

<https://doi.org/10.1590/2318-0331.252020190054>

A regional similarity-based approach for sub-daily rainfall nonparametric generation

Uma abordagem de similaridade regional para geração não paramétrica de precipitação subdiária

Milena Guerra de Aguiar¹  & Veber Afonso Figueiredo Costa¹ 

¹Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil

E-mails: milenaguerra.a@gmail.com (MGA), veber@chr.ufmg.br (VAFC)

Received: May 02, 2019 - Revised: September 05, 2019 - Accepted: October 25, 2019

ABSTRACT

Rainfall time series with high temporal resolution are required for estimating storm events for the design of urban drainage systems, for performing rainfall-runoff simulation in small catchments and for modeling flash-floods. Nonetheless, large and continuous sub-daily rainfall samples are often unavailable. For dealing with the limited availability of high-resolution rainfall records, in both time and space, this paper explored an alternative version of the k -nearest neighbors algorithm, coupled with the method of fragments (KNN-MOF model), which utilizes a state-based logic for simulating consecutive wet days and a regionalized similarity-based approach for sampling fragments from hydrologically similar nearby stations. The proposed disaggregation method was applied to 40 rainfall gauging stations located in the São Francisco and Doce river catchments. Disaggregation of daily rainfall was performed for the durations of 60, 180 and 360 minutes. Results indicated the model presented an appropriate performance to disaggregate daily rainfall, reasonably reproducing sub-daily summary statistics. In addition, the annual block-maxima behavior, even for low exceedance probabilities, was relatively well described, although not all expected variability in the quantiles was properly summarized by the model. Overall, the proposed approach proved a sound and easy to implement alternative for simulating continuous sub-daily rainfall amounts from coarse-resolution records.

Keywords: Sub-daily rainfall generation; Daily rainfall disaggregation; Hydrological and physical similarity.

RESUMO

Séries de precipitação de alta resolução temporal são necessárias para a modelagem de eventos regulares e extremos em projetos de sistemas pluviais urbanos, simulação de vazões em pequenas bacias, estudos de balanços hídricos e modelagem de inundações de curta duração, por exemplo. Não obstante, amostras longas e contínuas de precipitação subdiária são, via de regra, difíceis de ser encontradas. Para contornar o problema da limitada disponibilidade de dados de alta resolução temporal tanto na escala temporal como na espacial, esse artigo explorou uma versão alternativa do algoritmo dos k -ésimos vizinhos mais próximos, acoplada ao método dos fragmentos, (modelo KNN-MOF), que se baseia nos estados de chuva para simulação de dias chuvosos consecutivos e em uma abordagem de similaridade regional para amostragem dos fragmentos de estações similares próximas a estação de interesse. O método de desagregação proposto foi aplicado para um grupo de 40 postos pluviográficos localizados nas bacias dos rios Doce e São Francisco. A desagregação da precipitação diária foi realizada para as durações de 60, 180 e 360 minutos. Os resultados indicaram que o mesmo apresentou um desempenho apropriado para a desagregação de precipitação diária, reproduzindo as estatísticas como média, variância e assimetria para as durações analisadas. Adicionalmente, o comportamento dos máximos anuais, mesmo para baixas probabilidades de excedência, foi relativamente bem descrito, ainda que nem toda a variabilidade esperada dos quantis tenha sido abrangida pelo modelo. De forma geral, a abordagem proposta mostrou-se uma alternativa coerente e de fácil implementação para simular séries contínuas de precipitação subdiária a partir de registros de menor resolução temporal.

Palavras-chave: Geração de precipitação subdiária; Desagregação de precipitação diária; Similaridade hidrológica e fisiográfica.



INTRODUCTION

Rainfall time series with high temporal resolution are often required for estimating storm events for the design of urban drainage systems, for performing rainfall-runoff simulation in small catchments and for modeling flash-floods. Nonetheless, large and continuous sub-daily rainfall samples are, most often than not, unavailable. As compared to daily records, the fine-scale rainfall measurements are usually smaller in length and more affected by missing data, which may severely compromise the statistical inference of sub-daily precipitation block-maxima (Westra et al., 2012). In addition, the sub-daily pluviograph network is usually not sufficiently dense for properly describing the spatial variation of short-duration storm bursts and, as a result, it is often unfeasible to obtain a full picture on the time-space dependence structures of sub-daily rainfall events (Li et al., 2018).

As a means for dealing with the limited availability of high-resolution rainfall records, in both time and space, a number of techniques for obtaining continuous sequences of short-duration precipitation events from the disaggregation of daily rainfall amounts has been discussed in the literature. According to Sharma & Mehrotra (2010) and Diez-Sierra & del Jesus (2019), the most common models for this purpose encompass: (i) parametric point-process models using Poisson clustering, such as the Newman-Scott and the Bartlett-Lewis rectangular pulses (Koutsoyiannis & Onof, 2001; Yusop et al., 2014); (ii) the self-similarity approach, which resorts to random cascades (Kang & Ramirez, 2010; Müller & Haberlandt, 2018) or fractals/ multi-fractals (Serinaldi, 2010); and (iii) nonparametric models based on resampling techniques, such as the method of fragments (Westra et al., 2012; Li et al., 2018).

Point-process models are reported to preserve most summary statistics of sub-daily rainfall amounts and properly reproduce extreme events (Diez-Sierra & del Jesus, 2019). However, they require the estimation of a relatively large set of parameters (Pui et al., 2012), which should increase modeling uncertainty. In addition, such models were originally designed for generating fine-scale rainfall sequences and, for employing them for disaggregation, it becomes necessary to introduce “adjusting procedures” for correcting the simulated rainfall amounts (e.g., Koutsoyiannis & Onof, 2001). On the other hand, random cascade models, although parsimonious (Sivakumar & Sharma, 2008), are frequently unable to properly reproduce the observed variance, the extreme rainfall amounts and the temporal dependence structure of the full sequence of sub-daily rainfall, which, to some extent, constrains their application for flood risk assessment and design purposes (Diez-Sierra & del Jesus, 2019). Moreover, both approaches often require sub-daily rainfall information at the target site itself for parameter estimation, which makes it difficult to properly validate models and to use them for simulation in locations where only daily records are available.

Nonparametric models, in turn, may provide flexible means for simulating rainfall disaggregation, despite of their limited ability for extrapolation and reproduction of quantiles’ variability (Costa et al., 2015). First, such models are able to reasonably reproduce samples’ summary statistics and, to some extent, the block-maxima behavior for the finer-scale precipitation increments (Lu & Qin, 2014; Li et al., 2018; Diez-Sierra & del Jesus, 2019). In addition, the very nature of nonparametric models

avoids establishing assumptions regarding both the underlying distribution of the stochastic process from which sub-daily rainfall stems and the functional relationship between continuous and coarse-aggregated rainfall amounts, which may reduce modeling uncertainty (Pui et al., 2012; Sivakumar, 2017). Finally, once some degree of hydrological similarity between gauging stations is established, the aggregation of regional information for inference in the target site is straightforward (Westra et al., 2012; Li et al., 2018), which may be useful for circumventing the requirements of fine-scale information at the referred location.

One of the earliest nonparametric techniques comprises the so-called method of fragments, originally proposed by Svanidze, in 1961, for stochastic generation of streamflows (Svanidze, 1980). Since its introduction, the method of fragments has been widely used for precipitation and streamflow disaggregation (see, for instance, Li et al., 2018 and references therein). In the context of rainfall disaggregation, the method of fragments is often coupled with the k -nearest neighbors approach discussed in Lall & Sharma (1996), which resorts to the bootstrapping technique (Efron, 1979) for sampling rainfall amounts from the most similar events in a given sample (for the sake of brevity, the outlined coupled model is hereafter termed KNN-MOF).

Despite its widespread use, the KNN-MOF approach, as originally designed, presented two theoretical underpins: (i) the continuity at the boundaries, i.e., in successive days, is not warranted, which may entail an unsuitable reproduction of clustered rainfall events; and (ii) high-resolution rainfall information is required at the target site. For addressing these issues, in this paper we explore an alternative version of the KNN-MOF model, which utilizes a state-based logic for simulating consecutive wet days (Sharma & Srikanthan, 2006) and a regionalized similarity-based approach for sampling fragments from hydrologically similar nearby stations (Westra et al., 2012; Mehrotra et al., 2012), eliminating the needs of sub-daily records in the location of interest. By resorting to the outlined rationale, one may expect to obtain a well-suited disaggregation model, which could provide a reasonable framework for estimating short-duration design events from coarse-resolution records.

The remainder of the paper is organized as follows. The next section presents the material and methods, which encompass a brief description of the study area and the dataset, and describes with detail the simulation algorithm, the measure of hydrological similarity between rainfall stations and the strategy for model validation. Next, the case study is discussed. Finally, the conclusions and envisaged research developments are addressed.

MATERIAL AND METHODS

Study area and dataset

The proposed disaggregation method is applied to a set of 40 rainfall gauging stations in sub-basins 40 and 41, which are included in the São Francisco river catchment, and in sub-basin 56, which is a component of the Doce river catchment. Sub-daily rainfall data were made available by the Brazilian Geological Service (CPRM), which operates the selected gauging stations. The data set presents data resolution of 1 minute and the average size of

the utilized sub-daily time series is 6 years, with average yearly precipitation amounts varying from 800 to 1650 mm. Those days with missing values were simply discarded from the analysis.

The physical attributes of the catchments used in this study were obtained from the digital platform of the Brazilian Water National Agency (Agência Nacional de Águas, 2019). To identify the spatial distribution of those stations a GIS platform is employed.

