

## ZONAS HOMOGÊNEAS DE EVAPOTRANSPIRAÇÃO DE REFERÊNCIA PARA O ESTADO DE MINAS GERAIS

Franciele Alves Barbosa<sup>1</sup>, Patrícia de Oliveira e Lucas<sup>2</sup>, Marcelo Rossi Vicente<sup>3</sup>, Ronaldo Medeiros dos Santos<sup>4</sup>

**RESUMO:** A determinação da quantidade de água necessária para as culturas é um dos principais fatores para o correto planejamento, dimensionamento e manejo de qualquer sistema de irrigação. Uma forma de estimar a quantidade de água requerida por uma cultura é através da evapotranspiração de referência ETo. Porém, quando dados climáticos em um local são insuficientes para estimar o valor da ETo, uma estimativa regional pode ser realizada a partir de valores médios de zonas homogêneas, sendo a análise de agrupamento uma forma de obter esses valores médios. Portanto, esse trabalho teve como objetivo a criação de zonas homogêneas de ETo através do agrupamento de séries temporais de ETo para o estado de Minas Gerais. Para isso, foram realizados experimentos para a escolha do algoritmo e do número de *clusters* que melhor se ajustasse ao problema. Para a avaliação do número de *clusters*, utilizou-se uma segunda base de dados normais climatológicas com 62 estações. A partir dos experimentos, foi possível concluir que o algoritmo *K-means* e o modelo com 10 *clusters*, mostraram ser uma melhor opção para o agrupamento de zonas homogêneas para todo o estado de Minas Gerais.

**PALAVRAS-CHAVE:** K-means, *Ward*, *clusters*, irrigação.

## REFERENCE EVAPOTRANSPIRATION ZONES FOR THE STATE OF MINAS GERAIS

**ABSTRACT:** Determining the amount of water needed for crops is one of the main factors for the correct planning, dimensioning and management of any irrigation system. One way to

<sup>1</sup> Mestranda em Modelagem Computacional e Sistemas, UNIMONTES, Montes Claros, MG

<sup>2</sup> Prof. Mestre, IFNMG – campus Salinas, Salinas, MG

<sup>3</sup> Prof. Doutor, IFNMG – campus Salinas, CEP 39560-000, Salinas, MG. Fone: (38) 3841-7000. e-mail: marcelo.vicente@ifnmg.edu.br

<sup>4</sup> Prof. Doutor, IFNMG – campus Salinas, Salinas, MG

estimate the amount of water required by a crop is through the ETo reference evapotranspiration. However, when climatic data in a location are insufficient to estimate the ETo value, a regional estimate can be made from average values of homogeneous zones, with cluster analysis being a way to obtain these average values. Therefore, this work aimed to create homogeneous ETo zones by grouping ETo time series for the state of Minas Gerais. For this, experiments were carried out to choose the algorithm and the number of *clusters* that best fit the problem. To assess the number of *clusters*, a second normal climatological database with 62 stations was used. From the experiments, it was possible to conclude that the *K-means* algorithm and the model with 10 *clusters*, proved to be a better option for grouping homogeneous zones for the entire state of Minas Gerais.

**KEYWORDS:** K-means, *Ward*, *clusters*, irrigation.

## INTRODUÇÃO

Devido aos baixos índices de precipitação, temperaturas elevadas durante todo o ano e baixa capacidade de armazenamento de água no solo, algumas regiões do Brasil sofrem com o déficit hídrico e conseqüentemente não conseguem suprir a necessidade de água das culturas, recorrendo assim à irrigação.

A determinação da quantidade de água necessária para as culturas é um dos principais fatores para o correto planejamento, dimensionamento e manejo de qualquer sistema de irrigação. Uma forma de estimar a quantidade de água requerida por uma cultura é através da evapotranspiração de referência (ETo) e do coeficiente da cultura (Kc) (Carvalho et al., 2017).

Embora a ETo possa ser estimada a partir de dados climáticos, quando os dados em um determinado local são insuficientes para uma estimativa confiável, uma estimativa regional deve ser realizada para obter ETo em um local não medido. Nesse processo de estimativa regional, os locais devem ser atribuídos às regiões homogêneas (Chen, et al., 2013).

