ISSN 1807-1929
Revista Brasileira de Engenharia Agrícola e Ambiental
Brazilian Journal of Agricultural and Environmental Engineering
v.27, n.2, p.149-156, 2023
Campina Grande, PB – http://www.agriambi.com.br – http://www.scielo.br/rbeaa

DOI: http://dx.doi.org/10.1590/1807-1929/agriambi.v27n2p149-156

Gap-filling meteorological data series using the GapMET software in the state of Mato Grosso, Brazil[1]

Preenchimento de falhas em dados meteorológicos usando o programa GapMET no Estado de Mato Grosso, Brasil

Marlus Sabino[2] & Adilson P. de Souza[2*]

[1] Research developed at Universidade Federal de Mato Grosso, Setor de Física Ambiental, Cuiabá, MT, Brazil
[2] Universidade Federal de Mato Grosso, Programa de Pós-Graduação em Física Ambiental , Cuiabá, MT, Brazil

HIGHLIGHTS:

GapMET provides accurate gap-filling methods for meteorological data series.
Simple linear regression gap-filling had more accuracy for Mato Grosso's automatic weather station network density.
The use of satellite data as reference series reduces chances of fails to gap-fill, but also reduces the filling accuracy.

ABSTRACT: This paper aimed to introduce the GapMET software, developed by the authors, and evaluate the accuracy of its six methods for gap-filling the main meteorological variables monitored by weather station in the state of Mato Grosso, Brazil, using reference time series from neighbour weather station and/or remote sensing products. The methods were tested on seven different databases, with 25 to 80% artificial gaps, and their accuracy was given by the number of gaps left unfilled, the bias, the RMSE, and Pearson's correlation. The GapMET software showed good results in filling meteorological gaps regardless of the method applied. Methods that use only one neighbour weather station as a reference series showed better results because, in the state, the minimum distance for a weather station to have at least three neighbours as reference was 350 km, reducing the climatic similarity between them and consequently the accuracy when more than one reference series were needed. The use of satellite reference series reduced the probability of unfilled gaps; however, it showed higher bias and RMSE and lower correlations.

Key words: time series, missing data, automated weather stations, ERA5-Land

RESUMO: Este artigo objetiva introduzir o programa GapMET, desenvolvido pelos autores, e avaliar a precisão de seus seis métodos de preenchimento de dados nas principais variáveis meteorológicas monitoradas em estações do Estado de Mato Grosso, utilizando como séries de referência dados de estações vizinhas e/ou dos produtos de sensoriamento remoto. A precisão das estimativas foi aferida pela quantidade de falhas não preenchidas, viés, RMSE e correlação de Pearson, em agrupamentos de sete bases de dados com variações de 25 a 80% de falhas. O GapMET, independentemente do método, apresentou bons resultados no preenchimento de falhas meteorológicas. Métodos que utilizam apenas uma estação meteorológica vizinha como série referência apresentaram melhores resultados, visto que, no Estado, a distância mínima para uma estação meteorológica ter ao menos três estações vizinhas como referência foi de 350 km, reduzindo, assim, a semelhança climática entre elas e, consequentemente, a precisão do preenchimento quando mais de uma série de referência era necessária. O uso de séries obtidas por satélite reduziu a probabilidade de falhas não serem preenchidas, contudo, apresentou maiores erros de viés, RMSE e menores correlações.

Palavras-chave: séries temporais, dados ausentes, estações meteorológicas automáticas, ERA5-Land

• Ref. 265367 – Received 27 Jun, 2022
* Corresponding author - E-mail: pachecoufmt@gmail.com
• Accepted 28 Sept, 2022 • Published 10 Oct, 2022
Editors: Ítalo Herbet Lucena Cavalcante & Carlos Alberto Vieira de Azevedo

Introduction

Meteorological time series are essential to designing agrometeorological and environmental projects (Xiang et al., 2020; Sarnighausen et al., 2021). However, in Brazil, acquisition and storage of automated weather station (AWS) data are susceptible to instrumental and operational failures (Bier & Ferraz, 2017; Brubacher et al., 2020), affecting their use to assess climatic trends, stationarity, or seasonality (Box & Cox, 1964).

In the literature, there is a range of methods for gap-filling, e.g., the simple arithmetic means (MAS) (Bier & Ferraz, 2017), UK traditional method (MUK) (Tabony, 1983), inverse distance method (MID) (Hubbard, 1994), regional weighting method (MPR) (Paulhus & Kohler, 1952), simple linear regression (RLS) (Bleidorn et al., 2022), and multiple linear regression (RLM) (Coutinho et al., 2018), whose basic requirement is a reference time series, that can be extracted from nearby AWS or remote sensing datasets.

Although statistical analyses are used to calibrate and validate these methods, there is no consensus of which is the best methodology or reference dataset, since it depends on the desired accuracy and on the number and distance of nearby AWS. The state of Mato Grosso, despite its dimension and different climate types: Af (tropical rainforest climate), Aw (tropical savanna climate), and Cwa (tropical climate) (Souza et al., 2013; Jerszurki et al., 2019), has only 33 AWS of the National Institute of Meteorology (INMET) that meet the world meteorological organization (WMO) minimum period requirement (10 years) to estimate Provisional Climatological Normals (WMO, 1998).

Therefore, this paper aims to introduce the GapMET, a software program for gap-filling meteorological data series developed by the authors, and evaluate the accuracy of its six methods (MAS, MUK, MID, MPR, RLS, and RLM) for gap-filling the main variables monitored by AWS in the state of Mato Grosso, using reference time series from neighbour AWS and remote sensing.

