FILLING GAPS IN HISTORICAL RAINFALL AND AIR TEMPERATURE HISTORICAL SERIES IN THE STATES OF PERNAMBUCO AND ALAGOAS

Carla Sabrina da Silva[1], Ceres Duarte Guedes Cabral de Almeida[2], André dos Anjos Correia[1], Caio Sérgio Pereira de Araújo[3], Weslley Amaro da Silva[3], Géssica de Paula Alves Marinho[3]

ABSTRACT: The availability of historical climate series is essential for planning and decision-making in agricultural activities. This study aimed to compare the efficiency of four methods of filling gaps in historical series of rainfall and air temperature data. Data from 12 automatic meteorological stations (AWS) in the state of Pernambuco and six in the state of Alagoas belonging to the National Institute of Meteorology (INMET) were excluded so that the filling techniques could be applied and evaluated. The database was composed of monthly values of accumulated rainfall and average air temperature for the period 01/01/2011 to 12/31/2020. The data absence simulation was standardized to two months in summer and two months in winter for each year and locality. The results indicated that Regional Weighting was the best technique for filling gaps in mean air temperature and rainfall data in the states of Alagoas and Pernambuco. The Multiple Linear Regression method presented the worst result in the state of Alagoas however, it was the second method with the lowest error rate in the state of Pernambuco.

KEYWORDS: Agrometeorology, Regional Weighting, Multiple Linear Regression.

PREENCHIMENTO DE FALHAS EM SÉRIESHISTÓRICAS DE PRECIPITAÇÃOPLUVIOMÉTRICA E TEMPERATURA DO AR NOS ESTADOS DE PERNAMBUCO E ALAGOAS

RESUMO: A disponibilidade de series climáticas históricas é essencial para o planejamento e para a tomada de decisão nas atividades agrícolas. Este estudo objetivou comparar a

[1] Master student of Agricultural Engineering, Federal Rural University of Pernambuco (UFRPE), St Dom Manuel de Medeiros, s/n, Dois Irmãos, CEP 52171-900, Recife, PE. Fone (82) 996092190. e-mail: carlaabrina18@gmail.com
[2] Full Professor, AgriculturalCollege Dom Agostinho Ikas, CODAI, UFRPE, São Lourenço da Mata, PE
[3] PhD student of Agricultural Engineering, UFRPE, Recife, PE

proficuidade de quatro métodos de preenchimento de falhas em series históricas de dados pluviométricos e de temperatura do ar. Os dados de 12 estações meteorológicas automáticas (AWS) no estado de Pernambuco e seis no estado de Alagoas pertencentes ao Instituto Nacional de Meteorologia (INMET) foram excluídos, de modo que as técnicas de preenchimento, em estudo, fossem aplicadas e avaliadas. O banco de dados foi composto por valores mensais de precipitação pluvial acumulada e temperatura média do ano período de 01/01/2011 a 31/12/2020. A simulação de ausência de dados foi padronizada em dois meses no verão e dois meses no inverno para cada ano e localidade. Os resultados indicaram que Ponderação regional foi a melhor técnica para preenchimento de falhas nos dados de temperatura média do ar e precipitação pluviométrica nos estados de Alagoas e Pernambuco. O método de Regressão linear múltipla apresentou o pior resultado no estado de Alagoas, porém, foi o segundo método com menor índice de erro no estado de Pernambuco.

PALAVRAS-CHAVE: Agrometeorologia, Ponderação regional, Regressão Linear Múltipla.

INTRODUCTION

The availability of historical climate series is very important for planning and decision-making about agricultural activities since they are significantly affected at different stages of the production chain. To study the climate variability of a given region, a representative database is necessary, in which the observed values must be accurate to describe the study site precisely (ELY et al., 2021).

Climatic factors, especially rainfall and air temperature, directly influence plant growth, the total crop yield, the occurrence of pests and diseases, and the need for water and nutrients, among other agricultural activities from planting to post-harvest (BERGAMASCHI & MATZENAUER, 2014).

Rainfall precipitation is a component that presents greater variability in space and time (DIAZ et al., 2018) since it is a random natural phenomenon, with irregular distribution on the ground (MELLO et al., 2017). The knowledge of this information allows the preparation of studies on the use of water for agricultural activities. Air temperature, in turn, is an important meteorological variable for weather forecasting and climate studies (YAGUCHI et al., 2016). Historical temperature series are widely used however, they have a lack of data that prevents them from being used in some studies.

