RBRH

Revista Brasileira de Recursos Hídricos
Brazilian Journal of Water Resources
Versão On-line ISSN 2318-0331
RBRH, Porto Alegre, v. 23, e29, 2018
Scientific/Technical Article

https://doi.org/10.1590/2318-0331.231820170171

Alternative methodology to gap filling for generation of monthly rainfall series with GIS approach

Metodologia alternativa ao preenchimento de falhas para a geração de séries de precipitação mensal média de forma automatizada em ambiente SIG

Claudio Bielenki Junior[1], Franciane Mendonça dos Santos[1], Silvia Cláudia Semensato Povinelli[2] and Frederico Fábio Mauad[1]

[1]Escola Engenharia São Carlos, Universidade de São Paulo, São Carlos, SP, Brasil

[2]Universidade Federal São Carlos, São Carlos, SP, Brasil

E-mails: claudio@ana.gov.br (CBJ), fran.emilia@gmail.com (FMS), silviaclaudia@ufscar.br (SCSP), mauadffm@sc.usp.br (FFM)

Received: November 27, 2017 - Revised: May 14, 2018 - Accepted: June 03, 2018

ABSTRACT

As an alternative to Gap filling in monthly average rainfall series, we attempted to present a methodology for the generation of series only with the observed data available in the rainfall stations present in the study area and its surroundings. For this, a computational tool was developed with a GIS approach, using scripts in the Python language, to automate the study steps. Two calculation alternatives for the mean precipitation, variable Thiessen polygons or variable inverse distance weights (IDW), were considered. Random gaps were imposed from a series of data without gaps allowing us to evaluate the presented methodology. The results of the series calculated according to this methodology were compared to two methods of Gap filling. The behavior of the series was evaluated through the analysis of position and dispersion measurements as well as the temporal behavior by the evaluation of the correlograms and periodograms. The results are found to be satisfactory, which demonstrates the equivalence of the proposal with results found with the gap filling methods under the tested conditions. The differences found between the series were small, which was reflected in the Nash-Sutcliffe Indexes. There were no significant differences between the calculation alternatives by Thiessen polygons or IDW weights.

Keywords: Hydrology; Average rainfall; Geoprocessing.

RESUMO

Como alternativa ao preenchimento de falhas em séries de precipitação média mensal buscou-se desenvolver uma metodologia baseada apenas em dados observados disponíveis nas estações pluviométricas presentes na área de estudo e seu entorno. Para isso desenvolveu-se uma ferramenta computacional em ambiente de sistema de informações geográficas, com uso de scripts na linguagem Python. A partir de uma série de dados sem falhas impô-se falhas aleatórias para avaliação da metodologia apresentada. Posteriormente, os valores obtidos foram comparados às precipitações médias obtidas por polígonos de Thiessen variáveis e por interpoladores pelo inverso da distância (IDW) variáveis. Avaliou-se o comportamento das séries obtidas por meio da análise de medidas de dispersão e pelo comportamento dos correlogramas e periodogramas. Foram obtidos resultados satisfatórios que demonstram a equivalência da proposta com resultados encontrados com os dois métodos de preenchimento de falhas de referência, com pequenas diferenças entre as séries obtidas, conforme proximidade dos Índices de Nash-Sutcliff. Não foram observadas diferenças significativas entre as precipitações médias obtidas por polígonos de Thiessen e interpoladores IDW.

Palavras-chave: Hidrologia; Precipitação média; Geoprocessamento.

INTRODUCTION

Rainfall, characterized by its spatial and temporal variation, is one of the most important data for hydrological studies.

Hydrological monitoring, capable of promoting sufficient reliable data, is a preponderant part of a water resources information system, and it is a previous and fundamental step without which one cannot effectively execute the managing of these resources.

Pluviometers traditionally physically measure the amount of precipitation in a determined space and, generally, provide data to a small area. These rainfall measurements are used in rain-flow models (JAYAKRISHNAN; SRINIVASAN; ARNOLD, 2004). However, problems with rain gauge measurements were documented in several studies (LEGATES; DELIBERTY, 1993; FINNERTY et al., 1997; ANA, 2011).

Given the difficulties found in the monitoring of rainfall, gaps can be expected in the historical series. These gaps are due to problems such as the lack of an observer, mistakes in the register mechanisms, loss of notes and data or in the transcription of the registers made by operators and also closure of observations points. However, for most applications exists the need of continuous series analysis, demanding therefore, gap filling (STRECK et al., 2009; BERTONI; TUCCI, 2013).

Oliveira et al. (2010) and Mello, Kohls and Oliveira (2017) compared several methods of gap filling in historical series of rainfall. The gap filling process requires a thorough study of each available station, as well as its correlation to other stations. Besides being a time-consuming process, when there is a high percent of gaps the generation of synthetic data can take place resulting in little conformity to reality.

Alternatively to gap filling in monthly average rainfall series in a hydrographic basin we propose to carry this out in a automatized manner by means of geoprocessing tools.

Moreover, the present research intended to automatize the generation of monthly average rainfall series from data without gap filling through the variable Thiessen polygons and variable inverse distance weighting (IDW) methods by means of a computational tool, coupled to a geographical information system.

Also, to access and confront the results from monthly average rainfall series calculations by the Thiessen polygons and by the IDW method.

Afterwards compare the results reached by the proposed methodology to the ones obtained by the classic methods of gap filling by Regional Weighting calculated in function of average and correlation.

METHODOLOGY

Research presentation

With this study we tried to avoid the generation of synthetic data, using only available data at the rain gauges to generate monthly average rainfall series. Thus, for each month, a different combination of rain gauges was used.