Material and Methods

The time series, with data from 01/01/2008 to 12/31/2020, were obtained from automatic weather stations (AWS) belonging to the National Institute of Meteorology (INMET), located in 33 municipalities in Mato Grosso state, Brazil (Table 1). In each AWS, datasets of the meteorological variables were collected: downward solar radiation (SRD - MJ m^{-2} per day), relative air humidity (RH - %), maximum air temperature (Tmax - °C), minimum air temperature (Tmin - °C), and mean wind speed at 2 m (WS - m s^{-1}).

The gap-filling was done by GapMET 1.0 program, developed by the authors in MATLAB and 'R' language, and

Table 1. INMET automated weather stations (AWS) in state of Mato Grosso, Brazil

AWS code	AWS name	Lat.	Lon.	Alt.	Percentage of original gaps (%)				
					SRD	RH	Tmax	Tmin	WS
A-901	Cuiabá	-15.56	-56.06	242	41.27	14.02	15.25	16.59	26.38
A-902	Tangará da Serra	-14.65	-57.43	440	21.79	18.87	19.14	24.28	28.24
A-903	São José do Rio Claro	-13.45	-56.68	340	59.57	22.53	14.95	20.38	31.65
A-904	Sorriso	-12.56	-55.72	379	16.93	14.05	15.12	17.44	31.80
A-905	Campo Novo do Parecis	-13.79	-57.84	525	17.54	29.02	16.40	22.93	24.93
A-906	Guarantã do Norte	-9.95	-54.90	284	10.02	10.25	10.51	18.34	12.23
A-907	Rondonópolis	-16.46	-54.58	290	8.61	9.18	9.35	11.12	6.76
A-908	Água Boa	-14.02	-52.21	440	15.16	10.32	9.31	15.83	12.00
A-910	Apiacás	-9.56	-57.39	218	20.93	10.44	17.73	23.54	25.56
A-912	Campo Verde	-15.53	-55.14	748	11.69	13.98	12.80	15.14	13.50
A-913	Comodoro	-13.71	-59.76	577	16.32	15.46	11.18	14.23	13.43
A-914	Juara	-11.28	-57.53	263	17.10	14.63	15.03	23.96	26.17
A-915	Paranatinga	-14.42	-54.04	477	35.88	37.65	35.35	41.97	43.12
A-916	Querência	-12.63	-52.22	361	19.88	18.74	12.23	22.76	17.22
A-917	Sinop	-11.98	-55.57	367	25.67	18.57	16.36	22.68	17.20
A-918	Confresa	-10.64	-51.57	233	38.11	33.52	0.00	0.00	37.36
A-919	Cotriguaçu	-9.91	-58.57	265	28.28	16.19	16.99	25.58	15.69
A-920	Juína	-11.38	-58.77	365	8.78	10.49	8.53	0.00	12.30
A-921	São Felix do Araguaia	-11.62	-50.73	201	62.06	42.20	0.00	0.00	30.64
A-922	Vila Bela da Santíssima Trindade	-15.06	-59.87	213	11.41	13.52	14.09	16.55	37.31
A-924	Alta Floresta	-10.08	-56.18	292	44.49	16.30	17.27	26.07	32.11
A-926	Carlinda	-9.97	-55.83	294	10.63	6.93	5.54	8.17	11.22
A-927	Brasnorte (Novo Mundo)	-12.52	-58.23	426	31.84	23.08	22.87	26.26	33.29
A-928	Nova Maringá	-13.04	-57.09	334	8.00	7.73	9.43	18.30	20.13
A-929	Nova Ubiratã	-13.41	-54.75	466	52.41	15.12	15.96	19.77	17.33
A-930	Gaúcha do Norte	-13.18	-53.26	376	17.69	14.76	15.86	19.75	21.08
A-931	Santo Antônio do Leste	-14.93	-53.88	664	10.78	10.25	8.99	13.71	11.64
A-932	Guiratinga	-16.34	-53.77	525	14.80	5.12	6.44	9.27	31.31
A-933	Itiquira	-17.17	-54.50	593	21.94	21.58	21.50	25.18	25.52
A-934	Alto Taquari	-17.84	-53.29	862	6.28	6.00	7.20	9.50	24.36
A-935	Porto Estrela	-15.32	-57.23	148	24.41	22.74	24.89	27.52	28.85
A-936	Salto do Céu	-15.12	-58.13	301	11.50	12.68	12.95	19.48	11.31
A-937	Pontes de Lacerda	-15.23	-59.35	273	7.64	6.15	8.55	13.33	10.07

* Lat. – Latitude (decimal degrees); Log. – Longitude (decimal degrees); Alt. – Altitude (m); SRD - Downward solar radiation (MJ m^{-2} per day); RH - Relative air humidity (%); Tmax - Maximum air temperature (°C); Tmin - Minimum air temperature (°C); WS - Mean wind speed at 2 m (m s^{-1})

available at (https://github.com/Marlus-Sabino/GapMET). GapMET allows one to choose among six methods of gap-filling that are commonly used in Brazil, as well as setting the reference time series between neighbour AWS or external datasets.