The lack of data measured in meteorological stations is a recurring problem, which can be generated by equipment error or human error, depending on whether the station is automatic or conventional (BIER & FERRAZ, 2017). Several methodologies for filling gaps in climate data are available. For data on rainfall and air temperature, the methods of Multiple Linear Regression (MLR), Simple Linear Regression (SLR), Arithmetic Mean (AM), and Regional Weighting (RW) are judged to be the most efficient among the various statistical methods of data estimates (BIER & FERRAZ, 2017; SILVA et al., 2021; JUNQUEIRA et al., 2018).

Thus, the objective of this study was to compare the effectiveness of four gap-filling methods in historical series of rainfall and air temperature data from 18 automatic meteorological stations (AWS) distributed in the states of Pernambuco and Alagoas, northeast of Brazil, in the period from 01/01/2011 to 12/31/2020.

MATERIAL AND METHODS

To evaluate the efficiency of the four methods of gap filling, some data were excluded from the historical series of 12 automatic weather stations in the state of Pernambuco and six in the state of Alagoas, all belonging to the network of the National Institute of Meteorology - INMET.

The database used has monthly values of accumulated rainfall and average air temperature for the period January 2011 to December 2020. The data absence simulation was standardized in two months in summer and two months in winter for each year and location. The calculations of the statistical index and estimation of variables, as well as the preparation of graphs, were performed in Microsoft Excel® software.

Figure 1 shows the spatial distribution of the AWS located in the states of Alagoas (Arapiraca, Coruripe, Maceió, Palmeira dos Índios, Pão de Açúcar and São Luís do Quitunde) and Pernambuco (Arcoverde, Cabrobró, Caruaru, Floresta, Garanhuns, Ibimirim, Ouricuri, Palmares, Petrolina, Recife, Serra Talhada e Surubim).

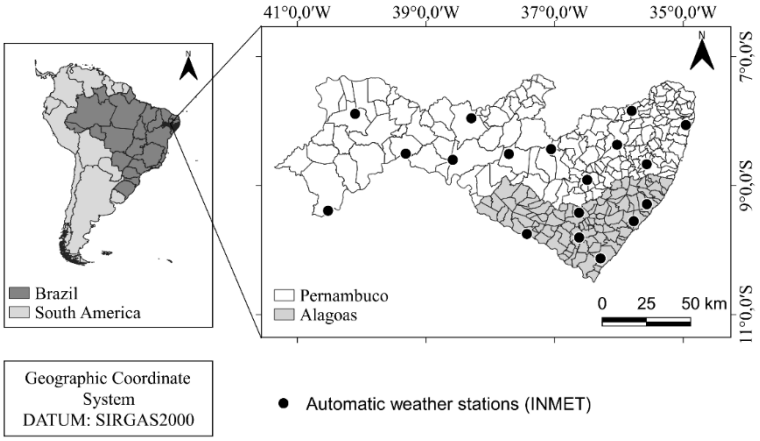

Figure 1. Geographical distribution of automatic weather stations in the states of Alagoas and Pernambuco. Source: INMET (2022).

Table 1 provides information about the meteorological stations studied, including the respective Word Meteorological Organization (WMO) codes, geographic data, and climate classification.

Table 1. Characterization of automatic weather stations (AWS), considered in this study, located in the states of Alagoas and Pernambuco.