For the analysis (validation) of the average rainfall series by the alternative methodology, without gap filling, these series were compared to the ones obtained from the series originated from data without gaps and to the ones obtained from data with gap filling by the Regional Weighting Method.

According to Bertoni and Tucci (2013), the regional weighting method is a simplified method normally used for filling monthly or annually rainfall series, aiming the homogenization of the information period and the statistical analysis of precipitation, and also being used to the extension of pluviometric series. The method can be classified according to the kind of statistic used in the weighting of stations: average or correlation.

Furthermore, the Thiessen polygon method and the inverse distance weighting (IDW) method was used for the calculation of average rainfall.

The Thiessen method is one of the most common methods for average rainfall determination and it consists on attributing a weight factor to total rainfall in each pluviometer, proportional to the influence area of each one (VILLELA; MATTOS, 1975).

IDW is a deterministic estimator of non-sampled values from a linear combination of values from known points, weighted in function of distance. On it, it is considered that the points closest to the non-sampled locals are more representative than those further away. Thus, weighting changes according to the linear distance of the samples to the non-sampled points (PHILIP; WATSON, 1982).

The use of available spatialization techniques in the Geographical Information Systems (GIS) facilitates the verification of how the variables observed in the historical series are distributed in space. Therefore, this tool has been widely used as a support to hydrological studies. Furthermore, through the facility that these systems have of incorporating commands and structured functions by means of scripts, it allows the optimization and automatization of several tasks and analysis.

Different statistical treatments were considered for the validation of results aiming to verify the equivalence of average series, the autocorrelation structure behavior and error measures.

Hydrographical basin selection and rainfall data attainment

The hydrographical basin chosen for the average rainfall series calculation aiming to compare the proposed methodology and the conventional methods of gap filling is the Rio das Cinzas Basin located in the state of Paraná-Brazil. The Cinzas river is 240 km long and its basin covers a total drainage area of 9,645 km2.

Thirty pluviometric stations were selected to carry out the proposed tests in the area of the hydrographical basin and in its surroundings, with rainfall data series accumulated in the monthly scale for the period between January 1976 to December 2003, totalizing 28 years of measurements that result in 336 months.

These stations were selected since they did not present gaps in the data for this period. The data concerning these stations were obtained already consisted in the hydrological data base of the National Water Agency (ANA, 2017). Figure 1 presents the localization of the stations in the basin area.

Pluviometry data of the selected points were verified for homogeneity according to a cluster analysis that delimits the homogenous regions. The Euclidian distance methodology by the Ward method (WARD JUNIOR, 1963) was used to determine the groups. The set of stations can be divided in three groups where is was observed that the variation takes place in the north-south direction of the basin.

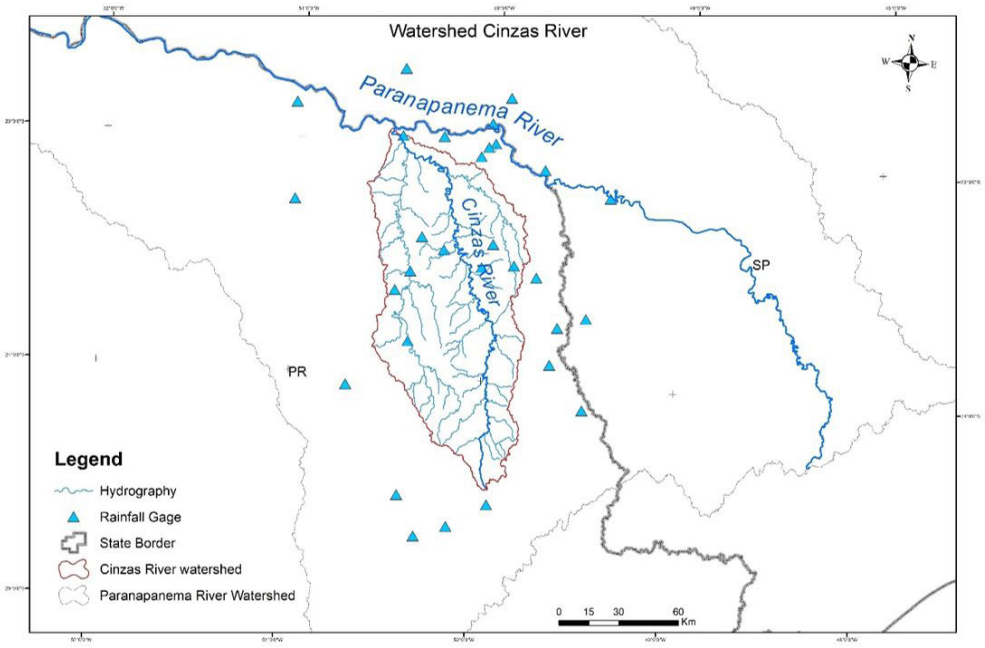

Figure 1. Localization of the pluviometric stations in the region of the Cinzas River hydrographical basin.

Calculation of average rainfall series from data without gaps

Based on the original data without gaps the monthly average rainfall series were calculated for the region of the hydrographical basin for study by the Thiessen polygon and the IDW methods. These series were adopted as reference rainfall for further comparison. Table 1 presents the nomenclature used for these series.

Gap generation in the original data series

Random gaps were added to the original data. Gap percentage values of 30% and 50% were used to compare the proposed methodology to the conventional filling methods, generating two new sets of data.