Como apresentado em Bork (2018), uma forma de dividir uma área de estudo em regiões homogêneas é através da análise de agrupamentos de séries temporais de ETo. Essa análise pode ser dividida em agrupamento hierárquico e não hierárquico, sendo sua principal finalidade agrupar objetos de acordo com suas características e com base em uma medida de similaridade. O método hierárquico de *Ward* e o método não hierárquico *K-means* são métodos de análise de agrupamento simples e altamente utilizados em estudos climatológicos com resultados satisfatórios (Araújo et al., 2015).

Nesse contexto, esse trabalho tem como principal objetivo a criação de zonas homogêneas de ETo através do agrupamento de séries temporais de ETo para o estado de Minas Gerais.

## MATERIAL E MÉTODOS

Para a execução do trabalho, utilizaram-se duas bases de dados climáticos. A primeira com propósito de gerar os modelos de agrupamento e a segunda para a avaliação do melhor modelo de agrupamento.

Para gerar os modelos de agrupamento, utilizou-se um banco de dados de evapotranspiração de referência (ETo) disponibilizados por Xavier et al. (2015) para o estado de Minas Gerais, obtido pelo método de interpolação pelo inverso da distância ponderada para 5285 pontos, em grade (0,5° x 0,5°). O período analisado foi de 1980 – 2016, onde optou-se por trabalhar com as médias mensais de ETo do período, ou seja, com 12 valores para cada uma das 5285 séries temporais.

Já a segunda base de dados climáticos foi utilizada para validação do número de *clusters* das zonas homogêneas de evapotranspiração. Utilizou-se a base de dados das Normais Climatológicas Brasileira (INMET, 2018) de 62 estações do período de 1981 a 2010 (Figura 1). De posse das variáveis meteorológicas (temperaturas máxima e mínimas, umidade relativa, velocidade do vento e insolação) calculou-se a ETo mensal pelo método FAO Penman–Monteith (Allen et al., 2006).

Para a escolha do número de *clusters* e do algoritmo que melhor se ajuste ao problema, foram realizados 2 experimentos. No Experimento 1, a métrica da silhueta, proposta por Rousseeuw (1987), foi utilizada para a escolha do algoritmo e para a definição da quantidade de *clusters*, como sugerido por Faceli et al. (2011). Foram testados o intervalo de 3 a 15 *clusters* para o algoritmo *K-means* e para o algoritmo de *Ward*, escolhendo-se aquele com o maior valor de silhueta. Por se tratar de um problema não supervisionado, não há um elemento externo para guiar o aprendizado, ou seja, os dados não são rotulados. Além disso, de acordo com Araújo et al. (2015), para definir regiões homogêneas é indicado a utilização de vários métodos de análise de agrupamento, optando-se por aqueles cujos resultados sejam mais coerentes com o fenômeno em estudo.

