENREICH, W. A Model for Hourly Solar Radiation Data Generation from Daily Solar Radiation Data Using a Generalized Regression Artificial Neural Network. **International Journal of Photoenergy**, vol. 2015, 1 – 13, 2015.
- KHANMOHAMMADI, N.; REZAIE, H.; MONTASERi, M. et al. The application of multiple linear regression method in reference evapotranspiration trend calculation. **Stochastic Environmental Research and Risk Assessment**, v. 32, pp. 661-673, 2018.
- LAABOUDI, A.; MOUHOUCHE, B. Neural network approach to reference evapotranspiration modeling from limited climatic data in arid regions. **International Journal of Biometeorology**, v. 56, pp. 831-841, 2012.

- LAM, J.C.; WAN, K. K. W.; YANG, L. Solar radiation modeling using ANNs for different climates in China. **Energy Conversion Management**, v. 49, 1080-1090, 2008.
- LANDERAS, G.; ORTIZ-BARREDO, A.; LÓPEZ, J. J. Comparison of artificial neural network models and empirical and semi-empirical equations for daily reference evapotranspiration estimation in the Basque Country (Northern Spain). **Agricultural Water Management**, v. 95, pp. 553-565, 2008.
- LANDERAS, G.; BEKOE, E.; AMPOFO, J. et al. New alternatives for reference evapotranspiration estimation in West Africa using limited weather data and ancillary data supply strategies. *Theoretical and Applied Climatology*, DOI 10.1007/s00704-017-2120-y, 2017.
- LADLANI, I.; HOUICHI, L.; DJEMILI, L.; et al. Modeling daily reference evapotranspiration (ET<sub>o</sub>) in the north of Algeria using generalized regression neural networks (GRNN) and radial basis function neural networks (RBFNN): a comparative study. **Meteorology and Atmospheric Physics**, v. 118, pp. 163-178, 2012.
- LADLANI, I.; HOUICHI, L.; DJEMILI, L. et al. Estimation of Daily Reference Evapotranspiration (ET<sub>o</sub>) in the North of Algeria Using Adaptive Neuro-Fuzzy Inference System (ANFIS) and Multiple Linear Regression (MLR) Models: A Comparative Study. **Arabian Journal for Science and Engineering**, DOI 10.1007/s13369-014-1151-2, 2014.
- LORENA, A. C.; JACINTHO, L. F. O.; SIQUEIRA, M. F. et al. Comparing machine learning classifiers in potential distribution modelling. **Expert Systems with Applications**, v. 38, pp. 5268-5275, 2011.
- LYRA, W. S.; SILVA, E. C.; ARAÚJO, M. C. U. et al. Classificação periódica: um exemplo didático para ensinar análise de componentes principais. **Theoretical Applied Climatology**, DOI 10.1007/s00704-015-1541-8, 2010.
- MAHIDA, H. R.; PATEL, V. N. Impact of Climatological Parameters on Reference Crop Evapotranspiration using Multiple Linear Regression Analysis. **SSRG International Journal of Civil Engineering (SSRG-IJCE)**, v. 2, pp. 21-24, 2015.
- MAKKINK, G. F. Testing the Penman formula by means of lysimeters. **Journal of the Institution of Water Engineers**, v. 11, n. 3, pp. 277-288, 1957.
- MALLIKARJUNA, P.; JYOTHY, S. A.; REDDY, K. C. S. Daily Reference Evapotranspiration Estimation using Linear Regression and ANN Models. **Journal of The Institution of Engineers (India)**, v. 97, n. 4, pp. 215-221, 2013.
- MANIKUMARI, N.; VINODHINI, G. Regression Models for Predicting Reference Evapotranspiration. **International Journal of Engineering Trends and Technology (IJETT)**, v. 38, n. 3, pp. 134-139, 2016.
- MAHRINGER, W. Verdunstungsstudien am Neusiedler See. **Arch Met Geoph Biokl Ser B**, v. 18, pp. 1–20, 1970.
- MARTÍ, P.; GASQUE, M. Ancillary data supply strategies for improvement of temperature-based ET<sub>o</sub> ANN models. **Agricultural Water Management**, v. 97, pp. 939-955, 2010.
- MAROFI, S., TABARI, H., ZARE ABYANEH, H. Predicting spatial distribution of snow water equivalent using multivariate non-linear regression and computational intelligence methods. *Water Resources Manage*, v. 25, pp. 1417–1435, 2011.
- Mathworks (EUA). MATLAB. 2017. Disponível em: <https://www.mathworks.com/products/matlab.html>. Acesso em 22 de ago. 2017.
- MARTÍ, P.; GONZÁLEZ-ALTOZANO, M. Reference evapotranspiration estimation without local climatic data. **Irrigation Science**, v. 29, pp. 479-495, 2011.
- MARTÍ, P.; ZARZO, M. Multivariate statistical monitoring of ET<sub>o</sub>: A new approach for estimation in nearby locations using geographical inputs. **Agricultural and Forest Meteorology**, v. 152, pp. 125 - 134, 2012.

- MELO PRADO, B. Q.; FERNANDES, H. R.; ARAÚJO, T. G. et al. Avaliação de variáveis climatológicas da cidade de Uberlândia (MG) por meio da análise de componentes principais. **Engenharia Sanitária e Ambiental**, v. 21, n. 21, pp. 407 - 413, 2016.
- MEYER, A. Über einige Zusammenhänge zwischen Klima und Boden in Europa. **Chemie de Erde**, v. 2, pp. 209 - 347, 1926.
- MICROCAL SOFTWARE, INC. Data analysis and technical graphics software. **Origin Professional Version 6.0**. Northampton, MA: Copyright©, 772p, 1999.
- MINGOTI, S. A. Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada. **Editora UFMG**, p. 300, 2017.
- MITCHELL, T. M. Machine Learning. Boston. **McGraw Hill Companies Inc.**, p. 424, 1997.
- MORRISON, D.F. *Multivariate statistical methods*. 2nd ed. Singapore: McGraw Hill. 415p., 1976.
- MOHAN, S.; ARUMUGAM, N. Relative Importance of Meteorological Variables in Evapotranspiration: Factor Analysis Approach. **Water Resources Management**, v. 10, pp. 1 - 20, 1996.
- NOBRE, P.; ALMEIDA, R. A.; MALAGUTTI, M.; GIAROLLA, E. Coupled Ocean–Atmosphere Variations over the South Atlantic Ocean. **American Meteorological Society**, DOI: 10.1175/JCLI-D-11-00444.1, 2012.
- ORACLE CORPORATION. **Chapter 1 General Information**. MySQL 5.7 Reference Manual, 2019. Disponível em: <https://dev.mysql.com/doc/refman/5.7/en/introduction.html>. Acesso em: 08 abr. 2019a.
- ORACLE CORPORATION. **Chapter 12 Functions and Operators**. MySQL 5.7 Reference Manual, 2019. Disponível em: <https://dev.mysql.com/doc/refman/5.7/en/functions.html>. Acesso em: 08 abr. 2019b.
- OZBAYOGLU, G.; OZBAYOGLU, M. E. A new approach for the prediction of ash fusion temperatures: a case study using Turkish lignites. **Fuel**, v. 85, pp. 545–552, 2006.
- PAPOULIS, P. Probability, **Random Variables, and Stochastic Processes**. 4th ed. McGraw Hill, 864 p., 1984.
- PANDORFI, H.; BEZERRA, A.C.; ATARASSI, R. T. et al. Artificial neural networks employment in the prediction of evapotranspiration of greenhouse-grown sweet pepper. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 20, n. 6, pp. 507-512, 2016.
- PENMAN, H. L. Natural evaporation from open water, bare soil, and grass. **Proceedings of the Royal Society**, London, v. 193, n. 1, p. 120-146, 1948.
- PEREIRA, A. R.; VILLA NOVA, N.; PEREIRA, A. S. et al. A model for the class-A pan coefficient. **Agricultural Water Management**, v. 75, pp. 75-82, 1995.
- PEREIRA, A. R.; ALLEN, R.; SMITH, M.; RAES, D. Crop evapotranspiration estimation with FAO56: Past and future. **Agricultural Water Management**, v. 147, pp. 4-20, 2015.
- PEREIRA, J. A. M.; CARVALHO, E. A. Determinação do tamanho de lote mínimo em amostragem aleatória para materiais, empregando a distribuição de Weibull. **Cerâmica**, v. 61, pp. 77-87, 2015.
- PETKOVIC, D.; GOCIC, M.; TRAJKOVIC, S. et al. Determination of the most influential weather parameters on reference evapotranspiration by adaptive neuro-fuzzy methodology. **Computers and Electronics in Agriculture**, v. 114, pp. 277-284, 2015.
- PIMENTEL-GOMES, F. **A estatística moderna na pesquisa agropecuária**. Piracicaba: Associação Brasileira para Pesquisa da Potassa e do Fosfato, 1984. 160 p.
- PIRI, J.; SHAMSHIRBAND, S.; PETKOVIC, D. et al. Prediction of the solar radiation on the Earth using support vector regression technique. **Infrared Physics e Technology**, v. 68, pp. 179-185, 2015.
- PLATT, J. Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines. **Redmond: Microsoft Research**, Tech Report. pp. 21, 1998.

- PRIESTLEY, C. H. B., TAYLOR, R. J. On the assessment of surface heat flux and evaporation using large-scale parameters. **Monthly Weather Review**, v. 100, n. 2, pp. 81-92, 1972.
- RENCHE, A.C. Methods of Multivariate Analysis. **A JOHN WILEY e SONS, INC. PUBLICATION**. pp. 727. 2ed. 2002.
- ROLIM, G. S.; CAMARGO, M. B. P.; LANIA, D. G.; et al. Classificação climática de Köppen e de Thornthwaite e sua aplicabilidade na determinação de zonas agroclimáticas para o estado de São Paulo. **Bragantia**, Campinas, v. 66, n. 4, p. 711-720, 2007.
- R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/2018>.
- ROHWER, C. Evaporation from free water surface. **USDA Tech Null**, v. 217, 1-96, 1931.
- ROMANENKO, V. A. Computation of the autumn soil moisture using a universal relationship for a large area. In: **Proceedings, Ukrainian Hydrometeorological Research Institute**, n. 3. Kiev, 1961.
- ROSA FILHO, C.D.; PARISI, S. A.; SOARES, W. A. Statistical distribution evaluation and modeling of multiple linear regression of average wind speed. *Journal of Environmental Analysis and Progress* v. 03, p.24-36, 2018.
- SANSIGOLO, C. A. Distribuições de probabilidade de velocidade e potência do vento. *Revista Brasileira de Meteorologia*, v.20, n.2, 207-214, 2005.
- SANTOS, C. M.; SOUZA, J. L.; FERREIRA JÚNIOR, R. A. et al. On modeling global solar irradiation using air temperature for Alagoas State, Northeastern Brazil. **Energy**, v. 71, 338-398, 2014.
- SANTOS, M. A.; ESCOBEDO, J. F.; TERAMOTO, E. T. et al. Assessment of ANN and SVM models for estimating normal direct irradiation (H<sub>b</sub>). *Energy Conversion and Management*, v. 126, p. 826-836, 2016. **Energy Conversion and Management**, v. 126, pp. 826-836, 2016.
- SANTOS, M. A.; SOUZA, J. L.; LYRA, G. B. et al. Observed and modelled solar radiation components in sugarcane crop grown under tropical conditions. **Agrophysics**, v. 31, pp. 231-241, 2017.
- SATYAMURTY, P.; NOBRE, A. C.; SILVA, P. L. D. Topics: South America. **Meteorological Monographs**, v.27, n.49, pp.119-139, 1998.
- SENTELHAS, P. C.; GILLESPIE, T. J.; SANTOS, E. A. Evaluation of FAO Penman-Monteith and alternative methods for estimating reference evapotranspiration with missing data in Southern Ontario, Canada. **Agricultural Water Management**, v. 97, pp. 635-644, 2010.
- SILVA, I. N.; SPATTI, D. H.; FLAUSINO, R. A. **Redes Neurais Artificiais para engenharia e ciências aplicadas – fundamentos teóricos e aspectos práticos**. 2nd ed. Artiliber 431 p., 2016a.
- SILVA, H. J. F.; SANTOS, M. S.; CABRAL JÚNIOR, J. B. et al. Modeling of reference evapotranspiration by multiple linear regression. **Journal of Hyperspectral Remote Sensing**, v. 6, n. 1, pp. 44-58, 2016b.
- SILVA, P.F.; LIMA, J. R. S.; ANTONINO, A. C. D. et al. Seasonal patterns of carbon dioxide, water fluxes over the Caatinga and grassland in the semi-arid region of Brazil. **Journal Arid Environments**, v. 147, pp. 71-82, 2017a.
- SILVA, M. B. P.; ESCOBEDO, J. F.; SANTOS, C. M. et al. Performance of the Angstrom-Prescott Model (A-P) and SVM and ANN techniques to estimate the daily Global Solar Irradiation in Botucatu/SP/Brazil. **Journal of Atmospheric and Solar-Terrestrial Physics**, v. 160, pp. 11-23, 2017b.
- SILVA, J. R. Valor da produção agropecuária do Estado de São Paulo: resultado preliminar de 2018. **Análises e Indicadores do Agronegócio**, São Paulo, v. 13, n. 11, p. 1-6, nov. 2018. Disponível em: <http://www.iea.sp.gov.br/ftpiea/aia/AIA-74-2018.pdf>. Acesso em: 21 de fevereiro de 2019.

