ISSN 1808-8546 (ONLINE) 1808-3765 (CD-ROM)

AVALIAÇÃO DO DESEMPENHO DE GERADORES DE DADOS EM SÉRIES METEOROLÓGICAS PARA O MUNICÍPIO DE ITACOATIARA-AM*

ARISTÓTELES DE JESUS TEIXEIRA FILHO[1]; LUCAS MELO VELLAME[2] E AUREO SILVA DE OLIVEIRA[3]

[*] Parte 1: Trabalho retirado do Relatório de Pesquisa de Pós-Doutorado "Preenchimento de falhas em observações meteorológicas importantes na estabilidade da modelagem de evapotranspiração potencial de referência para Município de Itacoatiara, AM".

[1]Professor Titular do Curso de Agronomia, Instituto de Ciências Exatas e Tecnologia, Universidade Federal do Amazonas, Campus Itacoatiara; Brasil, aristoteles@ufam.edu.br; ORCID: 0000-0002-0123-7515.
[2]Professor Titular do Curso de Agronomia, Centro de Ciências Agrárias, Ambientais e Biológicas, Núcleo de Engenharia de Água e Solo, Universidade Federal do Recôncavo da Bahia, Campus de Cruz das Almas; Brasil, lucasvellame@ufrb.edu.br; ORCID: 0000-0001-7649-5773.
[3]Professor Titular do Curso de Agronomia, Centro de Ciências Agrárias, Ambientais e Biológicas, Núcleo de Engenharia de Água e Solo, Universidade Federal do Recôncavo da Bahia, Campus de Cruz das Almas, Brasil, aureo@ufrb.edu.br; ORCID: 0000-0001-7661-551X.

1 RESUMO

O escopo desta investigação foi avaliar o desempenho de geradores de dados de séries meteorológicas de 2019 a 2023 usando o método de interpolação linear, os softwares Easy-Fit e o WeatherMan, para o município de Itacoatiara, Amazonas. No Brasil, a principal fonte de dados oriundos de estações meteorológicas é administrada e disponibilizada pelo Instituto Nacional de Meteorologia, e a série histórica disponibilizada, normalmente, apresenta falhas em escala temporal horária, diária ou mensal. A falta de continuidade nas séries de dados pode levar a conclusões errôneas em processos que dependem diretamente do clima, como a agricultura e o planejamento do uso dos recursos naturais. O estudo foi desenvolvido com os dados meteorológicos da região centro-leste do Estado do Amazonas. As falhas foram preenchidas usando interpolação linear e os geradores de dados climatológicos Easy-Fit e WeatherMan, sendo os resultados comparados com a normal climatológica disponível. A correlação entre a normal climatológica e a normal provisória do parâmetro temperatura máxima mensal média mostrou a existência de regressão, de modo que os três métodos podem ser sugeridos para o preenchimento de falhas, quando não se dispõe de dados completos. O método de interpolação linear e os geradores Easy-Fit e WeatherMan apresentaram potencial para o preenchimento de falhas em séries meteorológicas mensais da região de Itacoatiara-AM, mesmo diante da limitada disponibilidade de dados completos.

Palavras-chave: variáveis meteorológicas, modelagem estatística, agrometeorologia.

Recebido em 23/10/2025 e aprovado para publicação em 03/12/2025
DOI: http://dx.doi.org/10.15809/irriga.2025v30p303-326

TEIXEIRA FILHO, A. de J.; VELLAME, L. M. E.; OLIVEIRA, A. S. de

PERFORMANCE EVALUATION OF WEATHER DATA GENERATORS BASED ON TIME SERIES FOR THE MUNICIPALITY OF ITACOATIARA-AM

2 ABSTRACT

The purpose of this investigation was to evaluate the performance of weather data generators for the municipality of Itacoatiara, Amazonas, Brazil, on the basis of time series from 2019 to 2023 by the following methods: linear interpolation, easy-fit, and WeatherMan. In Brazil, INMET (National Institute of Meteorology) is the main source of meteorological data, but gaps are very common at hourly, daily, and monthly time scales. The lack of continuity in data series may lead to erroneous conclusions in studies that depend on it. on these data, such as agricultural and natural resource management planning. In this study, monthly meteorological data from the central-eastern region of the state of Amazonas were used. Missing data were filled using linear interpolation, and the climatological data generator Easy-Fit and WeatherMan Estimated values were then compared with the available climatological normal. The correlation between the climatological normal and the provisional normal of the mean monthly maximum temperature parameter demonstrated the existence of regression, indicating that the three methods may be recommended for gap filling. when incomplete datasets are available. The linear interpolation, easy-fit, and WeatherMan methods showed potential for filling the gaps in monthly meteorological series in the Itacoatiara-AM region.

Keywords: weather variables, statistical modeling, agrometeorology.

3 INTRODUCTION

In the northern region, particularly in the state of Amazonas, the number of meteorological stations (conventional and automatic) is small compared with the state's territorial extension, and observations often fail because of operational malfunctions, resulting in the absence of accurate information. Cities are geographically distant from one another, which often necessitates the search for alternatives to generate such information, making the database robust for use in agrometeorological planning, which is fundamental for decision-making. Although several methods exist for filling these gaps, uncertainties still exist concerning the performance of these approaches in humid tropical regions with low observational density, such as the Amazon.

To ensure the reliability of many climate studies, especially those analyzing climate variability, reliable and homogeneous data are necessary since the results of analyses with heterogeneous data can lead to contradictory and erroneous conclusions (Santos, 2012). Data gaps occur when data collection is deficient at some point, either because of the absence of the collector or because of errors and overlapping data from the collection instruments, which makes it necessary to apply equations to fill these gaps (Diaz; Pereira; Nóbrega, 2018). Some methods for analyzing these data may be suitable for this situation, but most of them require that the series be complete (Guzmán, 2018). Several methods can be used to fill in gaps in meteorological data, including the use of averages of observed data (climatological normals) or synthetic data from data generators (Oliveira *et al*., 2010).

The estimation of faulty meteorological data can be performed using methods organized into three basic categories: (i) within the station; (ii) between stations; and (iii) regression-based (Allen; Degaetano, 2001). The first category was the most suitable for the purposes of the investigation. The second is based on the hypothesis that the difference between the daily temperatures of nearby stations is equal to the difference between the monthly averages of the temperatures of these neighboring stations; however, no adjacent stations with complete data were found in the region. Finally, the third category assumes the existence of available data from one or more neighboring stations, which also demands the completeness of this information. In this scenario, it is difficult to obtain access to a robust dataset from the meteorological stations available at the time of the research, which is the responsibility of the managing body. Therefore, this difficulty led to the need to resort to data generators.

Most agricultural production simulation models require daily climate data. Typically, these data are available for model development and validation; however, in many cases, meteorological data for long-term analysis across different locations are unavailable, necessitating the use of meteorological data-generating models (Meinke). *et al.,* 1995). Recently, several data generators have been developed to simulate daily meteorological data (Bristow; Campbell, 1984; Richardson; Wright, 1984), such as the WGEN model (Richardson; Wright, 1984). In the case of the models used in DSSAT (Decision Support System for Agrotechnology (Transfer)), the minimum set of climate data required includes daily values of solar radiation, maximum and minimum temperatures, and precipitation (Hoogenboom; Jones; Boote, 1992). Solar radiation is important for estimating photosynthesis and evapotranspiration, temperature is important for determining crop growth and development rates, and precipitation is important for meeting water needs (Ritchie; Godwin; Singh, 1990).

The failure to collect or record this meteorological information is caused by several factors. Above all, those that most affect the cities in the interior of the state of Amazonas are the bureaucracy in the maintenance of the equipment and the lack of awareness among the population about the importance of keeping a meteorological station in perfect working order, since certain equipment ends up being targets of vandalism or theft (Nascimento *et al*., 2009). In modeling, research is needed to evaluate ways of filling in the missing data gaps in historical series to improve estimates of potential reference evapotranspiration and, consequently, of the processes in which this variable is used (Gionanella). *et al*., 2021).