For validating the disaggregation model, we utilized the rainfall gauging station with the longest record of sub-daily rainfall in the study area, namely, the Caeté gauging station (code ANA 01943010), located in the sub-basin 40. High-resolution rainfall data in such a station encompasses 28 years, spanning from 1990-2017, with average yearly precipitation amount of 1260 mm. However, 5 years of this data were removed due to high number of missing values. Figure 1 presents the study area, highlighting the target site of Caeté.

Stochastic simulation of sub-daily precipitation with the KNN-MOF model

In this paper, we utilize a variant of the *k*-nearest neighbors resampling method developed by Lall & Sharma (1996), which includes a regional similarity-based approach, as discussed in Westra et al. (2012), for synthetically generating sub-daily rainfall sequences. The rationale behind the proposed approach is that, for the set of wet days, one may consider that the joint distribution of some attribute of sub-daily rainfall, associated to a particular duration, and the daily rainfall amounts is the same at the target site and at a set of nearby stations, entailing hydrologic similarity. Rainfall attributes must be defined as to uncover as much information as possible regarding the full sequence of sub-daily rainfall increments. Once hydrologic similarity is established, logistic regression models are employed in order to identify the neighboring gauging stations with higher probability of similarity with the (ungauged) target site, on the basis of the catchments’

physical attributes, which, in turn, allows information transfer for rainfall disaggregation.

The resampling algorithm is based on the random selection of sub-daily rainfall “fragments” from hydrologically similar nearby gauging stations. The so-called “fragments” are defined as vectors containing the full sub-daily rainfall sequence, aggregated with respect to a particular duration and normalized by the total rainfall amount observed in a given day. For defining hydrological similarity, we employ the following hydrological attributes, as computed for the durations of 60, 180 and 360 minutes, for each wet day (Westra et al., 2012):

- The maximum rainfall intensity, for each duration, normalized by the total rainfall amount for that day;
- The fraction of zeros, i.e., for a particular duration, the percentage of intervals with no rainfall; and
- The timing of maximum rainfall intensity, which accounts for the time of the day in which the rainfall burst of maximum intensity occurs.

For assessing the hydrological similarity between two stations, we resort to the two-dimensional two-sample Kolmogorov Smirnov test (2D2S KS), as proposed by Fasano & Franceschini (1987), at the significance level of 5%. Such a test provides evidence as to whether the joint distributions of each of the hydrological attributes and the daily rainfall amounts time series are statistically similar in a pair of gauging stations. The 2D2S KS test statistic is given by the following equation (Press & Teukolsky, 1988):

$$P(D > observed) = Q_{KS} \left(\frac{D\sqrt{N}}{1 + \sqrt{1 - r^2 \left(0.25 - \frac{0.75}{\sqrt{N}} \right)}} \right) \tag{1}$$

in which,

$$N = \frac{N_1 N_2}{N_1 + N_2} \tag{2}$$

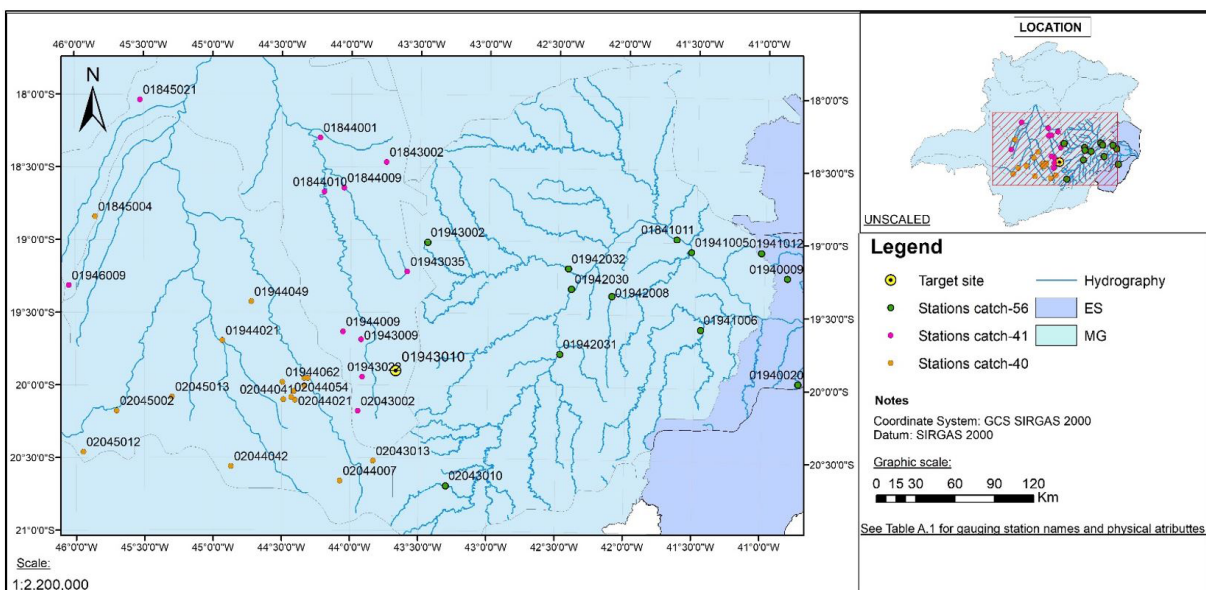


Figure 1. Location of the gauging stations for this case study.

N_1 and N_2 denote, respectively, the number of sample points in samples 1 and 2, D is the maximum difference of the integrated probabilities along the four quadrants defined over a given sample point (see Press et al., 1992 for details), and r is the complement of the average Pearson correlation coefficients, as estimated from both joint distributions. For computing the term in the right-hand side of Equation 1, one must numerically approximate the following expression (Westra et al., 2012):

$$Q_{KS}(x) = 2 \sum_{j=1}^{\infty} (-1)^{j-1} e^{-2j^2x^2} \quad (3)$$

For additional information on the 2D2S KS test, the reader is kindly referred to Press & Teukolsky (1988) and Press et al. (1992). The similarity between two stations is summarized by the 2D2S KS test result, which comprises a binary response u , with value of 1 being attributed to statistically similar stations and a value of 0 otherwise.

For modeling the binary response, as obtained from the hydrological attributes' similarity evaluation, a logistic regression model, which utilizes physical catchments' attributes as predictors, is established. Here, we consider a set of 4 potential predictors, namely, the absolute differences in latitude, in longitude, in elevation and in the product between the differences in latitude and longitude, which is intended to be a proxy for the Euclidian distance between the two gauging stations. In formal terms:

$$P(u=1) = \text{logit}(z) = \frac{e^z}{e^z + 1} \quad (4)$$

in which,

$$z = \beta_0 + \beta_1 v_1 + \beta_2 v_2 + \beta_3 v_3 + \beta_4 v_4 \quad (5)$$

with vector β encompassing the regression coefficients and vector v the previously outlined predictors.

Equation 4 provides the probability that two rainfall gauging stations are similar, conditioned on the physical attributes of the catchments. For disaggregating the daily rainfall amounts in the target site, a set of nearby stations with higher values of $P(u=1)$ must be retained. It should be noted that no formal procedures for variable selection are employed in regression analysis since we intend to evaluate the influence of each of these predictors with respect to the considered rainfall durations.

Following Westra et al. (2012), we select the neighbors on the basis of the average probability of similarity stemming from the three hydrological attributes. Seasonality is dealt with by assessing similarity for different periods defined along the water year (Oct.-Sep.), which, for our purposes, were defined as to coincide with the 4 seasons in the southern hemisphere. After identifying the most similar nearby stations, the disaggregation algorithm can be applied to the daily rainfall time series at the target site. In short, the algorithm works as follows:

1. The sequence of daily rainfall amounts at the target site is derived from the historical record;
2. Sequences of daily rainfall are obtained in the nearby stations identified with the previously described regional similarity-based approach. Dimensionless fragments (fr),

for each season, day and duration, are then estimated with Equation 6:

$$fr_{i,m}^s = \frac{X_{i,m}^s}{\sum_m X_{i,m}^s} \quad (6)$$

in which $X_{i,m}^s$ denotes the sub-daily rainfall amount in the gauging station s , aggregated along a duration m , in a given day i ;

3. For each wet day, similar daily rainfall depths are sought at the nearby stations. For accounting for seasonal characteristics in sub-daily sequences, the referred daily amounts are only evaluated within a moving window of ± 15 days (Pui et al., 2012; Westra et al., 2012), centered at the day of interest. Furthermore, for allowing the proper simulation of successive wet days, the sampling strategy is also conditioned on the previous and next days' states (wet or dry);
4. For a given day, all sample points along the period of record located inside the moving window and with similar wetness previous and next-day states are extracted from the set of neighbors. Such sample points are then ranked with respect to the absolute deviation from the daily amount at the target site, and the k -nearest neighbors are kept for resampling;
5. A random number is drawn from a standard uniform distribution and compared to the probability expressed by Equation 7:

$$P(j) = \frac{1/(j)}{\sum_{i=1}^k 1/i} \quad (7)$$

in which j denotes the rank of each of the k retained fragments. This allows the random selection of a fragment; and

6. The sub-daily rainfall amount is then computed by multiplying the fragment, as selected in previous step, by the daily precipitation depth at the day of interest.

After identifying the disaggregation models, for each season and duration, we utilize a rainfall gauging station with measurements encompassing a long period-of-record for assessing their predictive abilities. At each sub-daily scale, a set of 100 model runs (Westra et al., 2012) is simulated at the target site with the disaggregation algorithm. The performance of the models is then evaluated by comparing the mean values of a set of summary statistics for wet days, which encompasses the mean, the variance, the coefficient of skewness and the quantile curve, for the sub-daily temporal scales, with the observed counterparts. Also, for the quantile curves, the model ability in synthesizing the variability of the maximum rainfall amounts is assessed through the computation of the 95% confidence intervals.