- SHIRI, J.; KISI, O.; LANDERAS, G. et al. Daily reference evapotranspiration modeling by using genetic programming approach in the Basque Country (Northern Spain). **Journal of Hydrology**, v. 414-415, pp. 302-316, 2012.
- SHIRI, J.; NAZEMI, A. H.; SADRADDINI, A. A.; et al. Comparison of heuristic and empirical approaches for estimating reference evapotranspiration from limited inputs in Iran. **Computers and Electronics in Agriculture**, v. 108, pp. 230-241, 2014.
- SHIRI, J.; SADRADDINI, A. A.; NAZEMI, A. H. et al. Independent testing for assessing the calibration of the Hargreaves-Samani equation: New heuristic alternatives for Iran. **Computers and Electronics in Agriculture**, v. 117, pp. 70-80, 2015.
- SHEVADE, S. K.; KEERTHI, S. S.; Bhattacharyya, C. Improvements to the SMO Algorithm for SVM Regression. **IEEE Transactions on Neural Networks**, v. 11, n. 5. pp. 1188-1193, 2000.
- SMOLA, A. J.; SCHÖLKOPF, B. "A Tutorial on Support Vector Regression". Royal Holloway College, London, U. K., **Neuro COLT Tech. Rep.** TR 1998-030, 1998.
- SMOLA, A. J.; SCHÖLKOPF, B. A tutorial on Support Vector Regression. **Statistics and Computing**, v.14, pp. 199-222, 2004.
- SNYDER, R. L. Equation for evaporation pan to evapotranspiration conversions. **Journal of Irrigation and Drainage Engineering**, v. 118, n. 6, pp. 977-980, 1992.
- SNYDER, R. Climate Change Impacts on Water Use in Horticulture. **Horticulturae**, v. 3, n. 27, pp. 1-9, 2017.
- SRIRAM, A. V.; RASHMI, C. N. Estimation of Potential Evapotranspiration by Multiple Linear Regression Method. **IOSR Journal of Mechanical and Civil Engineering**, v. 11, pp. 65-70, 2014.
- STONE, R. J. Improved statistical procedure for the evaluation of solar radiation estimation models. **Solar Energy**, v. 51, n. 4, pp. 289-291, 1993.
- STRAPASSON, E.; VENCovsky, R.; Batista, L.A.R. Seleção de Descritores na Caracterização de Germoplasma de Paspalum sp. por meio de Componentes Principais. **Revista Brasileira Zootecnia**, v. 29, n. 2, pp. 373-381, 2000.
- SWINBANK, W. C. The measurement of vertical transfer of heat and water vapor by eddies in the lower atmosphere. **Journal of Meteorology**, v. 8, p. 135-145, 1951.
- TABARI, H.; KISI, O.; EZANI, A. et al. SVM, ANFIS, regression and climate based models for reference evapotranspiration modeling using limited climatic data in a semi-arid highland environment. **Journal of Hydrology**, v. 444-445, pp. 78-89, 2012.
- TABARI, H.; GRISMER, M. E. Comparative analysis of 31 reference evapotranspiration methods under humid conditions. **Irrigation Science**, v. 31, pp. 107-117, 2013.
- TAKAGI, T.; SUGENO, M.. Fuzzy identification of systems and its applications to modeling and control. **IEEE Transactions System, Man, Cybernetics**, vol. 15, p. 116-132, 1985.
- TANGUNE, B. F.; ESCOBEDO, J. F. Reference evapotranspiration in São Paulo State: Empirical methods and machine learning techniques. **International Journal of Water Resources and Environmental Engineering**, v. 10, n. 4, pp. 33-44, 2018.
- TURC, L. Évaluation des besoins en eau d'irrigation, évapotranspiration potentielle. **Annales Agronomiques**. V. 12, n. 1, pp. 13-49, 1961.
- THORNTHWAITE, C. W.; WILM, H. G. Report of the committee on evapotranspiration and transpiration, 1943-1944. *Transactions of the American Geophysical Union*, Washington, DC, v. 25, n. 5, p. 686-693, 1944.
- THORNTHWAITE, C. W. An approach toward a rational classification of climate. **Geographical Review**, v. 55-94, pp. 38,1948.
- TRABERT, W. Neue Beobachtungen über Verdampfungsgeschwindigkeiten. **Meteorologische Zeitschrift**, v. 13, pp. 261-263, 1896.
- TRAORE, S.; WANG, Y.; KERH, T. Artificial neural network for modeling reference evapotranspiration complex process in Sudano-Sahelian zone. **Agricultural Water Management**, v. 97, pp. 707-714, 2010.

- ULIANA, E. M.; SILVA, D. D.; SILVA, J. G. F.; FRAGA, M. S.; LISBOA, L. Estimate of reference evapotranspiration through continuous probability modelling. *Eng. Agríc., Jaboticabal*, v.37, n.2, p.257-267, 2017.
- University Of Waikato (New Zeland) (Org.). **Weka: Data Mining Software in Java**. 2018. Disponível em: <<http://www.cs.waikato.ac.nz/ml/weka/>>. Acesso em: 22 de agosto de 2018.
- VANDERLINDEN, K.; GIRÁLDEZ, J. V.; MEIRVENNE, M. V. Spatial Estimation of Reference Evapotranspiration in Andalusia, Spain. **Journal of Hydrometeorology**, DOI: 10.1175/2007JHM880.1, 2008.
- VAPNIK, V. N. The nature of Statistical learning theory. **Springer-Verlag**, New York, 1995.
- VAPNIK V, N. Statistical learning theory. New York: **Wiley**, 1998.
- WEIBULL, W. A statistical distribution function of wide applicability. *Journal of Applied Mechanics*, New York, v.18, n.3, p.293-297, 1951.
- WEN, X.; SI, J.; HE, Z. et al. Support-Vector-Machine-Based Models for Modeling Daily Reference Evapotranspiration With Limited Climatic Data in Extreme Arid Regions. **Water Resources Management**, DOI 10.1007/s11269-015-0990-2, 2015.
- WILLMOTT, C. J. On the validation of models. **Physical Geography**, v. 2, pp.184 - 194, 1981.
- WILLMOTT, C. J. On the validation of models. **International Journal of Climatology**, v. 32, pp.2088 - 2094, 2012.
- WITTEN, I. H.; FRANK, E.; HALL, M. A. **Data Mining: Practical Machine Learning Tools and Techniques**. 3 rd ed. pp. 630, 2011.
- XU, J.; PENG, S.; WANG, W. et al. Prediction of daily reference evapotranspiration by a multiple regression method based on weather forecast data. **Archives of Agronomy and Soil Science**, v. 59, n. 11, pp. 1487-1501, 2013.
- YASSIN, M. A.; ALAZBA, A. A.; MATTAR, M. A. Artificial neural networks versus gene expression programming forestimating reference evapotranspiration in arid climate. **Agricultural Water Management**, v. 163, pp. 110-124, 2016.
- YE, X. A.; LI, X.; LIU, J. et al. Variation of reference evapotranspiration and its contributing climatic factors in the Poyang Lake catchment, China. **Hydrological Processes**, v. 28, pp. 6151-6162, 2014.
- ZADEH, L. A. Fuzzy algorithms, **Info. e Ctl.** vol. 12, p. 94-102, 1968.
- ZANETTI, S. S.; SOUSA, E. F.; Oliveira, V. et al. Estimating evapotranspiration using artificial neural network and minimum climatological data. **Journal of Irrigation and Drainage Engineering**, v. 33, n. 2, pp. 83-89, 2007.

## CAPÍTULO 2

### COMPARAÇÃO DE DIFERENTES MÉTODOS DE INTERPOLAÇÃO ESPACIAL DA EVAPOTRANSPIRAÇÃO DE REFERÊNCIA ( $ET_{o_{FAO-56}}$ ) NO PLANALTO OCIDENTAL PAULISTA

#### RESUMO

Neste estudo é utilizada uma base de dados climáticos medida no período (2013-2017) para estimativa da Evapotranspiração de Referência ( $ET_{o_{FAO-56}}$ ) em 30 localidades do Planalto Ocidental Paulista (integração das regiões intermediárias de São José do Rio Preto e Ribeirão Preto) com o objetivo de auxiliar no gerenciamento de recursos hídricos com a caracterização da demanda atmosférica na região do Planalto Ocidental Paulista (região de grande destaque por sua produção agrícola no Estado de São Paulo). Assim, é realizada uma análise do desempenho estatístico de diferentes interpoladores espaciais (KO e MC) por meio dos seguintes indicadores estatísticos: *Mean Bias Error* (MBE), *Mean Square Error* (MSE), *Root Mean Squared Error* (RMSE), Coeficiente de determinação ( $R^2$ ), Coeficiente de correlação ( $r$ ) e Intervalo de Tolerância (IT). Os resultados obtidos mostraram que o método da Mínima Curvatura (MC) apresentou melhor desempenho estatístico:  $ET_{o_{MC}}$  (MBE = -0,01 mm dia<sup>-1</sup>; MSE = 0,01 mm dia<sup>-1</sup>; RMSE = 0,11 mm dia<sup>-1</sup>).

**Palavras-chave:** Evapotranspiração de Referência ( $ET_{o_{FAO-56}}$ ); interpoladores espaciais; Planalto Ocidental Paulista.

#### 2.1 INTRODUÇÃO

O conhecimento da evapotranspiração é de grande importância no planejamento de sistemas de irrigação em regiões de escassez hídrica, na melhoria das estimativas de requerimento de água das culturas agrícolas (ADELOYE et al., 2012; SHIRI et al., 2012). E no planejamento e gerenciamento de recursos hídricos o conhecimento prévio das condições climáticas de uma determinada região é de muito interesse no cenário agrícola, pois estes fazem partes dos fatores determinantes do sucesso e produção satisfatória das culturas (SANTOS et. al., 2011).

A evapotranspiração é um fenômeno físico que promove a transferência de água para a atmosfera por meio da evaporação da água e transpiração das culturas

agrícolas (ALLEN et al., 1998; SHIRI et al., 2015). É um fenômeno multidimensional, pois depende bastante da interação entre as variáveis climáticas (radiação solar, temperatura do ar, precipitação, umidade relativa do ar e velocidade do vento) e das condições fitossanitárias das culturas agrícolas (ALLEN et al., 1998; SENTELHAS et al., 2010; SNYDER, 2017).

Entretanto, geralmente a obtenção de medidas diretas e confiáveis de evapotranspiração é bastante difícil em razão dos altos custos para implantação de técnicas micrometeorológicas e lisimétricas, além das dificuldades técnicas de operação e manutenção dos equipamentos de medidas, bem como do processamento de dados. Em face disso, modelos estatísticos de estimativa de evapotranspiração por meio de variáveis climáticas são bastante difundidos na literatura (THORNTHWAITE, 1948; BLANEY e CRIDDLE, 1950; PRIESTLEY e TAYLOR, 1972; BENEVIDES e LOPEZ, 1970; HARGREAVES e SAMANI, 1985; ALLEN et al., 1998).

Modelos de estimativa de evapotranspiração necessitam apenas de variáveis climáticas como variável de entrada e são capazes de estimar evapotranspiração com bom desempenho estatístico como o modelo de referência de Penman – Monteith ( $ET_{OFAO-56}$ ) relatado pelo boletim FAO-56. Entretanto, sua principal desvantagem está relacionada com as dificuldades que muitos países em desenvolvimento têm em manter estações automáticas medindo rotineiramente as variáveis climáticas por longos períodos (ALLEN et al., 1998; MARTÍ e ZARZO, 2012).

Atualmente, em razão das aplicações hidrológicas e gerenciamento dos recursos hídricos, muitos pesquisadores vêm optando pela espacialização de dados climatológicos para investigar a variabilidade espacial de  $ET_{OFAO-56}$  (DALEZIOS et al., 2002; MARDIKIS et al., 2005; VILANOVA et al., 2012). Em que essa espacialização, a depender do tipo de estudo, pode ser realizada por meio de técnicas de interpolação espacial como a geoestatística ou o Sensoriamento Remoto (VANDERLINDEN et al., 2008; FERREIRA SILVA et al., 2018). Mardikis et al. (2005) destaca que os parâmetros climáticos são afetados pela forte variabilidade espacial e a natureza dos próprios parâmetros climáticos e, por isso, uma rede densa de Estações Meteorológicas Automáticas (EMA's) monitorando variáveis climáticas se faz necessário.

A interpolação espacial pode ser realizada por alguns métodos como a Krigagem Ordinária (KO), Inverso Quadrático da Distância (IDW), Mínima Curvatura (MC), Triangulação (Tri) entre outros (MAZZINI e SCHETTINI, 2009; LIMA ALVES e



VECCHIA et al., 2011; SANTOS et al., 2016). Em que a Krigagem Ordinária (KO) é classificada como uma técnica estocástica que utiliza análise variográfica para efetuar interpolações espaciais (MAZZINI e SCHETTINI, 2009; ALVES e VECCHIA, 2011; DAI et al., 2014), diferentemente das técnicas IDW, MC e Tri, que são classificadas como determinísticas (ALVES e VECCHIA, 2011; VAROUCHAKIS e HRISTOPULOS, 2013; DAI et al., 2014).

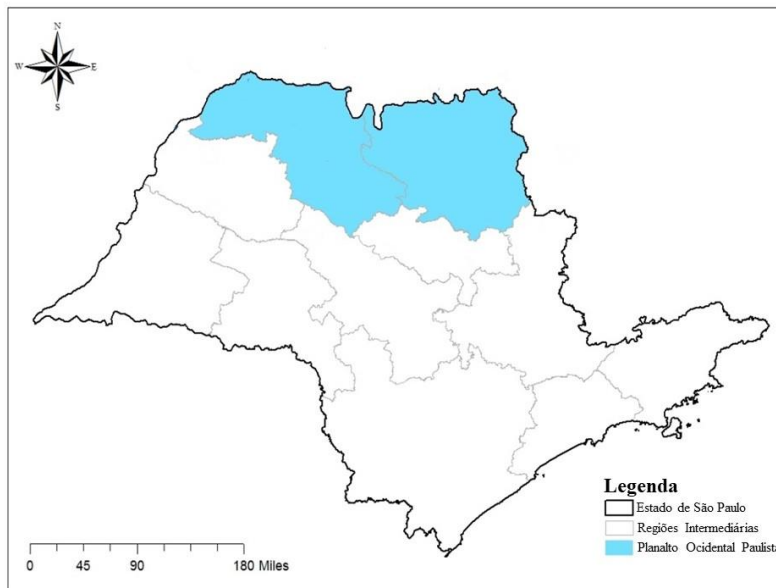
O objetivo deste estudo foi auxiliar no gerenciamento de recursos hídricos com a caracterização da demanda atmosférica na região do Planalto Ocidental Paulista (região de grande destaca por sua produção agrícola no Estado de São Paulo), analisando o desempenho de diferentes interpoladores espaciais: Krigagem Ordinária (KO) e Mínima Curvatura (MC).

## **2.2 MATERIAL E MÉTODOS**

### **2.2.1 Região de estudo e informações climáticas**

Na Figura 1 é apresentada a região de estudo (parte do Planalto Ocidental Paulista – situada mais a noroeste do Estado de São Paulo), o qual é composta pela integração de duas regiões intermediárias do Estado de São Paulo: São José do Rio Preto e Ribeirão Preto (latitude  $19^{\circ}32' - 21^{\circ}57'S$ , longitude  $46^{\circ}38' - 51^{\circ}28'O$  e altitude 335,00 – 1026,00 m). Essas regiões são denominadas de intermediárias desde o ano de 2017 pelo Instituto Brasileiro de Geografia e Estatística (IBGE, 2018) como forma de melhor delimitar as diferentes regiões do Estado de São Paulo em função de características locais distintas.

**Figura 2.** Localização da região de estudo no Planalto Ocidental Paulista.



A área total destas regiões integradas neste estudo é de aproximadamente 53.492,00 km<sup>2</sup>, abrangendo 164 municípios, ou seja, 25,43% dos municípios do Estado de São Paulo, com população de aproximadamente 4.188,425 de habitantes (9,29% da população total do estado) (IBGE, 2018).