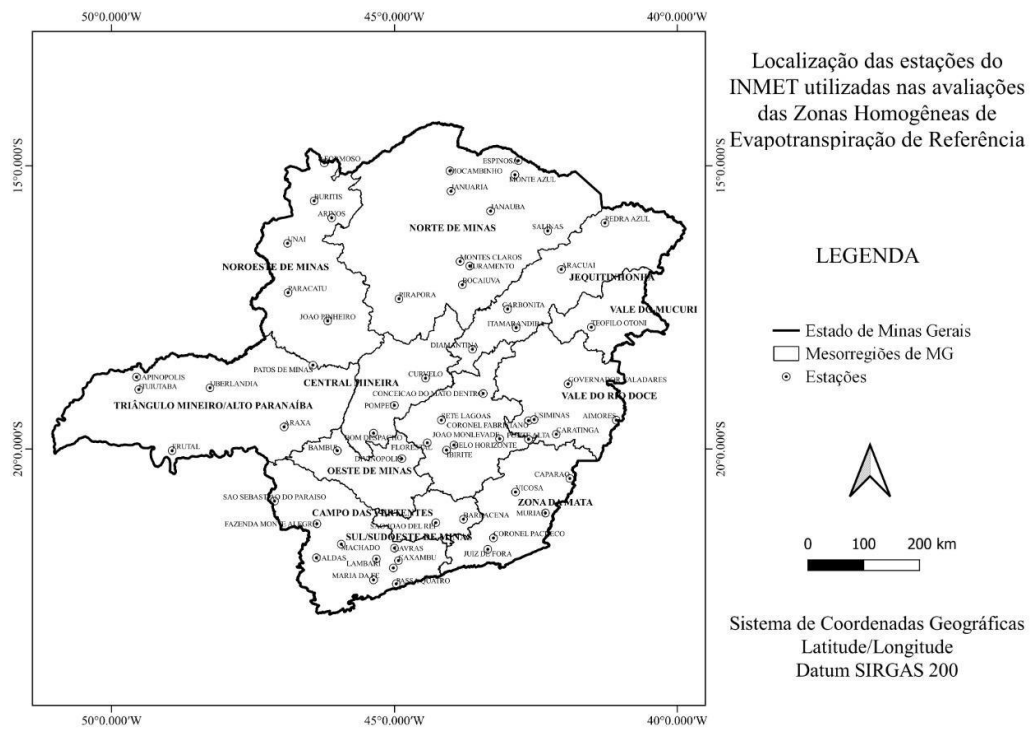


Figura 1. Mapa com a localização das 62 estações para avaliação.

Dessa forma, realizou-se o Experimento 2 para certificar se houve uma escolha correta do melhor número de *clusters* no Experimento 1. Para isso, escolheu-se os 5 melhores modelos do algoritmo que obteve melhor silhueta do Experimento 1 e empregou-se a base de dados de validação para verificar qual desses agrupamentos (número de *clusters*) apresentam melhor estimativa de ETo usando o centroide dos *clusters* como rótulo. Como métrica, utilizou-se a raiz quadrática média (RMSE) (Equação 1).

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

Em que,

RMSE - raiz quadrática média *intra-cluster*.

$y_i$  - ETo das estações usadas para a validação, em mm dia<sup>-1</sup>;

$(y_i)^{\wedge}$  - ETo estimada pelo agrupamento do algoritmo, em mm dia<sup>-1</sup>.

Após a definição do número de zonas homogêneas, foram utilizados diferentes indicadores estatísticos que correlacionaram os valores mensais de ETo (mm dia<sup>-1</sup>) observados pelas 62 estações meteorológicas convencionais, localizadas dentro da área de interesse, com os valores médios mensais de ETo (mm dia<sup>-1</sup>).

Os indicadores utilizados para a validação das zonas homogêneas de ETo foram: o coeficiente de correlação de Pearson (r); o coeficiente de determinação ( $R^2$ ); o índice de concordância (d) de Willmott (1981) e o índice de confiança (c) proposto por Camargo & Sentelhas (1997), além do RMSE.

O coeficiente de determinação ( $R^2$ ) é utilizado para avaliar a associação entre dois fatores, através da proporção da variabilidade em uma variável, sendo expresso pela Equação 2.

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

Em que,

$\bar{Y}$  - a média da ETo das estações usadas para a validação, em mm dia<sup>-1</sup>;

O Coeficiente de correlação de Pearson (r) ( $r = \sqrt{R^2}$ ) foram classificados segundo a metodologia de Hopkins (2009) (Tabela 1).

**Tabela 1.** Coeficiente de correlação de Pearson (r) e a sua respectiva classificação (Hopkins, 2009)

Coeficiente de correlação de Pearson (r)	Classificação
0,0 – 0,1	Muito Alta
0,1 – 0,3	Baixa
0,3 – 0,5	Moderada
0,5 – 0,7	Alta
0,7 – 0,9	Muito Alta
0,9 – 1,0	Quase Perfeita

O índice de concordância (d) varia entre zero e um, em que  $d = 1$  para a concordância perfeita,  $d = 0$  para nenhuma concordância. O valor d é determinado pela Equação:

$$d = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (|y_i - \bar{y}| + |\hat{y}_i - \bar{y}|)^2} \quad (3)$$

O Índice de Confiança (c) foi proposto por Camargo & Sentelhas (1997), pois permite analisar conjuntamente a precisão e a exatidão dos resultados obtidos, através do produto do coeficiente de correlação (r) pelo índice de concordância (d) (Castro et al., 2010). Os critérios de avaliação do desempenho de modelos quanto ao índice de confiança, estão apresentados na Tabela 2.

**Tabela 2.** Análise do desempenho do modelo com base no índice de confiança.

Valor de c	Desempenho
> 0,85	Ótimo
0,76 a 0,85	Muito bom
0,66 a 0,75	Bom
0,61 a 0,65	Mediano
0,51 a 0,60	Sufrível
0,41 a 0,50	Mau
≤ 0,4	Péssimo

## RESULTADOS E DISCUSSÃO

No Experimento 1, o algoritmo *K-means* apresentou, na maioria das execuções (3 a 15 *clusters*), valores de silhueta mais próximos de 1, quando comparado ao algoritmo *Ward*, indicando que as amostras estão melhores agrupadas.