The GapMET flowchart is shown in Figure 1. In the program, the user needs to input five parameters: 1) Type of reference series (neighbour AWS or external datasets); 2) Gap-filling method (simple arithmetic means - MAS, UK traditional method - MUK, inverse distance method - MID, regional weighting method - MPR, simple linear regression - RLS, or multiple linear regression - RLM); 3) Maximum standard deviations (limits the filling values between "D" standard deviations of the mean observed data); 4) Maximum distance between AWS (a limit, in km, for an AWS to be included as reference time series); and 5) Minimum reference AWS (defines the minimum amount of reference AWS required for the method). In addition, the user must provide the original dataset to be gap-filled, along with the geographic coordinates of the AWS (in methods that use neighbour AWS as reference) or the external reference time-series datasets.

In the selection of neighbour reference AWS, a less than 350 km distance was established as inclusion criteria. This distance was chosen because it was the minimum distance at which there were always at least three neighbours as reference time series. In the methods in which gap-filling is performed using only 1 AWS as a reference (RLS and MUK), within the AWS that passes the distance criteria, the one with the highest absolute value of Pearson's correlation coefficient was selected. Finally, since the wind speed measurements did not show a Gaussian distribution, a maximum of five standard deviations (D = 5) was adopted.

In the methods that used remote sensing datasets as reference time series, data from the meteorological reanalysis products from the land component of the fifth generation of European ReAnalysis (ERA5-Land) were used. These products, according to Muñoz-Sabater et al. (2021), are generated by the assimilation of data from surface and atmospheric observations from various sources into models that are physically and dynamically consistent. ERA5-Land products are freely available by Muñoz-Sabater (2019) in partnership with the independent intergovernmental organization Copernicus Climate Change Service (C3S/ECMWF) and are developed in hourly or monthly temporal resolution and 9 x 9 km spatial resolution.

The six methods of gap-filling tested were separated into two groups: 1) methods that use only neighbour AWS for reference time series, denoted with a (v) after the method initials; 2) methods that use as reference time series both neighbour AWS (v) and external datasets of the ERA5-Land product, denoted with an (s) after the method initials.

The methods belonging to group 1 were: i) simple arithmetic mean (MAS) (Eq. 1), in which the missing data are obtained by the mean of the values observed in the neighbour AWS; ii) inverse distance method (MID) (Eq. 2), whose gap-filling is performed by the mean values of the neighbour AWS weighted by their distance from the gap-filled AWS; iii) regional weighting method (MPR) (Eq. 3), which uses mean values of the neighbour AWS weighted by their historical time series mean; iv) multiple linear regression (MLR) (Eq. 4), in which the estimate is obtained by the linear and angular coefficients of a linear regression between the gap-filled AWS and the neighbour reference AWS.

The methods belonging to group 2 were: i) the traditional method of the United Kingdom (MUK) (Eq. 5), in which the estimate is made assuming a constant difference value between the historical mean on the reference series and the gap-filled AWS; ii) the simple linear regression (RLS) method (Eq. 6),

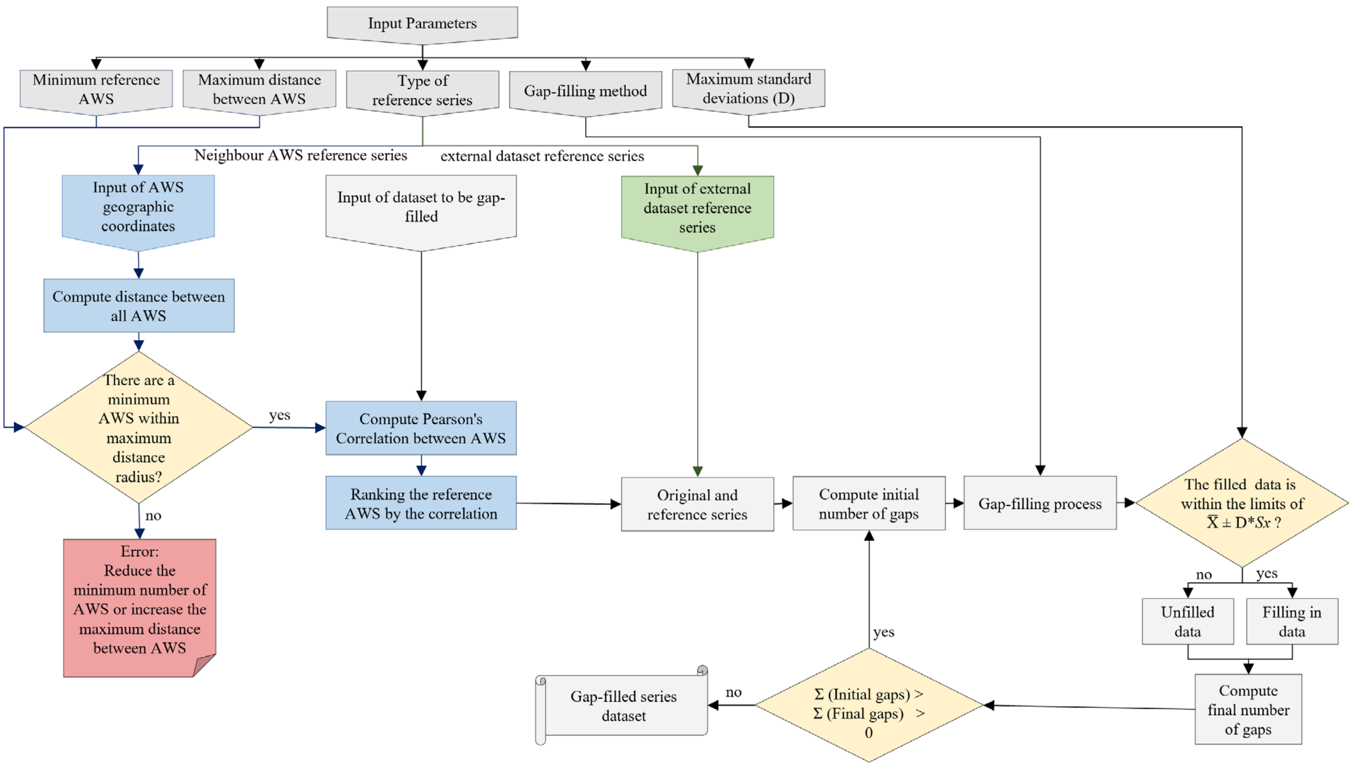


Figure 1. Flowchart of the GapMET gap-filling meteorological data procedure

which uses the coefficients obtained from the linear regression between the gap-filled AWS and the neighbour reference AWS.