O clima predominante na região de estudo é do tipo Aw (CEPAGRI, 2018; ROLIM, 2007; TANGUNE e ESCOBEDO, 2018), tropical (megatérmico), com verão chuvoso e inverno seco (KÖPPEN, 1928). Com temperatura média do ar variando de 19,40 a 24,20°C e intervalo de variação dos meses chuvosos de 9,0 – 300,0 mm de agosto a janeiro. O regime de chuvas (ou período úmido) da Região Sudeste do Brasil (em especial, no Estado de São Paulo) é normalmente influenciado pelo transporte de umidade em níveis mais baixos da atmosfera (850 hPa) por meio de Jatos de Baixos Níveis (JBN) que transportam umidade do Sul da região Amazônica até o Oceano Atlântico (NOBRE et al., 2012). Entretanto, em anos atípicos, o regime de chuvas é drasticamente interrompido, como entre os anos 2013/2014 e 2014/2015, que foram classificados como anos: excepcionalmente seco e muito seco, respectivamente (COELHO et al, 2016a). Nesses anos o padrão de chuvas na região Sudeste foi interrompido por uma anomalia da circulação anti-ciclônica, no Oceano Atlântico próximo à costa da região Sudeste do Brasil. Este tipo de circulação inibiu o transporte para esta região, refletindo baixa quantidade de chuvas precipitadas. Simultaneamente foi verificado um aumento da temperatura do Oceano Atlântico que contribuiu para a redução da precipitação na região (COELHO et al., 2016b).

No caso do período seco, é conhecido que o regime de chuvas é do tipo frontal, originada do encontro das massas frias e secas vindas da Região Sul com as massas quentes e úmidas, da Região Sudeste. Sua distribuição gera aumento na nebulosidade, com ocorrência de chuvas generalizadas na região, com intensidade de precipitação variando de fraca a moderada (SATYAMURTY et al., 1998).

### **2.2.2 Economia regional**

Com as condições climáticas favoráveis às atividades da agricultura, o Planalto Ocidental Paulista tem sua economia predominantemente destinada ao agronegócio. De maneira direta e indireta, responde pela geração de empregos e movimentação da economia local, sobretudo nos períodos de safra das principais culturas agrícolas.

Silva et al. (2019) relata em seu relatório técnico do Instituto de Economia Agrícola (IEA), vinculado à Secretaria de Agricultura e Abastecimento do Estado de São Paulo, a variação dos preços no mercado das principais *commodities* do Estado de São Paulo. As principais *commodities* agrícolas produzidas no Estado são: amendoim (água/seca), café, cana-de-açúcar, eucalipto, feijão (água/seca/com irrigação/sem irrigação), laranja, mandioca (indústria/mesa), milho (com irrigação/sem irrigação), seringueira e soja. No Apêndice A são mostradas informações detalhadas, referentes à safra 2017, do percentual da área de produção agrícola e produção relativa, da região de estudo. Pode-se verificar que a região lidera o *ranking* de produção em culturas como amendoim das águas e do período seco, cana-de-açúcar, laranja e seringueiras, além de possuir grandes áreas de cultivo dentro do Estado de São Paulo. A região também figura entre as primeiras na produção de café, feijão irrigado de inverno com irrigação, mandioca de mesa, milho de sequeiro e irrigado e soja.

Contudo, é importante destacar que o número de EMA's operando no Estado de São Paulo pode ser ainda maior, pois outras instituições como a Universidade de São Paulo (USP), usinas de setor sucroalcooleiro, associações de produtores de café, também operam com esse tipo de estação no Estado (AGRITEMPO, 2018).

### **2.2.3 Base de dados**

A base de dados climáticos monitorada na região de estudo é composta das variáveis climáticas: irradiação solar global ( $H$ ,  $\text{MJ m}^{-2} \text{ hora}^{-1}$ ); temperaturas máxima e mínima do ar ( $T_{\text{max}}$  e  $T_{\text{min}}$ ,  $^{\circ}\text{C}$ ), umidade relativa do ar ( $UR$ , %) e velocidade do vento ( $U_2$ ,  $\text{m s}^{-1}$ ) medida a 10 metros de altura da superfície em partição horária –

convertida posteriormente para 2 metros de altura conforme metodologia descrita por Allen et al. (1998) no boletim FAO-56.

A variável astronômica Irradiação solar global no topo da atmosfera ( $H_0$ , MJ m<sup>-2</sup> hora<sup>-1</sup>) também compôs a base de dados climáticos deste estudo. Para obtenção desta variável foi utilizada a metodologia descrita por Iqbal (1983) e detalhada no Apêndice B.

Na Tabela 1 são mostradas as informações geográficas (localidade, latitude, longitude, altitude e distância do mar), classificação climática e fonte de aquisição da base climática de dados das 30 localidades situadas na região de estudo do Planalto Ocidental Paulista.

**Tabela 1.** Informações geográficas das EMA's [localidades, latitudes (°), longitude (°), altitude (m)], classificação climática (clss. climática), fontes dos dados.

Localidades	Latitude	Longitude	Altitude	Dist. do mar	Clss. Climática	Fontes
	(°)	(°)	(m)	(km)		
Campina Verde*	19°32'S	49°32'W	555,2	587,0	Aw	INMET
Paranaíba**	19°41'S	51°10'W	424,0	668,0	Aw	INMET
Populina	19°52'46"S	50°28'13"W	394,0	610,0	Aw	UNESP
Sacramento*	19°53'S	47°26'W	912,0	465,0	Csa	INMET
Conceição das Alagoas*	19°59'S	48°09'W	568,0	494,0	Aw	INMET
Paranapuã	20°01'29"S	50°33'57"W	436,1	604,0	Aw	UNESP
Jales	20°09'S	50°35'W	457,0	581,0	Aw	INMET
Ituverava	20°21'S	47°46'W	600,0	440,0	Aw	INMET
Ilha Solteira	20°25'24"S	51°21'13.1"W	337,0	622,0	Aw	UNESP
Marinópolis	20°26'48"S	50°48'26.1"W	370,0	585,0	Aw	UNESP
Pereira Barreto I	20°31'42"S	51°14'58"W	426,0	589,0	Aw	UNESP
Barretos	20°33'S	48°32'W	533,0	444,0	Aw	INMET
Franca	20°34'S	47°22'W	1026,0	387,0	Aw	INMET
Itapura	20°38'28"S	51°28'29.9"W	335,0	608,0	Aw	UNESP
Pereira Barreto II	20°40'23"S	51°2'2.1"W	357,0	579,0	Aw	UNESP
Sud Mennucci	20°43'42"S	50°57'30"W	350,0	565,0	Aw	UNESP
Passos*	20°44'S	46°38'W	784,0	328,0	Cwa	INMET
Monte Azul Paulista	20°54'26"S	48°38'29"W	611,0	410,0	Aw	IAC
Brodowsky	20°59'03"S	47°39'33"W	863,0	359,0	Cwa	IAC
José Bonifácio	21°02'S	49°41'W	405,0	465,0	Aw	INMET
Arianha	21°07'S	48°50'W	525,0	398,0	Aw	INMET
Sertãozinho	21°08'16"S	47° 59' 24" W	579,0	362,0	Aw	IAC
Cássia dos Coqueiros	21°16'58"S	47°10'11"W	890,0	303,0	Cwb	IAC
Valparaíso	21°19'S	50°55'W	374,0	526,0	Aw	INMET
Pradópolis	21°20'S	48°06'W	544,0	339,0	Aw	INMET
São Simão	21°28'44"S	47° 33' 3" W	620,0	304,0	Aw	IAC
Lins	21°39'S	49°44'W	459,0	415,0	Aw	INMET
Tambaú	21°42'18"S	47° 16' 26" W	698,0	268,0	Aw	IAC
Ibitinga	21°51'S	48°48'W	492,0	344,0	Aw	INMET
São Carlos	21°57'S	47°52'W	863,0	267,0	Cwa	INMET

\*Estado de Minas Gerais, \*\*Estado de Mato Grosso do Sul.

Entre as 30 (trinta) EMA's, 5 (cinco) estão situadas fora dos limites territoriais do Estado de São Paulo e da região de estudo: Campina Verde, Conceição das Alagoas, Sacramento e Passos (Estado de Minas Gerais); e Paranaíba (Estado de Mato Grosso do Sul). Essas localidades foram selecionadas em razão de sua proximidade com os limites extremos do Estado de São Paulo. Outras localidades de

regiões intermediárias paulistas foram inseridas neste estudo (Ibitinga, Ilha Solteira, Itapura, Lins, Pereira Barreto I e II, São Carlos, Sud Mennucci, Valparaíso) devido à proximidade com a região de estudo.

Contudo, é importante destacar que o número de EMA's operando no Estado de São Paulo pode ser ainda maior, pois outras instituições como a Universidade de São Paulo (USP), usinas de setor sucroalcooleiro, associações de produtores de café, também operam com esse tipo de estação no Estado (AGRITEMPO, 2018).

#### **2.2.4 Processamento da base de dados climáticos**

Em razão do grande volume de informações, a necessidade e a demanda específica de alto desempenho de processamento, foi utilizada a tecnologia de gerenciamento de banco de dados MySQL 5.7 (*Relational Database Management System*) da Oracle em sua versão de código aberto (licença GNU), atuando como um servidor local (ORACLE CORPORATION, 2019a). Todas as operações computacionais em formato de algoritmos e funções foram desenvolvidas através da linguagem de alto nível de manipulação de dados SQL (*Structured Query Language*).

O MySQL 5.7 é um servidor de banco de dados *multithread* (encadeamento de execução de processos) e multiusuário (acesso simultâneo de diversos usuários). Seus recursos são planejados para aplicações em grandes volumes de dados, softwares de missão crítica e sistemas computacionais implementados em massa (ORACLE CORPORATION, 2019a). Sua principal função neste estudo foi armazenar, processar e padronizar o extenso volume de parâmetros meteorológicos de 30 Estações Meteorológicas Automáticas (EMA's) espalhadas pelo Planalto Ocidental Paulista.

A função *average* (AVG) da linguagem SQL foi utilizada para se chegar aos valores médios da radiação solar global, temperatura do ar, umidade relativa do ar e velocidade do vento a 10 metros de altura. As funções de máximo (MAX) e mínimo (MIN) foram utilizadas para se obter os valores máximos e mínimos referentes a temperatura e umidade relativa do ar. A função de soma (SUM) foi utilizada para se obter os valores de precipitação pluviométrica acumulados para o dia (ORACLE CORPORATION, 2019b).

### 2.2.5 Estatística descritiva

Para análise e caracterização do conjunto de dados climáticos e da  $ET_{OFAO-56}$  calculada, foi feita a estatística descritiva da base de dados, na qual foi calculada: Média ( $\bar{x}$ ); Erro Padrão (SE); Desvio Padrão (SD); Máximos (Max); Mínimos (Min); Coeficiente de Variação (CV); Kurtose (K) e assimetria (As). Para obtenção destes valores utilizou-se as seguintes equações:

- Média aritmética (Equação 1):

$$\bar{x} = \frac{\sum x_i}{n} \quad (1)$$

- Desvio Padrão (Equação 2)

$$SD = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} \quad (2)$$

- Erro Padrão (SE) (Equação 3):

$$SE = \frac{SD}{\sqrt{n}} \quad (3)$$

- Coeficiente de Variação (CV, %) (Equação 4):

$$CV = \frac{SD}{\bar{x}} \quad (4)$$

n representa o número de indivíduos e  $x_i$  os valores medidos das variáveis.

Para auxiliar na interpretação dos resultados obtidos em estatística descritiva com o Coeficiente de Variação (CV, %), é utilizado neste estudo a escala classificatória elaborada por Pimentel Gomes (1984):

- Baixo: < 10%
- Médio: 10 – 20%
- Alto: 20 – 30%
- Muito Alto: > 30

para auxiliar a analisar o nível de dispersão dos dados em relação ao seu valor médio.

Foram utilizadas planilhas eletrônicas da Microsoft Excel® e a versão 2018 do software Minitab® para a realização de toda estatística descritiva deste estudo.

### 2.2.6 Métodos de Interpolação Espacial (IE)

Castro et al. (2010) destaca que os interpoladores espaciais são ferramentas matemáticas importantes que atribuem valores relativos à variáveis já existentes e são capazes de transformar dados discretos em contínuos.

A escolha dos interpoladores espaciais depende da distribuição dos pontos utilizados nos cálculos dos modelos utilizados por esses interpoladores e da

correlação desses modelos com os fenômenos estudados (ARANOFF, 1989; CASTRO et al., 2010). Somente assim, que pode ser avaliado o desempenho dos interpoladores por cada variável estudada (MARDIKIS et al., 2005).

Para a realização das interpolações espaciais com os métodos citados acima, foi utilizado o *software* Surfer versão 9 (Golden Software, 2010).

#### - Mínima Curvatura (MC)

O método da Mínima Curvatura (MC) tem essa denominação derivada de uma ferramenta de desenho técnico e atualmente é bastante difundida como interpolador espacial (ALVES et al., 2012; SANTOS et al., 2016).

Esse método pode ser definido como um interpolador não exato capaz de suavizar superfícies onde os dados são atravessados (MEDINA e QUINTAS, 2008; SANTOS et al., 2016). Assim, novos grids de pontos são gerados e sucessivamente são aplicados modelos de regressão de mínimos quadrados sobre os grids iterativamente até a obtenção de melhores níveis de suavização.

Para tanto, o algoritmo de MC gera uma superfície que interpola todas as observações disponíveis de solução da equação diferencial modificada bi-harmônica (SMITH e WESSEL, 1990; SANTOS et al., 2016). Conforme mostram as Equação diferencial modificada bi-harmônica com tensão T (7):

$$(1 - T_i)\nabla^2(\nabla^2 Z)(T_i)\nabla^2 Z = 0 \quad (7)$$

e desenvolvida até sua solução única (8 - 10):

$$(1 - T_i)\frac{\partial^2 Z}{\partial n^2} + T_i\frac{\partial Z}{\partial n} = 0 \quad (8)$$

$$\frac{\partial(\nabla^2 Z)}{\partial n} = 0 \quad (9)$$

$$\frac{\partial Z}{\partial x \partial y} = 0 \quad (10)$$

Nessas condições a Equação 7 tem solução única com derivada segunda contínua chamada de spline bicúbica natural (SANTOS et al., 2016). Que acarreta na divisão do conjunto de dados em subconjuntos e polinômios de pequena ordem são

utilizados em cada subconjunto para interpolação (MAZZINI e SCHETTINI, 2009; ALVES et al., 2012).

Nessas interpolações ocorre uma série de iterações especificadas pelo usuário do software utilizado até certo limite de tolerância da superfície de mínima curvatura que por padrão esses limites são 100 iterações e 99% dos pontos (SANTOS et al., 2016).

#### - Krigagem Ordinária (KO)

Landim (2006) destaca que os primeiros conceitos fundamentais da geoestatística foram relatados inicialmente por Krige na segunda metade do século XX e posteriormente por Matheron na década seguinte.

Krige (1951) elaborou empiricamente uma série de estudos em reservas minerais de ouro na África do Sul, no intuito de analisar a relação entre as localizações geográficas e as jazidas de ouro e concluir se há dependência espacial entre as jazidas. Após a divulgação dos estudos iniciais de Krige (1951) e Matheron (1962, 1963, 1965) investigou as hipotéticas relações entre as localizações geográficas e a dependência espacial e divulgou a teoria das variáveis regionalizadas com um conhecimento mais fundamentado.

De acordo com Landim (2006), a geoestatística moderna é baseada nos seguintes conceitos: funções aleatórias, variáveis regionalizadas e estacionariedade; que abordam probabilisticamente uma gama de técnicas para análise e mapeamento de valores de variáveis regionalizadas distribuídas no espaço e no tempo.