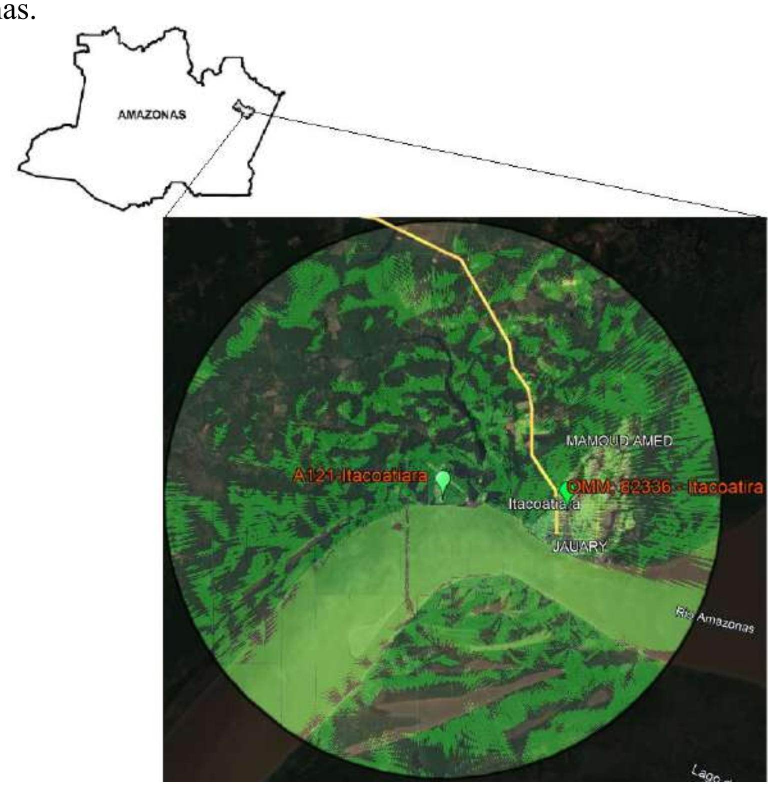
Filling in gaps in meteorological data in the Amazon Basin is limited by the low density of meteorological stations and the large distance between them compared to those in the southern and southeastern regions of the country, which leads to a reduction in the accuracy of statistical methods. The quality of modeling is compromised by missing data from the Itacoatiara and AM meteorological stations. Therefore, this study aimed to evaluate the performance of meteorological data series generators for the municipality of Itacoatiara, Amazonas, from 2019 to 2023 using the easy-fit and WeatherMan linear interpolation methods.

4 MATERIALS AND METHODS

4.1 Experimental Location

The study was developed using meteorological data from the central-eastern region of the state of Amazonas. The region has two meteorological stations (one conventional and one automatic) belonging to the network of the National Institute of Meteorology, located in the municipality of Itacoatiara, AM (Figure 1). Table 1 shows the locations of the World Meteorological Organization (WMO: 82336) station and the automatic station (A121). The investigations were carried out using a provisional climatological normal obtained from the period 2009 to 2023 from the Meteorological Database for Teaching and Research (BDMEP) of the National Institute of Meteorology (INMET). In particular, the aforementioned provisional climatological normal conforms to the WMO (1989), which was established as “a data series with a minimum of 10 years and less than 30 years”.

Figure 1. Spatial distribution of meteorological stations A121 and OMM 82336 in Itacoatiara, Amazonas.



Source: Authors (2025).

Table 1. Location of conventional meteorological stations.

Name	UF	Latitude	Longitude	Altitude	Installation Date	Code
		°		m		
Itacoatiara	AM	-3.13694444	-58.44277777	19,16	31/12/1926	82336
	AM	-3.13333333	-58.48277777	41.54	16/04/2008	A121

Source: Authors (2025)

4.2 Experimental Methodology

In the meteorological series from 2009 to 2023, 12.8%, 13.9%, 8.9%, 8.3%, 11.1%, 14.4%, 12.2%, and 39.4% of the missing data were identified (Table 2 and Table 3) for the variables maximum temperature, minimum temperature, maximum relative humidity, minimum relative humidity, average relative humidity, wind speed, atmospheric pressure, and global solar radiation, respectively. The data series from neighboring stations (Table 2) were analyzed. After it was found that complete data were missing from stations located up to 150 km from the Itacoatiara Meteorological Station (A121), filling the gaps using linear regression became unfeasible, leading to the need to change the method for completing the missing data.

Table 2. Months with missing data for meteorological variables for 2009 (09) to 2023 (23)

Months	Minimum air temperature						Maximum air temperature					
	09	12	13	19	20	21	09	12	13	19	20	21
J												
F												
M												
A												
M												
J												
J												
A												
S												
A												
N												
D												

Months	Minimum Relative Humidity of the air					Maximum Relative Humidity of the air				
	16	18	19	20	21	16	18	19	20	21
J										
F										
M										
A										
M										
J										
J										
A										
S										
A										
N										
D										

Source: Authors (2025)

Table 3. Months with missing data for meteorological variables for 2009 (09) to 2023 (23)

Months	Average Atmospheric Air Pressure									
	09	10	12	13	16	18	19	20	21	23
J										
F										
M										
A										
M										
J										
J										
A										
S										
A										
N										
D										
Months	**Average air wind speed**									
	09	**10**	**12**	**13**	**16**	**18**	**19**	**20**	**21**	**23**
J										
F										
M										
A										
M										
J										
J										
A										
S										
A										
N										
D										

Source: Authors (2025)

Table 2. Location of the stations near A121

Name	UF	Latitude	Longitude	Distance	Altitude	Code
		°		km	m	
Itacoatiara	AM	-3.13694444	-58.44277777	-	41.54	A121
Maues	AM	-3.39888888	-57.67388888	94.34	25.49	A122
Parintins	AM	-2.63916666	-56.75611111	198.26	18.72	A123
Urucará	AM	-2.53472221	-57.75805555	105.86	18.49	A124
Urubu River	AM	-2.6336111	-59,60055555	139.32	112.84	A125

Source: Authors (2025)

Data gaps were filled using the in-station method (Allen; Degaetano, 2001), which is based on hourly and daily information that is converted into monthly averages. According to Fante and Sant'Anna Neto (2016), from a statistical perspective, fill-in techniques are also more suitable for estimating monthly and annual values and gaps; therefore, they are not recommended for application to daily data gaps. The advantage of the in-station method is that it does not depend on data from other stations (Kemp *et al*., 1983). Filling was performed using the linear interpolation method, Easy-Fit, and the WeatherMan/DSSAT generator with the WGEN model (Richardson; Wright, 1984; Hoogenboom, 1984). *et al*., 2024). It is important to choose the methodology that best suits the research needs and that is in accordance with the geographical and climatological reality in which the meteorological stations studied are located (Fante; Sant'Anna Neto, 2016).

WGEN (Weather Generator) is an algorithmic tool designed to generate synthetic climate series spanning hundreds of years, preserving the essential statistical characteristics (mean, variance, autocorrelation, and transition probabilities) of local observed data. In the case of the Easy-Fit software, the Kolmogorov–Smirnov test was used to complete the data fitting, as the data were nonparametric. Easy-Fit was chosen because of its ability to handle uncertainties and make decisions on the basis of the probability of the information from the data analysis; thus, the best distribution fit was selected. The software allows for easy fitting of a large number of distributions to the available data in seconds, saving time and avoiding analysis errors, which allows for greater confidence in the generated data.

The arithmetic mean method (Equation 1) was composed of the arithmetic mean of the incomplete series, and the linear interpolation method in the interval (X0, X1) (Equation 2) was composed of the maximum and minimum values of the series, as well as the maximum and minimum values of the annual averages of the missing data, respectively. Linear interpolation (Equation 2) can be seen as a weighted average, whose weights are inversely proportional to the distance between the arrival point and the unknown point. The choice of interpolation was due to the possibility of comparing the filling process with more sophisticated software so that it could serve as a comparison or alternative in situations where there is no other way to complete the gaps in meteorological series. However, it is assumed that the variation between the known points is linear, which may not be true in all the cases. When the data exhibit nonlinear behavior, linear interpolation may result in inaccurate estimates.

$$\overline{m} = \frac{m_1 + m_2 + m_3 + \cdots + m_n}{n} \quad (1)$$

$$Y_x = Y_0 + \frac{\overline{X} - X_0}{X_1 - X_0}(Y_1 - Y_0) \quad (2)$$

where $\overline{m}$ is the arithmetic mean of the observations; (m1 + m2 + m3 + ... + mn) are the measured values; n is the number of years observed; Yx is the value of the fault to be found; Y0 is the minimum value of the variable Y; Y1 is the maximum value of the variable Y; X is the average between the maximum and minimum values of the variable Y; X0 is the minimum value of the variable X; and X1 is the maximum value of the variable X.

Easy-Fit software consists of a database containing models, data, and results, as well as underlying numerical algorithms to solve the parameter estimation problem. Depending on the mathematical structure, the

MODFIT. The EXE module performs parameter estimation in explicit functions, Laplace transforms, steady-state systems, ordinary and differential algebraic equations, and PDEFIT. EXE for parameter estimation in one-dimensional time-dependent partial algebraic equations. The maximum (Tx), minimum (Tn), and average (Tm) temperature data and the average relative humidity (RHm) and global solar radiation (Qg) for the municipality of Itacoatiara were converted to the WeatherMan/DSSAT v. 8.5.0 format. WGEN models (Richardson; Wright; Wgen, 1984; Hoogenboom) were used to generate the climate data. *et al., 2024*) and easy-fit methods. From the processed information, the models generate Tx, Tn, Tm, URm, and Qg data on the basis of internally calculated statistical parameters.