City – Code WMO	Coordinates		Elevation (m)	Climate Classification
	Latitude	Longitude		
ALAGOAS				
Arapiraca - A353	-10° 11 ' 44"	-37° 22 ' 51"	236.84	As
Coruripe - A355	-11° 52 ' 17"	-37° 42 ' 49"	82.39	As
Maceió - A303	-10° 26 ' 56"	-36° 13 ' 47"	84.12	As
Palmeira dos Índios - A327	-10° 34 ' 47"	-37° 22 ' 47"	278.01	As
Pão de Açúcar - A323	-10° 15 ' 3"	-38° 34 ' 9"	20.83	BSh
São Luís do Quitunde - A356	-10° 42 ' 45"	-36° 26 ' 3"	13.69	As
PERNAMBUCO				
Arcoverde - A309	-9° 33 ' 59"	-38° 56 ' 40"	683.95	Aw
Cabrobó - A329	-9° 29 ' 46"	-40° 41 ' 5"	342.74	BSh
Caruaru - A341	-9° 38 ' 5"	-37° 58 ' 18"	852.00	Aw
Floresta - A351	-9° 24 ' 4"	-39° 24 ' 57"	327.42	BSh
Garanhuns - A322	-9° 5 ' 21"	-37° 30 ' 24"	827.78	Aw
Ibimirim - A349	-9° 29 ' 26"	-38° 17 ' 18"	434.23	BSh
Ouricuri - A366	-8° 6 ' 51"	-41° 53 ' 50"	457.85	BSh
Palmares - A357	-9° 20 ' 0"	-36° 25 ' 55"	164.01	Aw
Petrolina - A307	-10° 36 ' 42"	-41° 28 ' 36"	372.72	BSh
Recife - A301	-9° 56 ' 27"	-35° 2 ' 27"	11.30	Am
Serra Talhada - A350	-8° 2 ' 45"	-39° 42 ' 18"	499.02	BSh
Surubim - A328	-8° 9 ' 37"	-36° 11 ' 56"	421.44	Aw

WMOcode - World Meteorological Organization; Am: humid or subhumid tropical climate; As: tropical and hot with autumn/winter rains; Aw: tropical climate with dry winter; BSh: hot semi-arid climate. Source: INMET (2022); Hugo et al. (2012); Alvares et al. (2013).

Table 2 presents the stations used to fill in faults, according to the requirement of the methods applied. The criterion for choosing the stations to be used in the filling was the proximity between them.

Table 2. Stations used for applying fault fills in each method.

Number	City	Stations			
		RW	AM	SLR	MLR
		ALAGOAS			
1	Arapiraca	2, 4, 5	2, 4, 5	4	2, 4
2	Coruripe	1, 3, 4	1, 3, 4	3	1, 3
3	Maceió	2, 4, 6	2, 4, 6	6	2, 6
4	Palmeira dos Índios	1, 2, 6	1, 2, 6	9	1, 2
5	Pão de Açúcar	1, 4, 4	1, 4, 4	2	1, 2
6	São Luís do Quitunde	2, 3, 4	2, 3, 4	3	3, 4
		PERNAMBUCO			
7	Arcoverde	9, 11, 12	9, 11, 12	12	11,12
8	Cabrobó	10, 13, 17	10, 13, 17	10	10, 13
9	Caruaru	11, 14, 18	11, 14, 18	14	14, 18
10	Floresta	7, 8, 17	7, 8, 17	8	8, 17
11	Garanhuns	4, 7, 9	4, 7, 9	4	4, 9
12	Ibimirim	7, 10, 17	7, 10, 17	7	10, 17
13	Ouricuri	8, 10, 15	8, 10, 15	8	8, 15
14	Palmares	6, 9, 18	6, 9, 18	9	6, 9
15	Petrolina	8, 10, 13	8, 10, 13	8	9, 13
16	Recife	9, 14, 18	9, 14, 18	14	13, 14
17	Serra Talhada	8, 10, 12	8, 10, 12	10	10, 12
18	Surubim	9, 14, 16	9, 14, 16	9	9, 16

RW-Regional Weighting; AM-Arithmetic Mean; SLR-Simple Linear Regression; and MLP-Multiple Linear Regression.

The methods of Arithmetic Mean (AM), Simple Linear Regression (SLR), Multiple Linear Regression (MLR), and Regional Weighting (RW) were evaluated. To analyze their efficiency the mean error (ME) and the mean percentage error (MPE) were measured, which quantify underestimates and overestimates, as well as the mean absolute error (MAE) and the square root of the mean error (RMSE), which are measures of the precision of the estimates.

RESULTS AND DISCUSSION

The lowest statistical errors were found in the estimation of the air temperature in all AWS, independently of the filling methods used. This result is explained by the fact that the temperature had a lower percentage of failures, which resulted in increased accuracy in the estimate.