Dentre os métodos de interpolação geoestatísticos conhecidos destacam-se: Krigagem Indicativa (KI), Krigagem Ordinária (KO), Krigagem Simples (KS), Krigagem Universal (KU) e Co-Krigagem (CK). O método adotado neste estudo foi o da Krigagem Ordinária (KO) (WEBSTER e OLIVER, 2007). Segundo Landim (2006), a Krigagem é o melhor estimador linear não tendencioso. Diferentemente dos demais interpoladores, esse método necessita de informações advindas de variogramas. Para Vieira et al. (1983), a teoria da geoestatística leva em consideração alguns aspectos como a suposição de estacionariedade da hipótese intrínseca.

Para a construção dos variogramas dos dados de *ET<sub>0FA0-56</sub>* e posterior interpolação pela KO, inicialmente realizou-se a construção de histograma dos dados, análise estatística descritiva e teste de normalidade pelo método de Shapiro-Wilk. Mesmo que a normalidade dos dados não seja uma exigência na geoestatística, sua



assimetria deve ser avaliada (LANDIM, 2006; OLIVER, 2010). Em seguida, foi realizada a construção um variograma experimental isotrópico para cada conjunto de dados de  $ET_{FAO-56}$ . Para isso, os valores de semivariância foram calculados pela Equação (8) (JOURNEL e HUIJBREGTS, 1989):

$$\gamma * (h) = \frac{1}{2 N(h)} \sum [(Z(x_i) - Z(x_i + h))^2] \quad (8)$$

em que:  $\gamma * (h)$ : semivariância para um vetor  $h$ ;

$Z(x)$  e  $Z(x + h)$ : são pares de observações;

$N(h)$ : número de pares de valores medidos e  $h$ : distância entre os pares de observações  $[x, (x + h)]$ .

Os variogramas experimentais foram então ajustados pelo melhor modelo teórico escolhido entre o esférico, exponencial e gaussiano, com base nos resultados da validação cruzada, para aquisição dos parâmetros, patamar (C), efeito pepita (Co+C) e alcance da dependência espacial (A). Também foi determinado o Índice de Dependência Espacial (IDE, %) por meio do método proposto por Zimback (2001), que faz a relação entre patamar (C) e efeito pepita (Co+C), como mostra a Equação (9):

$$\left( GDE = \frac{C}{Co+C} * 100 \right) \quad (9)$$

Zimback (2001) destaca uma escala classificatória do desempenho de GDE:

- valores  $\geq 75\%$  a dependência espacial dos dados é considerada forte;
- valores entre 25% e 75% a dependência espacial dos dados é considerada moderada;
- valores  $\leq 25\%$  a dependência espacial dos dados é considerada fraca.

Posteriormente a aquisição e verificação dos parâmetros semivariográficos dos dados, foram realizadas interpolações por KO dos dados de  $ET_{FAO-56}$  para estimativa dos mesmos nas localidades não medidas ao longo da cobertura da região de estudo.

### 2.2.7 Análise da precisão dos modelos de IE

Para análise da base de dados climáticos e Evapotranspiração de Referência ( $ET_{\text{FAO-56}}$ ), foram utilizadas técnicas de estatística descritiva e indicativos estatísticos.

A variável calculada  $ET_{\text{FAO-56}}$  é utilizada neste estudo como a referência para os processos de validação cruzada das técnicas: Krigagem Ordinária (KO) e Mínima Curvatura (MC). Para tanto, são utilizados os indicativos estatísticos: coeficiente de determinação ( $R^2$ ), coeficiente de correlação ( $r$ ), *Mean Bias Error* (MBE), *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE) e Intervalo de Tolerância (IT). Além das versões em porcentagem dos indicativos: *Relative Mean Bias Error* (rMBE) e *Relative Root Mean Squared Error* (rRMSE).

A seguir, são descritas detalhadamente as técnicas utilizadas para caracterização do banco de dados e validação cruzada dos modelos de interpolação espacial de  $ET_{\text{FAO-56}}$ .

#### 2.2.7.1 Validação cruzada dos métodos de IE

Para a avaliação do desempenho estatístico dos modelos de estimativa de  $ET_{\text{FAO-56}}$  foram utilizados em conjunto os seguintes indicativos estatísticos: coeficiente de determinação ( $R^2$ ), coeficiente de correlação ( $r$ ), *Mean Bias Error* (MBE), *Root Mean Square Error* (RMSE).

A utilização conjunta desses indicativos auxilia na análise dos modelos de melhor desempenho na estimativa de  $ET_{\text{FAO-56}}$ .

Em que o coeficiente de determinação ( $R^2$ ) indica o nível de variação da variável dependente  $ET_{\text{FAO-56}}$  atribuída às variáveis climáticas (variáveis independentes) (Equação (10)):

$$R^2 = \frac{[\sum_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X})]^2}{\sum_{i=1}^n (Y_i - \bar{Y}) \sum_{i=1}^n (X_i - \bar{X})} \quad (10)$$

O coeficiente de correlação ( $r$ ) será utilizado para revelar o grau de relação entre  $ET_{\text{FAO-56}}$  e estimativas de  $ET_{\text{FAO-56}}$  por meio de outras técnicas estatísticas (Equação 11):

$$r = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X})}{\sqrt{\sum_{i=1}^n (Y_i - \bar{Y}) \sum_{i=1}^n (X_i - \bar{X})}} \quad (11)$$

O *Mean Bias Error* (MBE) (Equação x) indica por meio da diferença de valores estimados e medidos, o desvio da média para indicar se existe superestimação ou subestimação na estimativa de  $ET_{\text{FAO-56}}$  (JACOVIDES et al., 1995). E quanto mais próximo de zero, melhor o ajuste do modelo (Equação 12).

$$MBE = \frac{\sum_{i=1}^n (Y_i - X_i)}{n} \quad (12)$$

O *Mean Square Error* (MSE) indica o erro médio entre os valores estimados e medidos. Este indicativo permite analisar a proporção do valor real do erro produzido pelo modelo (13):

$$MSE = \left[ \frac{\sum_{i=1}^n (Y_i - X_i)^2}{n} \right] \quad (13)$$

O *Root Mean Square Error* (RMSE) (Equação 14) fornece informações sobre o desempenho de modelos por meio da raiz quadrada do erro médio. Seu valor é sempre positivo e quanto menor os valores obtidos de *RMSE* (espalhamento), melhor desempenho dos modelos (STONE, 1993; JACOVIDES et al., 1995; SILVA et al., 2017). Ou seja:

$$RMSE = \left[ \frac{\sum_{i=1}^n (Y_i - X_i)^2}{n} \right]^{\frac{1}{2}} \quad (14)$$

$Y_i$  representa os valores estimados,  $\bar{Y}_i$  são os valores médios estimados,  $X_i$  são os valores medidos,  $\bar{X}$  são os valores médios medidos e  $n$  é o número de observações.

E por fim o Intervalo de Tolerância (IT), como mostra a Equação (15):

$$IT = 1 \pm \sqrt{\frac{2}{n}} \quad (15)$$

em que  $n$  é o número de observações. Mais detalhes sobre essa metodologia (YOUNG, 2010; KRISHNAMOORTHY e LIAN, 2011).

### 2.3 RESULTADOS E DISCUSSÃO

Com base nos dados de Evapotranspiração de Referência ( $ET_{OFAO-56}$ ) obtidos do Planalto Ocidental Paulista no período (2013-2017), é realizado um estudo de caracterização de  $ET_{OFAO-56}$  nesta região. Com alguns interpoladores espaciais para analisar o desempenho estatístico destes e sua distribuição espacial sobre a região de estudo.

Para este estudo, foram considerados os interpoladores espaciais: Krigagem Ordinária (KO) e Mínima Curvatura (MC). Os resultados das simulações desses interpoladores são apresentados abaixo.

É importante destacar que o interpolador estocástico Krigagem Ordinária (KO) foi testado nas mesmas condições dos demais interpoladores (nas partições anuais e sazonais) e não obteve dependência espacial para em seguida passar pela validação cruzada. Foi verificado, após análise dos variogramas obtidos com a técnica KO, efeito

pepita puro e os gráficos desta análise estão inseridos no Apêndice I. Castro et al. (2010) destaca que a alta variabilidade espacial da temperatura do ar resultante do relevo pode ser apontada como possível causa para o insucesso de tentativa de utilização da técnica KO.

A seguir, são sequencialmente apresentadas uma análise descritiva da base de dados de  $ET_{OFAO-56}$ . E conseqüentemente a validação cruzada dos modelos de interpolação espacial e seus mapas confeccionados mostrando a distribuição de  $ET_{OFAO-56}$  na região de estudo.

### 2.3.1 Análise da estatística descritiva

A estatística descritiva apresentada na Tabela 2 mostra os valores máximo, mínimo, médio desvio padrão (SD,  $\pm$ ), Coeficiente de Variação (CV, %), Curtose (K) e Assimetria (As) da  $ET_{OFAO-56}$  obtida para as 30 localidades contempladas com EMA's na região de estudo para os anos (2013 - 2017).

**Tabela 2.** Indicativos da estatística descritiva utilizados: Desvio Padrão (SD), Coeficiente de Variação (CV), Coeficiente de Curtose (K), Coeficiente de Assimetria (As).

Sazonalidade	$ET_{OFAO-56}$ (mm dia <sup>-1</sup> )						
	Máximo	Mínimo	Média	SD ( $\pm$ )	CV (%)	K	As
Verão	9,43	0,84	4,99	1,00	19,98	-0,08	-0,36
Outono	8,68	0,87	3,55	0,80	22,62	-0,14	0,08
Inverno	8,62	0,82	3,99	1,04	25,97	-0,18	0,23
Primavera	8,82	0,93	5,21	1,11	21,24	-0,09	-0,51
Anual	9,43	0,82	4,40	1,21	27,46	-0,64	0,10

Nível de confiança: 95%.

A Tabela 3 mostra que os resultados obtidos de SD pelos períodos sazonais verão, inverno e primavera apresentaram maior variabilidade em relação as suas médias do que o outono. E que o período anual obteve o maior valor de SD em relação ao seu valor médio, quando comparado com os demais períodos sazonais.

Os valores médios obtidos entre os diferentes períodos sazonais analisados apresentaram valores mínimos (0,822 mm dia<sup>-1</sup>) no inverno e máxima (9,432 mm dia<sup>-1</sup>) no verão. E valores médios com intervalo de variação entre 3,55 e 5,21 mm dia<sup>-1</sup> em decorrência dos efeitos sazonais sobre a demanda atmosférica ao longo do ano. Os maiores valores do SD indicam maior variabilidade em relação aos seus valores médios. Mardikis et al. (2005) obteve intervalo de variação de SD inferior aos obtidos neste estudo.

Tangune (2017) em seu estudo, interpolando a Evapotranspiração de Referência ( $ET_{\text{FAO-56}}$ ) por meio da técnica krigagem ordinária (método estocástico) no Estado de São Paulo (Brasil), observou que os maiores e menores valores de  $ET_{\text{FAO-56}}$  ocorrem respectivamente nos períodos sazonais de transição verão-primavera e inverno respectivamente, com valores médios de  $ET_{\text{FAO-56}}$  em torno de  $3,82 \text{ mm dia}^{-1}$ ; Vilanova et. al. (2012) verificou, ao analisar dados de algumas localidades do Estado de Minas Gerais, que os maiores e menores valores de  $ET_{\text{FAO-56}}$  ocorrem respectivamente na transição entre os meses dezembro-janeiro (verão) e junho (inverno), com média anual de  $4,40 \text{ mm dia}^{-1}$  de  $ET_{\text{FAO-56}}$  na região; Mardikis et al. (2005) na Grécia (hemisfério Norte) verificou que os maiores e menores valores de  $ET_{\text{FAO-56}}$  ocorrem em agosto (transição do verão para o outono) e dezembro (inverno), com valores médios variando ao longo do ano entre  $1,20$  e  $6,02 \text{ mm dia}^{-1}$ .

Os valores do Coeficiente de Variação (CV) obtido na região de estudo variaram entre  $19,98$  e  $27,46\%$  entre a partição sazonal verão e anual, respectivamente. Esses valores podem ser classificados em termos de variação, de acordo com a classificação de Pimentel Gomes (1984), como de médio a alto. Esses resultados tiveram variabilidade de CV inferior aos encontrados por Mardikis et al. (2005) na Grécia.

### **2.3.2 Validação cruzada dos métodos: KO e MC**

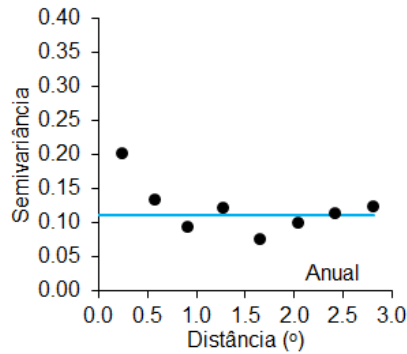
Nesta seção são mostradas sequencialmente as validações cruzadas e as distribuições espaciais dos métodos de interpolação espacial com  $ET_{\text{FAO-56}}$  utilizados [mínima curvatura; na região do Planalto Ocidental Paulista (integração das regiões intermediárias de São José do Rio Preto e Ribeirão Preto)].

As validações cruzadas dos métodos utilizados neste estudo são apresentadas por meio de tabelas que mostram o desempenho e a comparação dos modelos sazonais e anuais. Foi considerado o Intervalo de Tolerância (IT) de  $0,74$  a  $1,26$  para o número de amostras ( $n = 30$ ) deste estudo.

A Figura 2 mostra o variograma da partição anual de  $ET_{\text{FAO-56}}$  na região de estudo e verifica-se a ocorrência de efeito pepita puro por haver inexistência de patamar no variograma teórico. Landim (2006) afirma que o interpolador geoestatístico krigagem não pode ser utilizado quando os dados não apresentarem dependência espacial, sendo recomendada a utilização de outros tipos de interpoladores. Os

variogramas sazonais (verão, outono, inverno e primavera) estão localizados no Apêndice I.

**Figura 2.** Variogramas das partições: sazonais [(a) primavera, (b) verão, (c) outono e (d) inverno] e anual (e).



A Tabela 3 mostra os parâmetros variográficos da técnica KO. Esses parâmetros corroboram com as informações verificadas na Figura 2. Os parâmetros variográficos sazonais estão inseridos no Apêndice I.