To evaluate the gap-filling methods, the linear correlation between the climatological normal from 1991 to 2020 (INMET, 2025) was used as an alternative, which partially incorporates and partially extrapolates the provisional climatological normal from 2009 to 2023 (15 years), as evaluated in this investigation. Thus, the methods were evaluated using linear correlation and the coefficient of determination (R^2) with a meteorological series from 2009 to 2023 and the climatological normal from 1991 to 2020 (INMET, 2025). The statistical hypotheses were tested using the p value; if the p value is less than 0.05, the null hypothesis is rejected (H0: 1 β= 0); that is, the parameter is statistically significant, and the climatological normal has a significant effect on the provisional climatological normal. If the p value is greater than or equal to 0.05, the null hypothesis (H0) is not rejected; that is, there is not enough evidence to assert, at the 5% probability level, that the climatological normal has a significant effect on the provisional climatological normal.

5 RESULTS AND DISCUSSION

Owing to flaws in the meteorological database of stations A122 (Maués), A123 (Parintins), A124 (Urucará), and A125 (Rio Urubu), which are “neighboring” A121 and WMO: 82336, it was impossible to apply simple linear regression (SLR), multiple linear regression (MLR), and stepwise multiple linear regression (SML), which were intended to estimate the missing data for the variables of air temperature, relative humidity, mean atmospheric pressure, and mean wind speed. This limitation corroborates what Ritchie, Godwin, and Singh (1990) state, indicating that one of the greatest deficiencies limiting the use of large-scale models is the lack of soil and climate data.

For the hourly datasets of air temperature, relative humidity, atmospheric pressure, and average wind speed, no statistical techniques were applied to fill in the gaps; however, techniques were applied to the daily data, even considering that Fill (1987) and Bertoni and Tucci (2007) state that none of the various gap-filling methods “are suitable for filling daily gaps, being more recommended for filling monthly or annual gaps." The problem is that the WeatherMan/DSSAT v. 4.8.5.0 program requires daily input data for the generation of climate data by the WGEN model. Therefore, data generation must be adjusted to the research assumptions given the observed situation, given that the meteorological stations (Table 2) did not have complete data.

Given the impossibility of applying the previously discussed methods, linear interpolation was applied, and the Easy-Fit and WeatherMan/DSSAT v. 4.8.5.0 software were used to complete the missing data for the aforementioned variables (Table 2, Table 3). In the case of the Easy-Fit software, the Kolmogorov–Smirnov test was used for adjustment because of the nonparametric nature of the data. Chibana *et al*. (2005) reported that several methods can be used to fill in gaps in meteorological data, including the use of averages of observed data (climatally normal) or synthetic data obtained from data generators.

Table 3. Comparative performance of gap-filling methods

Ps	Met	r_S	R^2	d	w	Desemp	RMSE	EQM	RD	IN	CMR
T_x	I	0.99	0.98	1.00	0.99	Excellent	1.07	3.30	0.53	1.11	0.02
	EF	0.98	0.97	1.00	0.98	Excellent	1.07	3.30	0.57	0.90	0.02
	WM	0.99	0.98	1.00	0.99	Excellent	1.07	3.30	0.51	1.04	0.02
T_n	I	0.93	0.86	1.00	0.93	Excellent	0.27	1.15	0.00	-2.51	-0.02
	EF	0.94	0.88	1.00	0.94	Excellent	0.27	1.15	0.00	-2.32	-0.02
	WM	0.93	0.87	1.00	0.93	Excellent	0.27	1.15	0.00	-2.04	-0.02
UR_m	I	0.97	0.94	1.00	0.97	Excellent	3.27	3.93	0.00	0.90	0.04
	EF	0.99	0.97	1.00	0.98	Excellent	3.27	3.93	0.00	0.80	0.03
	WM	0.98	0.96	1.00	0.98	Excellent	3.27	3.93	0.00	0.98	0.03
U_2	I	0.79	0.62	0.98	0.78	Very good	0.09	8.30	0.00	-5.15	0.20
	EF	0.80	0.63	0.99	0.79	Very good	0.09	8.30	0.00	-1.82	0.14
P_{atm}	I	0.98	0.96	1.00	0.98	Excellent	0.10	0.09	0.00	-16.87	0.00
	EF	0.98	0.95	1.00	0.98	Excellent	0.10	0.09	0.00	-17.22	0.00
n	I	0.98	0.97	0.99	0.98	Excellent	1.31	25.24	0.00	1.43	-0.17
	EF	0.99	0.98	1.00	0.99	Excellent	1.31	25.24	0.00	1.09	-0.12

Note: Values correspond to the monthly averages of the maximum (T_x) and minimum (T_n) air temperatures, average relative air humidity (UR_m), wind speed (U_2), atmospheric pressure (P_{atm}) and insolation (n). Variables: Spearman correlation coefficient (r_S), coefficient of determination ($R^{2)}$, Willmott index (d), root mean square error (RMSE), mean square error (MSE), deviation ratio (RD), modeling efficiency (EM), residual mass coefficient (CMR), interpolation (I), easy fit (EF), WeatherMan (WM), parameters (Ps), performance (Desemp) and methods (Met). Source: Authors (2025)

The literature indicates that the acceptable threshold for the number of missing data points in a month is determined by the climatological variable in question. As recommended by Krusche, Saraiva, and Reboita (2002), in the case of temperatures, the monthly average can be calculated if there are up to 3 consecutive missing values or 5 alternating values in the same month. However, given the need to complete the series and the scarcity of complete meteorological data for the region analyzed, it was not possible to follow the recommendation of Krusche, Saraiva, and Reboita (2002), considering that Nascimento *et al*. (2009) report that, when dealing with climatological studies for the Amazon, scientific publications show a scarcity in the production and organization of data. Although the number of automatic weather stations in northern Brazil has increased, it is still small compared with that in other Brazilian regions.

