Distribution of rainfall probability in the Tapajos River Basin, Amazonia, Brazil

Vanessa Conceição dos Santos[1] [iD]**; Claudio Blanco[2*]** [iD]**;
José Francisco de Oliveira Júnior[3]** [iD]

[1]Universidade Federal do Pará (UFPA), Belém, PA, Brasil
Programa de Pós-Graduação em Engenharia Civil (PPGEC).
E-mail: vanessasantos.esa@gmail.com
[2]Universidade Federal do Pará (UFPA), Belém, PA, Brasil
Faculdade de Engenharia Sanitária e Ambiental (FAESA).
E-mail: blanco@ufpa.br
[3]Universidade Federal de Alagoas (UFAL), Maceió, AL, Brasil
Instituto de Cieências Atmosféricas (ICAT).
E-mail: junior_inpe@hotmail.com
[*]Corresponding author

ABSTRACT

Studies on the probability of rainfall and its spatiotemporal variations are important for the planning of water resources and optimization of the calendar of agricultural activities. This study identifies the occurrence of rain by first-order Markov Chain (MC) and by two states in the Tapajos River Basin (TRB), Amazon, Brazil. Cluster analysis (CA), based on the *Ward* method, was used to classify homogeneous regions and select samples for checking the probability of rainfall occurrence by season. The historical series of daily rainfall data of 80 stations were used for the period 1990-2014. The CA technique identified 8 homogeneous regions and their probability of occurrence of rainfall, helping to determine which regions and periods have greater need of irrigation. Results of the probability of occurrence of dry and rainy periods in the TRB were used to define the dry (May thru September) and rainy seasons (October thru April). Elements of the matrix transition probabilities showed variability in relation to time and, in addition, the influence of geographical position of seasonal rainfall in determining dry and rainy periods at specific sites in the TRB.

Keywords: cluster analysis, dry and rainy days, Markov chains.

Distribuição da probabilidade de chuva na bacia do rio Tapajós, Amazônia, Brasil

RESUMO

Estudos sobre a probabilidade de chuvas e suas variações espaço-temporais são importantes para o planejamento de recursos hídricos e otimização do calendário de atividades agrícolas. O presente estudo identifica a ocorrência de chuvas pela cadeia de Markov de primeira ordem (CM) e por dois estados da bacia hidrográfica do rio Tapajós (BHRT), Amazônia, Brasil. A análise de agrupamento (AC), baseada no método de *Ward*, classificou

regiões homogêneas e selecionou a amostra para verificar a probabilidade de ocorrência de chuvas por estação. A série histórica de dados diários de precipitação de 80 estações compreendendo o período de 1990-2014 foi usada. A técnica AC identificou 8 regiões homogêneas e suas probabilidades de ocorrência de chuvas, auxiliando assim no estabelecimento de períodos com maior necessidade de agricultura irrigada de acordo com a região. Os resultados sobre a probabilidade de ocorrência de períodos secos e chuvosos na bacia definiram as estações seca (maio a setembro) e chuvosa (outubro a abril). Elementos das probabilidades da matriz de transição mostraram variabilidade em relação ao tempo e, além disso, a influência da posição geográfica da precipitação sazonal na determinação dos períodos seco e chuvoso em locais específicos na bacia do rio Tapajós.

Palavras-chave: análise de agrupamentos, cadeias de Markov, dias secos e chuvosos.

1. INTRODUCTION

Rainfall is a climatic factor of major environmental importance, especially in relation to floods and prolonged periods of drought (Marengo *et al.*, 2010). Rainfall regime impacts environmental conditions and almost all productive activities of society, since it is the main water supply for human and economic activities (Almeida *et al.*, 2011). An understanding of the potential occurrence of dry and rainy days can contribute decisively to the decision-making process regarding planting times, water deficit risk assessment and irrigation system design (Keller Filho *et al.*, 2006; Marcelino *et al.*, 2012; Carvalho *et al.*, 2017). According to Vela *et al.* (2007), historical rainfall data are highly relevant for monitoring the impacts caused by their excess or prolonged shortage within a specific region. Despite the great importance of pluviometric data in the planning and design of engineering works, prolonged rainy weather can be a limiting factor for the use of agricultural equipment, affecting previously established schedules (Fernandes *et al.*, 2002).