**Tabela 3.** Parâmetros variográficos de  $ET_{\text{FAO-56}}$ .

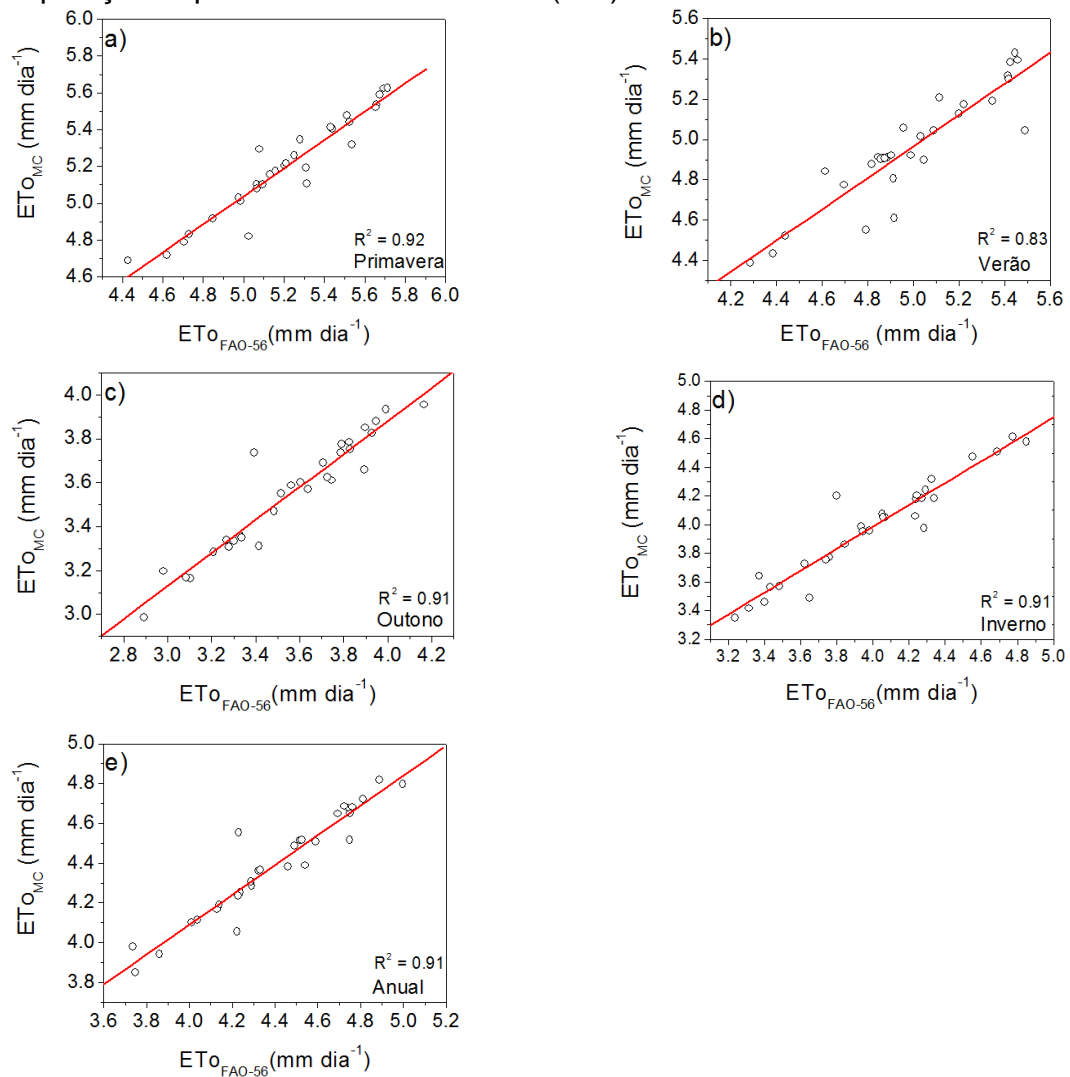
Distância (°)	Semivariância	Pares	Efeito pepita
0,25	0,20	14,00	0,11
0,58	0,13	43,00	0,11
0,92	0,09	58,00	0,11
1,28	0,12	47,00	0,11
1,66	0,07	43,00	0,11
2,06	0,10	44,00	0,11
2,43	0,11	34,00	0,11
2,82	0,12	36,00	0,11

Em seguida, são apresentados mapas que mostram a distribuição espacial de  $ET_{\text{FAO-56}}$  e a sua variação na região de estudo sendo representada por meio de diferentes cores nos mapas. Essas distribuições espaciais são realizadas por meio dos diferentes interpoladores determinísticos utilizados que se distinguem de acordo com o ajuste revelado pela validação cruzada.

A Figura 3 (a - e) mostra os valores obtidos pelo coeficiente de determinação ( $R^2$ ) da validação cruzada do método de interpolação espacial da Mínima Curvatura (MC) para  $ET_{\text{FAO-56}}$  anual e sazonais (primavera, verão, outono e inverno). Esses resultados revelam variação do coeficiente de determinação ( $R^2$ ) entre 0,83 e 0,92, indicando que existe bom ajuste matemático da predição de  $ET_{\text{FAO-56}}$  por meio do

método MC. Com destaque para o modelo Primavera, que obteve numericamente o maior valor de  $R^2$ .

**Figura 3 (a - e).** Dispersão dos valores estimados de  $ET_{FAO-56}$  utilizando o método de interpolação espacial Mínima Curvatura (MC) no Planalto Ocidental Paulista.



Na Tabela 4, estão apresentados os coeficientes angulares da reta de regressão linear simples (a) e o coeficiente linear (b). Além dos indicativos estatísticos da validação cruzada: *Mean Bias Error* (MBE), *Mean Squared Error* (MSE) e *Root Mean Squared Error* (RMSE) e coeficiente de correlação (r).

**Tabela 4.** Validação cruzada do método da Mínima Curvatura (MC) dos valores estimados de MC.

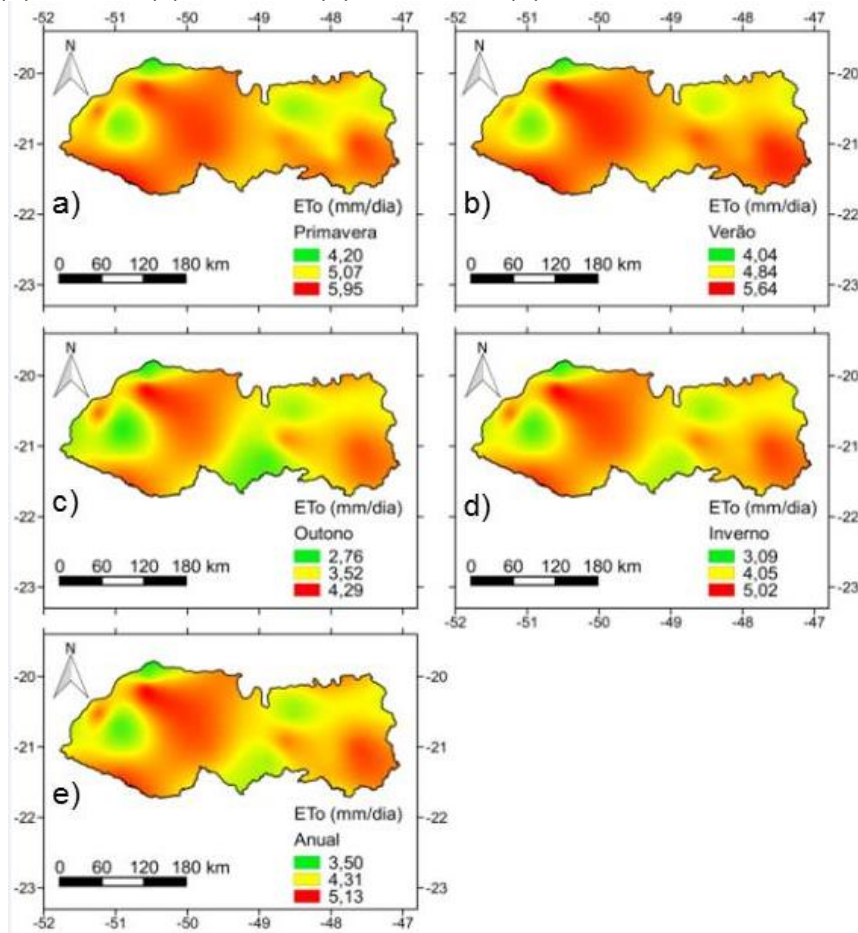
Modelos	MC					
	<i>a</i>	<i>b</i>	MBE	MSE	RMSE	r
	(-)	(-)	(mm dia <sup>-1</sup> )	(mm dia <sup>-1</sup> )	(mm dia <sup>-1</sup> )	(-)
<b>Primavera</b>	1,22	0,76	-0,01	0,01	0,11	0,96
<b>Verão</b>	4,13	0,17	-0,03	0,02	0,14	0,91
<b>Outono</b>	0,88	0,75	-0,01	0,01	0,11	0,95
<b>Inverno</b>	0,93	0,76	-0,01	0,02	0,14	0,96
<b>Anual</b>	1,09	0,75	-0,01	0,01	0,11	0,95

Analisando-se a Tabela 4, verifica-se que os indicativos estatísticos possuem valores similares. Com baixo espalhamento e coeficiente de correlação (*r*) acima de 0,90.

Nas Figuras 4 (a - e) são apresentados os mapas da interpolação espacial de  $ET_{FAO-56}$  (média de 4 anos) para as 30 localidades da região de estudo utilizando o método MC. Nesses mapas, a distribuição espacial de  $ET_{FAO-56}$  é representada em partição sazonal (primavera, verão, outono e inverno) e anual e as cores vermelho-alaranjado representam localidades com altos valores de  $ET_{FAO-56}$ , ou seja, alta demanda atmosférica e as cores amarelo-verde representam as localidades com menores valores de  $ET_{FAO-56}$  (menor demanda atmosférica). As diferenças dos valores de  $ET_{FAO-56}$  podem ser atribuídas ao efeito combinado dos fatores e elementos de clima locais que influenciam essa variável.



**Figura 4 (a - e).** Mapa da distribuição espacial de  $ET_{FAO-56}$  no Planalto Ocidental Paulista utilizando o método de interpolação espacial Mínima Curvatura (MC): (a) Primavera, (b) Verão, (c) Outono, (d) Inverno e (e) anual.



A coloração dos mapas varia conforme os níveis de demanda atmosférica ( $ET_{FAO-56}$ ) em cada época do ano. Em que se pode verificar elevada demanda atmosférica na estação Primavera ( $5,95 \text{ mm dia}^{-1}$ ; Figura 4a), provavelmente no período de transição com o verão, e menor demanda atmosférica na transição outono-inverno ( $2,76 \text{ mm dia}^{-1}$ ; Figura 4b). Nas partições sazonais de verão e primavera são verificadas as menores amplitudes de valores de  $ET_{FAO-56}$  enquanto que gradualmente a amplitude desses valores vai aumentando nas outras partições sazonais. Com a partição anual (Figura 4e) representando os valores médios dos valores observados na região. Esses resultados indicam alta variabilidade espacial entre essas diferentes partições.

As regiões que englobam os municípios de Ituverava (SP), Franca (SP), Passos (MG) e Sacramento (MG) possuem altitudes variando entre 600 e 1026 m de altura que na parte mais ao nordeste das Figuras 6 (a - e) mostram que pode haver influencia

espaço-temporal sobre os níveis de  $ET_{FAO-56}$  ao longo do ano nas partições anual e sazonais. Mais ao centro das Figuras 4 (a - e) o relevo dessa parte da região de estudo pode ter influenciado a variabilidade dos valores de  $ET_{FAO-56}$  local e razão dos municípios de Monte Azul Paulista e Brodowsky possuem altitudes variando entre 611 e 863 m de altitude. À Oeste da região de estudo é verificado que isoladamente existem pequenas partes do território com menores valores de  $ET_{FAO-56}$  com possíveis influencias microclimáticas.

Não é possível verificar e afirmar que exista uma amplitude significativa dos valores de  $ET_{FAO-56}$  entre as latitudes mais extremas da região de estudo. Ou que a demanda atmosférica da região possa diminuir em razão do aumento das latitudes da região. A análise dos mapas revela que o comportamento de  $ET_{FAO-56}$  pode ser explicado pela variação do clima em razão da mudança das estações do ano e do microclima local. Este é influenciado por particularidades do relevo local que afetam o ciclo hidrológico da regional.

Dentro do âmbito climatológico, outros pesquisadores testaram e analisaram o desempenho de diferentes métodos de interpolação espacial: Alves et al. (2012) constataram que o interpolador mais eficiente também foi o da Mínima Curvatura (MC) para interpolar a precipitação pluvial no estado de Goiás (Brasil); para Attorre et al. (2007) IDW e KO obtiveram resultados satisfatórios na Itália; para Silva et al. (2013) avaliando o método do IDW no sertão do Ceará com a utilização da variável evapotranspiração, os resultados foram satisfatórios; Alves et al. (2012) testando diversos métodos de interpolação espacial, verificaram que o método de melhor desempenho foi o de Krigagem Ordinária (KO) para avaliar a temperatura do ar; Mardikis et al. (2005) e Santos et al. (2017) também concluíram que técnica a Krigagem Ordinária (KO) obteve melhor desempenho estatístico em seus estudos analisando a precipitação pluvial e evapotranspiração. Sugere-se, portanto, que não há um método de interpolação espacial de consenso na literatura, o mais indicado é o que se ajusta melhor às condições climáticas locais, quantidade de estações meteorológicas disponíveis e à distância entre estas.

## **2.4 CONCLUSÃO**

Dos resultados apresentados e discutidos neste estudo, conclui-se que:

- O método da Mínima Curvatura se ajustou ao conjunto de dados e obteve validação cruzada com desempenho superior aos dos demais métodos utilizados;

- A tentativa de interpolação espacial da base dados de  $ET_{OFAO-56}$  com a técnica Krigagem Ordinária (KO) não resultou em dependência espacial das 30 localidades do Planalto Ocidental Paulista, com análise dos variogramas indicando efeito pepita puro. O insucesso de tentativa de utilização da técnica KO ocorreu em razão do número de estações meteorológicas instaladas na região ser considerado aquém do ideal para utilização dessa técnica, mesmo que classificação de Koppen regional (verificada como Aw) indique similaridade climática entre a maior parte das localidades estudadas.