To carry out this task, a macro, based on the function of random numbers, was implemented to the Excel software using Visual Basic for Application. This macro calculates the numbers of gaps that need to be generated multiplying the number of series of months by the number of stations and by the adopted percent of gaps and distributes it randomly in the set of original data erasing the registered value.

Gap filling

The series (with N months) with gaps were filled by the regional weighting methods calculated based on the average (MPRM) according to Equation 1 and based on correlation (MPRC) according to Equation 2 (ANA, 2014).

Table 1. Nomenclature used for the series originated from data without flaws.

Abbreviation	Adopted flaws %	Filling method	Estimation method
*d*_C T	-	-	Thiessen Polygons
*d*_C IDW	-	-	IDW

$$y = \frac{1}{N-1}.\left[\left(\frac{\bar{y}}{\bar{x}_1}\right).x_1 + \left(\frac{\bar{y}}{\bar{x}_2}\right).x_2 + ... \left(\frac{\bar{y}}{\bar{x}_{N-1}}\right).x_{N-1}\right] \quad (1)$$

y = total monthly rainfall, estimated (filled) for the "Y" station, in the referred month; $\bar{y}$ = total average rainfall in the "Y" station, in the referred month, corresponding to the common observation period; $\bar{x}_i$ = total average rainfall for the "Xi" station of the homogenous group, in the reference month, corresponding to the common observation period; x_i = total monthly observed in the "Xi" station, in the month that the total rainfall in "Y" station must be filled.

$$y = \frac{s_y}{N-1}.\left[r_1.\frac{x_1-\bar{x}_1}{s_{x_1}} + r_2.\frac{x_2-\bar{x}_2}{s_{x_2}} + .. + r_{N-1}.\frac{x_{N-1}-\bar{x}_{N-1}}{s_{x_{N-1}}}\right] + \bar{y} \quad (2)$$

y = total monthly rainfall, estimated (filled or extended) for "Y" station, in the referred month; $\bar{y}$ = total average rainfall for "Y" station, in the reference month, corresponding to the common observation period; Sy= standard deviation of the total rainfall for "Y" station, in the reference month, corresponding to the common observation period; $\bar{x}_i$ = total average rainfall for "Xi" station of the homogenous group, in the referred month, corresponding to the common observation period; s_{x_i} = standard deviation of the total

rainfall for "Xi" station, of the homogenous group, in the referred month, corresponding to the common observation period; x_i = total monthly observed in the "Xi" station, in the month that the total rainfall for "Y" station must be filled or extended; r_i = correlation coefficient between the total rainfall series in the "Y"station and the corresponding series in "Xi" station, considering the common observation period in the month of reference.

A macro was implemented in the Excel software with Visual Basic for Application for gap filling.

For the selection of stations to be used in the filling of each month the implemented algorithm calculates the distance matrix between the stations based on its coordinates and a correlation matrix between the stations. With these two matrixes an index matrix directly proportional to correlation and inversely proportional to distance is calculated, selecting as support stations (up to five of them) to gap fill those stations with the highest rates.

After gap filling each of the series of the pluviometric stations were verified again, by means of accumulated double mass curves, in relation to consistency.

Average rainfall calculation from filled data

Based on the filled data, eight series of average monthly rainfall for the basin of the Cinzas River according to the Thiessen polygons and the IDW methods were calculated.

Table 2 presents the nomenclature used for these series.

Alternative methodology

None of the gaps present in the entry series were filled to calculate the average rainfall series for the basin varying Thiessen polygons and IDW. Only the available data was used.

The availability of data was assessed for each month of the series, where those containing gaps in the respective month were excluded from the selection of pluviometric stations. The average rainfall from that month was then calculated from the selected stations using the respective Thiessen polygons and IDW.

For the following month a new assessment of data availability was made and a new selection of stations was configured. This way, a new combination was used for the calculation of average rainfall of this month.

Successively, for each month of the series, the station combinations can vary according to data availability. Due to the random characteristic of the gaps, we do not know, a priori, the number of combinations necessary for the calculation of average rainfall.

In long series, and with a significant number of stations, the number of combinations can be excessively high, making this procedure unviable to be performed manually. Thus, it was necessary for this recursive procedure to be implemented in an automatized manner.

A tool that permitted the selection of stations in function of data availability and the respective calculation of average precipitation iteratively to each month of the series was developed in Python language in a geoprocessing platform (ArcGIS) (Toms, 2015).

In general, the algorithm for average monthly rainfall calculation according to the variable Thiessen polygon method comprises the following steps:

Let the data series with t months and n stations in the study area

For months varying from 1 to t:

For station varying from 1to n:

If station(n) has observed data => VectorSelecti on(month):=station(n)

Polygons:=Thiessen Polygon(VectorSelection(month))

PolygonsInterest:=Cut(Polygons, Basin)

For each feature in PolygonsInterest:

PolygonsInterest.weight ratio:=PolygonsInterest.area/Basin.area

PolygonsInterest.AvgRainfall:=PolygonsInterest.Staiton(Recover (Precipitation)) * PolygonsInterest.weight ratio

AvgRainfall(month):=PolygonsInterest.AvgRainfall(Summarize).

Figure 2 presents the window of the variable Thiessen polygons tool developed for the ArcGIS.

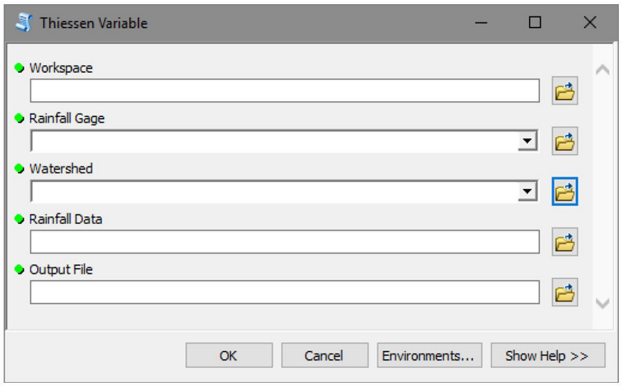

Figure 2. Variable Thiessen polygons tool window.