$$X_i = \frac{1}{n}\sum_{r=1}^{n} X_r \tag{1}$$

$$X_i = \frac{\sum_{r=1}^{n}\left(\frac{X_r}{d_r}\right)}{\sum_{r=1}^{n}\left(\frac{1}{d_r}\right)} \tag{2}$$

$$X_i = \frac{1}{n}\sum_{r=1}^{n}\left(\frac{X_r}{Xm_r}\right)Xm_i \tag{3}$$

$$X_i = \beta_0 + \sum_{r=1}^{n}\left(\beta_{1r} \times X_r\right) \tag{4}$$

$$X_i = X_r + \left(Xm_r - Xm_i\right) \tag{5}$$

$$X_i = \beta_0 + \beta_1 \times X_r \tag{6}$$

where:

X_i - variable to be gap-filled in AWS "i";

n - number of AWSs considered;

X_r - corresponding observed data in the reference AWS "r";

d_r - distance between the AWS to be filled in and the reference AWS "r";

Xm_r - historical averages in the reference AWS "r";

Xm_i - historical average in the AWS to be filled;

β_0 - linear coefficient of the regression;

β_1 - angular coefficients; and,

β_{1r} - angular coefficients to be estimated at reference stations "1" to "n".

The preliminary analysis of the 33 automatic weather stations in the state of Mato Grosso showed that each AWS had, on average, 18.61% gaps in the time series of each variable reaching a maximum of 62.06% (Table 1). Hence, to test the potential of each gap-filling method to estimate the variable for different possible amounts of gaps, from the original database of each meteorological variable, seven databases containing artificially and randomly inserted gaps were generated, so that in each AWS there were 25, 30, 40, 50, 60, 70 and 80% gaps in the time series.

The gap-filling accuracy was obtained by comparing the observed with the filled data of the artificially inserted gaps employing the bias, the root mean squared error (RMSE and normalized RMSE), Pearson's correlation coefficient, and the percentage of data not filled. Finally, the filling methods were ranked in a way that the methods whose results of Bias, RMSE, and percentage of unfilled gaps were closer to zero and whose Pearson coefficients, in absolute values, were close to one received better positions (lower rank value).

Results and Discussion

The effects of the percentage of gaps on the accuracy of the different gap-filling methods (Table 2) indicate that the errors rise with the increase of gaps in the time series, reaching, on average, 2.42% of unfilled gaps, 1.37% Bias, 10.08% normalized RMSE and minimum correlation of 0.6851, in the time series with up to 80% missing data.

Among the methods, those belonging to group 1 (MAS, MID, MPR, RLM) showed, in general, the worst results with an increase in the error magnitude as more gaps were added to the time series. This aspect is more explicit in the ability to fill all gaps, which shows that in the methods of group 1, approximately 0.05 to 9.94% of the gaps were left unfilled, while in the methods of group 2 (MUK (v) and RLS (v)), the remaining unfilled gaps do not exceed 0.1%. This is mainly because the methods of group 1 require at least three neighbour AWS with data to fill the gaps; consequently, the increase in missing data in the time series of all AWS reduces the probability of meeting this condition.

In addition, as GapMET allows the estimated data in an AWS to be used as reference when gap-filling a neighbour AWS, error propagation can occur when not using only original measure of the variable. This process, however, does not occur in the methods that use external series (MUK (s) and RLS (s)), since these series, a priori, do not have gaps, consequently allowing the errors to remain relatively constant regardless of the number of gaps in the time series to be filled. Therefore, in methods with external series, the accuracy depends only on the quality of the external datasets.

Nevertheless, it is important to note that the performance errors are similar among the methods in the tests with up to 40% missing data. Therefore, in the real circumstances of the AWS time series of Mato Grosso state, whose number of missing data is approximately 20%, all gap-filling methods showed good results.