## REFERÊNCIAS

- ADELOYE, A. J.; RUSTUM, R.; KARIYAMA, I. D. Neural Computing Modeling of the reference crop evapotranspiration. **Environmental Modelling e Software**, v. 29, pp. 61-73, 2012.
- AGRITEMPO – Sistema de Monitoramento Agrometeorológico. Estações meteorológicas para o Estado de São Paulo. Disponível em: <<http://www.agritempo.gov.br/agritempo/jsp/Estacao/index.jsp?siglaUF=SP>>. Acesso em 28 de dezembro de 2018.
- ALLEN, R. G.; PEREIRA, L. S.; RAES, D. et al. Crop evapotranspiration – guidelines for computing crop water requirements. pp. 50. **Food and Agriculture Organization of the United Nations (FAO)** - Irrigation and drainage paper 56, Rome, 1998.
- ALVES, E.D.L.; BIUDES, M.S.; VECCHIA, F.A.S. Interpolação espacial na climatologia: Análise dos critérios que precedem sua aplicação. **Revista Geonorte**, Edição Especial 2, V.1, N.5, pp.606-618, 2012.
- ARANOFF, S. **Geographic information systems: a management perspective**. Ottawa: WDL Publications, 294p., 1989.
- BENEVIDES, J. G; LOPEZ, D. Formula para el caculo de la evapotranspiracion potencial adaptada al tropico (15° N - 15° S). **Agronomia Tropical**, v. 20, n. 5, pp. 335-345, 1970.
- BLANEY, H. F.; CRIDDLE, W. D. **Determining Water Requirements in Irrigated Areas from Climatological and Irrigation Data**. USDA. SCSTP- 96, p. 48, 1950.
- CASTRO, F. S.; PEZZOPANE, J. E. M.; CECÍLIO, R. A. Avaliação do desempenho dos diferentes métodos de interpoladores para parâmetros do balanço hídrico climatológico. **Revista Brasileira de Engenharia Agrícola**, Campina Grande, v. 14, n. 2, p. 871-880, 2010.
- CENTRO DE PESQUISAS METEOROLÓGICAS E CLIMÁTICAS APLICADAS A AGRICULTURA - **CEPAGRI**. Clima dos Municípios Paulistas. Disponível em: <<http://www.cpa.unicamp.br/outras-informacoes/clima-dos-municipios-paulistas.html>>. Acesso em: 10 de fev. 2018.
- COELHO, C. A. S.; CARDOSO, D. H. F.; FIRPO, M. A. Precipitation diagnostics of an exceptionally dry event in São Paulo, Brazil. **Theoretical and Applied Climatology**, DOI 10.1007/s00704-015-1540-9, 2016a.
- COELHO, C. A. S.; OLIVEIRA, C. P.; AMBRIZZI, T. et al. The 2014 southeast Brazil austral summer drought: regional scale mechanisms and teleconnections. **Climate Dynamics**, DOI 10.1007/s00704-015-1540-9, 2016b.
- DAI, F.; ZHOU, Q.; LV, Z. et al. Spatial prediction of soil organic matter content integrating artificialneural network and ordinary kriging in Tibetan Plateau. **Ecological Indicators**, v. 45, pp. 184-194, 2014.

- DALEZIOS, N. R.; LOUKAS, A.; BAMPZELIS, D. Spatial variability of reference evapotranspiration in Greece. **Physics and Chemistry of the Earth**, v. 27, pp. 1031-1038, 2002.
- DAVIS, J.C. 1986. **Statistics and data analysis in geology**. Second Edition. John Wiley and Sons Inc., New York, 646 p.
- FERREIRA SILVA, C. O.; MANZIONE, R. L.; ALBUQUERQUE FILHO, J. L. Large-Scale Spatial Modeling of Crop Coefficient and Biomass Production in Agroecosystems in Southeast Brazil. **Horticulturae**, doi: 10.3390/horticulturae 4040044, 2018.
- Golden Software. Surfer versão 9. Tutorial. **User's guide**, 2010.
- HARGREAVES, G. H, SAMANI, Z. A. Reference crop evapotranspiration from temperature. **Applied Engineering in Agriculture**, v. 1, n. 2, pp. 96-99, 1985.
- Instituto Brasileiro de Geografia e Estatística (IBGE)**. Acessado em 2018.
- Instituto de Economia Agrícola do Estado de São Paulo (**IEA**). Disponível em: <<http://www.iea.agricultura.sp.gov.br/out/Bancodedaos2.html>>. Acesso em: 21 de dezembro de 2018.
- IQBAL, M. **An introduction to solar radiation**. New York: Academic Press, 1983. 390 p.
- HODAM, S.; SARKAR, S.; MARAK, A. G. R. et al. Spatial Interpolation of Reference Evapotranspiration in India: Comparison of IDW and Kriging Methods. **Journal of The Institution of Engineers (India): Series A**, 2017. DOI 10.1007/s40030-017-0241-z.
- JACOVIDES, C. P; KONTOYIANNIS, H. Statistical procedures for the evaluation of evapotranspiration computing models. **Agricultural Water Management**, v. 27, pp. 365-371, 1995.
- JOURNEL, A. G. e HUIJBREGTS, C.J. **Mining Geostatistics**. London: Academic Press. 600p., 1978.
- KÖPPEN, W.; GEIGER, R. *Klimate der Erde*. Gotha: **Verlag Justus Perthes**. Wall map 150cmx200cm, 1928.
- KRIGE, D. G. A statistical approach to some basic mine valuation problems on the Witwatersrand. **Journal of the Chemical, Metallurgical and Mining Society of South Africa**, Dec. 1951:119-139, 1951.
- KRISHNAMOORTHY, K.; LIAN, X. Closed-form approximate tolerance intervals for some general linear models and comparison studies. **Journal of Statistical Computation and Simulation**, v. 82, pp. 547-563, 2012.
- LANDIM, P.M.B. Sobre Geoestatística e mapas. **Terra e Didática**, v.2, n.1, pp.19–33, 2006.
- LIMA ALVES, E.D.; VECCHIA, F.A.S. Análise de diferentes métodos de interpolação para a precipitação pluvial no Estado de Goiás. **Acta Scientiarum. Human and Social Sciences**. v. 33, n. 2, pp. 193-197, 2011. DOI: 10.4025/actascihumansoc.v33i2.13815.
- MARDIKIS, M. G.; KALIVAS, D. P.; KOLLIAS, V. J. Comparison of Interpolation Methods for the Prediction of Reference Evapotranspiration—An Application in Greece. **Water Resources Management**, v. 19, pp. 251-278, 2005.
- MARTÍ, P.; ZARZO, M. Multivariate statistical monitoring of ETo: A new approach for estimation in nearby locations using geographical inputs. **Agricultural and Forest Meteorology**, v. 152, pp. 125 - 134, 2012.
- MATHERON, G. *Traité de Géostatistique Appliquée*, Tome I : Mémoires du Bureau de Recherches Géologiques et Minières. Paris: **Editions Technip**, v. 14, 1962.
- MATHERON, G. *Traité de Géostatistique Appliquée*, Tome II : Mémoires du Bureau de Recherches Géologiques et Minières. Paris: **Editions Technip**, v. 24, 1963.
- MATHERON, G. *Les variables régionalisées et leur estimation: une application de la théorie des fonctions aléatoires aux sciences de la nature*. Paris: **Masson et Cie**, 1965.
- MAZZINI, P. L. F.; SCHETTINI, C. A. F. Avaliação de metodologias de interpolação espacial aplicadas a dados hidrográficos costeiros quase sinóticos. **Brazilian Journal of Aquatic Science and Technology**, v. 13, n. 1, pp. 53-64, 2009.

- MEDINA, A. S.; QUINTAS, M.C.L. "O Método de Interpolação de Curvatura Mínima", **II Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação**. Recife, 8 - 11 de setembro de 2008.
- NOBRE, P.; ALMEIDA, R. A.; MALAGUTTI, M.; GIAROLLA, E. Coupled Ocean Atmosphere Variations over the South Atlantic Ocean. **American Meteorological Society**, DOI: 10.1175/JCLI-D-11-00444.1, 2012.
- ORACLE CORPORATION. **Chapter 1 General Information**. MySQL 5.7 Reference Manual, 2019. Disponível em: <https://dev.mysql.com/doc/refman/5.7/en/introduction.html>. Acesso em: 08 abr. 2019a.
- ORACLE CORPORATION. **Chapter 12 Functions and Operators**. MySQL 5.7 Reference Manual, 2019. Disponível em: <https://dev.mysql.com/doc/refman/5.7/en/functions.html>. Acesso em: 08 abr. 2019b.
- PIMENTEL GOMES, F. **A estatística moderna na pesquisa agropecuária**. Piracicaba: Associação Brasileira para Pesquisa da Potassa e do Fosfato, 1984. 160 p.
- PRIESTLEY, C. H. B., TAYLOR, R. J. On the assessment of surface heat flux and evaporation using large-scale parameters. **Monthly Weather Review**, v. 100, n. 2, pp. 81-92, 1972.
- ROLIM, G. S.; CAMARGO, M. B. P.; LANIA, D. G.; et al. Classificação climática de Köppen e de Thornthwaite e sua aplicabilidade na determinação de zonas agroclimáticas para o estado de São Paulo. **Bragantia**, Campinas, v. 66, n. 4, p. 711-720, 2007.
- SANTOS, T. O.; NETO, M. L. P.; ALMEIDA, G. L. P. et al. **Análise temporal da variação da temperatura em uma região do agreste pernambucano utilizando-se a geoestatística**. Apresentado no XVII Congresso Brasileiro de Agrometeorologia – 18 a 21 de Julho de 2011.
- SANTOS, H. S.; CUNHA, G. N.; CASTRO, J. R. Avaliação Qualitativa das Técnicas de Interpolação de Mínima Curvatura, Krigagem e Bidirecional na Formação de Imagens. **Revista de Engenharias da Faculdade Salesiana**, n. 4, pp. 2-16, 2016.
- SANTOS, H. S.; CUNHA, G. N.; CASTRO, J. R. Space-time variability of evapotranspiration and precipitation in the State of Paraná, Brazil. **Ambiente e Água - An Interdisciplinary Journal of Applied Science**, n. 5, v. 12, pp. 743 - 759, 2017.
- SATYAMURTY, P.; NOBRE, A. C.; SILVA, P. L. D. Topics: **South America. Meteorological Monographs**, v.27, n.49, pp.119-139, 1998.
- SENTELHAS, P. C.; GILLESPIE, T. J.; SANTOS, E. A. Evaluation of FAO Penman Monteith and alternative methods for estimating reference evapotranspiration with missing data in Southern Ontario, Canada. **Agricultural Water Management**, v. 97, pp. 635-644, 2010.
- SILVA, M. G.; ARRAES, F. D. D.; LEDO, E. R. F. et al. Avaliação da evapotranspiração de referência por Penman-Monteith usando dados climáticos mínimos no sertão do Ceará. **Revista Agro@ambiente**, v. 7, n. 3, pp. 284-293, 2013.
- SILVA, M. B. P.; ESCOBEDO, J. F.; SANTOS, C. M. et al. Performance of the Angstrom-PreScott Model (A-P) and SVM and ANN techniques to estimate the daily Global Solar Irradiation in Botucatu/SP/Brazil. **Journal of Atmospheric and Solar-Terrestrial Physics**, v. 160, pp. 11-23, 2017.
- SILVA, J. R. Valor da produção agropecuária do Estado de São Paulo: resultado preliminar de 2018. **Análises e Indicadores do Agronegócio**, São Paulo, v. 13, n. 11, p. 1-6, nov. 2018. Disponível em: [http://www.iea.sp.gov.br/ftpiea/aia/AIA-74\\_2018.pdf](http://www.iea.sp.gov.br/ftpiea/aia/AIA-74_2018.pdf). Acesso em: 21 de fevereiro de 2019.
- SHIRI, J.; KISI, O.; LANDERAS, G. et al. Daily reference evapotranspiration modeling by using genetic programming approach in the Basque Country (Northern Spain). **Journal of Hydrology**, v. 414-415, pp. 302-316, 2012.

- SHIRI, J.; SADRADDINI, A. A.; NAZEMI, A. H. et al. Independent testing for assessing the calibration of the Hargreaves-Samani equation: New heuristic alternatives for Iran. **Computers and Electronics in Agriculture**, v. 117, pp. 70-80, 2015.
- SMITH, W. H. F.; WESSEL, P. "Gridding with continuous curvature splines in tension". **Geophysics**, v.55, n.3, pp. 293-305, 1990.
- SNYDER, R. Climate Change Impacts on Water Use in Horticulture. **Horticulturae**, v. 3, n. 27, pp. 1-9, 2017.
- STONE, R. J. Improved statistical procedure for the evaluation of solar radiation estimation models. **Solar Energy**, v. 51, n. 4, pp. 289-291, 1993.
- TANGUNE, B. F., **Evapotranspiração de referência no estado de são paulo: métodos empíricos, aprendizado de máquina e geoespacial**. 2017, 126 p. Tese (Agronomia) – Universidade Estadual Paulista, Botucatu-SP, 2017.
- TANGUNE, B. F.; ESCOBEDO, J. F. Reference evapotranspiration in São Paulo State: Empirical methods and machine learning techniques. **International Journal of Water Resources and Environmental Engineering**, v. 10, n. 4, pp. 33-44, 2018.
- THORNTHWAITE, C. W. An approach toward a rational classification of climate. **Geographical Review**, v. 55–94, pp. 38,1948.
- VANDERLINDEN, K.; GIRÁLDEZ, J. V.; MEIRVENNE, M. V. Spatial Estimation of Reference Evapotranspiration in Andalusia, Spain. **Journal of Hydrometeorology**, DOI: 10.1175/2007JHM880.1, 2008.
- VAROUCHAKIS, E. A.; HRISTOPULOS, D. T. Comparison of stochastic and deterministic methods for mapping groundwater level spatial variability in sparsely monitored basins. **Environmental Monitoring and Assessment**. v. 185, pp. 1-19, 2013.
- VIEIRA, S.R; HATFIELD, J.L.; NIELSEN, D.R; BIGGAR, J. W. Geostatistical theory and application to variability of some agronomical properties. **Hilgardia**, v.51, n.3, pp.1-75, 1983.
- VILANOVA, M. R. N.; SIMÕES, S. J. C.; TRANNIN, I. C. B. Interpolação geoespacial da evapotranspiração de referência (ET<sub>o</sub>) em regiões com escassez de dados. **Revista Ambiente e Água**. v. 7, n.2, 2012.
- WEBSTER, R.; OLIVER, M.A. **Geostatistics for Environmental Scientists**. 2. ed. Chichester: Wiley and Sons, 2007. 315p.
- OLIVER, M.A. **An Overview of Geostatistics and Precision Agriculture**. In: ed. Geostatistical Applications for Precision Agriculture. Dordrecht: Springer, p.1–34, 2010.
- YOUNG, D. S. tolerance: An R Package for Estimating Tolerance Intervals. **Journal of Statistical Software**, v. 36, pp. 1-39, 2010.
- ZIMBACK, C. R. L. **Análise espacial de atributos químicos de solos para fins de mapeamento da fertilidade do solo**. 2001. 114 f. Tese (Livre – Docência) – Faculdade de Ciências Agrônômicas, Universidade Estadual Paulista, Botucatu, 2001.

## CONSIDERAÇÕES FINAIS

Com base nos resultados apresentados nos dois capítulos deste estudo, conclui-se que:

- Entre as técnicas utilizadas para estimativa da Evapotranspiração de Referência: os modelos de Regressões Múltiplas (RM) da quarta combinação se revelaram mais práticos e acessíveis para predição desta variável em razão de sua maior simplicidade e menor esforço computacional, quando comparados aos modelos de Aprendizado de Máquinas (AM). E que a técnica multivariada de Componentes Principais (CP) revelou-se bastante interessante para preenchimento de valores faltantes da base de dados, análise exploratória das variáveis climáticas e predição de  $ET_{O_{FAO-56}}$ . Esta técnica analisa o efeito conjunto das variáveis e determina gradualmente a variável mais importante para a construção de modelos de estimativa de  $ET_{O_{FAO-56}}$ . E também é capaz de realizar estimativa de  $ET_{O_{FAO-56}}$  com base na proximidade entre as EMA's com boa precisão e exatidão.
- Que o método determinístico da Mínima Curvatura (MC) é o mais indicado para utilizar a base de dados de Evapotranspiração de Referência ( $ET_{O_{FAO-56}}$ ) na região do Planalto Ocidental Paulista. Este método mostrou desempenho estatístico satisfatório após a validação cruzada.