Após a escolha do algoritmo K-means, realizou-se o Experimento 2. Os modelos com 10 e 11 *clusters* apresentaram melhores resultados, com valor médio do RMSE de 0,26 mm d<sup>-1</sup>. Os resultados encontrados estão similares aos encontrados por Junior et al. (2017) e por Santos et al. (2020). Apesar do modelo de 10 e 11 *clusters* possuírem a mesma média de RMSE, escolheu-se, por parcimônia, como melhor agrupamento o modelo com 10 *clusters*. A partir da escolha do modelo de 10 zonas, foram geradas as séries temporais médias mensais e anual de cada zona homogênea, conforme apresentadas na Tabela 3.

**Tabela 3.** Média de evapotranspiração de referência mensal e anual (mmd-1) das zonas homogêneas.

Zona	JAN	FEV	MAR	ABR	MAI	JUN	JUL	AGO	SET	OUT	NOV	DEZ	MÉDIA
1	4,23	4,50	4,01	3,77	3,25	3,01	3,34	4,13	4,74	4,82	4,23	4,03	4,01
2	4,16	4,38	3,74	3,20	2,54	2,25	2,48	3,14	3,65	4,04	3,96	3,96	3,46
3	4,74	5,05	4,52	4,23	3,74	3,50	3,80	4,64	5,35	5,49	4,66	4,47	4,52
4	4,67	4,86	4,25	3,61	2,96	2,63	2,84	3,51	4,22	4,63	4,30	4,39	3,91
5	4,55	4,84	4,30	3,95	3,40	3,14	3,43	4,24	4,95	5,13	4,45	4,28	4,22
6	4,15	4,38	3,78	3,38	2,75	2,47	2,73	3,45	3,99	4,33	4,11	3,98	3,62
7	3,93	4,09	3,49	2,98	2,28	2,00	2,23	2,87	3,32	3,76	3,84	3,82	3,22
8	4,25	4,35	3,96	3,68	3,03	2,82	3,19	4,07	4,56	4,80	4,64	4,36	3,98
9	4,60	4,78	4,14	3,43	2,75	2,41	2,58	3,19	3,72	4,15	4,11	4,27	3,68
10	4,16	4,40	3,89	3,58	3,00	2,75	3,07	3,87	4,43	4,64	4,26	4,05	3,84

Na Tabela 4 observa-se os índices de validação das zonas homogêneas. Analisando conjuntamente a precisão e a exatidão dos resultados obtidos, através do índice de confiança (c), o desempenho foi considerado Muito Bom e Ótimo, segundo critério de Camargo & Sentelhas (1997). Os resultados das estimativas de ETo obtidos pelo algoritmo K-means, baseando-se nos indicadores expostos na Tabela 4, podem ser considerados satisfatórios.

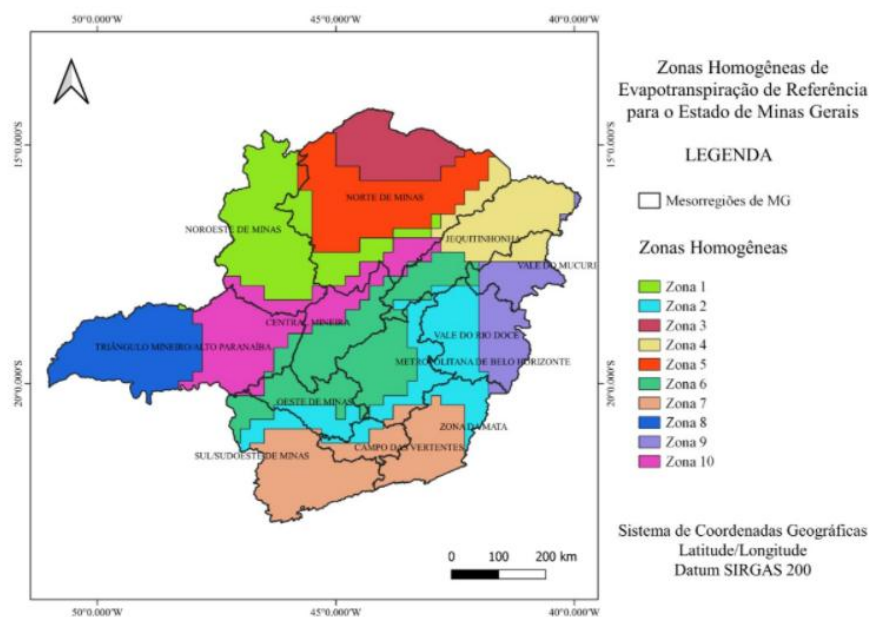
**Tabela 4.** Indicadores utilizados para a validação das zonas homogêneas de ETo.