Table 2. Nomenclature used for the series originated from filled data.

Abbreviation	adopted gap %	Filling method	Estimation method
*d*_30F P MPRM T	30	Average Regional Weighting	Thiessen Polygons
*d*_30F P MPRC T	30	Regional Weighting by Correlation	Thiessen Polygons
*d*_30F P MPRM IDW	30	Average Regional Weighting	IDW
*d*_30F P MPRC IDW	30	Regional Weighting by Correlation	IDW
*d*_50F P MPRM T	50	Average Regional Weighting	Thiessen Polygons
*d*_50F P MPRC T	50	Regional Weighting by Correlation	Thiessen Polygons
*d*_50F P MPRM IDW	50	Average Regional Weighting	IDW
*d*_50F P MPRC IDW	50	Regional Weighting by Correlation	IDW

In general, the algorithm for average monthly rainfall calculation according to the variable IDW method comprises the following steps:

Let the data series with t months and n stations in the study area

For month varying from 1 to t:

For station varying from 1 to n:

If station(n) has observed data => VectorSelection (month):=station(n)

RainfallImage:=IDW(VectorSelection(month))

InterestImage:=Cut(RainfallImage,Basin)

AvgRainfall(month):=InterestImage(Zonal(Basin,Average)).

Figure 3 presents the window of the variable IDW tool developed for the ArcGIS.

Average rainfall calculation by the alternative method

Four series were calculated by the proposed alternative method. Table 3 presents the nomenclatures used for these series.

Statistical comparisons between average rainfall series

A series of statistical analysis was used in order to carry out comparisons between the results. Figure 4 shows the methodology flowchart for comparison between average rainfall series.

The analyses correspond, initially, to graphical analysis and parametric and nonparametric statistical tests to analyze data normality and series stationarity.

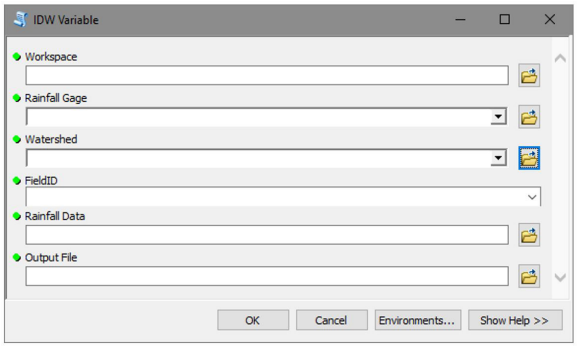

Figure 3. Variable IDW tool window.

Table 3. Nomenclature used for the series originated from data without filling.

Abbreviation	% adopted gaps	Filling method	Estimation method
*d*_30F TV	30	-	Variable Thiessen Polygons
*d*_30F IDW	30	-	Variable IDW
*d*_50F TV	50	-	Variable Thiessen Polygons
*d*_50F IDW	50	-	Variable IDW

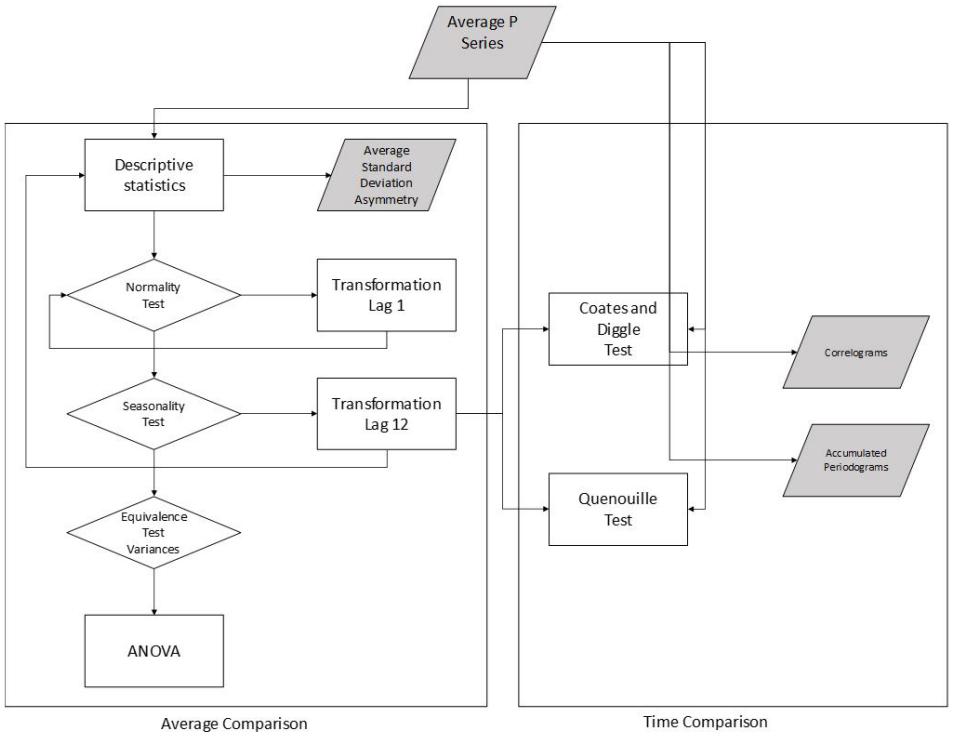

Figure 4. Methodology flowchart for comparison between average rainfall series.