RESULTS AND DISCUSSION

After obtaining the sub-daily rainfall data in the 40 rainfall gauging stations, for the durations of 60, 180 and 360 minutes, the hydrological attributes – maximum intensity, fraction of zeros and timing of the maximum intensity – were estimated for each site.

Table 1. Regression model coefficients for the duration of 60 minutes (quantities inside brackets denote the p -values for the estimates).

Maximum Intensity					
Season	Intercept (β_0)	Latitude (β_1)	Longitude (β_2)	LatxLong (β_3)	Elevation (β_4)
DJF	-0.3816 [0.00]	-0.1007 [0.00]	-0.2219 [0.00]	0.0426 [0.42]	-0.0004 [0.02]
MAM	-1.0302 [0.00]	-0.2184 [0.65]	-0.2250 [0.00]	0.0841 [0.41]	0.0001 [0.42]
JJA	0.4919 [0.00]	-0.1374 [0.11]	-0.3586 [0.00]	-0.0528 [0.12]	-0.0007 [0.00]
SON	-0.0017 [0.25]	0.0060 [0.11]	-0.1897 [0.00]	0.1301 [0.21]	-0.0009 [0.58]
Timing of the Maximum Intensity					
Season	Intercept (β_0)	Latitude (β_1)	Longitude (β_2)	LatxLong (β_3)	Elevation (β_4)
DJF	-4.7091 [0.00]	-0.4951 [0.42]	-0.2871 [0.75]	0.0463 [0.98]	-0.0005 [0.013]
MAM	-3.2334 [0.00]	0.1207 [0.84]	-0.6331 [0.04]	-0.1007 [0.17]	0.0001 [0.90]
JJA	-0.5423 [0.34]	-0.1544 [0.00]	-0.3661 [0.00]	-0.1239 [0.18]	-0.0008 [0.01]
SON	-1.9019 [0.00]	-0.3080 [0.19]	-0.0247 [0.55]	-0.1183 [0.47]	-0.0011 [0.12]

DJF = December to February; MAM = March to May; JJA = July to June; SON = September to November.

Next, the 2D2S KS test was employed for assessing the similarity of pairwise stations with respect to each of the hydrological attributes and each season.

Results from the 2D2S KS test demonstrated that, for the maximum intensity and time of maximum intensity attributes, the model presented an appropriate performance, identifying similarity for a number of stations. On the other hand, for the remaining attribute, i.e., the fraction of zeros, similarity was not readily established, except for the period spanning from June to August, which may be partially ascribed to the small number of wet days in these months and to the relatively small precipitated amounts in such days. A similar problem was reported by Westra et al. (2012) when disaggregating daily rainfall amounts in stations spread across Australia, which may, to some extent, point to a flawed proposal of hydrological attributes for defining similarity. As a result, the fraction of zeros attribute was not considered for subsequent analyses.

The next step was identifying the logistic regression models, which relate the binary response of each hydrological attribute to the catchments' physical characteristics. Table A1 (Appendix A) presents a summary of the elevation, latitude, longitude for the gauging stations, including those of the target site. These attributes were used as predictors in the regression model (which does not consider the Caeté station). Regression coefficients were then estimated by means of the iteratively reweighted least squares approach. Table 1 presents the regression coefficients for each of the hydrological attributes for the duration of 60 minutes and for each season. Results for the remaining durations are presented in Tables A2 and A3 also in Appendix A. One may notice that the difference in longitude is the physical characteristic with higher coefficient values (significant for most durations at the 5% level), which suggests that, for the study area, it is the most influencing physical factor to identify similarity.

For each hydrological attribute, the probability of a given nearby station of being similar to the target site was computed with the obtained regression model. For identifying the donor sites, the mean value of the two probabilities was estimated in each season, and the best ranked stations were kept for rainfall disaggregation. Table A4 (Appendix A) presents the averaged probabilities, for the duration of 60 minutes, with respect to the Caeté station. As it can be seen, the probabilities may vary

Table 2. Relative errors between the observed summary statistics and the predictions' ensemble mean value for distinct numbers of neighbors.

Statistics	k = 10	Maximum number of k
Mean of the annual maximum	4.94%	4.21%
Mean	0.50%	0.43%
Variance	0.16%	-0.38%
Skewness	4.47%	3.64%

depending on the season, presenting higher values in less rainy periods, making it easier to correlate the attributes. Also, Table A4 indicates the need to separate the model by seasons, as different stations were selected for each one.

After defining potential donor sites, we proceeded to the calibration of the disaggregation model. This process requires the specification of the k -nearest neighbors (k), the number of donor gauging stations (S) and the maximum absolute deviation of the daily rainfall amounts with respect to that observed in the target site. The summary statistics for the wet days were calculated to evaluate model performance.

The first parameter to be estimated was the number of neighbors, k . For assessing the impact of this parameter in the model, we fixed the remaining parameters, using $S = 10$ and the deviation = 10%. Results are depicted in Table 2, which also presents the relative error between the observed records and the mean simulated value (i.e., the mean value of the predictions' ensemble), for each statistic, for the 60 minutes duration. Positive values indicate overestimation and negative values underestimation.

From Table 2, one may observe that, in this case study, the usage of the number maximum of neighbors does not have a substantial impact in the model, due to the difficulties to obtain more than 10 days that are within the maximum absolute deviation of 10%. Considering the number of data and gauging stations that Westra et al. (2012) had available for their work, this parameter would be essential in order to limit the number of days to resample. Therefore, considering the mentioned aspects, this parameter will not be employed for further analyses and durations. The second parameter to be estimated was the maximum absolute deviation of the daily rainfall amounts. For the duration of 60 minutes and

keeping $S = 10$, we tested deviations of 5% and 10%, in order to assess which one would entail a better performance. Results are presented by Table 3. One may notice that decreasing the deviation to 5% did not result in substantial changes in the statistical properties and also caused a reduction of model variability, as there are less fragments selected to be resampled. In view of this result, we defined the deviation of 10% for the remaining simulations.

Lastly, to estimate the number of gauging stations (S) to be used for disaggregation of the target site, we considered the values of

Table 3. Relative errors between the observed summary statistics and the predictions' ensemble mean value for distinct deviations with respect to daily rainfall amount in the target site.

Statistics	Deviation = 5%	Deviation = 10%
Mean of the annual maximum	2.87%	4.21%
Mean	-0.44%	0.43%
Variance	-1.81%	-0.38%
Skewness	3.40%	3.64%

5, 10 and 15. Results for the 60 minutes duration are presented in Table 4 and those for the remaining durations are located in Tables A5 and A6 in Appendix A. One may observe that, for the duration of 60 minutes, the model was more efficient in reproducing the observed values when 15 stations were used in the disaggregation. However, when increasing the duration to 180 or 360 minutes, the model was able to disaggregate the series at the target site with less neighboring stations.

Considering the results obtained from the calibration step, which are the maximum number for the k -nearest neighbors, the maximum absolute deviation of 10% and number of donor gauging stations (S) of 15 for 60 minutes duration and 5 for durations of 180 and 360 minutes, the final simulations were carried out. 100 model runs were performed and, on each run, the observed daily time series was disaggregated. As the fragments are randomly chosen (step 5 of the disaggregation algorithm), one may expect different outcomes at each run, which, in turn, should converge for the observed summary statistics on average terms. Figure 2 presents the results of each one of the 100 model runs for the 60 minutes duration, considering the assessed statistical

Table 4. Comparison between the observed summary statistics and the predictions' ensemble mean value for distinct numbers of rainfall gauging stations "donors" (S) – 60 minutes duration.

Statistics	Records	S = 5		S = 10		S = 15	
		Simulated	Deviation	Simulated	Deviation	Simulated	Deviation
Mean of the annual maximum	36.98	38.90	5.20%	38.53	4.21%	38.38	3.79%
Mean	2.40	2.45	2.11%	2.41	0.43%	2.41	0.51%
Variance	17.58	18.21	3.58%	17.51	-0.38%	17.35	-1.33%
Skewness	3.98	4.06	2.05%	4.13	3.64%	4.04	1.61%

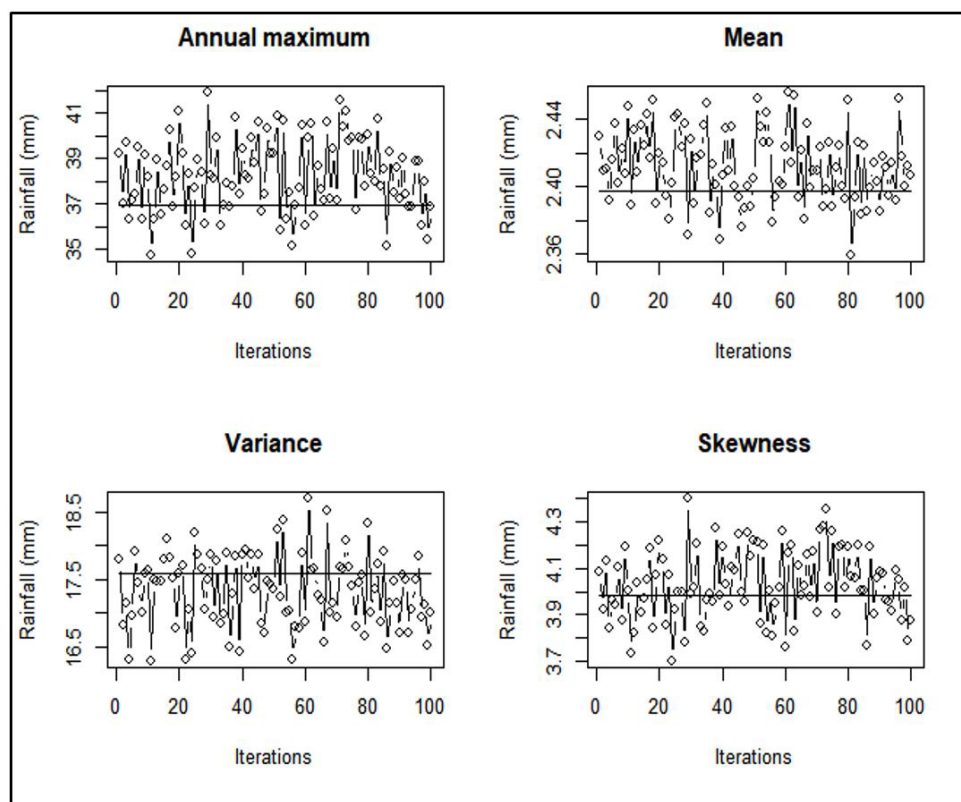


Figure 2. Annual maximum, mean, variance and skewness results for each one of the 100 model runs for the 60 minutes duration. Dots denote model predictions and the solid horizontal line the mean value of the observed statistics.