Zona	ETo NORMAIS (mm d <sup>-1</sup> )	ETo ZONA (mm d <sup>-1</sup> )	RMSE (mm d <sup>-1</sup> )	R <sup>2</sup>	d	r	c	Classificação
1	4,21	4,01	0,31	0,84	0,93	0,916	0,85	Muito bom
2	3,49	3,46	0,22	0,90	0,97	0,951	0,93	Ótimo
3	4,58	4,52	0,31	0,77	0,93	0,875	0,81	Muito bom
4	3,89	3,91	0,21	0,93	0,98	0,965	0,94	Ótimo
5	4,19	4,22	0,23	0,86	0,96	0,926	0,89	Ótimo
6	3,76	3,63	0,36	0,77	0,92	0,878	0,81	Muito bom
7	3,48	3,22	0,37	0,88	0,94	0,936	0,88	Ótimo
8	4,14	3,98	0,23	0,94	0,97	0,968	0,94	Ótimo
9	3,69	3,68	0,18	0,95	0,99	0,975	0,96	Ótimo
10	3,69	3,84	0,19	0,96	0,97	0,981	0,95	Ótimo

O mapa das 10 zonas homogêneas encontradas pelo agrupamento é apresentado na Figura 2. Observar-se que, a mesorregião do Norte de Minas possui majoritariamente a Zona 3 e a Zona 5 e ambas possuem a maior média de ETo.

Já a mesorregião Noroeste de Minas possui majoritariamente a Zona 1. A mesorregião Jequitinhonha possui majoritariamente as Zonas 3 e 4. A mesorregião Vale do Mucuri possui majoritariamente a Zona 9. A mesorregião do Vale do Rio Doce possui majoritariamente as Zonas 2 e 9.

A mesorregião Central Mineira possui majoritariamente a Zona 6. A mesorregião Triângulo Mineiro/Alto Paranaíba possui majoritariamente as Zonas 8 e 10. As mesorregiões Metropolitana de Belo Horizonte e Oeste de Minas possuem majoritariamente as Zonas 6 e 2. E as mesorregiões Zona da Mata, Sul/Sudoeste de Minas e Campo das Vertentes possuem majoritariamente a Zona 7, que tem uma menor média de ETo.

**Figura 2.** Zonas Homogêneas de ETo para o estado de Minas Gerais.

## CONCLUSÕES

Concluiu-se que para o agrupamento de séries temporais de evapotranspiração de referência, o *K-means* demonstrou ser uma melhor opção com o agrupamento de 10 zonas homogêneas para todo o estado de Minas Gerais.

## REFERÊNCIAS BIBLIOGRÁFICAS

ALLEN, R. G.; PEREIRA, L. S.; RAES, D.; SMITH, M. Evapotranspiración del cultivo: Guías para la determinación de los requerimientos de agua de los cultivos. **Estudio FAO Riego y Drenaje**, n. 56. Roma: FAO, 2006. URL <https://openknowledge.fao.org/server/api/core/bitstreams/f3660258-d07f-487e-b3c1-01661c83cb16/content>.

ARAÚJO, J. M. S.; MAIA, A. G.; NEVES, J. A. Uso dos índices PCP e PCD na determinação de regiões com precipitação pluvial homogênea (The use of PCP and PCD indices to determining precipitation homogeneous areas). **Revista Brasileira de Climatologia**, 17, Dec. 2015. doi: 10.5380/abclima.v17i0.38903. URL <https://doi.org/10.5380/abclima.v17i0.38903>.

BORK, C. K. Regionalização de vazões mínimas para o estado do rio grande do sul. Master's thesis, Universidade Federal de Pelotas, 2018.