The graphical analysis corresponded to the use of correlograms and periodograms of the calculated series. The used statistical tests were: the Shapiro-Wilk test (SHAPIRO; WILK, 1965), to examine series normality; the Mann-Kendall test (MANN, 1945; KENDALL, 1975), to examine tendency presence in the series; and the Kruskal-Wallis test (KRUSKAL; WALLIS, 1952), to examine series stationarity.

The successive difference series with a 1-month Lag was used for data normalization. A second transformation with a successive difference series with a 12-month Lag for the removal of seasonal effects was also applied to the prior.

These transformed series were then retested and once the normality and stationarity conditions were satisfied they are analyzed by analysis of variance (ANOVA) to test the equality of the averages. Beyond that, the series were compared, pair to pair among themselves, by the Tukey Multiple Comparison test (TUKEY, 1953).

The autocorrelation structures were analyzed to verify the behavior between the series throughout time. To do so, the temporal series comparison methodologies were applied as presented by Coates and Diggle (1986) whose test is called the Accumulated sum test, and the methodology presented by Quenouille (1958) and extended to analysis of several series by Rosenhead (1968) was also applied, and which we will call the Autocorrelation Functions Equality Test (F.A.).

For the accumulated sum test ratio values between the periodograms between each possible pair between series were calculated. Then we obtained the Oj statistic and its distribution was compared to the uniform distribution U(0, 1) by the nonparametric Kolmogorov-Smirnov test.

Autocorrelation values of each series were obtained for the equality test between autocorrelation functions, where the average autocorrelation values for the 1-month Lag determined the coefficient of the auto regressive adjustment model of order 1. The residual series were obtained from this model, and from them, partial autocorrelation values were obtained. T statistic values were determined according to the method proposed by Rosenhead (1968), which were tested according to the $\chi 2$ distribution.

There are many ways to estimate error measurement for the methods or models selection described in literature. In this work the comparison between series, in terms of error estimation, was performed by statistical parameters root-mean-square error (RMSE), mean absolute percentage error (MAPE) and the Nash-Sutcliffe Efficiency Coefficient.

RMSE is commonly used to express numerical results accuracy with the advantage that is presents error values in the same dimensions of the analyzed variable. RMSE is calculated according to Equation 3.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_{obs,i} - X_{sim,i})^2}{n}} \quad (3)$$

where Xobs is the observed value and Xsim is the simulated value in time i.

MAPE measures the error in percentage. It is calculated as the percentage error average according to Equation 4.

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{X_{obs,i} - X_{sim,i}}{X_{obs,i}}\right| \quad (4)$$

where Xobs is the observed value and Xsim is the simulated value in time i.

For Machado and Vettorazzi (2003), one of the most important statistical criteria to evaluate the hydrological models' adjustment is the Nash-Sutcliffe Efficiency Coefficient, which is calculated according to equation 5. The coefficient can vary from negative infinite to 1, where 1 is the indication of a perfect adjustment (ASCE, 1993). In accordance with Silva et al. (2008), when the coefficient value result is bigger than 0.75, the model performance is considered good. For values between 0.36 and 0.75, it is considered acceptable, while values under 0.36 are considered unacceptable.

$$E = 1 - \frac{\sum_{i=1}^{n}(X_{obs,i} - X_{sim})^2}{\sum_{i=1}^{n}(X_{obs,i} - \overline{X_{obs}})^2} \quad (5)$$

where Xobs is the observed event; Xsim, the event simulated by the model; $\overline{X}$ obs, the average of the observed even in the period; and n, the number of events.

RESULTS AND DISCUSSION

Variance analysis

Analyzing the tests that were carried out it was possible to verify that the transformed series (by the first and second differences, 1 and 12 months Lags) presented data with a normal distribution, they have variance equivalence and do not present tendency or seasonality. Thus, a comparison, regarding average, of the transformed series was carried out to examine if the methodologies tested for the average monthly rainfall calculation of the hydrographical basin present significant statistical differences.

A single factor variance analysis study for the transformed series was performed. Table 4 presents the ANOVA results for the transformed series.

According to the statistical test that was made there is no evidence to reject the H0 hypothesis of equality for the series average.

Tukey multiple comparisons

The equality of the averages, pair to pair between series, was also verified by the Tukey Multiple Comparisons test. In Tables 5 and 6 Tukey Q statistical values are present in the

Table 4. ANOVA for transformed series.

Analysis	*F Statistic*	*P-Value*	*F critical*
30% gaps	0.000023	1	2.013
50% gaps	0.000027	1	2.013

cells under the main diagonal and P-Value for average equality is present in the cells above the main diagonal.

It can be seen that no average distinguishes itself from the others, indicating that the series are equivalent and that the processes that generate them can be used with similar results to the level of significance of 5%.

Autocorrelation structures analysis

In terms of average and variance the series present equivalent values, however their behaviors throughout time were analyzed to verify if they presented a same autocorrelation structure. For this, we decided to analyze the temporal series comparison proposals of the Accumulated Sum Test (COATES; DIGGLE, 1986) and an extension of the Autocorrelation Functions Equality Test (QUENOUILLE, 1958) adapted by Rosenhead (1968).

Accumulated sum test

P-Values from Kolmogorov-Smirnov tests carried out when applying the Accumulated Sum Test are presented in Tables 7 and 8. The H0 hypothesis of equality of functions of spectral density cannot be rejected for the values that were encountered, to the level of significance of 5% for each pair of tested series. This indicates that the series come from a same stochastic process, therefore being the methods that generate the series that are considered equivalent.