Table 3 presents the available data and their respective percentage of failures in the AWS evaluated, in which there is a greater amount of data available in the AWS of the state of Pernambuco and consequently lower percentages of failures.

The Regional Weighting (RW) method represented more satisfactorily the actual data of mean air temperature and rainfall (Tables 4 and 5) for the data collected in the AWS of

Pernambuco and Alagoas with mean absolute errors (MAE) of 0.33 and 0.14°; and 26.26 and 6.53 mm, respectively. Silva et al. (2021) corroborate this result, in which the estimated rainfall data were similar or closer to the actual data estimated by the Regional Weighting method. The MAE values in the estimates of the RW method are lower in the AWS of Pernambuco due to the greater amount of data available.

Table 3. Quantitative of available data and respective percentage of failures in the AWS studied.

City – Code WMO	Available data		Failures (%)	
	Rainfall	Temperature	Rainfall	Temperature
ALAGOAS				
Arapiraca - A353	60,801	72,094	30.65	17.77
Coruripe - A355	59,703	83,958	31.90	4.24
Maceió - A303	79,578	81,776	9.23	6.73
Palmeira dos Índios - A327	75,112	82,955	14.33	5.38
Pão de Açúcar - A323	73,504	76,927	16.16	12.26
São Luís do Quitunde - A356	76,763	80,174	12.44	8.55
PERNAMBUCO				
Arcoverde - A309	73,590	84,461	16.06	3.66
Cabrobó - A329	83,708	83,989	4.52	4.20
Caruaru - A341	75,112	81,077	14.33	7.52
Floresta - A351	82,494	83,988	5.91	4.20
Garanhuns - A322	83,127	83,136	5.18	5.17
Ibimirim - A349	81,210	85,464	7.37	2.52
Ouricuri - A366	71,837	77,450	18.06	11.66
Palmares - A357	64,782	79,517	26.11	9.30
Petrolina - A307	84,425	85,273	3.70	2.74
Recife - A301	85,883	85,885	2.04	2.04
Serra Talhada - A350	81,327	84,882	7.24	3.18
Surubim - A328	69,730	81,821	20.46	6.67

The arithmetic average in the state of Alagoas was presented as the second method, behind the Regional Weighting method, with lower statistical errors for the temperature variable (MAE, 0.78° and RMSE, 1.82°) and for the variable rainfall (MAE, 27.88 mm and RMSE, 69.27 mm), these errors are greater than those that showed the worst performance in the state of Pernambuco, Simple Linear Regression and Arithmetic Mean.

The Multiple Linear Regression (MLR) method had the lowest performance for the state of Alagoas, while in the state of Pernambuco, it proved to be the second most efficient method. In the state of Alagoas, the MLR method presents a maximum value of MAE of 27.8° of air temperature and 296.84 mm of rainfall (Table 4), while the in the state of Pernambuco, 2.52° and 33.77 mm (Table 5), respectively. According to Ruezzene et al. (2020), techniques used to fill gaps in precipitation database, in some cases, are not as efficient due to the low correlation of the data available at neighboring stations, thus generating very different estimates of the original series and, consequently, greater errors in the results, over or underestimating them.

The Arithmetic Mean method was the least accurate in the estimates of air temperature and rainfall in the AWS of Pernambuco with average MAE values of 0.78° and 31.65 mm, respectively (Table 5).

The Simple Linear Regression method had the second worst performance in the accuracy of rainfall estimates in the data of the two states studied with averages of 48.01 mm (Alagoas) and 21.93 mm (Pernambuco). Silva et al. (2021) performed tests in the state of Pernambuco and concluded that the Simple Linear Regression showed better efficiency in estimating missing data of average air temperature and rainfall, however, the adopted period of one year can corroborate the divergence of result since the present study considered 10 years.

Table 4. Maximum, minimum, and absolute mean values of the statistical indexes calculated for each filling method evaluated in this study with data from the AWS in the state of Alagoas.