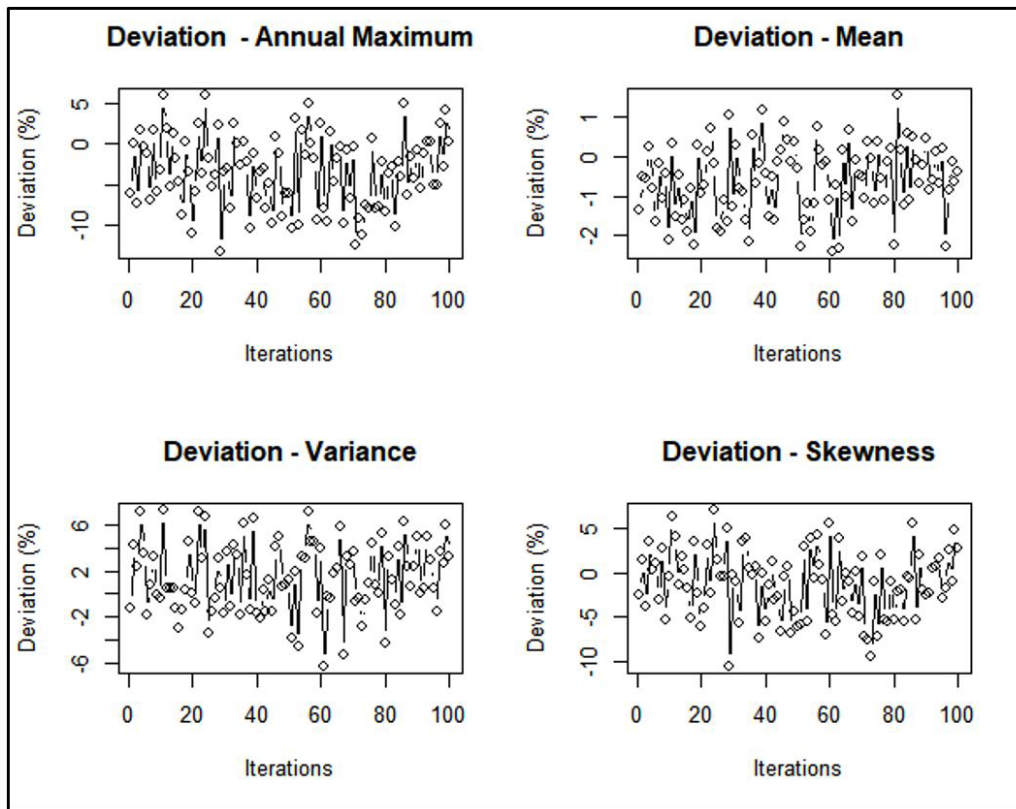


Figure 3. Deviations of model predictions from the observed summary statistics of rainfall amounts for the 60 minutes duration.

properties. The continuous horizontal line represents the mean values of the observed statistics and each point represent the model predictions for each iteration.

Regarding Figure 3, it represents the deviations from the observed values for each iteration, for the duration of 60 minutes. For other durations, the reader is kindly referred to Figures A1 to A4 located in Appendix A. For all statistical properties and durations, the model resulted on deviations of $\pm 10\%$, which indicates an appropriate performance. It is important to highlight that, for the duration of 60 minutes, the model can better reproduce the observed mean and variance, as compared to the other durations, which tend to overestimate them. Other important aspect is that this work considered considerably less data than Westra et al. (2012) to resample the fragments for each simulation and yet presented reasonable results. The mentioned authors had available about 250 years of data among the selected gauging stations for resampling, while, in this work, the selected number of stations resulted in approximately 90 years of data for 60 minutes duration (15 stations with 6 years each) and 30 years of data for 180 and 360 minutes duration (5 gauging stations with 6 years each).

To evaluate the model ability in synthetizing the variability of the annual maximum rainfall amounts, the quantile curves were constructed considering the 95% confidence intervals (CI), as shown in Figures 4 to 6. The hollow points indicated the recorded values from the gauging station, whereas the continuous line represents the median of the 100 simulations and the dashed lines represent the confidence bands.

Overall, the results indicate an appropriate model performance to estimate the annual maximum rainfall amounts, mainly for smaller durations and higher exceedance probabilities.

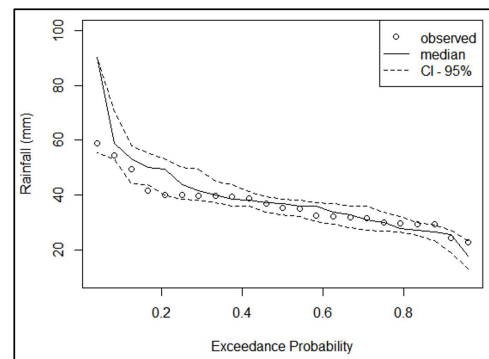


Figure 4. Quantile curve of annual maximum rainfall amounts – 60 minutes duration, including the observed values, median and confidence interval (CI) of 95%.

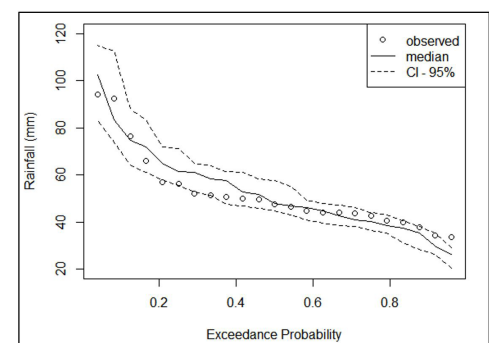


Figure 5. Quantile curve of annual maximum rainfall amounts – 180 minutes duration, including the observed values, median and confidence interval (CI) of 95%.

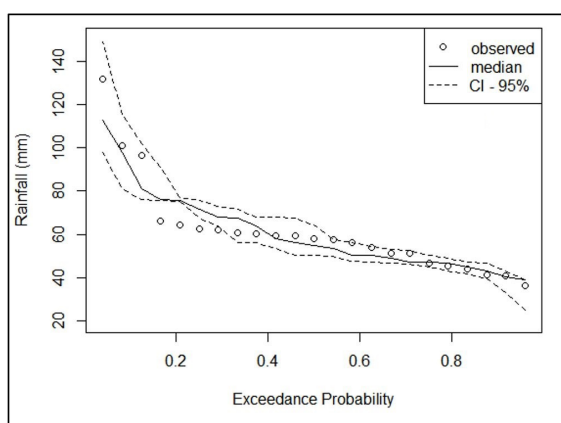


Figure 6. Quantile curve of annual maximum rainfall amounts – 360 minutes duration, including the observed values, median and confidence interval (CI) of 95%.

However, the model tends to overestimate the rainfall amounts with lower exceedance probability for the 60 minutes duration and underestimate them for the 360 minutes duration.

It is worth highlighting that, for the durations of 60 and 180 minutes, most of the observed values fall inside the confidence bands, which suggests that, to some extent, the resampling algorithm is able to capture the expected estimation variability that stems from sampling errors. This was also observed by Westra et al. (2012) and Li et al. (2018). For the 360 minutes duration, despite enclosing most observed points, the model worsens its performance, with narrow confidence intervals, which are not being able to resample with variability. Further analysis must be made in order to define a method to compensate and deal with this aspect.

In order to visualize the spatial distributions of the selected gauging's obtained from the similarity probabilities, Figures A5 and A6 (Appendix A) were produced. The stations employed to disaggregate the target site daily time series for the duration of 60 minutes are indicated for each season, with the rank of each station located previous to its code. As the observed behavior is similar for the other durations, we presented in this paper the figures only for the mentioned duration.

As it can be seen, for the period of December to February (DJF) the model selects the geographically nearest stations, which are probably affected by similar meteorological systems, as the most similar - even though the climate conditions are not expected to vary considerably due to the relatively small size of the study area. The behavior for the periods between March to May (MAM) and September to November (SON) are quite similar and, in these cases, the model selects more distant stations - yet still in the same catchment of the target site. For the period between June and August (JJA), which is the driest season of the year, it is easier to identify similarity due to the small number of rainy days. This fact explains the reason for the selection of distant stations instead of the close ones. As such, regarding the rank of similarity probability, it is interesting to observe that the highest ranked stations are often distant from the target site, indicating the hydrological attributes have higher influence over the physical counterparts for some stations.

CONCLUSIONS

This article presented a method for disaggregating daily rainfall, which is based on an alternative version of the KNN-MOF model that utilizes a state-based logic for simulating consecutive wet days (Sharma & Srikanthan, 2006) and a regionalized similarity-based approach for sampling fragments from hydrologically similar nearby stations (Westra et al., 2012). Similarity identification was made by considering hydrological attributes, such as maximum intensity and timing of the maximum intensity, and physical attributes, such as absolute difference in latitude, in longitude, and in the product between the differences in latitude and longitude.