Table 5. Tukey test between transformed series (30% gap analysis).

Series	*d_CT*	*d_CIDW*	*d_30FPMPRMT*	*d_30FPMPRCT*	*d_30FPMPRMIDW*	*d_30FPMPRCIDW*	*d_30FTV*	*d_30FIDW*
d_CT	-	1	1	1	1	1	1	1
d_CIDW	0.007	-	1	1	1	1	1	1
d_30FPMPRMT	0.001	0.008	-	1	1	1	1	1
d_30FPMPRCT	0.002	0.009	0.001	-	1	1	1	1
d_30FPMPRMIDW	0.006	0.013	0.005	0.004	-	1	1	1
d_30FPMPRCIDW	0.005	0.013	0.004	0.004	0.000	-	1	1
d_30FTV	0.001	0.0061	0.002	0.003	0.007	0.007	-	1
d_30FIDW	0.008	0.015	0.006	0.006	0.002	0.002	0.009	-

Table 6. Tukey test between transformed series (50% gap analysis).

Series	*d_CT*	*d_CIDW*	*d_50FPMPRMT*	*d_50FPMPRCT*	*d_50FPMPRMIDW*	*d_50FPMPRCIDW*	*d_50FTV*	*d_50FIDW*
d_CT		1	1	1	1	1	1	1
d_CIDW	0.007		1	1	1	1	1	1
d_50FPMPRMT	0.005	0.002		1	1	1	1	1
d_50FPMPRCT	0.004	0.003	0.001		1	1	1	1
d_50FPMPRMIDW	0.001	0.006	0.004	0.003		1	1	1
d_50FPMPRCIDW	0.001	0.008	0.006	0.005	0.002		1	1
d_50FTV	0.015	0.008	0.010	0.011	0.014	0.016		1
d_50FIDW	0.003	0.004	0.002	0.001	0.002	0.004	0.012	

Table 7. P-Value values for the Kolmogorov-Smirnov test (30% gap analysis).

Series	*d_CT*	*d_CIDW*	*d_30FPMPRMT*	*d_30FPMPRCT*	*d_30FPMPRMIDW*	*d_30FPMPRCIDW*	*d_30FTV*	*d_30FIDW*
d_CT	1.0000							
d_CIDW	0.4963	1.000						
d_30FPMPRMT	0.7574	0.329	1.000					
d_30FPMPRCT	0.6256	0.242	0.626	1.000				
d_30FPMPRMIDW	0.5925	0.223	0.789	0.789	1.000			
d_30FPMPRCIDW	0.6256	0.205	0.757	0.757	0.725	1.000		
d_30FTV	0.6589	0.306	0.626	0.757	0.757	0.819	1.000	
d_30FIDW	0.6589	0.205	0.725	0.789	0.725	0.757	0.354	1.000

Table 8. P-Value values for the Kolmogorov-Smirnov test (50% gap analysis).

Series	*d_CT*	*d_CIDW*	*d_50FPMPRMT*	*d_50FPMPRCT*	*d_50FPMPRMIDW*	*d_50FPMPRCIDW*	*d_50FTV*	*d_50FIDW*
d_CT	1.000							
d_CIDW	0.496	1.000						
d_50FPMPRMT	0.120	0.380	1.000					
d_50FPMPRCT	0.144	0.408	0.819	1.000				
d_50FPMPRMIDW	0.306	0.144	0.528	0.408	1.000			
d_50FPMPRCIDW	0.329	0.205	0.659	0.592	0.874	1.000		
d_50FTV	0.039	0.626	0.954	0.874	0.874	0.847	1.000	
d_50FIDW	0.592	0.283	0.789	0.659	0.938	0.789	0.592	1.000

Autocorrelation function equality test

Table 9 presents the T statistic values and the respective p-values for a 5% trust level when carrying out autocorrelation function equality test.

It can be seen that in both cases the H0 hypothesis that the autocorrelation functions of the series are equal cannot be rejected to the considered trust level. Thus, the series calculated by the different methods present the same autocorrelation structures.

Graphically the existing equality of the autocorrelation structures are presented by the dispersion between the accumulated normalized periodograms between the analyzed series. When the autocorrelation structures are equivalent the points adjust themselves in a straight line, corroborating with the behavior seen in the accumulated normalized periodograms for the transformed series plotted in relation to d_CT series (Figures 5 and 6).

A test series was included in these graphs (randomly generated from the average normal distribution of 80 and standard deviation of 60, which are the average values and standard deviation closest to the ones presented in the studied series) aiming to highlight the behavior of a series that does not contain the same correlation structure.

It can be seen that the series adjusted well to the straight, except for the N(80,60) test series, which comes from a different stochastic process.

Table 9. P-Value values for the autocorrelation function equality test (30% and 50% gap analysis).