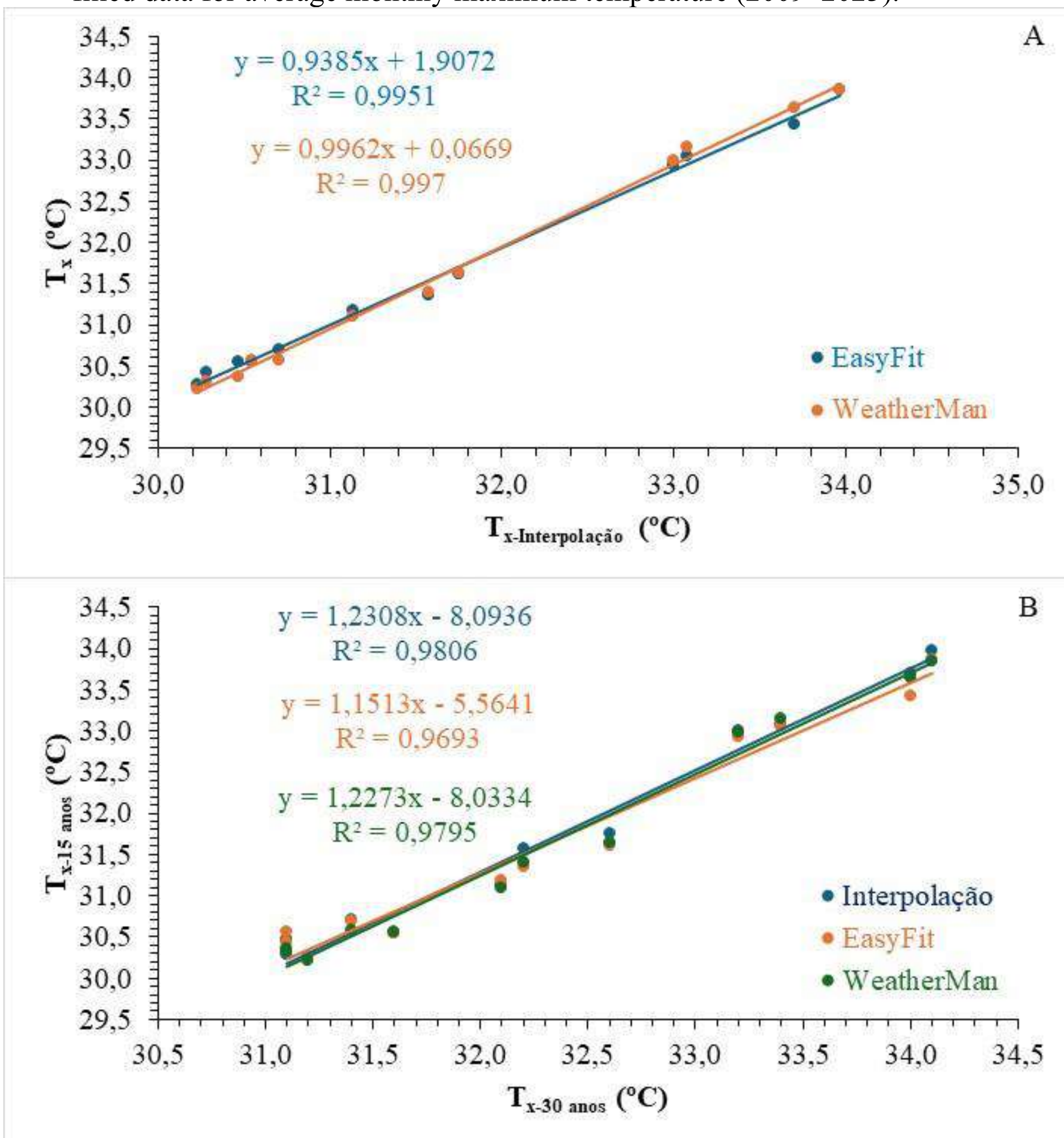
The gaps in the data were filled using the linear interpolation method and the Easy-Fit and WeatherMan/DSSAT software (Table 3). The monthly averages were subsequently linearly correlated to verify the degree of association between the variables (Figures 1 to 8). The methods fit adequately, with coefficients of determination ($R^{2)}$ ranging from 0.9951 to 0.997, which suggests that the Easy-Fit and WeatherMan generators performed well in filling the gaps in the monthly maximum air temperature data (Figure 2A). The climatological normal from 1991 to 2020 was correlated with the filled data from 2009 to 2023 (Figure 2B), with R^2 values ranging from 0.9693 to 0.9806. With a p value < p(0.05) (Table 4), the null hypothesis (H_0, i.e., $\beta_{1} \neq 0$) was rejected, confirming the existence of regression. Thus, all three methods can be adopted to fill such gaps when complete data from neighboring meteorological stations are not available for comparisons between stations or when regression-based methods are used (Allen; Degaetano, 2001). Research conducted by Nascimento *et al*. (2009) and applied to the Amazonas State area also presents other ways to fill gaps. Although these approaches differ from those used in this research, they have a similar objective: to minimize the factors that lead to data gaps, such as formatting errors, unperforming measurements, and eliminating data due to collection errors (Pickering et al., 1994).

Table 4. Regression statistics for the analysis of maximum temperature (T_x) obtained by the simple linear regression method.

Method		**$T_{x\text{-}15\ years}$**		
		R^2 - EasyFit	**Stat t**	**p value**
$T_{x\text{-}30\ years}$	Interpolation	0.98	-4.57	0.000
	EasyFit	0.97	-2.66	0.000
	WeatherMan	0.98	-4.42	0.000

Source: Authors (2025)

Figure 2. A – Comparative performance of gap-filling methods for monthly maximum air temperature; B – Comparison between the climatological normal (1991–2020) and the filled data for average monthly maximum temperature (2009–2023).

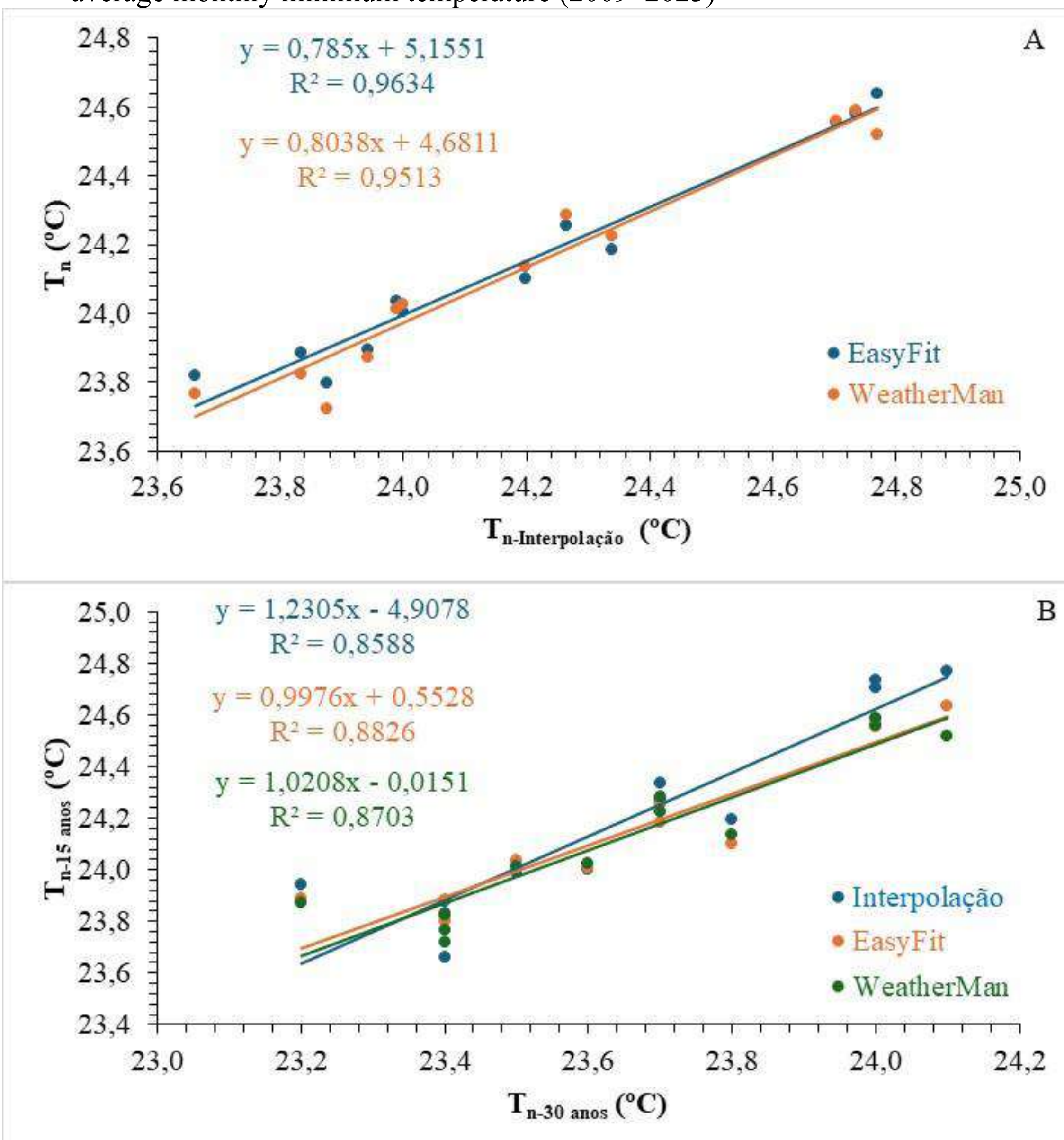


Source: Authors (2025)

An analysis of the data in Figure 3A reveals that the monthly minimum air temperature, which is filled in using the linear interpolation method, adequately fits the WeatherMan/DSSAT and Easy-Fit data, with R^2 values varying between 0.9513 and 0.9634, respectively. In Figure 3B, correlating the 30-year climatological normal data with the provisional 15-year climatological normal data (Table 5), the data generated by Easy-Fit, WeatherMan/DSSAT, and linear interpolation presented a coefficient of determination (R^2) varying between 0.8588 and 0.8703. Indeed, a p value

< 0.05 rejects hypothesis H0, indicating that the 30-year normal data explain the relationship with the 15-year data; in other words, regression exists and that the generated data can be explained by the climatological normal.