Considering the hydrological attributes for identifying stations' similarity, the maximum intensity and timing of the maximum intensity performed appropriately. However, due to difficulties regarding the fraction of zeros, such an attribute was not employed in this study. Concerning the physical attributes, the one with higher influence was the longitude, which was identified through the logistic regression model coefficients.

The similarity probabilities, which are the results of combining both hydrological and physical attributes, presented different values depending on the season of the year, with higher values in less rainy periods, making it easier to correlate the attributes. Also, it is important to consider seasonality, as different stations are selected for each season.

After selecting potential donor sites, we proceeded to the calibration of the disaggregation model, identifying which parameters resulted in better model performance, considering the statistics summary for the wet days. The "optimum" model was achieved considering the non-specification of a number for the k -nearest neighbors, the maximum absolute deviation of 10% and number of gauging stations (S) of 15 for 60 minutes duration and 5 for durations of 180 and 360 minutes.

The simulations indicated an appropriate model performance when comparing to the statistics obtained for the validation gauging station of Caeté. Regarding the annual maximum rainfall amounts, the quantile curves constructed show the most part of observed values are between the 95% interval. However, the model presents narrow intervals for higher rainfall durations, which implies difficulties in representing the expected quantiles' variability for such durations.

Also, the model is able to appropriately reproduce the quantiles of low exceedance probability for the 180 minutes duration, indicating its capacity to reproduce extreme rainfall amounts which are important for estimating storm events for the design of drainage systems and performing rainfall-runoff simulations. For the 60 minutes duration there is an overestimation and for the 360 minutes duration there is an underestimation of such quantiles.

We acknowledge that further analyses must be made in order to compensate and deal with the aforementioned difficulties and a broader comparison with other methods with distinct theoretical foundations and complexity levels, such as that of the disaggregation coefficients (Teodoro et al., 2014), is necessary for allowing some generalization of the proposed approach. However, the obtained results are deemed promising for addressing the problem of rainfall disaggregation, particularly in what concerns extreme events, when only daily records are available.

ACKNOWLEDGEMENTS

The authors wish to thank the agencies CAPES (“Coordenação de Aperfeiçoamento de Pessoal de Nível Superior”) and CNPq (“Conselho Nacional de Desenvolvimento Científico e Tecnológico”) for their support to this research. The authors also wish to acknowledge the anonymous reviewers and editors for the valuable comments and suggestions, which helped improving the paper.

REFERENCES

- Agência Nacional de Águas – ANA. (2019). *HidroWeb v3.1.0*. Brasília. Retrieved in 2019, May 2, from www.snirh.gov.br/hidroweb/publico/mapa_hidroweb.jsfr
- Costa, V., Fernandes, W., & Naghettini, M. (2015). A Bayesian model for stochastic generation of daily precipitation using an upper-bounded distribution function. *Stochastic Environmental Research and Risk Assessment*, 29(2), 563-576. <http://dx.doi.org/10.1007/s00477-014-0880-9>.
- Diez-Sierra, J., & del Jesus, M. (2019). Subdaily rainfall estimation through daily rainfall downscaling using random forests in Spain. *Water*, 11(1), 125. <http://dx.doi.org/10.3390/w11010125>.
- Efron, B. (1979). Bootstrap methods: another look at the Jackknife. *Annals of Statistics*, 7(1), 1-26. <http://dx.doi.org/10.1214/aos/1176344552>.
- Fasano, G., & Franceschini, A. (1987). A multidimensional version of the Kolmogorov-Smirnov test. *Monthly Notices of the Royal Astronomical Society*, 225(1), 155-170. <http://dx.doi.org/10.1093/mnras/225.1.155>.
- Kang, B., & Ramirez, J. A. (2010). A coupled stochastic space-time intermittent random cascade model for rainfall downscaling. *Water Resources Research*, 46(10), W10534. <http://dx.doi.org/10.1029/2008WR007692>.
- Koutsoyiannis, D., & Onof, C. (2001). Rainfall disaggregation using adjusting procedures on a Poisson cluster model. *Journal of Hydrology*, 246(1-4), 109-122. [http://dx.doi.org/10.1016/S0022-1694\(01\)00363-8](http://dx.doi.org/10.1016/S0022-1694(01)00363-8).
- Lall, U., & Sharma, A. (1996). A nearest neighbor bootstrap for resampling hydrological time series. *Water Resources Research*, 32(3), 679-693. <http://dx.doi.org/10.1029/95WR02966>.
- Li, X., Meshgi, A., Wang, X., Zhang, J., Tay, S. H. X., Pijcke, G., Manocha, N., Ong, M., Nguyen, M. T., & Babovic, V. (2018). Three resampling approaches based on method of fragments for daily-to-subdaily precipitation disaggregation. *International Journal of Climatology*, 38, e1119-e1138. <http://dx.doi.org/10.1002/joc.5438>.
- Lu, Y., & Qin, X. (2014). Multisite rainfall downscaling and disaggregation in a tropical urban area. *Journal of Hydrology*, 509, 55-65. <http://dx.doi.org/10.1016/j.jhydrol.2013.11.027>.
- Mehrotra, R., Westra, S., Sharma, A., & Srikanthan, R. (2012). Continuous rainfall simulation: 2. A regionalized daily rainfall generation approach. *Water Resources Research*, 48(1), W01536. <http://dx.doi.org/10.1029/2011WR010490>.
- Müller, H., & Haberlandt, U. (2018). Temporal rainfall disaggregation using a multiplicative cascade model for spatial application in urban hydrology. *Journal of Hydrology*, 556, 847-864. <http://dx.doi.org/10.1016/j.jhydrol.2016.01.031>.
- Press, W. H., & Teukolsky, S. A. (1988). Kolmogorov-Smirnov Test for Two-Dimensional Data. *Computers in Physics*, 2(4), 74-77. <http://dx.doi.org/10.1063/1.4822753>.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical recipes in Fortran: the art of scientific computing* (2nd ed.). Cambridge: Cambridge University Press.
- Pui, A., Sharma, A., Mehrotra, R., Sivakumar, B., & Jeremiah, E. (2012). A comparison of alternatives for daily to sub-daily rainfall disaggregation. *Journal of Hydrology*, 470, 138-157. <http://dx.doi.org/10.1016/j.jhydrol.2012.08.041>.
- Serinaldi, F. (2010). Multifractality, imperfect scaling and hydrological properties of rainfall time series simulated by continuous universal multifractal and discrete random cascades. *Nonlinear Processes in Geophysics*, 17(6), 697-714. <http://dx.doi.org/10.5194/npg-17-697-2010>.
- Sharma, A., & Srikanthan, R. (2006). Continuous rainfall simulation: a nonparametric alternative. In *30th Hydrology and Water Resources Symposium*. Launceston, Tasmania: Conference Design.
- Sharma, A., & Mehrotra, R. (2010). Rainfall generation. In M. Gebremichael (Ed.), *Rainfall: state of the science* (pp. 215-246). Washington: American Geophysical Union. <http://dx.doi.org/10.1029/2010GM000973>.
- Sivakumar, B. (2017). Section 3.6: nonparametric methods. In B. Sivakumar, *Chaos in hydrology: bridging determinism and stochasticity*. Switzerland: Springer. <http://dx.doi.org/10.1007/978-90-481-2552-4>.
- Sivakumar, B., & Sharma, A. (2008). A cascade approach to continuous rainfall data generation at point locations. *Stochastic Environmental Research and Risk Analysis*, 22(4), 451-459. <http://dx.doi.org/10.1007/s00477-007-0145-y>.
- Svanidze, G. G. (1980). *Mathematical modeling of hydrologic systems*. Fort Collins: Water Resources Publications.
- Teodoro, P. E., Neivock, M. P., Marques, J. R. F., Flores, A. M. F., & Rodrigues, C. C. B. (2014). Influência de diferentes coeficientes de desagregação na determinação de equações IDF para Aquidauana/MS. *Revista Eletrônica de Engenharia Civil*, 9(2), 1-9.
- Westra, S., Mehrotra, R., Sharma, A., & Srikanthan, R. (2012). Continuous rainfall simulation: 1. A regionalized subdaily

disaggregation approach. *Water Resources Research*, 48(1), 1535. <http://dx.doi.org/10.1029/2011WR010489>.

Yusop, Z., Nasir, H., & Yusof, F. (2014). Disaggregation of daily rainfall data using Bartlett Lewis Rectangular Pulse Model: a case study in central Peninsular Malaysia. *Environmental Earth Sciences*, 71(8), 3627-3640. <http://dx.doi.org/10.1007/s12665-013-2755-7>.

Authors contributions

Milena Guerra de Aguiar: Literature review, generation and discussion of results and paper writing.

Veber Afonso Figueiredo Costa: Contribution in the paper conception, results discussion and paper writing.

Appendix A. Results' tables and figures.

Table A1. Summary of the catchments' physical characteristics.