Parameters	REGIONAL WEIGHTING							
	ME	MPE	MAE	RMSE	ME	MPE	MAE	RMSE
	Average air temperature				Rainfall			
Maximum	1.02	4.07	0.33	3.28	25.93	19.39	26.26	115.29
Minimum	-0.26	-1.04	0.05	0.53	-5.12	-8.22	1.09	29.47
Absolute Average	0.13	0.52	0.29	1.44	4.50	4.03	8.02	62.75
	ARITHMETIC MEAN							
Maximum	1.35	5.65	1.35	3.78	40.48	112.34	74.90	114.66
Minimum	-1.11	-4.40	0.10	0.68	-74.90	-45.63	13.99	43.71
Absolute Average	0.09	0.42	0.78	1.82	5.65	17.81	27.88	69.27
	SIMPLE LINEAR REGRESSION							
Maximum	2.60	10.70	3.60	5.21	7.13	113.57	111.47	143.72
Minimum	-3.60	-14.62	1.04	1.34	-55.58	-20.85	12.31	36.76
Absolute Average	0.74	3.13	2.09	2.87	25.38	17.86	48.01	78.02
	MULTIPLE LINEAR REGRESSION							
Maximum	24.33	0.04	27.80	89.78	52.52	19.68	296.84	1088.11
Minimum	-1.74	0.00	1.05	3.02	-253.60	0.06	43.97	85.28
Absolute Average	5.76	0.01	7.27	20.77	60.63	1.99	100.26	278.36

ME: Mean error; MPE: Mean percentage error; MAE: Mean Absolute error; and RMSE: Square root of the mean error.

The Regional Weighting method was the most efficient, followed by the arithmetic mean in the state of Alagoas presenting the lowest statistical errors for the temperature variable and for the rainfall variable. In the same way, the Regional Weighting method was the most efficient to fill gaps in Pernambuco's database. Therefore, the Multiple Linear Regression method, in general, can be considered the second most efficient method, due to the state of Pernambuco having a greater number of AWS and a lower percentage of failures in the stations. Several studies have indicated the Multiple Linear Regression method as the most advantageous in estimating rainfall data (OLIVEIRA et al., 2010; MELLO et al., 2017; SILVA & JARDIM, 2017; JUNQUEIRA et al., 2018).

Table 5. Maximum, minimum, and absolute mean values of the statistical indexes calculated for each filling method evaluated in this study with data from the AWS in the state of Pernambuco.

Parameters	REGIONAL WEIGHTING							
	ME	MPE	MAE	RMSE	ME	MPE	MAE	RMSE
	Average air temperature				Rainfall			
Maximum	0.07	0.24	0.14	0.90	6.17	13.04	6.53	89.66
Minimum	-0.10	-0.40	0.00	0.19	-3.19	-2.55	0.00	10.51
Absolute Average	0.01	0.05	0.03	0.50	0.50	2.36	2.03	31.30
	ARITHMETIC MEAN							
Maximum	2.48	11.84	2.48	2.53	71.11	193.74	125.40	140.48
Minimum	-1.90	-7.52	0.12	0.37	-125.40	-65.07	4.33	16.64
Absolute Average	0.07	0.04	0.78	1.00	6.64	40.54	31.65	47.95
	SIMPLE LINEAR REGRESSION							
Maximum	0.20	0.78	2.19	3.47	16.13	161.38	68.52	103.87
Minimum	-2.19	-7.98	0.04	0.43	-13.11	18.83	1.43	22.88
Absolute Average	0.38	1.46	0.63	1.39	2.64	71.11	21.93	46.53
	MULTIPLE LINEAR REGRESSION							
Maximum	0.59	2.12	0.88	2.52	8.57	64.90	33.77	107.61
Minimum	-0.55	-2.30	0.04	0.26	-31.86	-20.10	2.06	17.37
Absolute Average	0.19	0.79	0.26	0.95	5.55	18.01	7.73	41.36

ME: Mean error; MPE: Mean percentage error; MAE: Mean Absolute error; and RMSE: Square root of the mean error.

CONCLUSIONS

In general, the Regional Weighting method estimated more accurately the mean air temperature and rainfall values, followed by the Multiple Linear Regression and the Arithmetic Mean methods in both states (Alagoas and Pernambuco).

The greater or lesser accuracy of the methods in the estimates depends on the availability of data.

ACKNOWLEDGMENTS

To the Federal Rural University of Pernambuco (UFRPE) for the provision of structure to carry out the study.

To the Coordination for the improvement of Higher Education Personnel (CAPES) by promoting research.