Figure 3. A – Performance of gap-filling methods for monthly minimum air temperature; B – Comparison between the climatological normal (1991–2020) and the filled data for average monthly minimum temperature (2009–2023)



Source: Authors (2025)

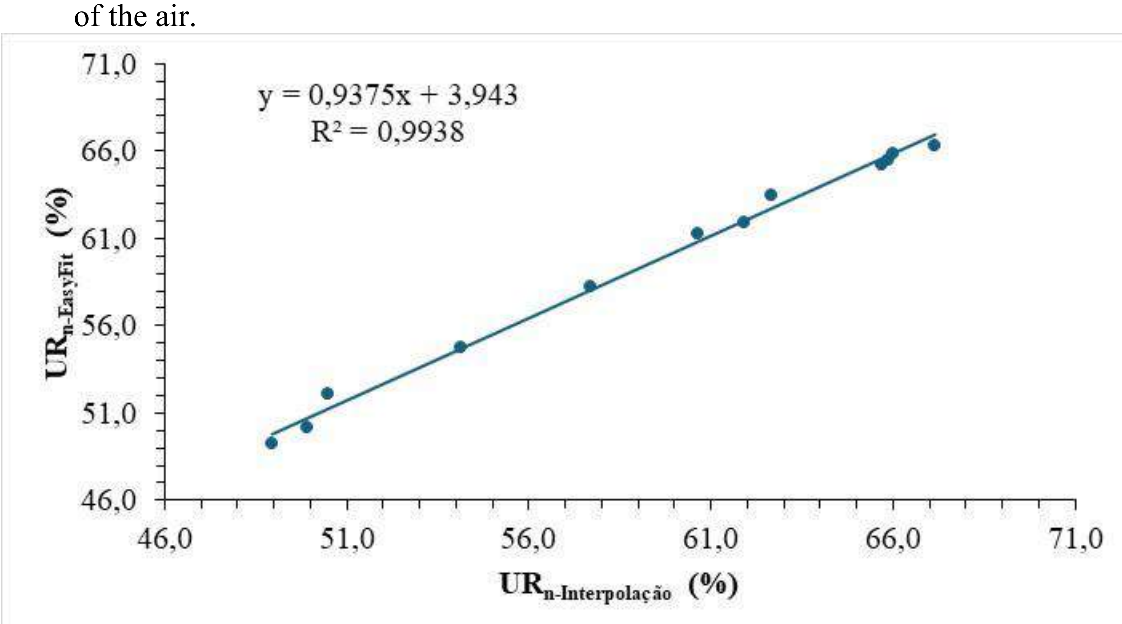
Table 5. Regression statistics for the analysis of the minimum temperature (Tn) obtained by the simple linear regression method.

Method		$T_{n\text{-}15\ years}$		
		R^2 - EasyFit	Stat t	p value
$T_{n\text{-}30\ years}$	Interpolation	0.8588	-1.32	0.000
	EasyFit	0.8826	0.20	0.000
	WeatherMan	0.8703	-0.01	0.000

Source: Authors (2025)

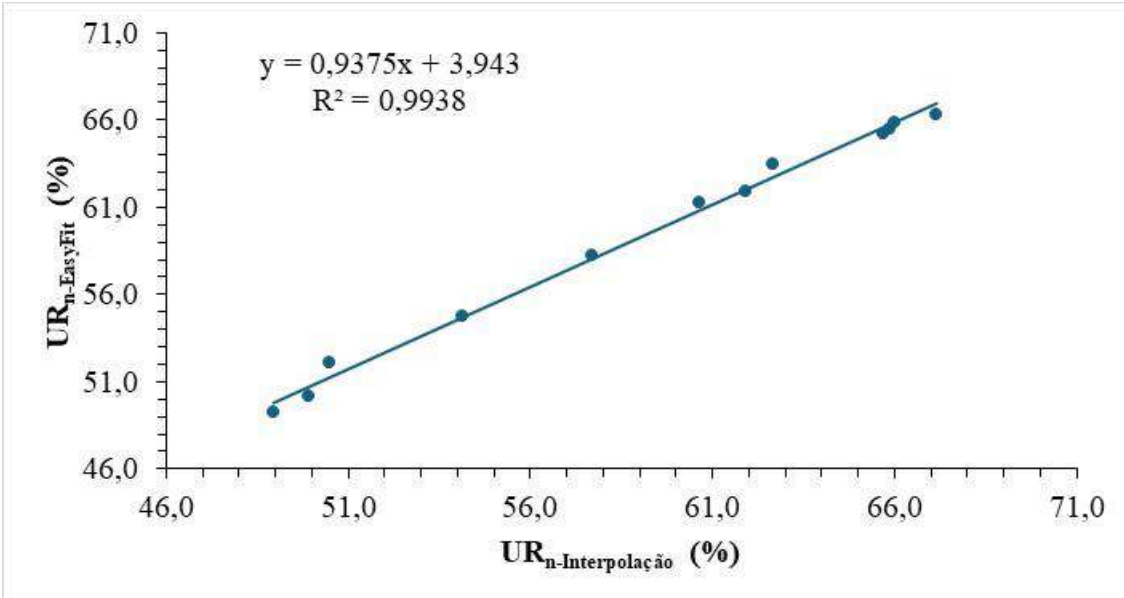
The data generated for the relative air humidity, monthly maximum (Figure 4) and monthly minimum (Figure 5), using the linear interpolation method and the Easy-Fit software, when correlated (Table 6), presented R^2 values of 0.9858 and 0.9938, respectively. These values indicate an excellent fit between the interpolation and the easy-fit software. Unfortunately, in terms of the climatological normal from 1991 to 2020, there were no available data for these variables, which made it impossible to analyze the correlation between them. However, the p value < 0.05 between these variables allows rejection of the null hypothesis (H0, i.e., $\beta_1 \neq 0$), confirming that the regression exists.

Figure 4. Correlation between methods for calculating the maximum monthly relative humidity of the air.



Source: Authors (2025)

Figure 5. Correlation between methods for calculating the minimum monthly relative humidity of the air.



Source: Authors (2025)

Table 6. Regression statistics for the analysis of maximum (RH $_x$) and minimum (RH $_n$) relative humidity obtained by the simple linear regression method.

Method		**UR $_{x\text{-}15\ years}$ (EasyFit)**		
		R ^{2}	**Stat t**	**p value**
UR $_{x\text{-}15\ years}$	Interpolation	0.9858	7.37	0.000
			UR $_{n\text{-}15\ years}$ (EasyFit)	
UR $_{n\text{-}15\ years}$	Interpolation	0.9938	2.83	0.000

Source: Authors (2025)

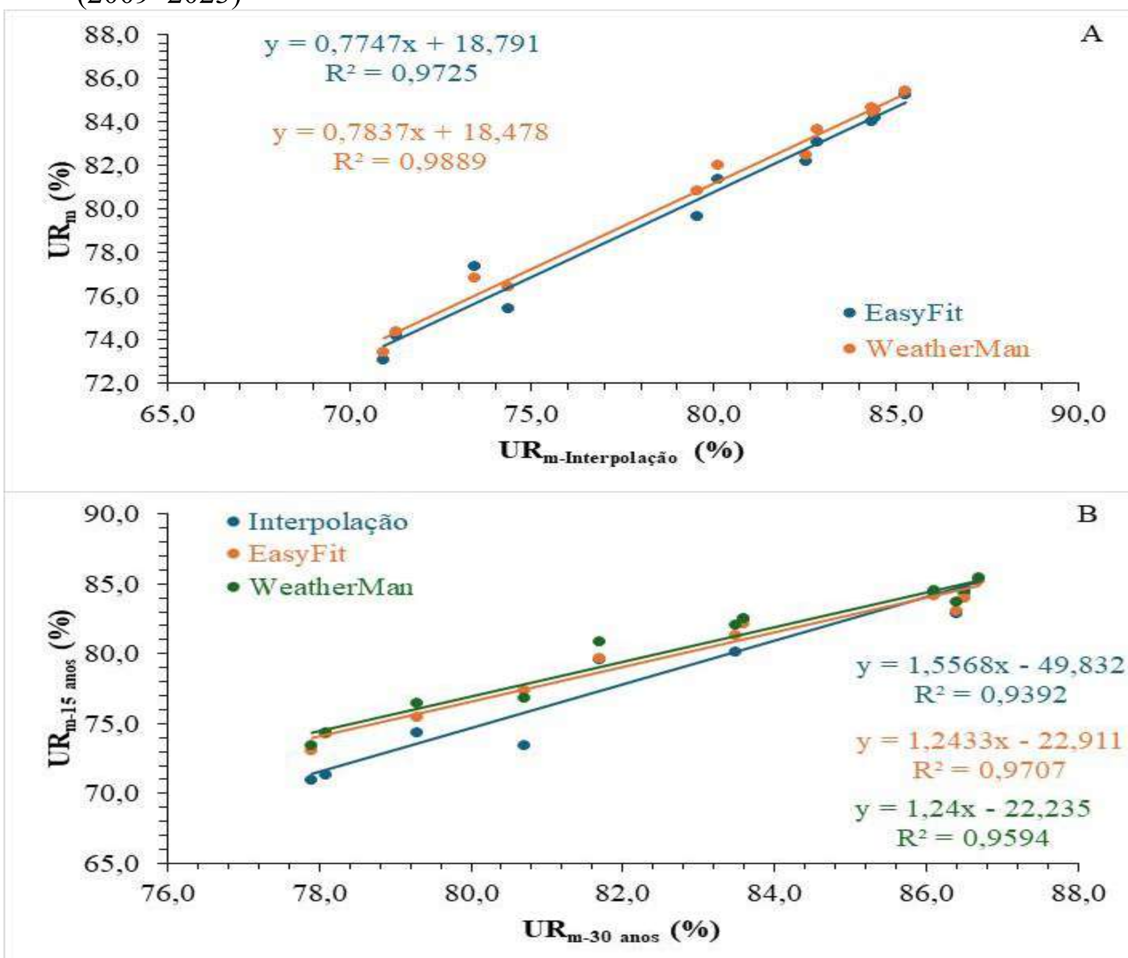
The average monthly relative humidity (Figure 6A), which was determined using the linear interpolation method, was correlated with data from Easy-Fit and WeatherMan/DSSAT, with R^2 values varying from 0.9725 to 0.9889. An analysis of Figure 6B, which illustrates the correlation between the climatological normal data from 1991 to 2020 and the provisional climatological normal data from 2009 to 2023, reveals that the methods present R^2 variations from 0.9392 to 0.9707. In Table 7, it is noted that the generators presented a p value < 0.05, allowing us to affirm that all the methods presented excellent fits and could be adopted.