Station code	Gauge Station Name	Latitude	Longitude	Elevation	Data Extent
01943010	CAETE (Estação de interesse)	19° 54' 10"	43° 39' 59"	825	1990-2017
02043002	LAGOA GRANDE (MMV)	20° 10' 45"	43° 56' 34"	1326	2000-2006
01943022	CAIXA DE AREIA	19° 57' 02"	43° 54' 10"	1159	2000-2006
01943009	VESPASIANO	19° 41' 13"	43° 55' 14"	677	2000-2006
01944009	PEDRO LEOPOLDO	19° 38' 04"	44° 03' 09"	730	2000-2006
01946009	SAO GOTARDO	19° 18' 51"	46° 02' 39"	1086	2000-2005
01943035	VAU DA LAGOA	19° 13' 05"	43° 35' 17"	1085	2000-2006
01844010	PONTE DO LICINIO-JUSANTE	18° 40' 20"	44° 11' 36"	547	2001-2005
01844009	PRESIDENTE JUSCELINO-JUSANTE	18° 38' 41"	44° 03' 02"	576	2001-2005
01843002	GOUVEA	18° 27' 56"	43° 44' 35"	1106	2001-2006
01844001	SANTO HIPOLITO (ANEEL/CEMIG)	18° 18' 21"	44° 13' 32"	530	2001-2005
01845021	CANOEIROS	18° 02' 17"	45° 31' 23"	796	2000-2005
02044007	ENTRE RIOS DE MINAS	20° 39' 37"	44° 04' 18"	871	2000-2006
02044042	CARMO DA MATA (ETA-COPASA)	20° 33' 45"	44° 52' 02"	854	2000-2006
02043013	CONGONHAS-LINIGRAFO	20° 31' 06"	43° 50' 08"	871	2000-2008
02045012	PIUM-I	20° 27' 33"	45° 56' 38"	809	2000-2006
02045002	IGUATAMA	20° 10' 12"	45° 42' 56"	639	2000-2006
02045013	SANTO ANTONIO DO MONTE	20° 05' 03"	45° 17' 48"	969	2000-2006
02044052	JARDIM	20° 02' 50"	44° 24' 32"	779	2000-2007
01944027	JUATUBA	19° 57' 20"	44° 20' 03"	724	2000-2008
01944004	PONTE NOVA DO PARAOPEBA	19° 56' 57"	44° 18' 19"	708	2002-2006
01944021	VELHO DA TAIPA	19° 41' 31"	44° 55' 56"	636	2000-2005
01944049	PAPAGAIOS	19° 25' 38"	44° 43' 11"	741	2000-2006
01845004	LAGOA DO GOUVEIA	18° 49' 59"	45° 50' 26"	1038	2000-2005
01944062	FAZENDA SANTA RITA	19° 58' 58"	44° 29' 32"	820	2000-2008
02044021	ALTO DA BOA VISTA	20° 06' 20"	44° 24' 04"	905	2000-2008
02044024	FAZENDA CURRALINHO	20° 00' 27"	44° 19' 52"	786	2000-2006
02044041	FAZENDA LARANJEIRAS	20° 06' 08"	44° 29' 05"	895	2000-2007
02044054	SERRA AZUL	20° 05' 12"	44° 25' 38"	817	2000-2008
01841011	TUMIRITINGA	18° 58' 15"	41° 38' 30"	137	2000-2006
01940009	PANCAS	19° 13' 51"	40° 50' 07"	112	2000-2006
01940020	CALDEIRAO	19° 57' 17"	40° 44' 30"	694	2000-2006
01941005	BARRA DO CUIETE-JUSANTE	19° 03' 42"	41° 31' 59"	143	2000-2006
01941006	ASSARAI-MONTANTE	19° 35' 39"	41° 27' 29"	156	2000-2006
01941012	ALTO RIO NOVO	19° 03' 29"	41° 01' 39"	535	2000-2006
01942008	DOM CAVATI	19° 22' 26"	42° 06' 07"	319	2000-2006
01942030	CENIBRA	19° 19' 40"	42° 23' 51"	225	2000-2006
01942031	CACHOEIRA DOS OCULOS-MONTANTE	19° 46' 36"	42° 28' 35"	248	2000-2006
01942032	NAQUE VELHO	19° 11' 17"	42° 25' 20"	205	2000-2006
01943002	CONCEICAO DO MATO DENTRO	19° 00' 51"	43° 26' 48"	624	2000-2006
02043010	PIRANGA	20° 41' 17"	43° 18' 02"	608	2000-2008

Table A2. Regression model coefficients for the duration of 180 minutes (quantities inside brackets denote the p -values for the estimates).

Maximum Intensity					
Season	Intercept (β_0)	Latitude (β_1)	Longitude (β_2)	LatxLong (β_3)	Elevation (β_4)
DJF	1.3692 [0.00]	-0.9454 [0.00]	-0.7300 [0.00]	0.3111 [0.02]	-0.0007 [0.03]
MAM	0.4315 [0.04]	-0.4245 [0.04]	-0.6882 [0.00]	0.3248 [0.02]	-0.0001 [0.80]
JJA	2.3737 [0.00]	-0.4146 [0.10]	-1.2252 [0.00]	0.3069 [0.05]	-0.0006 [0.10]
SON	0.9288 [0.00]	0.1989 [0.40]	-0.4499 [0.00]	-0.0438 [0.72]	0.0002 [0.60]
Timing of the Maximum Intensity					
Season	Intercept (β_0)	Latitude (β_1)	Longitude (β_2)	LatxLong (β_3)	Elevation (β_4)
DJF	-26.5661 [1.00]	0.0000 [1.00]	0.0000 [1.00]	0.0000 [1.00]	0.0000 [1.00]
MAM	-3.7642 [0.00]	-0.1715 [0.85]	-0.3550 [0.58]	0.0584 [0.93]	-0.0009 [0.62]
JJA	-0.2821 [0.20]	0.6195 [0.00]	-0.5347 [0.00]	-0.1660 [0.32]	-0.0007 [0.10]
SON	-3.4161 [0.00]	0.2574 [0.76]	-0.4279 [0.58]	-0.2286 [0.79]	-0.0027 [0.24]

DJF = December to February; MAM = March to May; JJA = July to June; SON = September to November.

Table A3. Regression model coefficients for the duration of 360 minutes (quantities inside brackets denote the p -values for the estimates).

Maximum Intensity					
Season	Intercept (β_0)	Latitude (β_1)	Longitude (β_2)	LatxLong (β_3)	Elevation (β_4)
DJF	1.4108 [0.00]	-1.0828 [0.00]	-0.7838 [0.00]	0.4526 [0.00]	-0.0009 [0.01]
MAM	0.3760 [0.07]	-0.2652 [0.20]	-0.6430 [0.00]	0.3037 [0.02]	-0.0003 [0.46]
JJA	2.0095 [0.00]	-0.2586 [0.28]	-0.9016 [0.00]	0.1726 [0.22]	-0.0007 [0.06]
SON	1.0502 [0.00]	0.0903 [0.68]	-0.4713 [0.00]	-0.0203 [0.87]	0.0001 [0.77]
Timing of the Maximum Intensity					
Season	Intercept (β_0)	Latitude (β_1)	Longitude (β_2)	LatxLong (β_3)	Elevation (β_4)
DJF	-26.5661 [1.00]	0.0000 [1.00]	0.0000 [1.00]	0.0000 [1.00]	0.0000 [1.00]
MAM	-4.8886[0.04]	0.8225 [0.72]	-2.2739 [0.70]	0.7356 [0.89]	-0.0032 [0.70]
JJA	-0.9126 [0.00]	-0.4354 [0.07]	-0.8159 [0.00]	0.1582 [0.43]	-0.0003 [0.51]
SON	-26.5661 [1.00]	0.0000 [1.00]	0.0000 [1.00]	0.0000 [1.00]	0.0000 [1.00]

DJF = December to February; MAM = March to May; JJA = July to June; SON = September to November.

Table A4. Similarity Probabilities for each season – Duration of 60 minutes.

DJF		MAM		JJA		SON	
Station Code	Probability	Station Code	Probability	Station Code	Probability	Station Code	Probability
01943009	0.320	01943035	0.199	01843002	0.685	01843002	0.418
02044024	0.315	02043002	0.198	02043013	0.637	01844009	0.382
01943022	0.314	01943022	0.197	01943002	0.624	01844001	0.379
01944009	0.314	01843002	0.191	01844009	0.623	01943002	0.379
01944004	0.311	01943009	0.189	02044007	0.618	02043013	0.378
01944027	0.311	02043013	0.188	01844001	0.611	01943035	0.372
01944062	0.309	01943002	0.188	01943035	0.610	02044007	0.369
02043013	0.306	02043010	0.182	01844010	0.589	01844010	0.365
02044054	0.306	01844009	0.180	02043010	0.588	02043010	0.360
02044052	0.305	01944009	0.179	01944009	0.564	01943009	0.341
02044021	0.296	01844010	0.176	01943009	0.563	01944009	0.340
02044041	0.292	01844001	0.175	02044054	0.532	02044054	0.321
02044007	0.278	02044007	0.175	02044024	0.530	02044024	0.321
01943035	0.276	01944004	0.164	02044052	0.522	01943022	0.320
02043002	0.272	01944027	0.162	02044021	0.520	02043002	0.319
01943002	0.256	02044024	0.160	01944062	0.517	02044021	0.318
02043010	0.254	02044021	0.158	02044041	0.512	02044052	0.317
01944049	0.251	02044052	0.156	01944027	0.512	01944004	0.314
01944021	0.243	02044054	0.154	01944004	0.511	01944027	0.314
02044042	0.235	02044041	0.153	01943022	0.507	02044041	0.314
02045013	0.224	01942032	0.153	02044042	0.498	01944062	0.313
01942031	0.210	01944062	0.150	01944049	0.496	01944049	0.308
01844009	0.207	01942030	0.149	02043002	0.487	02044042	0.308
01843002	0.204	01942031	0.147	01944021	0.431	01944021	0.280
01844010	0.198	01845021	0.144	01845021	0.420	01845021	0.273
02045002	0.185	01944049	0.144	02045013	0.397	01942032	0.267
02045012	0.174	02044042	0.138	01942032	0.361	01942030	0.263
01942030	0.169	01944021	0.134	01942030	0.356	02045013	0.262
01844001	0.166	01942008	0.133	01942031	0.352	01942031	0.258
01942008	0.164	01841011	0.129	02045012	0.347	01942008	0.249
01942032	0.159	01941005	0.123	02045002	0.340	02045002	0.237
01940020	0.145	02045013	0.116	01942008	0.336	02045012	0.235
01946009	0.140	01845004	0.116	01845004	0.317	01845004	0.229
01845021	0.133	01941006	0.108	01946009	0.284	01946009	0.214
01845004	0.132	02045002	0.102	01940020	0.264	01841011	0.212
01941006	0.125	01946009	0.099	01841011	0.247	01941005	0.205
01941012	0.113	01941012	0.098	01941012	0.242	01941006	0.202
01841011	0.104	01940009	0.097	01941005	0.235	01940020	0.199
01941005	0.104	02045012	0.096	01941006	0.232	01941012	0.194
01940009	0.081	01940020	0.069	01940009	0.162	01940009	0.164

DJF = December to February; MAM = March to May; JJA = July to June; SON = September to November.