## REFERÊNCIAS

- ABTEW, W. Evapotranspiration measurements and modeling for three wetland systems in South Florida. **Journal of the American Water Resources Association**, v. 127, n. 3, pp. 140-147, 1966.
- ALBRECHT, F. Die Methoden zur Bestimmung Verdunstung der natürlichen Erdoberfläche. **Arch Meteor Geoph Biokl**, Serie, B2, v. 1–38, 1950.
- ALLEN, R. G.; PEREIRA, L. S.; RAES, D. et al. Crop evapotranspiration - guidelines for computing crop water requirements. pp. 50. **Food and Agriculture Organization of the United Nations (FAO) - Irrigation and drainage paper 56**, Rome, 1998.
- ALLEN, R. G.; PRUITT, W. O. FAO-24 reference evapotranspiration factors. **Journal of Irrigation and Drainage Engineering**, v. 175, n. 5, pp. 758 - 773, 1991.
- ALTHOFF, D.; BAZAME, H. C.; FILGUEIRAS, R. et al. Heuristic methods applied in reference evapotranspiration modeling. **Ciência e Agrotecnologia**, v. 42, n. 3, pp. 314-324, 2018.
- BENEVIDES, J. G; LOPEZ, D. Formula para el caculo de la evapotranspiracion potencial adaptada al tropico (15° N - 15° S). **Agronomia Tropical**, v. 20, n. 5, pp. 335-345, 1970.
- BLANEY, H. F.; CRIDDLE, W. D. **Determining Water Requirements in Irrigated Areas from Climatological and Irrigation Data**. USDA. SCSTP- 96, p. 48, 1950.
- BROCKAMP, B.; WENNER, H. Verdunstungsmessungen auf den Steiner See bei Münster. **Dt Gewsserkuändl**, Mitt. 7, pp. 149–154, 1963.
- CUENCA, R. H. Irrigation system design: an engineering approach. **Prentice-Hall, Englewood Cliffs**, NJ, p. 133, 1989.
- DALTON, J. M. Experimental essays on the constitution of mixed gases; on the force of steam of vapour from waters and other liquids in different temperatures, both in a torricellian vacuum and in air on evaporation and on the expansion of gases by heat. **Memoirs and proceedings of the Manchester Literary e Philosophical Society**, v. 5, pp. 535-602, 1802.
- FERREIRA SILVA, C. O.; MANZIONE, R. L.; ALBUQUERQUE FILHO, J. L. Large-Scale Spatial Modeling of Crop Coefficient and Biomass Production in Agroecosystems in Southeast Brazil. **Horticulturae**, doi: 10.3390/horticulturae 4040044, 2018.
- HAMON, W. R. Estimating potential evapotranspiration. **Journal of the Hydraulics Division Proceedings of the American Society of Civil Engineers**, v. 87, pp. 107-120, 1961.
- HARGREAVES, G. H, SAMANI, Z. A. Reference crop evapotranspiration from temperature. **Applied Engineering in Agriculture**, v. 1, n. 2, pp. 96-99, 1985.
- IRMAK, S.; IRMAK, A.; ALLEN, R. G.; JONES, J. W. Solar and net radiation based equations to estimate reference evapotranspiration in humid climates. **Journal of Irrigation and Drainage Engineering**, v. 129, n. 5, pp. 336-347, 2003.

JENSEN, M. E.; HAISE, H. R. Estimating evapotranspiration from solar radiation. **Proceeding of the Journal of Irrigation and Drainage Division: American Society of Civil Engineers**, v. 89, pp. 15-41, 1963.

LANDERAS, G.; ORTIZ-BARREDO, A.; LÓPEZ, J. J. Comparison of artificial neural network models and empirical and semi-empirical equations for daily reference evapotranspiration estimation in the Basque Country (Northern Spain). **Agricultural Water Management**, v. 95, pp. 553-565, 2008.

MAHRINGER, W. Verdunstungsstudien am Neusiedler See. **Arch Met Geoph Biokl Ser B**, v. 18, pp. 1–20, 1970.

MAKKINK, G. F. Testing the Penman formula by means of lysimeters. **Journal of the Institution of Water Engineers**, v. 11, n. 3, pp. 277-288, 1957

MARTÍ, P.; ZARZO, M. Multivariate statistical monitoring of ETo: A new approach for estimation in nearby locations using geographical inputs. **Agricultural and Forest Meteorology**, v. 152, pp. 125 - 134, 2012.

MEYER, A. Über einige Zusammenhänge zwischen Klima und Boden in Europa. **Chemie de Erde**, v. 2, pp. 209 - 347, 1926.

PENMAN, H. L. Natural evaporation from open water, bare soil, and grass. **Proceedings of the Royal Society**, London, v. 193, n. 1, p. 120-146, 1948.

PEREIRA, A. R.; VILLA NOVA, N.; PEREIRA, A. S. et al. A model for the class-A pan coefficient. **Agricultural Water Management**, v. 75, pp. 75-82, 1995.

PRIESTLEY, C. H. B., TAYLOR, R. J. On the assessment of surface heat flux and evaporation using large-scale parameters. **Monthly Weather Review**, v. 100, n. 2, pp. 81-92, 1972.

ROHWER, C. Evaporation from free water surface. **USDA Tech Null**, v. 217, 1-96, 1931.

ROMANENKO, V. A. Computation of the autumn soil moisture using a universal relationship for a large area. In: **Proceedings, Ukrainian Hydrometeorological Research Institute**, n. 3. Kiev, 1961.

SANTOS, C. M.; SOUZA, J. L.; FERREIRA JÚNIOR, R. A. et al. On modeling global solar irradiation using air temperature for Alagoas State, Northeastern Brazil. **Energy**, v. 71, 338-398, 2014.

SENTELHAS, P. C.; GILLESPIE, T. J.; SANTOS, E. A. Evaluation of FAO Penman-Monteith and alternative methods for estimating reference evapotranspiration with missing data in Southern Ontario, Canada. **Agricultural Water Management**, v. 97, pp. 635-644, 2010.

SILVA, M. B. P.; ESCOBEDO, J. F.; SANTOS, C. M. et al. Performance of the Angstrom-Prescott Model (A-P) and SVM and ANN techniques to estimate the daily Global Solar Irradiation in Botucatu/SP/Brazil. **Journal of Atmospheric and Solar–Terrestrial Physics**, v. 160, pp. 11-23, 2017.

SNYDER, R. L. Equation for evaporation pan to evapotranspiration conversions. **Journal of Irrigation and Drainage Engineering**, v. 118, n. 6, pp. 977-980, 1992.

TANGUNE, B. F.; ESCOBEDO, J. F. Reference evapotranspiration in São Paulo State: Empirical methods and machine learning techniques. **International Journal of Water Resources and Environmental Engineering**, v. 10, n. 4, pp. 33-44, 2018.

THORNTHWAITE, C. W. An approach toward a rational classification of climate. **Geographical Review**, v. 55–94, pp. 38, 1948.

TRABERT, W. Neue Beobachtungen über Verdampfungsgeschwindigkeiten. **Meteorologische Zeitschrift**, v. 13, pp. 261-263, 1896.

TURC, L. Évaluation des besoins en eau d'irrigation, évapotranspiration potentielle. **Annales Agronomiques**. V. 12, n. 1, pp. 13-49, 1961.

VANDERLINDEN, K.; GIRÁLDEZ, J. V.; MEIRVENNE, M. V. Spatial Estimation of Reference Evapotranspiration in Andalusia, Spain. **Journal of Hydrometeorology**, DOI: 10.1175/2007JHM880.1, 2008.



## APÊNDICE A – Produção Agrícola do Planalto Ocidental Paulista

Área de produção relativa (Área/prod., %), área com período de cultivo abaixo de um ano (Área/<1 ano, %), área nova de cultivo (Área nova, %), área de corte (Área/corte, %), produção de sacas de 25 kg (Prod. Sc. 25 kg, %), produção de sacas de 60 kg (Prod. Sc. 60 kg, %), produção de caixas de 40,8 kg (Prod. cx. 40,8 kg, %), produção de coágulos 1k (Prod. coág. Kg, %), produção em toneladas (Prod. t, %), produção de novos pés em unidades (Pés/prod. un, %), produção de pés em unidades (Pés/prod. un, %), *ranking*.

Cultura	Área/prod, ha (%)	Área/ < 1 ano ha (%)	Área nova ha (%)	Área/corte ha (%)	Prod, sc,25kg (%)	Prod, sc,60kg (%)	Prod, cx,40,8kg (%)	Prod, coág, kg (%)	Prod, t (%)	Pés/novos un (%)	Pés/prod, un (%)	Ranking
Amendoim/água	34,39	-	-	-	32,16	-	-	-	-	-	-	1
Amendoim/seca	36,85	-	-	-	40,44	-	-	-	-	-	-	1
Café	34,48	-	34,28	-	-	31,26	-	-	-	-	-	2
Cana-de-açúcar	-	-	39,79	41,10	-	-	-	-	42,04	-	-	1
Eucalipto	-	6,71	5,71	-	-	-	-	-	-	-	-	6
Feijão/água	0,71	-	-	-	-	0,56	-	-	-	-	-	4
Feijão/seca	2,16	-	-	-	-	1,73	-	-	-	-	-	7
Feijão/inv./c, irrigação	30,88	-	-	-	-	27,20	-	-	-	-	-	2
Feijão/inv./s, irrigação	4,41	-	-	-	-	3,39	-	-	-	-	-	7
Laranja	-	-	-	-	-	-	23,72	-	-	34,86	25,36	1
Mandioca/indústria	0,49	-	0,36	-	-	-	-	-	0,32	-	-	9
Mandioca/feira	12,46	-	9,15	-	13,01	-	-	-	-	-	-	2
Milho/sequeiro	17,15	-	-	-	-	16,72	-	-	-	-	-	3
Milho/irrigação	23,81	-	-	-	-	22,82	-	-	-	-	-	3
Milho/safrinha	6,99	-	-	-	-	5,53	-	-	-	-	-	5
Seringueira	-	-	-	-	-	-	-	75,21	-	70,00	71,15	1
Soja	18,09	-	-	-	-	16,9	-	-	-	-	-	3

Fonte: IEA (2017).

## APÊNDICE B – Irradiação solar global no topo da atmosfera ( $H_o$ )

A variável astronômica irradiação solar global no topo da atmosfera ( $H_o$ , MJ m<sup>-2</sup>) é expressa da seguinte forma:

$$H_o = 1367 (24/\pi) E_o [(\pi/180) \omega_s \text{ sen } (\varphi) \text{ sen } (\delta) + \cos (\varphi) \cos (\delta) \text{ sen } (\omega_s)]$$

Em que ( $E_o$ ) é o fator de correção da excentricidade da órbita terrestre:

$$E_o = 1,000 + 0,034 \cos(r) + 0,0013 \text{ sen } (r) + 0,000719 \cos (2 r) + 0,000077 \cos (2 r)$$

Ângulo horário solar ( $\omega_s$ , °):

$$\omega_s = \cos^{-1}[-\text{tg } (\varphi) \text{ t } (\delta)]$$

Declinação solar ( $\delta$ , °):

$$\delta = 23,45 \text{ sen } [(360/365) (DJ + 284)]$$

Ângulo do dia ( $\Gamma$ , rad):

$$\Gamma = 2\pi(DJ - 1)/366$$

Fotoperíodo ( $N$ , horas):

$$N = (2\omega_s/15)$$

Latitude local - ( $\varphi$ , °)

Dia Juliano (DJ)

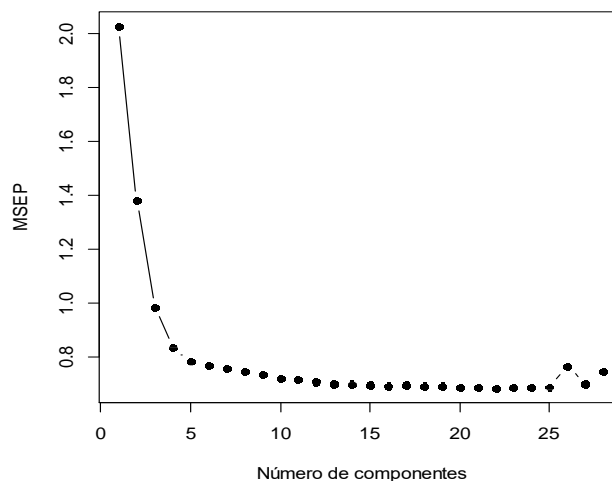
## APÊNDICE C - Percentual de dados presentes

Percentual de dados presentes após padronização do banco de dados.

Tmax	Tmin	H	UR	U <sub>2</sub>
(°C)	(°C)	(MJ m <sup>-2</sup> dia <sup>-1</sup> )	(%)	(m s <sup>-1</sup> )
46650 (94,87%)	46629 (94,83%)	46265 (94,09%)	46684 (94,94%)	42850 (87,15%)

## APÊNDICE D - Componentes Principais para cada variável climática

- Temperatura máxima:



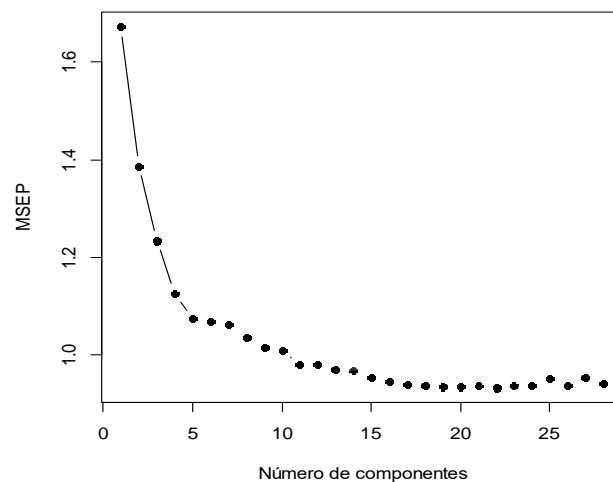
MSEP – Validação cruzada – Método *Kfold* (5% de retirada) - Ver Josse, J. and Husson, F. (2012):

Componentes Principais (CP), autovalores ( $\phi$ , °C), autovalores acumulados ( $\phi_{acm}$ , °C), variabilidade ( $\phi$ , %), variabilidade acumulada ( $\phi_{acm}$ , %).