To the Federal Institute of Alagoas (IFAL) for encouraging the dissemination of research.

BIBLIOGRAPHIC REFERENCES

ALVARES, C. A.; STAPE, J. L.; SENTELHAS, P. C.; GONÇALVES, J. L. D. M.; SPAROVEK, G. Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, v. 22, n. 6, p. 711-728, 2013. DOI: 10.1127/0941-2948/2013/0507

BERGAMASCHI, H.; MATZENAUER, R. **O milho e o clima**. Emater/RS, p. 84, 2014.

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, n. 2, p. 215-226, 2017. DOI: https://doi.org/10.1590/0102-77863220008

DIAZ, C. C. F., PEREIRA, J. A.S., NOBREGA, R. S. Comparação de dados estimados pelo método da ponderação regional (PR) e dados estimados pelo TRMM para o preenchimento de falhas de precipitação na bacia hidrográfica do rio Pajeú. **Revista Brasileira de Climatologia**, v. 22, n. 1, p. 347-354, 2018. DOI: http://dx.doi.org/10.5380/abclima.v22i0.46911

ELY, D., LIMBERGER, L., MANGILI, F., GAMERO, P., SCHMENGLER, M. Análise de métodos para o preenchimento de falhas aplicados em séries de dados pluviométricos do estado do Paraná (Brasil). **RAEGA - O Espaço Geográfico em Análise**, [S.l.], v. 51, p. 122-142, 2021. DOI: http://dx.doi.org/10.5380/raega.v51i0.72944

HUGO, A.; BARROS, C.; COELHO, J.; FILHO, A.; BARROS DA SILVA, A.; AYANE, G.; SANTIAGO, C. F. **Boletim de Pesquisa e Desenvolvimento 211**. Recife, 2012.

INMET - Instituto Nacional de Meteorologia (2022). **"Código, Coordenadas Geográficas e Relevo das Estações Meteorológicas Automáticas dos Estados de Pernambuco e Alagoas"**. Brasília.

JUNQUEIRA, R.; AMORIM, J. DA S.; OLIVEIRA, A. S. DE. Comparação entre diferentes metodologias para preenchimento de falhas em dados pluviométricos. **Periódico da Universidade Vale do Rio Verde**, v. 2, n. 1, 2018. DOI: http://dx.doi.org/10.5892/st.v2i1.4982

MELLO, Y. R. DE; KOHLS, W.; OLIVEIRA, T. M. N. DE. Uso de diferentes métodos para o preenchimento de falhas em estações pluviométricas. **Boletim de Geografia**, v. 35, n. 1, p. 112-121, 2017. DOI: https://doi.org/10.4025/bolgeogr.v35i1.30893

OLIVEIRA, L. F. C.; FIOREZE, A. P.; MEDEIROS, A. M. M.; SILVA, M. A. S. Comparison of gap filling methodologies of annual historical series of rainfall. **Revista Brasileira de**

Engenharia Agrícola e Ambiental, v. 14, n. 11, p. 1186–1192, 2010. DOI: https://doi.org/10.1590/S1415-43662010001100008

RUEZZENE, C. B.; MIRANDA, R. B.; TECH, A. R. B.; MAUAD, F. F. Revisão bibliográfica dos principais métodos de preenchimento de falhas em dados deprecipitação**. Brazilian Journals of Business, Curitiba**, v. 2, n. 3, p. 2057-2071, 2020. DOI: https://doi.org/10.34140/bjbv2n3-015

SILVA, A. A. F.; JARDIM, C. H. Aplicação de técnicas de preenchimentode falhas de dados de pluviosidade mensal e anual para o noroeste do estado de Minas Gerais - Brasil. **Geografias**, [S. l.], v. 15, n. 2, p. 83–106, 2017. DOI: https://doi.org/10.35699/2237-549X.2017.16058

SILVA, L. M. C.; BEZERRA, F. S.; VASCONCELOS, M. C. D.; SILVA, M. R. B. D.; SANTOS, A. C. D. D.; ALMEIDA, C. D. G. D. Gap filling procedures of climatological series in the state of Pernambuco. **IRRIGA**, [S. l.], v. 1, n. 4, p. 754–764, 2021. DOI: https://doi.org/10.15809/irriga.2021v1n4p754-764.