Table A5. Comparison between the observed summary statistics and the predictions' ensemble mean value for distinct numbers of rainfall gauging stations "donors" (S) – 180 minutes duration.

Statistics	Records	S = 5		S = 10		S = 15	
		Simulated	Deviation	Simulated	Deviation	Simulated	Deviation
Mean of the annual maximum	51.23	51.57	0.67%	53.44	4.31%	50.39	-1.63%
Mean	4.35	4.26	-1.97%	4.20	-3.32%	4.19	-3.72%
Variance	50.26	47.76	-4.97%	48.24	-4.03%	47.40	-5.69%
Skewness	3.20	3.23	0.85%	3.42	7.01%	3.26	1.91%

Table A6. Comparison between the observed summary statistics and the predictions' ensemble mean value for distinct numbers of rainfall gauging stations "donors" (S) – 360 minutes duration.

Statistics	Records	S = 5		S = 10		S = 15	
		Simulated	Deviation	Simulated	Deviation	Simulated	Deviation
Mean of the annual maximum	60.10	59.92	-0.31%	61.85	2.91%	57.92	-3.63%
Mean	6.09	5.93	-2.68%	5.87	-3.65%	5.85	-4.01%
Variance	86.61	83.61	-3.47%	85.43	-1.37%	82.61	-4.62%
Skewness	2.79	2.81	0.65%	2.94	5.45%	2.79	0.22%

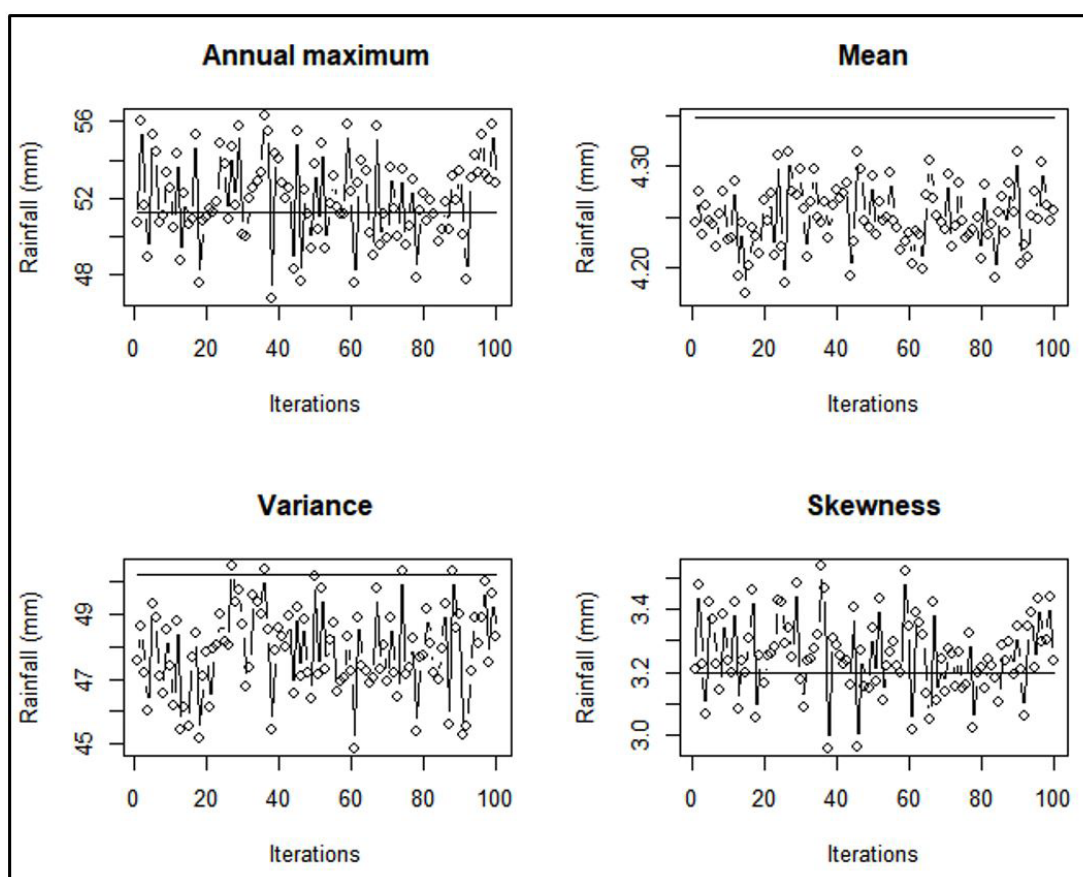


Figure A1. Annual maximum, mean, variance and skewness results for each one of the 100 model runs for the 180 minutes duration. Dots denote model predictions and the solid line the mean value of the observed statistics.

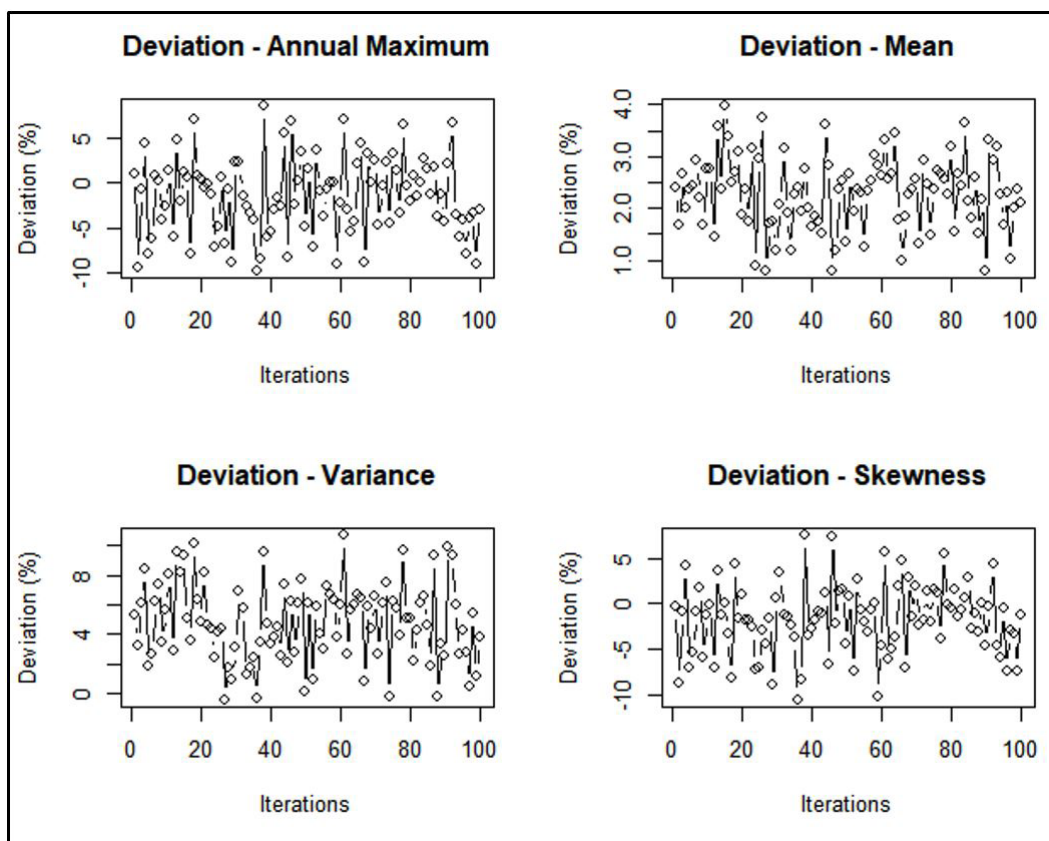


Figure A2. Deviations of model predictions from the observed summary statistics of rainfall amounts for the 180 minutes duration.

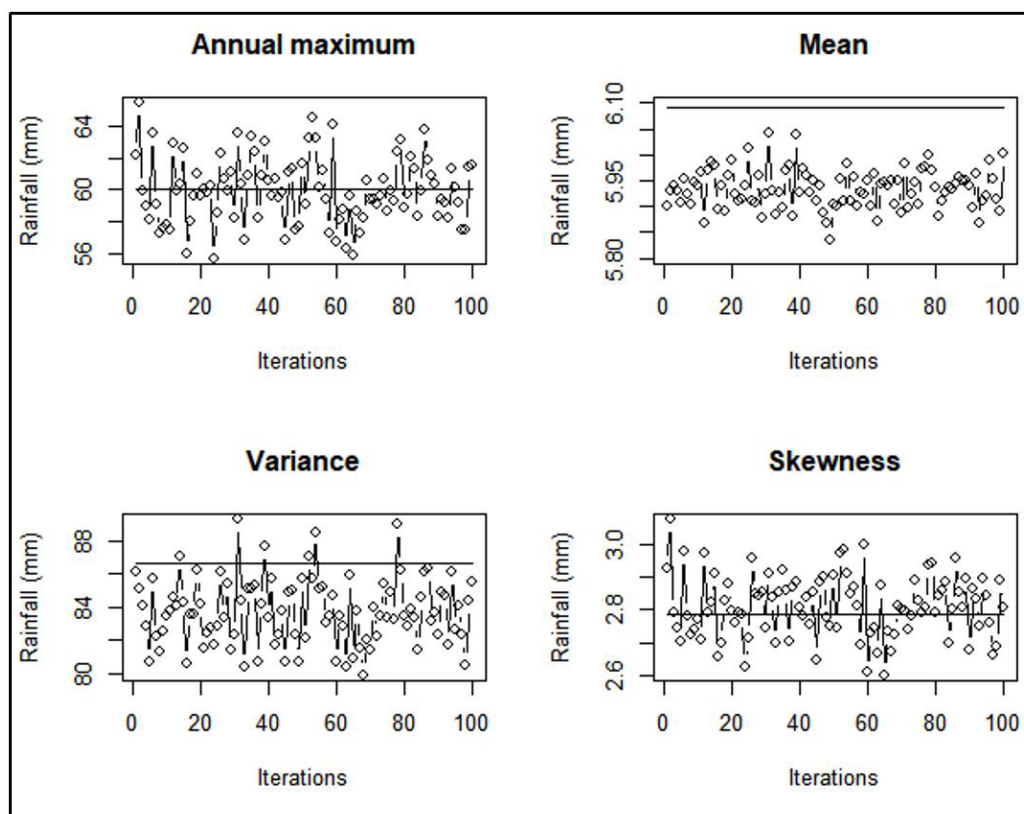


Figure A3. Annual maximum, mean, variance and skewness results for each one of the 100 model runs for the 360 minutes duration. Dots denote model predictions and the solid line the mean value of the observed statistics.