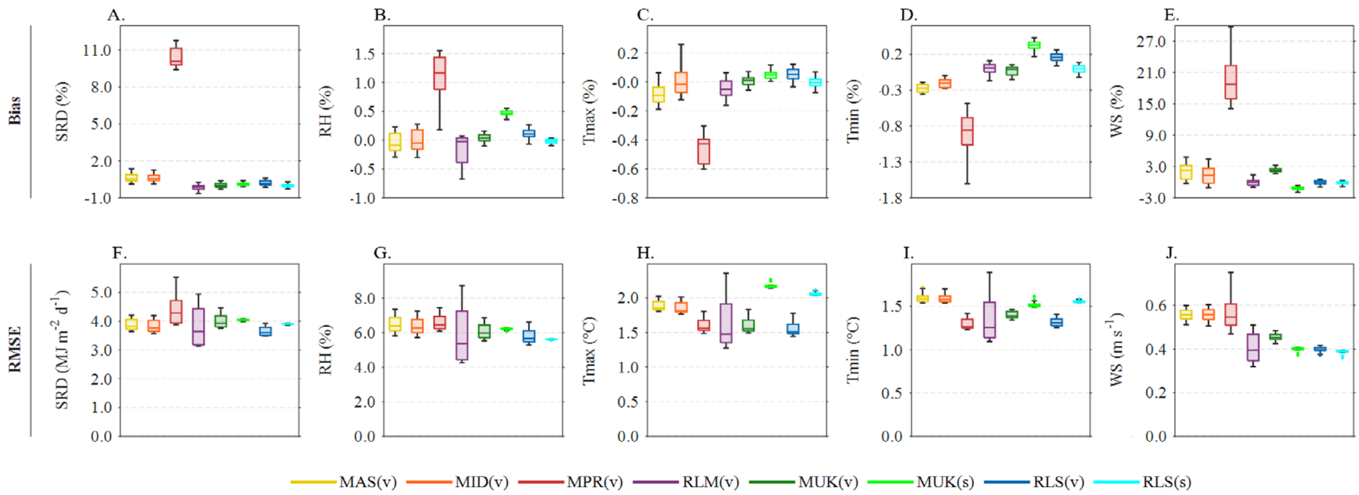
The results of Bias (Figures 2A to E), RMSE (Figures 2F to J), as well as the dispersion and correlation of observed and gap-filled data (Figure 3) as a function of the meteorological variable and method, are shown below. The variables relative air humidity (RH), maximum air temperature (Tmax), and minimum air temperature (Tmin) showed more accurate gap-filling estimates, with bias ranging between -1.0 and 1.5% and Pearson's correlations between 0.69 and 0.91. On the other hand, the variables incident global radiation (SRD) and wind speed (WS), despite generating bias between -3 and 5% (except for the MPR method), showed greater dispersion of the estimated data, with correlations lower than 0.60 for SRD and lower than 0.75 for WS.

Variables are shown on Y axis: SRD - Downward solar radiation; RH - Relative air humidity; Tmax - Maximum air temperature; Tmin - Minimum air temperature; WS - Wind speed. Gap-filling methods are shown on X axis: MAS - Simple arithmetic means; MUK - UK traditional method; MID - Inverse distance method; MPR - Regional weighting method; RLS - Simple linear regression; RLM - Multiple linear regression. Gap-filling methods followed by (v) used neighbour weather stations as reference time series, while methods followed by (s) used remote sense data as reference

Table 2. Mean performance of the six Gap-filling methods based on the percentage of artificial missing data generated in the time series of 33 weather stations in the state of Mato Grosso, Brazil

	Gaps (%)	MAS (V)	MID (V)	MPR (V)	RLM (V)	MUK (V)	MUK (S)	RLS (V)	RLS (S)	Mean
Unfilled gaps (%)	25	0.05	0.05	0.05	0.05	0.00	0.01	0.00	0.00	0.02
	30	0.05	0.05	0.06	0.05	0.00	0.01	0.00	0.00	0.03
	40	0.08	0.08	0.09	0.08	0.00	0.01	0.00	0.00	0.04
	50	0.11	0.11	0.13	0.11	0.01	0.01	0.00	0.00	0.06
	60	0.19	0.19	0.23	0.46	0.01	0.01	0.01	0.00	0.14
	70	0.50	0.50	0.55	3.22	0.03	0.02	0.02	0.00	0.61
	80	3.03	3.03	3.16	9.94	0.10	0.02	0.09	0.00	2.42
	Mean	0.57	0.57	0.61	1.99	0.02	0.01	0.02	0.00	
Bias (%)	25	0.35	0.25	5.05	-0.10	0.55	-0.06	-0.04	-0.01	0.75
	30	0.23	0.11	4.83	-0.10	0.55	-0.04	0.07	-0.03	0.70
	40	0.28	0.11	5.23	-0.14	0.34	-0.06	0.03	-0.05	0.72
	50	0.48	0.32	5.71	0.03	0.42	-0.07	0.12	-0.06	0.87
	60	0.58	0.47	6.51	0.00	0.44	-0.03	0.14	-0.01	1.01
	70	0.62	0.54	6.52	-0.18	0.51	-0.08	0.10	-0.06	1.00
	80	0.82	0.78	8.77	-0.12	0.49	0.02	0.18	-0.01	1.37
	Mean	0.48	0.37	6.09	-0.09	0.47	-0.05	0.09	-0.03	
Normalized RMSE (%)	25	10.24	10.08	9.71	7.39	9.11	9.95	8.43	9.39	9.29
	30	9.99	9.86	9.34	7.26	8.90	9.71	8.24	9.32	9.08
	40	9.66	9.55	9.16	7.34	8.57	9.28	7.93	8.90	8.80
	50	9.79	9.70	9.50	8.04	8.69	9.25	8.03	8.85	8.98
	60	9.86	9.79	9.89	8.77	8.82	9.12	8.13	8.78	9.15
	70	10.05	10.00	10.23	9.93	9.00	9.09	8.27	8.61	9.40
	80	10.73	10.70	11.60	11.67	9.44	9.19	8.66	8.64	10.08
	Mean	10.04	9.96	9.92	8.63	8.93	9.37	8.24	8.93	
Pearson's correlation coefficient	25	0.7034	0.7145	0.8131	0.8598	0.8033	0.7479	0.8137	0.7480	0.7755
	30	0.7007	0.7101	0.8101	0.8568	0.8005	0.7389	0.8104	0.7366	0.7705
	40	0.6862	0.6943	0.7974	0.8381	0.7931	0.7341	0.8033	0.7313	0.7597
	50	0.6723	0.6791	0.7837	0.8038	0.7841	0.7333	0.7948	0.7303	0.7477
	60	0.6551	0.6612	0.7677	0.7618	0.7725	0.7321	0.7832	0.7292	0.7328
	70	0.6317	0.6374	0.7467	0.7003	0.7562	0.7313	0.7666	0.7280	0.7123
	80	0.6025	0.6074	0.7144	0.6243	0.7340	0.7296	0.7420	0.7269	0.6851
	Mean	0.6646	0.6720	0.7761	0.7778	0.7776	0.7353	0.7877	0.7329	