The Central-West region, for example, ranks first nationally in corn and soybean production, with the State of Mato Grosso as the main producer of both crops, producing approximately 64% of maize and 60% of soybeans, with more than 4,498,000 and 9,518,000 hectares planted, respectively (CONAB, 2019). According to Garcia *et al.* (2013), in a study carried out in the Sinop region (MT), the best growing season indicated for maize crop corresponds to the wet season during spring and summer, which favors high water availability to the soil due to rainfall and less probability of occurrence of consecutive dry days. The authors also affirm that in order to obtain a good income from the agricultural activity, it is important to know the conditions of the environment and the culture, from before planting until post-harvest, as the rainfall and the temperature of the air, together with the photoperiod, are the main meteorological variables for productivity.

Semenov (2008) and Martin *et al.* (2007) state that stochastic models applied in hydrology are often used to complement daily climatological data. In addition, the models assess the effect of climate change on daily precipitation (Yoo *et al.*, 2016). Markov Chain (MC) models are often proposed to rapidly obtain weather forecasts (dry or rainy) and their transition throughout the year (Lennartsson *et al.*, 2008; Breinl *et al.*, 2013). In the present work, the MC method is used to model the occurrence of daily precipitation, as occurs often in the literature (Sharif *et al.*, 2007, Damé *et al.*, 2007, Selvaraj and Selvi, 2010, Sukla *et al.*, 2016, Yoo *et al.*, 2016). The emphasis on the application of the Markov chain derives from the use of the information from the previous day (dry or rainy) to provide a prognosis about the possible occurrence of a dry or rainy day for a given region (Carvalho *et al.*, 2017). The daily precipitation model, based on MC, presents several advantages, such as the ease of parameter estimation and data generation when compared to other models also used in rainfall modeling. For example, the Poisson model

IPABH

has more complex structures and greater difficulties in parameter estimation (Yoo *et al.*, 2016), while MC basically involves two components: the occurrence of binary precipitation (precipitation or absence of precipitation) and the quantity of precipitation on wet days (Breinl *et al.*, 2013). The MC are also the most-indicated models for the study of rainy- and dry-day sequences, since other models for probabilities have difficulty in describing the daily persistence of the occurrence conditions (Sukla *et al.*, 2016). Another factor is the rapid response made possible by the model, which may contribute to a greater economic return for rural producers, as it supports the identification of alternative sowing dates. This planning can be optimized if time distribution is associated with a spatial distribution of occurrences by characterizing regions with similar behaviors.

Thus, regions with hydrological similarities can be classified into groups. The literature reports that the multivariate cluster analysis (CA) method provides striking results to determine homogeneity (Yang *et al.*, 2010; Cabrera *et al.*, 2012; Lyra *et al.*, 2014). This study sought to determine the distribution of rainfall probability using first-order Markov chains and two states (dry and rainy days), grouped by Cluster analysis to identify pluviometrically homogeneous regions in the Tapajos River Basin, Amazonia, Brazil. Other authors have used rainfall (daily, monthly or annual) and other information in their studies (Raziei *et al.*, 2012; Gonçalves *et al.*, 2016).

2. MATERIALS AND METHODS

2.1. Study area and data

The Tapajos River Basin (TRB) drains an area of 493,200 km^2 and comprises 6% of Brazilian territory. It occupies areas in the states of Mato Grosso (MT), Pará (PA) and a small part of the Amazon (AM), and lies between 2° and 15° S and between 53° and 61° W. According to Kottek *et al.* (2006), the TRB presents two climatic typologies by Köppen-Geiger´s climatic classification. The predominant climate from the headwaters to the center of the basin is classified as *"Aw"*, that is, rainfall in the summer, the characteristic climate of the savannahs. From the center to the mouth, climate is classified *"Am"*, or rather, tropical monsoon climate, with a dry season and intense rains for the rest of the year. In the study by Santos *et al.* (2014) in the TRB, within the *"Am"* climate area, the months with the lowest rainfall are from May to October. In the *"Aw"* climate, the dry season runs from April to September, with July being the month with the lowest rainfall value in both climates, with mean values of 36 mm / year and 6 mm / year, respectively.