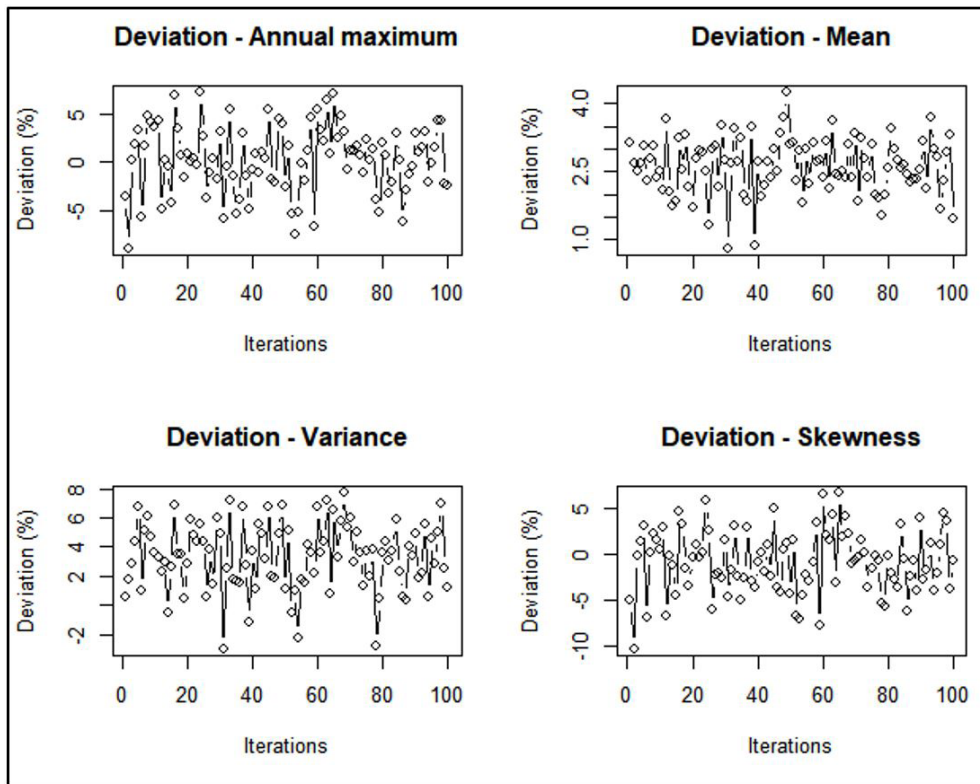


Figure A4. Deviations of model predictions from the observed summary statistics of rainfall amounts for the 360 minutes duration.

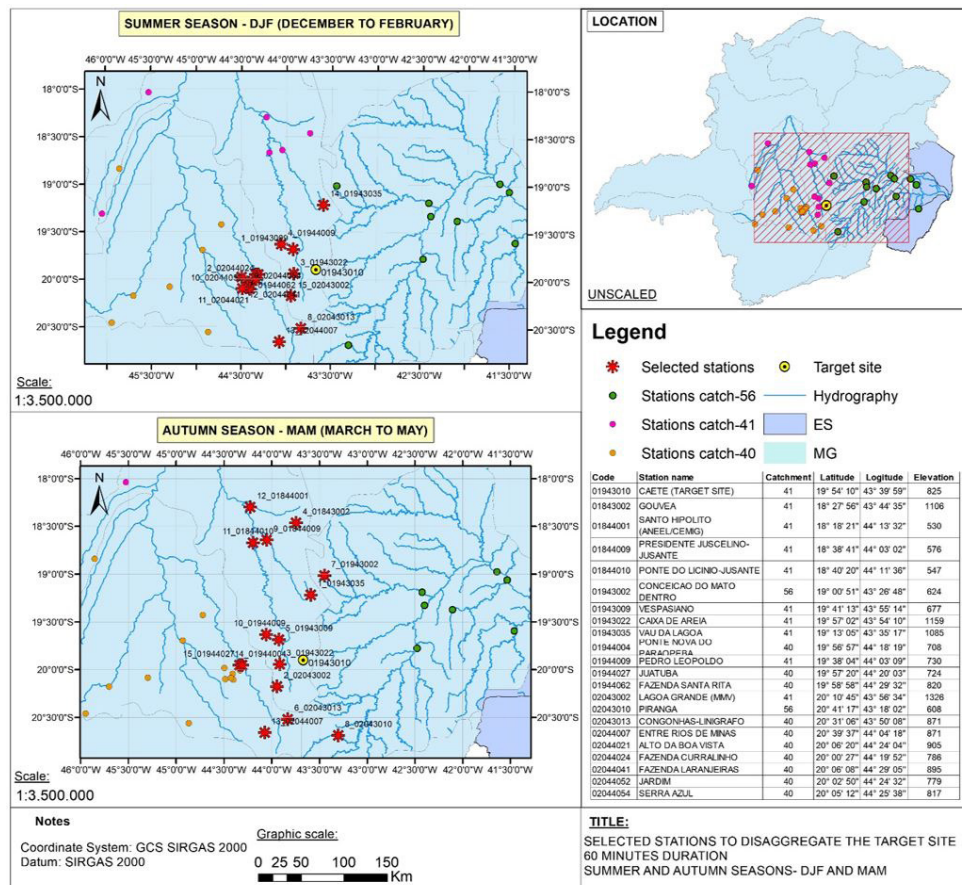


Figure A5. Selected stations to disaggregate daily rainfall amounts at the target site for the 60 minutes duration and the summer and autumn seasons.

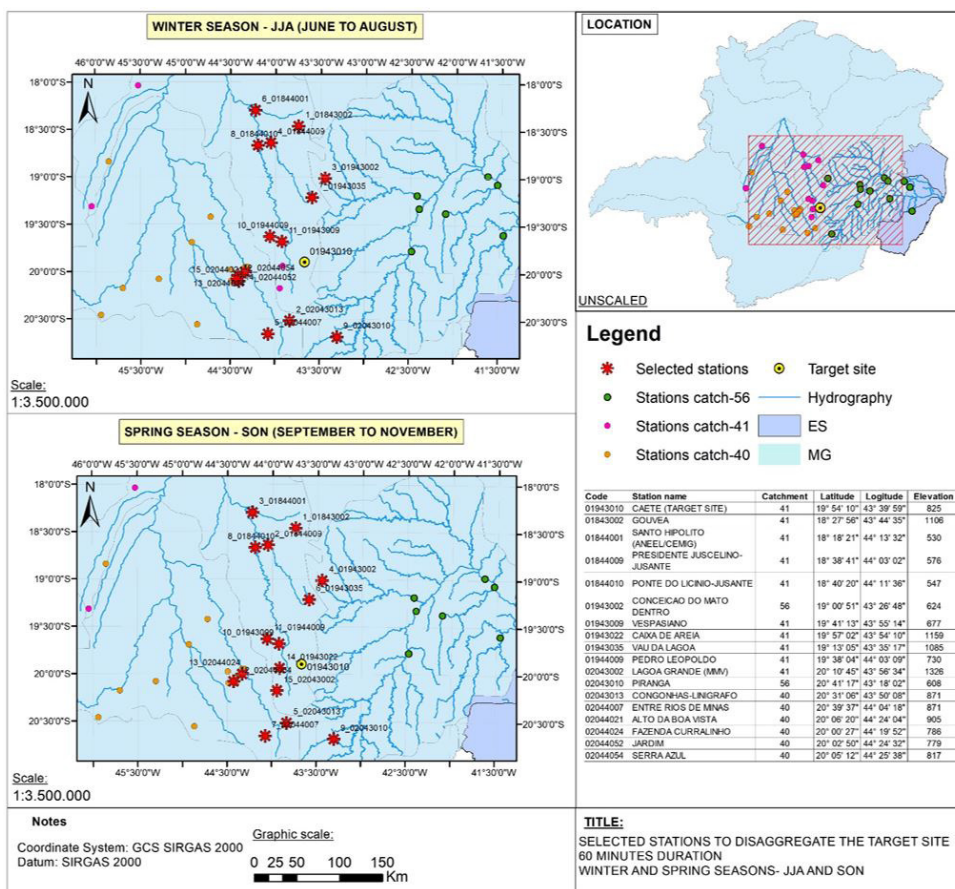


Figure A6. Selected stations to disaggregate daily rainfall amounts at the target site for the 60 minutes duration and the winter and spring seasons.