T	Degree of Freedom	30% gaps		50% gaps	
		Value	P-value	Value	P-value
1	7	0.016	1	0.107	1
2	7	0.126	1	0.162	1
3	7	0.010	1	0.241	1
4	7	0.063	1	0.026	1
5	7	0.012	1	0.030	1
6	7	0.169	1	0.154	1
7	7	0.150	1	0.312	1
8	7	0.106	1	0.021	1
Sum	56	0.652	-	1.052	-

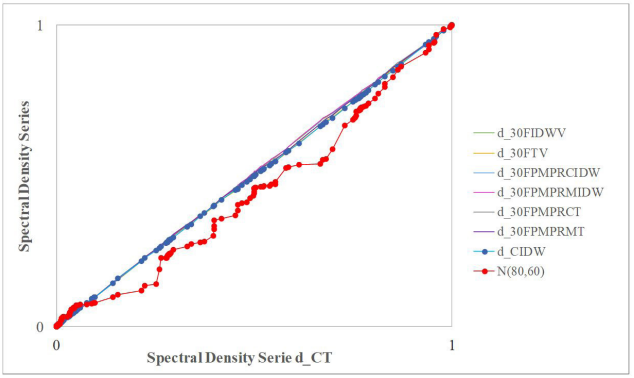

Figure 5. Accumulated normalized periodograms – transformed series (30% gap analysis).

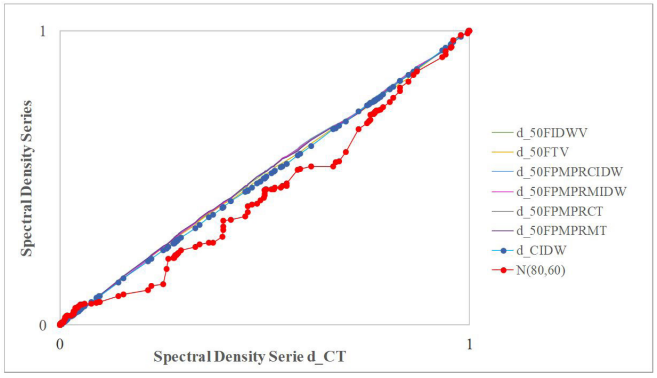

Figure 6. Accumulated normalized periodograms – transformed series (50% gap analysis).

Table 10. Error measurements for the 30% gap analysis.

Statistic	C IDW	30F P MPR M T	30F P MPR C T	30F P MPR M IDW	30F P MPR C IDW	30F TV	30F IDW V
RMSE	5.073	7.230	8.472	6.227	6.870	7.849	6.193
MAPE	3.962	6.108	55.313	4.877	49.354	5.940	4.777
NASH	0.995	0.991	0.987	0.993	0.991	0.989	0.993

Table 11. Error measurements for the 50% gap analysis.

Statistic	C IDW	50F P MPRM T	50F P MPRC T	50F P MPRM IDW	50F P MPRC IDW	50F TV	50F IDW V
RMSE	5.073	9.176	11.634	8.926	10.522	9.588	8.463
MAPE	3.962	8.024	11.223	7.924	105.416	7.651	7.366
NASH	0.995	0.985	0.976	0.986	0.980	0.983	0.987

The analysis of rainfall series periodograms allowed us to efficiently visualize the correspondence between series throughout time, and furthermore, the accumulated sum and autocorrelation function equality tests showed appropriate statistics for the comparison of temporal series.

Error estimation

Root-mean-square error (RMSE), mean absolute percentage error (MAPE) and the Nash-Sutcliffe Efficiency Coefficient (NASH) were calculated to compare the results of data obtained from the series. Tables 10 and 11 show the error measurements for the 30% and 50% gap analysis.

When comparing the results found using the alternative methodology – variable Thiessen polygons or variable IDW interpolation – to the results found using the regional weighting gap filling method, by the average or by correlation, the statistical tests pointed equivalence between the series, besides this, the error measurements are of the same order, bot for conditions tested (30% and 50% gaps).

Increasing the gap, from 30% to 50%, the proposed alternative methodology with variable IDW interpolation came even closer to the conventional methods.

From the tested filling methods, the regional weighting calculated based on correlations presented error measurements higher than regional weighting calculated based on averages, which is in conformity to the results obtained by Oliveira et al. (2010) and Mello, Kohls and Oliveira (2017).

GIS environment automatization

Geographical information systems are fundamental in any hydrological study, particularly in this work it was the tool that made possible the use of the presented alternative proposal, since without the automatization that these systems provided it would not be possible to carry out a sufficient number of tests, as it was, in feasible time.

The large number of combination of pluviometric stations avaialbe in each month demanded that the proposed procedure was calculated in a recursive manner.

The ArcGIS GIS, with support to programing by means of scripts in Python language, has a consolidated community of users, with a vast content of documentation and support that made its usage easier.

The developed tool presented two methods for the calculation of average monthly rainfall series, the Thiessen polygons and the IDW, however, because of the system's facility, other weighting algorithms can be implemented and tested.

CONCLUSIONS

Based on the statistical tests that were carried out we can conclude that the alternative methodology for gap filling for the generation of average monthly rainfall series for a hydrographical basin reached satisfactory results when compared to other filling methods.

This result is confirmed by the low error measurements when compared to average rainfall series calculated by Thiessen polygons originated from data without gaps.

The automatization of this methodology in a GIS environment also brought advantages, such as lower time demand in processing, thus assuring efficiency in the results.

Another advantage is the minimization of common typical mistakes associated to the edition several calculus spreadsheets in hydrological studies.

One of the difficulties encountered in this work was the acquisition of consisted pluviometry data. Besides the necessity of finding pluviometric stations for the region of the hydrographical basin for a relatively long period and without gaps, the access to this information, as well as the data research form, and the ways of acquisition through the Hydro Web platform was precarious and inefficient.

REFERENCES

ANA – AGÊNCIA NACIONAL DE ÁGUAS. *Recursos Hídricos e Ambientais (RHA)*: qualificação de dados hidrológicos e reconstituição de vazões naturais no país. Brasília: ANA, 2011. 442 p. Relatório final do contrato nº 016/ANA/2009

ANA – AGÊNCIA NACIONAL DE ÁGUAS. *Procedimentos para a consistência de dados pluviométricos*. Brasília: ANA/SGH, 2014. 30 p.