Table 7. Regression statistics for the analysis of relative air humidity (RH $_{m}$) and monthly average obtained by the simple linear regression method.

Method		UR $_{m\text{-}15\ years}$		
		R ^{2} - $_{EasyFit}$	Stat t	p value
UR $_{m\text{-}30\ years}$	Interpolation	0.9392	-4.79	0.000
	EasyFit	0.9707	-4.04	0.000
	WeatherMan	0.9594	-3.32	0.000

Source: Authors (2025)

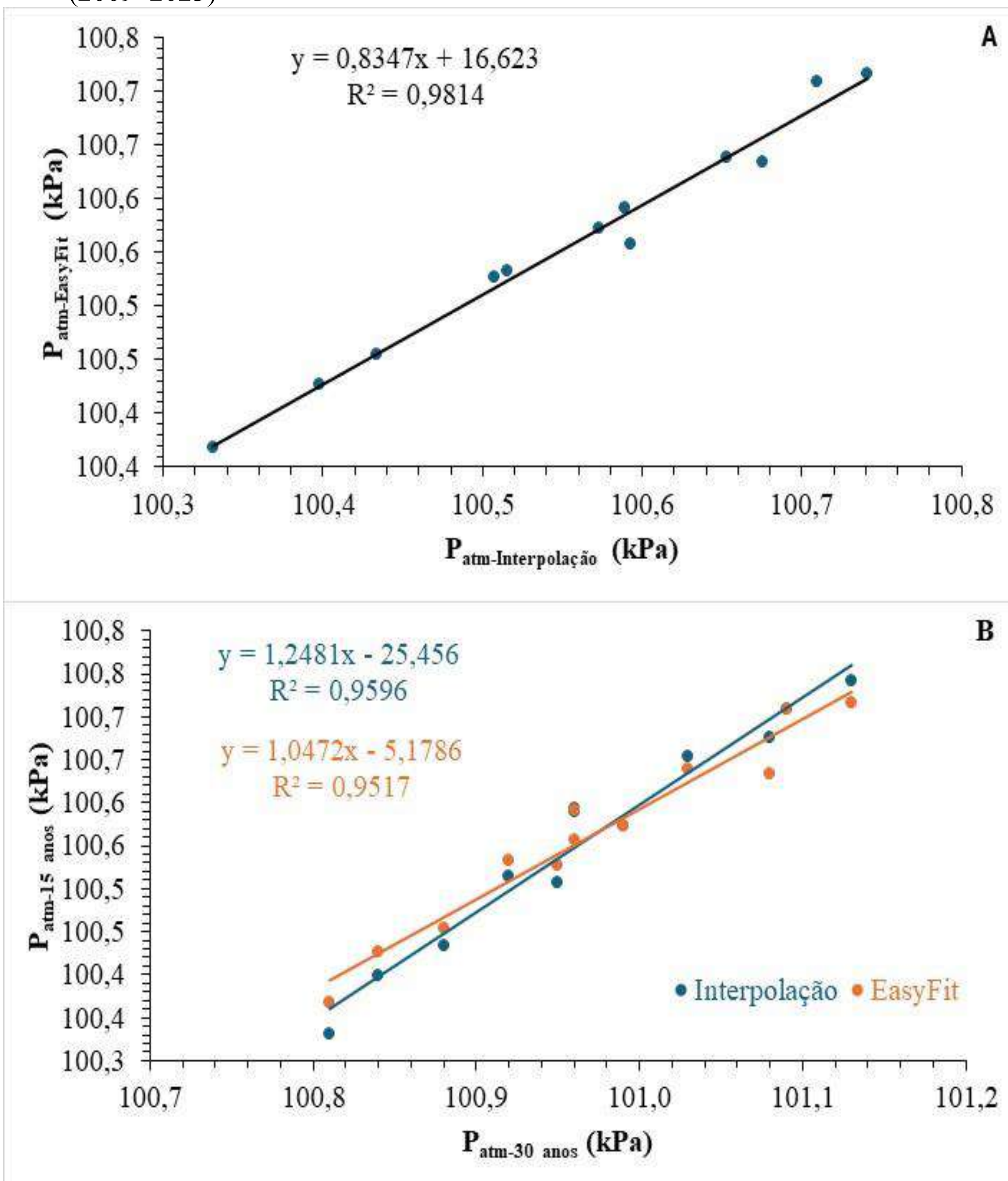
Figure 6. A – Performance of the methods for filling in the average monthly relative humidity; B – Comparison between the climatological normal (1991–2020) and the filled-in series (2009–2023)



Source: Authors (2025)

The average monthly atmospheric pressure variable (Figure 7A), when correlated, had an R^2 of 0.9814, demonstrating an excellent fit between the methods. The normal data from 1991 to 2020 were correlated with the provisional climatological normal data, which were filled in by linear interpolation and easy fit (Figure 7B), with the coefficient of determination varying between 0.9517 and 0.9596. As shown in Table 8, when the p value is < 0.05, the hypothesis H0 for both interpolation and easy fit is rejected, indicating that regression exists and that both methods can be adopted to fill in missing atmospheric pressure data.

Figure 7. A – Comparative performance between interpolation and easy-fit methods in terms of estimating monthly atmospheric pressure; B – Comparison between the climatological normal (1991–2020) and the filled data of average monthly atmospheric pressure (2009–2023)



Source: Authors (2025)

Table 8. Regression statistics for atmospheric pressure (P_{atm}) analysis obtained by the simple linear regression method.

Method		$P_{atm\text{-}15\ years}$		
		R^2 - EasyFit	**Stat t**	**p value**
$P_{atm\text{-}30\ years}$	Interpolation	0.9596	-3.11	0.000
	EasyFit	0.9517	-0.69	0.000

Source: Authors (2025)