CAMARGO, A. P.; SENTELHAS, P. C. Avaliação do desempenho de diferentes métodos de estimativa da evapotranspiração potencial no estado de São Paulo, Brasil. **Revista Brasileira de Agrometeorologia**, v. 5, n. 1, p. 89-97, 1997.

CARVALHO, L. G.; RIOS, G. F. A.; MIRANDA, W. L.; NETO, P. C. Evapotranspiração de referência: uma abordagem atual de diferentes métodos de estimativa. **Pesquisa Agropecuária Tropical**, 41(3), July 2011. doi: 10.5216/pat.v41i3.12760. URL <https://doi.org/10.5216/pat.v41i3.12760>.

CASTRO, F. da S.; PEZZOPANE, J. E. M.; CECÍLIO, R. A.; PEZZOPANE, J. R. M.; XAVIER, A. C. Avaliação do desempenho dos diferentes métodos de interpoladores para parâmetros do balanço hídrico climatológico. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 14, n. 8, p. 871–880, 2010. doi: 10.1590/S1415-43662010000800012. URL <https://doi.org/10.1590/S1415-43662010000800012>.

CHEN, C. T.; CHANG, Y. C.; WU, G. T. Determination of homogeneous regions for regional reference evapotranspiration estimation using the self-organizing map in western taiwan. **Paddy and Water Environment**, 12(1):173–179, 2013. doi: 10.1007/s10333-013-0374-2. URL <https://doi.org/10.1007/s10333-013-0374-2>.

FACELI, K.; LORENA, A. C.; GAMA, J.; CARVALHO, .; A. C. P. D. L. F. **Inteligência artificial: uma abordagem de aprendizado de máquina**. LTC, 2011.

HOPKINS, W. G.; MARSHALL, S. W.; BATTERHAM, A. M.; HANIN, J. Progressive statistics for studies in sports medicine and exercise science. **Medicine & Science in Sports & Exercise**, v. 41, n. 1, p. 3–13, 2009. doi: 10.1249/MSS.0b013e31818cb278. URL <https://doi.org/10.1249/MSS.0b013e31818cb278>.

INMET - Instituto Nacional de Meteorologia. **Normais Climatológicas do Brasil**, 1981-2010. Brasília: INMET, 2018. URL <http://www.inmet.gov.br/portal/index.php?r=clima/normaisClimatologicas>.

JUNIOR, F. A. R.; VIEIRA, L. B.; QUEIROZ, D. M.; SANTOS, N. T. Geração de zonas de manejo para cafeicultura empregando-se sensor SPAD e análise foliar. **Revista Brasileira de Engenharia Agrícola e Ambiental**, 15(8):778–787, Aug. 2011. doi: 10.1590/s1415-43662011000800003. URL <https://doi.org/10.1590/s1415-43662011000800003>.

OLIVEIRA, G. M.; SANTIAGO, E. J. P.; LEITÃO, M. V. B. R. ROCHA, R. C.; GONÇALVES, I. S.; SILVA, R. R. Estimativa da evapotranspiração de referência para ambiente protegido. **Irriga**, 1(1):21–30, Oct. 2017. doi: 10.15809/irriga.2017v1n1p21-30. URL <https://doi.org/10.15809/irriga.2017v1n1p21-30>.

ROUSSEEUW, P. J.; Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. **Journal of Computational and Applied Mathematics**, 20:53–65, Nov. 1987. doi: 10.1016/0377-0427(87)90125-7. URL [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).

SANTOS, T. A.; VICENTE, M. R.; LEITE, C. V.; SANTOS, R. M.; SOUZA, J. L. T. Zonas homogêneas de evapotranspiração de referência para o norte e noroeste de Minas Gerais. **Revista Brasileira de Agricultura Irrigada**, 13(4): 3540–3555, Mar. 2020. doi: 10.7127/rbai.v13n4001078. URL <https://doi.org/10.7127/rbai.v13n4001078>.

WILLMOTT, C. J. On the validation of models. **Physical Geography**, v. 2, p. 184–194, 1981.

XAVIER, A. C.; KING, C. W.; SCANLON, B. R. Daily gridded meteorological variables in brazil (1980-2013). **International Journal of Climatology**, 36(6):2644–2659, Oct. 2015. doi: 10.1002/joc.4518. URL <https://doi.org/10.1002/joc.4518>.