MAS - Simple arithmetic means; MUK - UK traditional method; MID - Inverse distance method; MPR - Regional weighting method; RLS - Simple linear regression; RLM - Multiple linear regression



Meteorological variables: SRD - Downward solar radiation; RH - Relative air humidity; Tmax - Maximum air temperature; Tmin - Minimum air temperature; WS - Wind speed. Gap-filling methods: MAS - Simple arithmetic means; MUK - UK traditional method; MID - Inverse distance method; MPR - Regional weighting method; RLS - Simple linear regression; RLM - Multiple linear regression. Gap-filling methods followed by (v) used neighbour weather stations as reference time series, while methods followed by (s) used remote sense data as reference time series

Figure 2. Performance of Bias (A to E) and root mean squared error (F to J) based on the meteorological variables and methods used to gap-fill the time series of 33 weather stations in the state of Mato Grosso, Brazil

time series. The red line on the diagram represents the Pearson correlation coefficient (r), and the colour variation represents the amount of data in the x and y coordinates

Similarly, Ventura et al. (2016) and Xavier et al. (2016) found better estimates in the variables air temperature and relative air humidity compared to the variables wind speed and incident solar radiation. The lower precision in gap-filling the variables SRD and WS can be explained by the fact that the daily values of these variables show greater spatial variations when compared to Tmin, Tmax, and RH, in addition to local

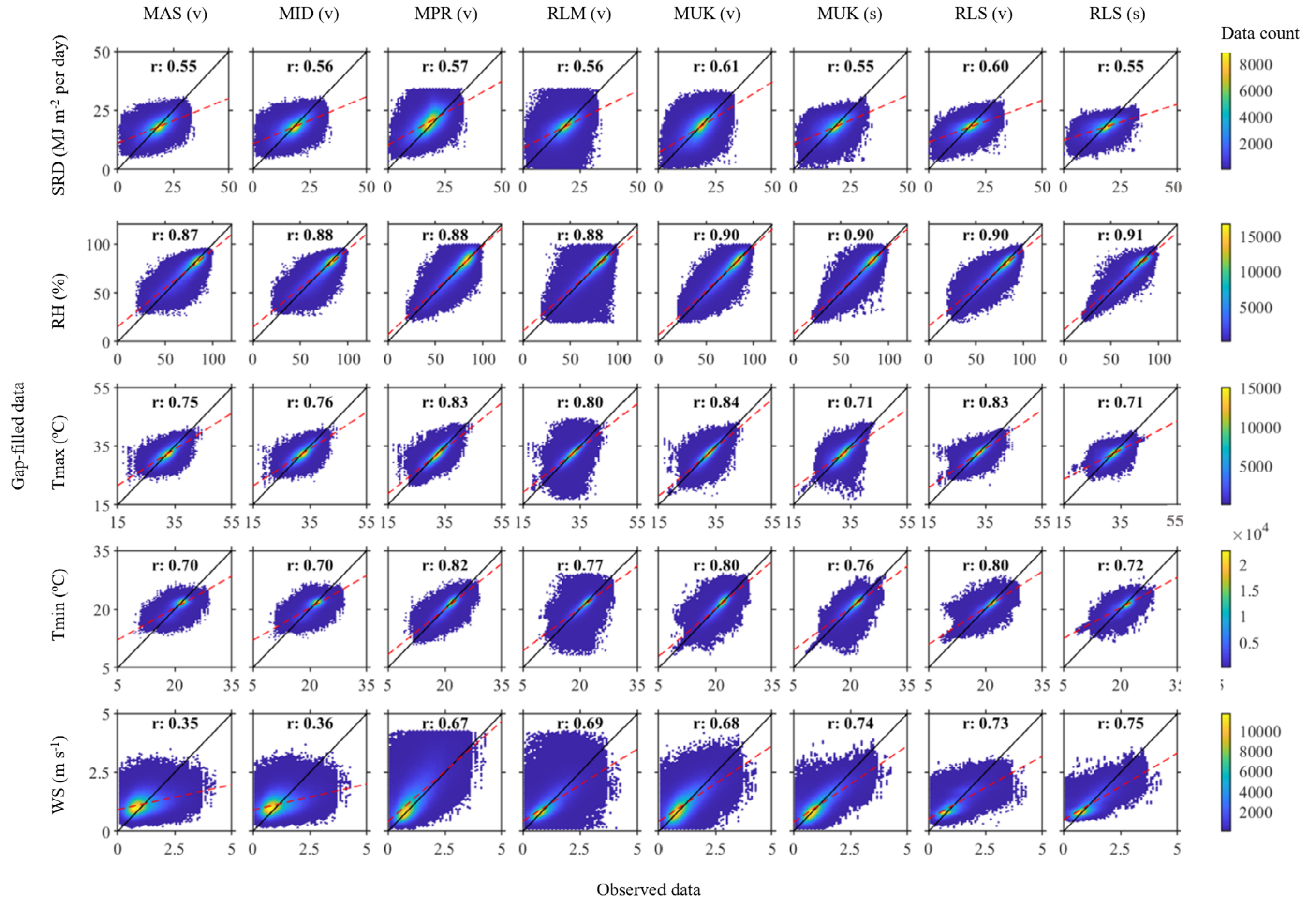
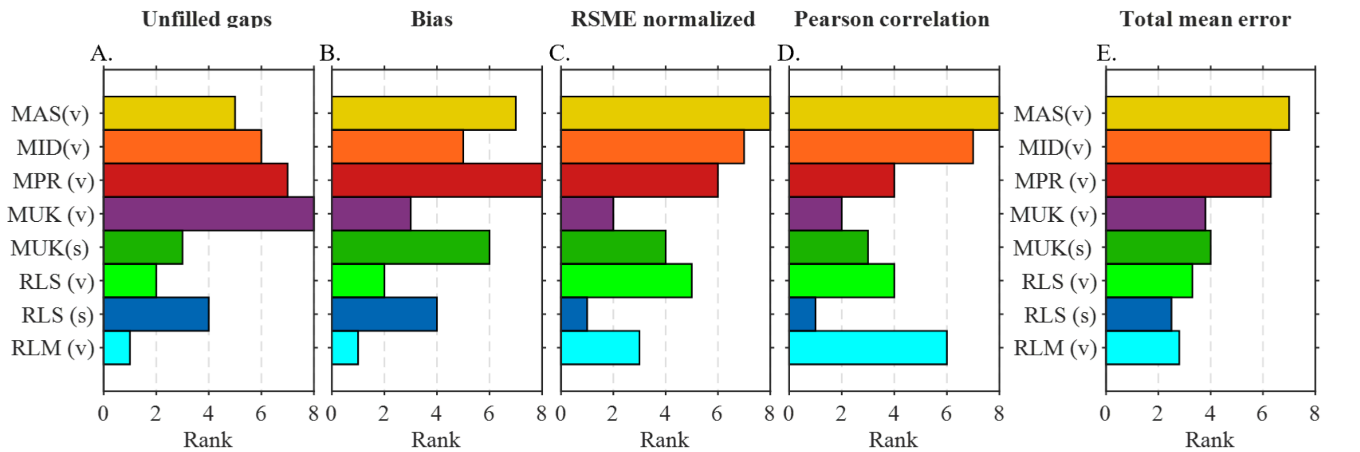