CP	$\phi$ (°C)	$\phi_{acm}$ (°C)	$\phi$ (%)	$\phi_{acm}$ (%)
1	2,02	2,02	9,04	9,04
2	1,38	3,40	6,16	15,20
3	0,98	4,39	4,40	19,60
4	0,83	5,22	3,72	23,32
5	0,79	6,01	3,51	26,83

Nº de componentes utilizados para estimação dos dados ausentes e realização do PCA: 5

- Temperatura mínima:



MSEP – Validação cruzada – Método *Kfold* (5% de retirada) - Ver Josse, J. and Husson, F. (2012):

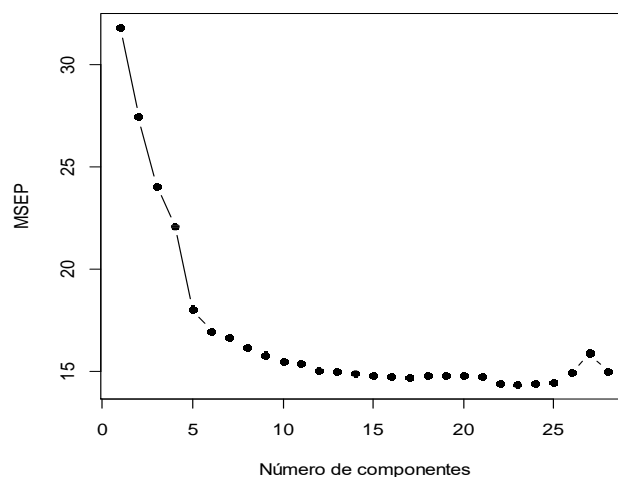
Componentes Principais (CP), autovalores ( $\phi$ , °C), autovalores acumulados ( $\phi_{acm}$ , °C), variabilidade ( $\phi$ , %), variabilidade acumulada ( $\phi_{acm}$ , %).

CP	$\phi$ (°C)	$\phi_{acm}$ (°C)	$\phi$ (%)	$\phi_{acm}$ (%)
1	1,67	1,67	5,82	5,82
2	1,39	3,06	4,83	10,65
3	1,23	4,29	4,29	14,94
4	1,13	5,42	3,92	18,86
5	1,07	6,49	3,74	22,59

Nº de componentes utilizados para estimação dos dados ausentes e realização do PCA: 5.



- Umidade relativa média:



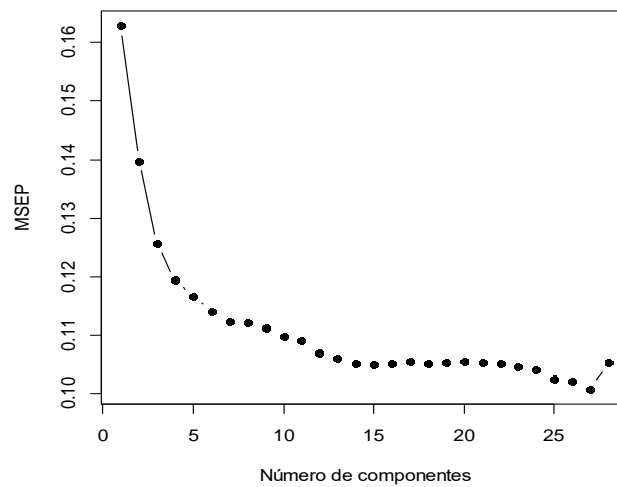
MSEP – Validação cruzada – Método *Kfold* (5% de retirada) - **Ver Josse, J. and Husson, F. (2012)**

Componentes Principais (CP), autovalores ( $\phi$ , °C), autovalores acumulados ( $\phi_{acm}$ , °C), variabilidade ( $\phi$ , %), variabilidade acumulada ( $\phi_{acm}$ , %).

CP	$\phi$ (%)	$\phi_{acm}$ (%)	$\phi$ (%)	$\phi_{acm}$ (%)
1	31,80	31,80	6,74	6,74
2	27,44	59,24	5,81	12,55
3	24,08	83,32	5,10	17,65
4	22,13	105,44	4,69	22,33
5	18,08	123,52	3,83	26,16
6	16,96	140,48	3,59	29,75

Nº de componentes utilizados para estimação dos dados ausentes e realização do PCA: 6.

- Velocidade dos Ventos:



MSEP – Validação cruzada – Método *Kfold* (5% de retirada) - Ver Josse, J. e Husson, F. (2012):

Componentes Principais (CP), autovalores ( $\phi$ , °C), autovalores acumulados ( $\phi_{acm}$ , °C), variabilidade ( $\phi$ , %), variabilidade acumulada ( $\phi_{acm}$ , %).

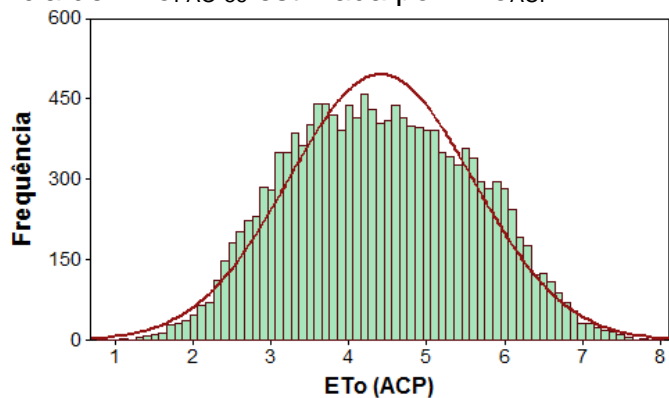
CP	$\phi$ (m s <sup>-1</sup> )	$\phi_{acm}$ (m s <sup>-1</sup> )	$\phi$ (%)	$\phi_{acm}$ (%)
1	0,16	0,16	5,23	5,23
2	0,14	0,30	4,49	9,72
3	0,13	0,43	4,03	13,75
4	0,12	0,55	3,83	17,59
5	0,12	0,66	3,75	21,33
6	0,11	0,78	3,66	25,00
7	0,11	0,89	3,61	28,60

Nº de componentes utilizados para estimação dos dados ausentes e realização do PCA: 7.

## APENDICE E – Distribuições de Frequência de EToFAO-56

O gráfico de distribuição de frequência de ETo<sub>FAO-56</sub> estimado pela técnica CP (Figura 6) mostra que, obedecendo uma distribuição normal dos dados, a maior parte dos valores de ETo<sub>CP</sub> varia no intervalo entre 3,5 – 5,3 mm dia<sup>-1</sup>. Que se mostrou adequada para a descrição dos modelos propostos.

Distribuição de frequência de ETo<sub>FAO-56</sub> estimada por ETo<sub>ACP</sub>.

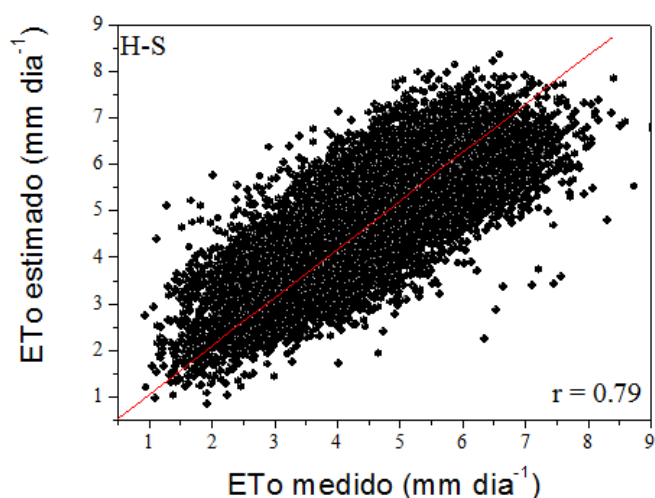


## APÊNDICE F – Matriz de pesos de (MLP)

b1 =	-2.7602	w1 =	1.2038	-0.5119	0.1204	-0.4424	-0.0790	-1.7199
	2.2876		0.4019	-3.4422	-0.0122	1.2667	0.5679	-1.6192
	3.0439		-1.4640	-0.5859	0.2015	-0.5121	-0.2323	1.0173
	-1.1854		0.2767	-0.8432	-0.1078	-0.0611	0.7311	-1.3818
	-1.4113		1.1185	0.2194	0.2053	-0.6913	0.2033	1.1754
	-1.5586		0.8486	-0.0948	-0.0418	-0.5242	-0.0668	-1.9835
	1.3487		-0.6799	0.4292	0.4030	1.1860	1.4413	-0.6650
	-0.1605		0.0200	-0.1828	-0.1914	-0.2464	0.1459	-0.4239
	0.6240		-1.5153	0.4799	-0.1134	0.5508	-0.9727	-0.6346
	-1.2046		0.4035	-0.4143	-0.4593	-1.4959	-1.7072	0.8844
	-1.9137		0.1746	-0.0191	-0.3755	0.2750	-0.1834	-0.5676
	0.6378		-0.4469	-0.0129	-0.5176	0.6360	0.2718	0.7792
	-0.6298		0.0490	0.2421	-0.0270	0.2708	-0.0490	-0.2031
	0.4043		-0.1768	-0.4211	0.1293	-0.1831	1.8372	0.2835
	-0.9232		0.9374	-0.4079	0.0858	-0.2022	0.3522	0.3873
	-0.2394		0.4008	-0.3222	-0.1122	-0.2616	-0.1825	0.3646
	-0.7775		-1.2050	0.8916	-0.1857	0.4973	2.0780	-0.9756
	1.8060		1.2412	-0.0062	0.9019	0.6453	1.0318	0.5556
	1.1272		0.3658	-0.8806	-0.0607	0.7407	0.0873	1.0949
	1.7680		0.2308	-0.8410	-0.1477	0.3294	0.3781	0.3220
1.3523	1.0775	-1.2153	-1.6678	0.1756	-0.7265	0.2189		
1.9208	1.4636	-0.3644	0.3002	0.4221	-1.0259	-1.2091		
-1.0203	-1.0906	0.5919	0.3777	0.5346	-0.6237	2.0047		
2.2353	-0.4237	0.6668	0.4935	0.9762	-1.6655	-0.8391		
1.1471	0.3387	-0.3947	-0.4093	-0.4465	0.5362	-1.5146		

b2 =	-0.78	w2 =	-0.5494	-0.0864	-0.1656	0.1092	-0.4345	-0.1079	-0.2186	-0.7693	0.24
			1.3567	-0.0409	1.5065	0.7072	0.0153	-0.0210	-0.0489	0.5676	-0.00

### APÊNDICE G – Modelo de Hargreaves-Samani (H-S)

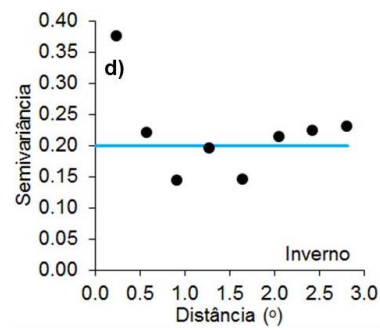
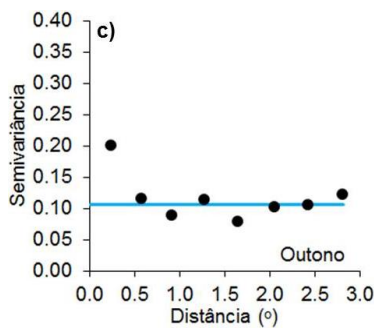
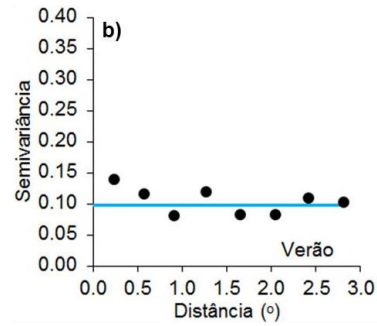
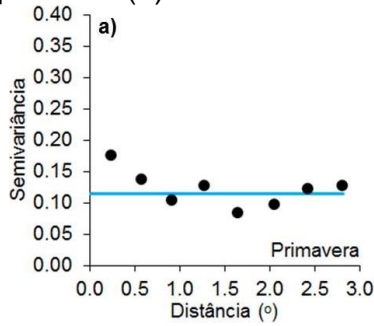


### APÊNDICE H – Performance do treinamento das técnicas de Aprendizado de Máquinas (AM)

Modelos	Arquitetura	Treinamento (70% dos dados)	
		rRMSE (%)	RMSE (mm dia <sup>-1</sup> )
MLP1	Tmax, Tmin, H <sub>0</sub>	16,51	0,73
MLP2	Tmax, Tmin, H <sub>0</sub> , Rn	13,43	0,59
MLP3	Tmax, Tmin, H <sub>0</sub> , Rn, UR	12,29	0,54
MLP4	Tmax, Tmin, H <sub>0</sub> , Rn, UR, U <sub>2</sub>	0,60	0,03
ANFIS1	Tmax, Tmin, H <sub>0</sub>	16,70	0,73
ANFIS2	Tmax, Tmin, H <sub>0</sub> , Rn	13,54	0,60
ANFIS3	Tmax, Tmin, H <sub>0</sub> , Rn, UR	12,50	0,55
ANFIS4	Tmax, Tmin, H <sub>0</sub> , Rn, UR, U <sub>2</sub>	0,75	0,03
GRNN1	Tmax, Tmin, H <sub>0</sub>	2,48	0,11
GRNN2	Tmax, Tmin, H <sub>0</sub> , Rn	0,18	0,01
GRNN3	Tmax, Tmin, H <sub>0</sub> , Rn, UR	0,01	0,00
GRNN4	Tmax, Tmin, H <sub>0</sub> , Rn, UR, U <sub>2</sub>	0,00	0,00
SVM1	Tmax, Tmin, H <sub>0</sub>	61,38	0,77
SVM2	Tmax, Tmin, H <sub>0</sub> , Rn	48,62	0,61
SVM3	Tmax, Tmin, H <sub>0</sub> , Rn, UR	45,29	0,57
SVM4	Tmax, Tmin, H <sub>0</sub> , Rn, UR, U <sub>2</sub>	4,18	0,05

## APÊNDICE I – Variogramas

Variogramas das partições: sazonais [(a) primavera, (b) verão, (c) outono e (d) inverno] e anual (e).



### Primavera

Distância (°)	Semivariância	Pares	Efeito pepita
0,25	0,17	14	0,12
0,58	0,14	43	0,12
0,92	0,10	58	0,12
1,28	0,12	47	0,12
1,66	0,08	43	0,12
2,06	0,10	44	0,12
2,43	0,12	34	0,12
2,82	0,13	36	0,12

### Verão

Distância (°)	Semivariância	Pares	Efeito pepita
0,25	0,14	14	0,10
0,58	0,12	43	0,10
0,92	0,08	58	0,10
1,28	0,12	47	0,10
1,66	0,08	43	0,10
2,06	0,08	44	0,10
2,43	0,11	34	0,10
2,82	0,10	36	0,10

**Outono**

<b>Distância (°)</b>	<b>Semivariância</b>	<b>Pares</b>	<b>Efeito pepita</b>
0,25	0,20	14	0,11
0,58	0,11	43	0,11
0,92	0,09	58	0,11
1,28	0,11	47	0,11
1,66	0,08	43	0,11
2,06	0,10	44	0,11
2,43	0,10	34	0,11
2,82	0,12	36	0,11

**Inverno**

<b>Distância (°)</b>	<b>Semivariância</b>	<b>Pares</b>	<b>Efeito pepita</b>
0,25	0,38	14	0,20
0,58	0,22	43	0,20
0,92	0,14	58	0,20
1,28	0,19	47	0,20
1,66	0,15	43	0,20
2,06	0,21	44	0,20
2,43	0,22	34	0,20
2,82	0,23	36	0,20