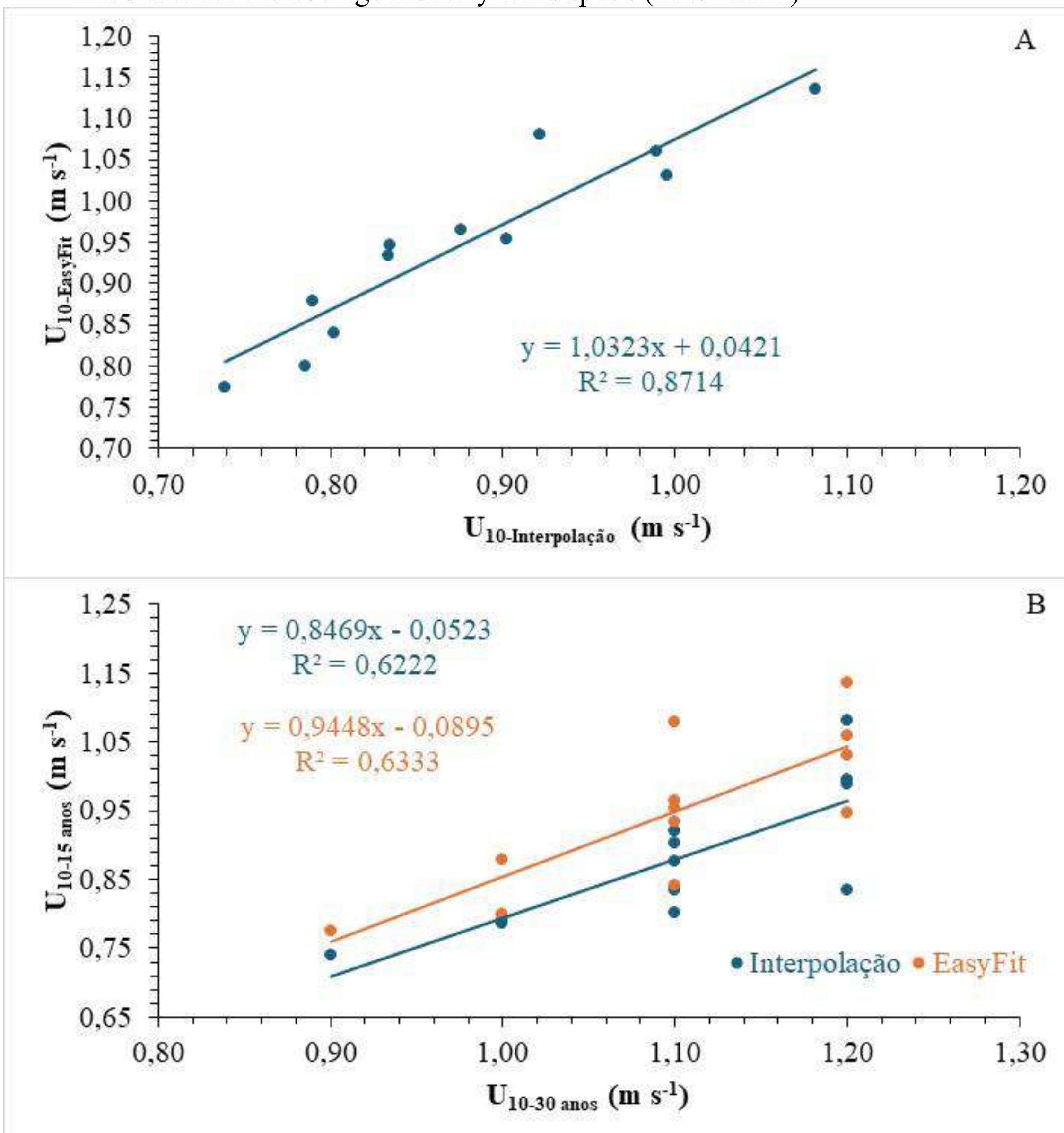
When correlated, the average monthly wind speed (Figure 8A), which was obtained using the linear interpolation method and the Easy-Fit software, had an R^2 of 0.8714. An analysis of the data in Figure 8B, which correlates the 30-year climatological normal data (1991 to 2020) with the provisional 15-year climatological normal data (2009 to 2023), revealed that both methods fit well, with R^2 values ranging from 0.6222 to 0.6333. See in Table 9 that the p value < 0.05, rejecting hypothesis H0, which shows that regression exists; therefore, its use can be recommended to fill in the missing data gaps.

Table 9. Regression statistics for wind speed analysis (U10) obtained by the simple linear regression method.

Method		$U_{\text{-15 years}}$		
		R^2 - EasyFit	**Stat t**	**p value**
$U_{\text{-30 years}}$	Interpolation	0.62	-0.23	0.002
	EasyFit	0.63	-0.36	0.002

Source: Authors (2025)

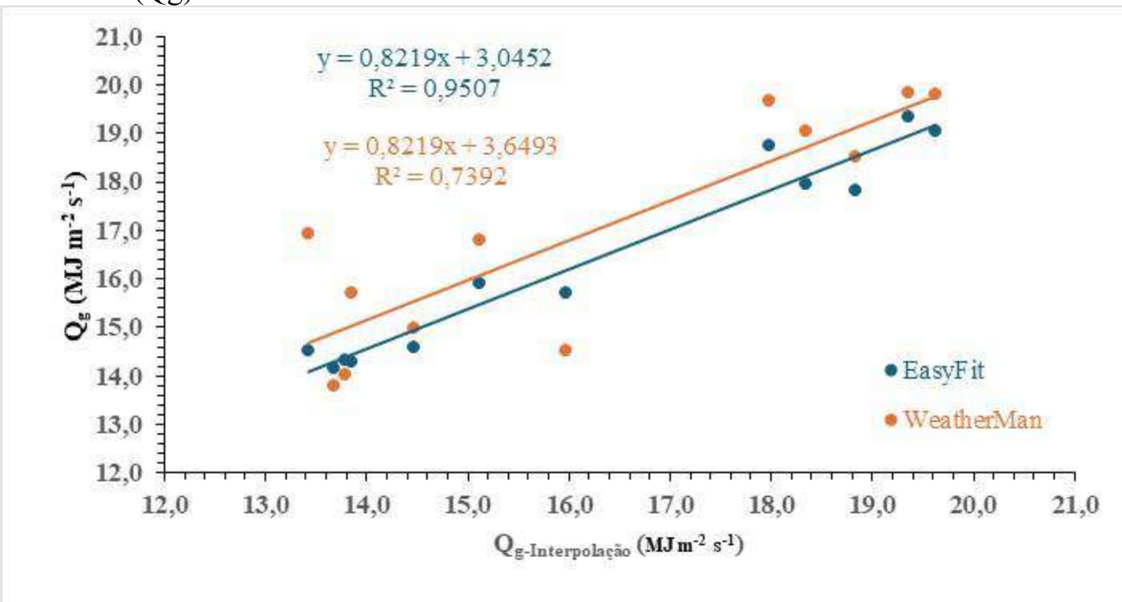
Figure 8. A – Linear fit between the interpolation and easy-fit methods for the average monthly wind speed; B – Comparison between the climatological normal (1991–2020) and the filled data for the average monthly wind speed (2009–2023)



Source: Authors (2025)

The average monthly global solar radiation (Figure 9), obtained using the linear interpolation method , was correlated with data from the Easy-Fit software and WeatherMan/DSSAT. Compared with the WeatherMan/DSSAT data, the easy-fit data better fit the interpolation data, with an R^2 of 0.9507. The evaluated methods presented p values < 0.05 (Table 10), showing that regression exists and can be used.

Figure 9. Correlation between gap-filling methods for average monthly global solar radiation (Qg).



Source: Authors (2025)

Table 10. Regression statistics for the analysis of global solar radiation (Qg) obtained by the simple linear regression method.

Method		**$Q_{g\text{-15 years}}$ (EasyFit)**		
		R^2	**Stat t**	**p value**
$Q_{g\text{-15 years}}$	Interpolation	0.9507	3.14	0.000
		$Q_{g\text{-15 years}}$ (WeatherMan)		
$Q_{g\text{-15 years}}$	Interpolation	0.7392	1.44	0.000

Source: Authors (2025)

The regression analyses of the quantitative variables measured on the same individuals were well fitted, since the coefficient of determination ranged between 62.22% (monthly average wind speed) and 99.70% (monthly maximum air temperature), thus reflecting the phenomenon studied. In this sense, the analysis of variance of the regression applied to the monthly average wind speed presented a p value of 0.002, which was lower than the probability level p = 0.05, demonstrating that the regression exists; that is, the 30-year climatological normal explains the provisional 15-year climatological normal, which was completed by the linear interpolation method.

Notably, Meinke *et al. (1995), when evaluating three different models for generating meteorological data, indicated that WGEN is the most suitable for generating long series of temperature and solar radiation data, providing good estimates of potential productivity for maize, wheat, and sunflower crops under Australian climatic conditions. In this research, for maximum and minimum air temperatures and solar radiation,* Meinke's findings were not confirmed. *et al.* . (1995), with the best data generators being interpolation, WeatherMan, and easy-fit, respectively. According to Hayhoe (1998), the WGEN version adapted for the EPIC (Erosion/Productivity Impact) program is not a good option for generating climate data because it does not maintain the relationships between meteorological variables because of the simplifications assumed in the program.

6 CONCLUSION

Under the conditions in which this work was carried out, the following conclusions can be drawn:

- In the literature consulted, no validation of filling gaps in meteorological series similar to this one was found because of the particularities of the region investigated. However, it was envisioned that a robust series could be generated that could provide applications in various segments focused on decision-making in studies of soil, plants, and the local atmosphere, as well as guaranteeing the stability of modeling aimed at local agriculture, given the gaps in the current series.
- The easy-fit and WeatherMan/DSSAT interpolation methods have shown potential for filling gaps in monthly meteorological series from the Itacoatiara-AM region, even in the face of limited availability of complete data.
- The interpolation method performed best for the average monthly maximum temperature and average monthly atmospheric pressure, whereas the easy-fit method yielded better fits for the average monthly minimum temperature, average monthly relative humidity, and average wind speed.
- For the average wind speed (p value = 0.02), the interpolation method and the easy-fit generator can also be suggested for generating meteorological data.
- Considering the data gaps in the meteorological stations in the middle and lower Amazon region and the impossibility of filling them with other more sophisticated methods, it is possible to use the generators discussed in this investigation; however, greater rigor should be employed by the researcher, thus avoiding errors in data collection at the origin station.