Figure 3. Dispersion between the variables measured and gap-filled data obtained by the different GapMET methods at the 33 weather stations of Mato Grosso state, Brazil

effects not considered in the methods, such as altitude and cloud cover (Xavier et al., 2016).

Based on the ranking results as a function of the different errors (Figure 4E), better performances are observed for the methods of group 2, especially the RLS method. In this group, methods based on ERA5-land data showed lower performance in terms of RMSE and Pearson's correlation (Figures 4C and D). However, these methods were able to reduce the probability of unfilled gaps and showed less dispersion of errors with an increase in missing data (Figures 4A and B), indicating that the test with other external databases besides ERA5-land could promote improvements in the estimates.

Meanwhile, the RLM method, despite obtaining good positions in the analysis of bias errors, RMSE, and correlation (Figures 4B, C and D), had more difficulty in filling in all missing data of any meteorological variables (Figure 4A).



Gap-filling methods are shown on Y axis: MAS - Simple arithmetic means; MUK - UK traditional method; MID - Inverse distance method; MPR - Regional weighting method; RLS - Simple linear regression; RLM - Multiple linear regression. Gap-filling methods followed by (v) used neighbour weather stations as reference time series, while methods followed by (s) used remote sense data as reference time series. Methods with lower rank value show better results

Figure 4. Error analysis ranking of the six methods to gap-fill meteorological data of 33 weather stations in the state of Mato Grosso, Brazil

These results disagree with those found by Fernandez (2007), Ventura et al. (2016), and Bier & Ferraz (2017), who suggested the RLM method as the most accurate among similar methods considered in this study. However, these authors did not consider the impact of the increase in the number of missing data, which were being tested in time series with a percentage of gaps below 30% of the datasets.

Bleidorn et al. (2022), for example, compared the MAS, RLS, and RLM methods with a percentage of missing data ranging from 5 to 40% and observed better performances with the RLS method. In this case, these results were justified by the algorithm used, which, similarly to GapMET, selects the reference AWS by the highest correlation coefficient, allowing the change of the reference AWS when finding missing data in it, which does not occur with the RLM method, which requires at least three reference AWS. Besides, it is also important to highlight that, when evaluating the performance in the datasets with less than 40% of missing values (Table 2), the errors were similar among all methods, and in the analyses of the normalized RMSE and Pearson's correlation coefficient the RLM method showed better performance than the RLS.

Moreover, the lower performance of group 1 methods, as well as the divergences between these results and the literature regarding the differences between the RLS and RLM methods, results from the fact that in this study the minimum distance for all AWS to have at least three reference AWS was 350 km, while in Fernandez (2007), Ventura et al. (2016), and Bier & Ferraz (2017) this distance was, usually, less than 100 km due to the higher AWS network density of the South and Southeast regions of Brazil. The use of nearby AWS as reference is justified because, in general, it has a higher correlation with the AWS to be filled, reducing the risk of using AWS that are in different climatic regions (Brubacher et al., 2020), although the use of 100 km criteria can limit the gap-filling in areas with low AWS network density, such as in the Mato Grosso state.