ANA – AGÊNCIA NACIONAL DE ÁGUAS. *HidroWeb*: sistemas de informações hidrológicas. Brasília: ANA, 2017.

ASCE – AMERICAN SOCIETY OF CIVIL ENGINEERS. Task Committee on Definition of Criteria for Evaluation of Watershed Models of the Watershed Management: Committee Irrigation and Drainage Division: criteria for evaluation of watershed models. *Journal of Irrigation and Drainage Engineering*, v. 119, p. 429-442, 1993.

BERTONI, J. C.; TUCCI, C. E. M. Precipitação. In: TUCCI, C. E. M. *Hidrologia*: ciência e aplicação. 4. ed. Porto Alegre: Ed. Universidade/UFRGS/ABRH, 2013. 943 p.

COATES, D. S.; DIGGLE, P. J. Test for comparing two estimated spectral densities. *Journal of Time Series Analysis*, v. 7, n. 7, p. 7-20, 1986. http://dx.doi.org/10.1111/j.1467-9892.1986.tb00482.x.

FINNERTY, B. D.; SMITH, M. B.; SEO, D. J.; KOREN, V. I.; MOGLEN, G. Sensitivity of the Sacramento soil moisture accounting model to space-time scale precipitation inputs from NEXRAD. *Journal of Hydrology*, v. 203, n. 1-4, p. 21-38, 1997.

JAYAKRISHNAN, R.; SRINIVASAN, R.; ARNOLD, J. G. Comparison of rain gage and WSR-88D Stage III precipitation data over the Texas-Gulf basin. *Journal of Hydrology*, v. 292, n. 1-4, p. 135-152, 2004. http://dx.doi.org/10.1016/j.jhydrol.2003.12.027.

KENDALL, M. G. *Rank correlation methods*. 4th ed. London: Charles Griffin, 1975.

KRUSKAL, W. H.; WALLIS, W. A. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, v. 47, n. 260, p. 583-621, 1952. http://dx.doi.org/10.1080/01621459.1952.10483441.

LEGATES, D. R.; DELIBERTY, T. L. Measurement biases in the United States rain gauge network. *Water Resources Bulletin*, v. 29, p. 855-861, 1993.

MACHADO, R. E.; VETTORAZZI, C. A. Simulação da produção de sedimentos para a microbacia hidrográfica do Ribeirão dos Marins, SP. *Revista Brasileira de Ciência do Solo*, v. 27, n. 4, p. 735-741, 2003. http://dx.doi.org/10.1590/S0100-06832003000400018.

MANN, H. B. Non-parametric tests against trend. *Econometrica*, v. 13, n. 3, p. 245-259, 1945. http://dx.doi.org/10.2307/1907187.

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

OLIVEIRA, L. F. C.; FIOREZE, A. P.; MEDEIROS, A. M. M.; SILVA, M. A. S. Comparação de metodologias de preenchimento de falhas de séries históricas de precipitação pluvial anual. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v. 14, n. 11, p. 1186-1192, 2010. http://dx.doi.org/10.1590/S1415-43662010001100008.

PHILIP, G. M.; WATSON, D. F. A precise method for determining contoured surfaces. *Australian Petroleum Exploration Association Journal*, v. 22, p. 205-212, 1982.

QUENOUILLE, M. The comparison of correlations in time-series. *Journal of the Royal Statistical Society. Series B. Methodological*, v. 20, n. 1, p. 158-164, 1958.

ROSENHEAD, J. An extension of Quenouille's test for the compatibility of correlation structures in time series. *Journal of the Royal Statistical Society. Series B. Methodological*, v. 30, p. 180-184, 1968.

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality: complete samples. *Biometrika*, v. 52, n. 3-4, p. 591-611, 1965. http://dx.doi.org/10.1093/biomet/52.3-4.591.

SILVA, P. M. O.; MELLO, C. R.; SILVA, A. M.; COELHO, G. Modelagem da hidrografia de cheia em uma bacia hidrográfica da região Alto Rio Grande. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v. 12, n. 3, p. 258-265, 2008. http://dx.doi.org/10.1590/S1415-43662008000300006.

STRECK, N. A.; BURIOL, G. A.; HELDWEIN, A. B.; GABRIEL, L. F.; PAULA, G. M. Associação da variabilidade da precipitação pluvial em Santa Maria com a oscilação decadal do Pacífico. *Pesquisa Agropecuária Brasileira*, v. 44, n. 12, p. 1553-1561, 2009. http://dx.doi.org/10.1590/S0100-204X2009001200001.

TOMS, S. *ArcPy and ArcGIS*: geospatial analysis with python. Birmingham: Packt Publishing, 2015. 224 p.

TUKEY, J. W. *The problem of multiple comparisons*. Princeton: Mimeographs Princeton University, 1953.

VILLELA, S. M.; MATTOS, A. *Hidrologia aplicada*. São Paulo: McGraw Hill, 1975. 245 p.

WARD JUNIOR, J. H. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, v. 58, n. 301, p. 236-244, 1963. http://dx.doi.org/10.1080/01621459.1963.10500845.

Authors contributions

Claudio Bielenki Junior: GIS and VBA programming, statistical analysis, literature review, writing.

Franciane Mendonça dos Santos: Data acquisition, manipulation and consistency, literature review.

Silvia Cláudia Semensato Povinelli: Review literature, proposition testing, article review.

Frederico Fábio Mauad: Orientation, supervision, review.