7 REFERENCES

ALLEN, RJ; DEGAETANO, AT Estimating Missing Daily Temperature Extremes Using an Optimized Regression Approach. **International Journal of Climatology** , United Kingdom, v. 21, n. 11, p. 1305-1319, 2001. DOI: https://doi.org/10.1002/joc.679. Available at: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.679. Accessed on: July 1, 2026.

BERTONI, JC; TUCCI, CEM Precipitation . *In* : Tucci, CEM (org.). **Hydrology** : science and application . 4th ed. Porto Alegre: UFRGS, 2007. p. 177-241.

BRISTOW, KL; CAMPBELL, GS On the relationship between incoming solar radiation and daily maximum and minimum temperature . **Agricultural and Forest Meteorology** , United Kingdom, v. 31, p. 159-166, 1984.

CHIBANA, EY; BRITO, LR de S.; FLUMIGNAN, D.; MOTA, RG; VIEIRA, A. de S.; FARIA, RT de. Filling in gaps in climate data. *In* : XIV BRAZILIAN CONGRESS OF AGROMETEOROLOGY, 14th, 2005, Campinas. **Proceedings** [...]. Campinas: SBA, 2005.

DIAZ, CCF; PEREIRA, JAS; NÓBREGA, RS. Comparison of data estimated by two different methods for filling rainfall gaps in the Pajeú River basin, Pernambuco, Brazil. **Brazilian Journal of Climatology** , Curitiba, v. 22, p. 324-339, 2018.

FILL, HD Hydrological information. *In* : TUCCI, CEM (org.). **Hydrology** : science and application. Porto Alegre: ABRH, 1987. p. 95-202.

FANTE, KP; SANT'ANNA NETO, JL. Statistical techniques for the homogenization of temperature data in climatological time series. **Brazilian Journal of Climatology** , Curitiba, year 12, v. 18, p. 140-156, 2016. DOI: https://doi.org/10.5380/abclima.v18i0.43202. Available at: https://revistas.ufpr.br/revistaabclima/article/view/43202. Accessed on: Oct. 16, 2025.

GIONANELLA, THG; OLIVEIRA, FC de; MARCHI, VA de A.; TLUSZCZ, J. Performance of Gap Filling Methods in Reference Evapotranspiration Data for the Western Region of Paraná. **Revista Brasileira de Meteorologia** , São José dos Campos, v. 36, n. 3, p. 415-422, 2021. DOI: http://dx.doi.org/10.1590/0102-77863630001. Available at: https://www.scielo.br/j/rbmet/a/ySBfCn8CbbRLQfnLWCXbwDd/?lang=pt&format=pdf. Accessed on: July 1, 2026.

GUZMÁN, NG. **Modeling for estimating missing data in meteorological data series** . 2018. Dissertation (Master's in Computational Modeling) – Polytechnic Institute, State University of Rio de Janeiro, Nova Friburgo, 2018.

HAYHOE, HN Relationship between weather variables observed and WXGEN generated data series. **Agricultural and Forest Meteorology** , Amsterdam , v. 90, p. 203-214, 1998.

HOOGENBOOM, G.; JONES, JW; BOOTE, KJ Modeling growth , development , and yield of grain legumes using Soygro , Pnutgro , and Beangro . **Transaction of the ASAE** , Michigan, v. 35, n. 6, p. 2043-2056, 1992. Available at: https://elibrary.asabe.org/abstract.asp?JID=3&AID=28833. Accessed on: July 2, 2026.

HOOGENBOOM, G.; PORTER, CH; SHELIA, V.; BOOTE, KJ; SINGH, U.; PAVAN, W.; OLIVEIRA, FAA; MORENO-CADENA, LP; FERREIRA, TB; WHITE, JW; LIZASO, JI; SMALL, DNL; KIMBALL, BA; ALDERMAN, PD; THORP, K.R.; CUADRA, SV; VIANNA, MS; VILLALOBOS, FJ; BATCHELOR, W.D.; ASSENG, S.; JONES, MR; HOPF, A.; DIAS, HB; JINTRAWET, A.; JAIKLA, R.; MEMIC, E.; HUNT, LA; JONES, JW Decision Support System for Agrotechnology Transfer (DSSAT). Version 4.8.5.0. Gainesville: DSSAT Foundation, 2024. Available at: www.dssat.net. Accessed on: January 27, 2025.

INMET. **Climatological Normals of Brazil** . Brasília, DF: internet, 2025. Available at: https://portal.inmet.gov.br/normais. Accessed on: July 20, 2025.

KEMP, W.P.; BURNELL, DG; EVERSON, DO; THOMSON, AJ Estimating Missing Daily Maximum and Minimum Temperatures Journal **of Climate and Applied Meteorology** , Boston, vol. 22, no. 9, p. 1587-1593, 1983.

KRUSCHE, N.; SARAIVA, MBJ; REBOITA, MS. **Provisional climatological normals for Rio Grande, RS, from 1991 to 2000.** 1st ed. Rio Grande: FURG, 2002.

MEINKE, H.; CARBERRY, PS; McCASKILL , MR; HILLS, MA; McLEOD , I. Evaluation of radiation and temperature data generators in the Australian tropics and subtropics using crop simulation models. **Agricultural and Forest Meteorology** , Amsterdam , v. 72, p. 295-316, 1995.

NASCIMENTO, TS do; SARAIVA, JMB; SENNA, R.; AGUIAR, FEO Filling gaps in rainfall databases based on CPC (climate) data (Prediction Center): a case study of the Solimões River - Amazonas. **Brazilian Journal of Climatology** , São José dos Campos, year 6, v. 7, p. 143-158, 2009. Available at: https://ojs.ufgd.edu.br/index.php/rbclima/article/view/13578#:~:text=Este%20trabalho%20tem%20como%20objetivo%20preencher%20falhas%20em, com%20dados%20 sint%C3%A9ticosadquiridos%20do%20CPC%20%28Climate%20Prediction%20Center%29. Accessed on: March 11, 2025.

OLIVEIRA, LFC; FIOREZE, AP; MEDEIROS, AMM; SILVA, MAS. Comparison of methodologies for filling gaps in historical series of annual rainfall. **Revista Brasileira de Engenharia Agrícola e Ambiental** , Campina Grande, v. 14, n. 11, p. 1186-1192, 2010.

WORLD METEOROLOGICAL ORGANIZATION. **Guidelines World Meteorological Organization on the calculation of Climate normals** . Geneva : WMO, 2017. (WMO, no. 1203). Available at: https://library.wmo.int/viewer/60338?medianame=1203_es_#page=1&viewer=picture&o=bookmark&n=0&q=. Accessed on: April 7, 2024.

PICKERING, NB; HANSEN, JW; JONES, JW; WELLS, CM; CHAN, VK; GODWIN, D.C. WeatherMan : A Utility for Managing and Generating Daily Weather Data. **Agronomy Journal** , Washington, DC, v. 86, n. 2, p. 332-337, Mar. 1994. DOI: http://dx.doi.org/10.2134/agronj1994.00021962008600020023x. Available at: https://acsess.onlinelibrary.wiley.com/doi/10.2134/agronj1994.00021962008600020023x. Accessed on: May 27, 2026.

RICHARDSON, C.W.; WRIGHT, FROM **WGEN** : a model for generating daily weather variable . Washington, DC: USDA-ARS. 1984. (ARS-8).

RITCHIE, J.T.; GODWIN, DC; SINGH, U. Soil and weather inputs for the IBSNAT crop models. *In* : IBSNAT SYMPOSIUM: DECISION SUPPORT SYSTEMS FOR AGROTECHNOLOGY TRANSFER, 1989, Honolulu. **Proceedings** [...]. Honolulu: University of Hawaii, 1989. p. 31-45. Available at: https://www.researchgate.net/publication/312737550_Soil_and_weather_inputs_for_the_IBSNAT_crop_models. Accessed on: August 11, 2025.

SANTOS, RS dos. **Homogeneity and reconstruction of climatological series for localities in the State of Minas Gerais** . 2012. Thesis (Doctorate in Agricultural Meteorology) – Federal University of Viçosa, Viçosa, 2012.

WMO **Calculation of monthly and annual 30-year standard normals** . Geneva: WMO, 1989. (WCDP, n. 10/WMO-TD, n. 341).