Conclusions

1. All six methods of the GapMET software had good performance for gap-filling meteorological datasets with less 40% of missing values.

2. In Mato Grosso state, Brazil, with minimum distance between AWS of 350 km, the best statistical performances are obtained with methods that use only one reference AWS, especially RLS.

3. The use of refence series obtained from remote sensing reduces the probability of unfilled gaps; however, an analysis of other satellite products sources is necessary to reduce RMSE and improve the correlations of the estimated data.

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) Finance Code 001. To the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for the productivity grant (Process 308784/2019-7).

Literature Cited

Bier, A. A.; Ferraz, S. E. T. Comparação de metodologias de preenchimento de falhas em dados meteorológicos para estações no Sul do Brasil. Revista Brasileira de Meteorologia, v.32, p.215-226, 2017. https://doi.org/10.1590/0102-77863220008

Bleidorn, M. T.; Pinto, W. de P.; Schmidt, I. M.; Mendonça, A. S. F.; Reis, J. A. T. dos. Methodological approaches for imputing missing data into monthly flows series. Ambiente e Água, v.17, p.1-27. 2022. https://doi.org/10.4136/ambi-agua.2795

Box, G. E. P.; Cox, D. R. An analysis of transformations. Journal of the Royal StatisticalSociety: Series B (Methodological), v.26, p.211-243, 1964. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x

Brubacher, J. P.; Oliveira, G. G. de; Guasselli, L. A. Preenchimento de falhas e espacialização de dados pluviométricos: desafios e perspectivas. Revista Brasileira de Meteorologia, v.35, p.615-629, 2020. http://dx.doi.org/10.1590/0102-7786354006

Coutinho, E. R.; Silva, R. M. da; Madeira, J. G. F.; Coutinho, P. R. de O. dos S.; Boloy, R. A. M.; Delgado, A. R. S. Application of Artificial Neural Networks (ANNs) in the Gap Filling of Meteorological Time Series. Revista Brasileira de Meteorologia, v.33, p.317-328, 2018. https://doi.org/10.1590/0102-7786332013

Fernandez, M. N. Preenchimento de Falhas em Séries Temporais. Rio Grande: FURG, 2007. 106p. Dissertação Mestrado

Hubbard, K. G. Spatial variability of daily weather variables in the high plains of the USA. Agricultural and Forest Meteorology, v.68, p.29-41, 1994. https://doi.org/10.1016/0168-1923(94)90067-1

Jerszurki, D.; Souza, J. L. M. de; Silva, L. de C. R. Sensitivity of ASCE-Penman–Monteith reference evapotranspiration under different climate types in Brazil. Climate Dynamics, v.53, p.943-956, 2019. https://doi.org/10.1007/s00382-019-04619-1

Muñoz-Sabater, J.; Dutra, E.; Agustí-Panareda, A.; Albergel, C.; Arduini, G.; Balsamo, G.; Boussetta, S.; Choulga, M.; Harrigan, S.; Hersbach, H.; Martens, B. ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. Earth System Science Data, v.13, p.4349-4383, 2021. https://doi.org/10.5194/essd-13-4349-2021

Muñoz-Sabater, J. ERA5-Land hourly data from 1981 to present. 2019. Available at: https://doi.org/10.24381/cds.e2161bac. Accessed on: Jun. 2021.

Paulhus, J. L. H.; Kohler, M. A. Interpolation of missing precipitation records. Monthly Weather Review, v.80, p.129-133, 1952. https://doi.org/10.1175/1520-0493(1952)080<0129:IOMPR>2.0.CO;2

Sarnighausen, V. C. R.; Gomes, F. G.; Dal Pai, A.; Rodrigues, S. A. Estimativa da evapotranspiração de referência para Botucatu-SP por meio de modelos de regressão. Revista Brasileira de Climatologia, v.28. p.766-787, 2021. http://dx.doi.org/10.5380/rbclima.v28i0.71569

Souza, A. P.; Mota, L. L.; Zamadei, T.; Martim, C. C.; Almeida, F. T.; Paulino, J. Classificação climática e balanço hídrico climatológico no estado de Mato Grosso. Nativa, v.1, p.34-43, 2013. https://doi.org/10.31413/nativa.v1i1.1334

Tabony, R. C. The estimation of missing climatological data. Journal of Climatology, v.3, p.297-314, 1983. https://doi.org/10.1002/joc.3370030308

Ventura, T. M.; Santana, L. L. R.; Martins, C. A.; Figueiredo, J. M. de. Análise da aplicabilidade de métodos estatísticos para preenchimento de falhas em dados meteorológicos. Revista Brasileira de Climatologia, v.19, p.168-177, 2016. http://dx.doi.org/10.5380/abclima.v19i0.44989

Xavier, A. C.; King, C. W.; Scanlon, B. R. Daily gridded meteorological variables in Brazil (1980–2013). International Journal of Climatology, v.36, p.2644-2659, 2016. https://doi.org/10.1002/joc.4518

Xiang, K.; Li, Y.; Horton, R.; Feng, H. Similarity and difference of potential evapotranspiration and reference crop evapotranspiration – a review. Agricultural Water Management, v.232, p.1-16, 2020. https://doi.org/10.1016/j.agwat.2020.106043

WMO - World Meteorological Organization. Technical Regulations: General Meteorological Standards and Recommended Practices, v.1, n.49. Geneva: Secretariat of the World Meteorological Organization, 1988, 49p.