Daily rainfall data provided by 80 gauge stations used in this study were retrieved from the National Water Agency (ANA) database, via HIDROWEB (http://hidroweb.ana.gov.br), and from the National Institute of Meteorology (INMET), via BDMET (http://www.inmet.gov.br/projetos/rede/pesquisa/). Because they were daily data, they were not submitted to any statistical treatment to fill in the flaws. Bertoni and Tucci (2007) present in their studies several methodologies used to fill in flaws and also comment that none is indicated to fill in daily failures, being more recommended to fill monthly or annual failures. In selecting the periods of the historical rainfall series, the admissible fault limit of up to 1.8% in each data series was applied as a criterion. This limit of failures was adopted based on the study by Baú *et al.* (2013), where the stochastic model chosen (Markov Chains) admitted, with good reasonableness, its application, but does not rule out the possibility of negative, albeit small, interference in the accuracy of the results. Figure 1 shows the site of the study area and the distribution of rainfall gauge stations. Table 1 provides information on the same.

Rev. Ambient. Água vol. 14 n. 3, e2284 - Taubaté 2019

IPABH
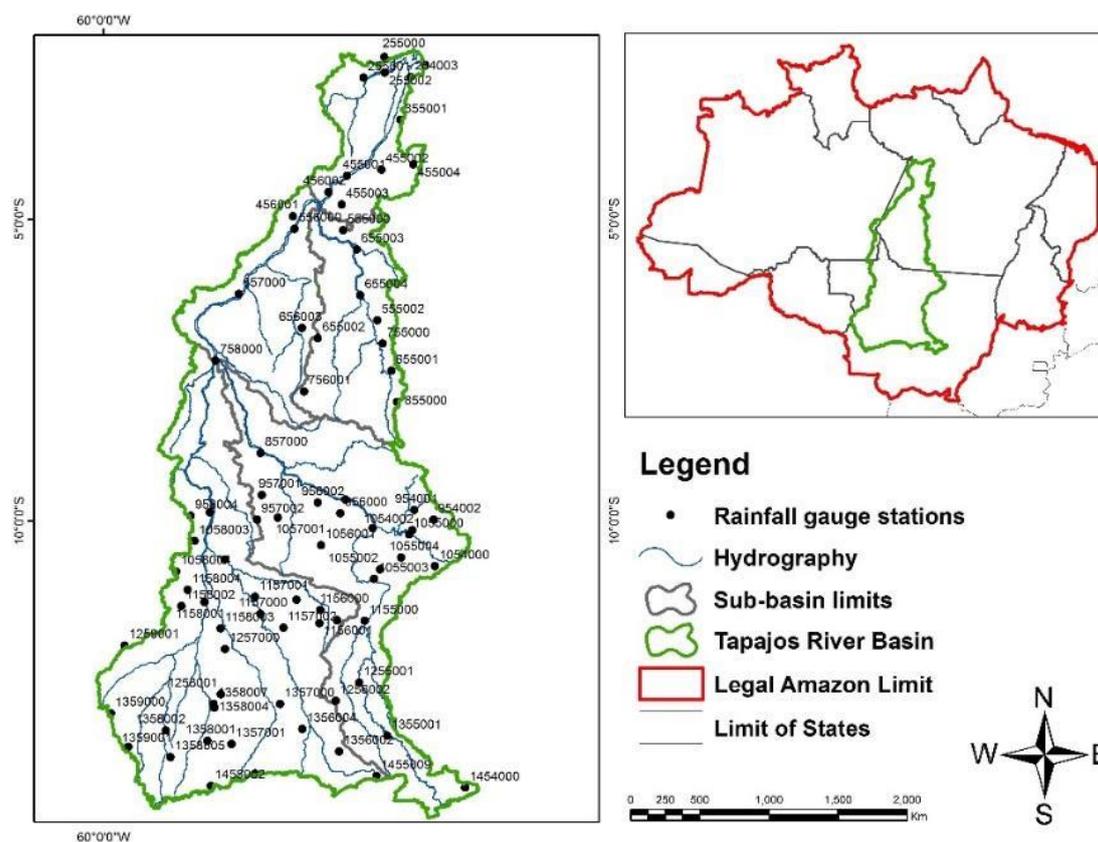
Figure 1. Location map of the area under analysis: Tapajos River Basin, with code of 80 rainfall gauge stations.

Table 1. Rainfall gauge stations in current study, with geographical coordinates (latitude and longitude, °), identifier (ID), code and start / end of time series.

ID	Code	Lat. (S)	Long. (W)	Start	End	Annual mean precipitation (mm)
01	756001	-7.8503	-56.7000	01/01/2000	31/12/2002	1707.3
02	255000	-2.2683	-55.4806	01/01/2002	31/12/2010	2578.3
03	855000	-8.1872	-55.1194	01/01/1992	28/02/2002	2607.8
04	1359001	-13.7781	-59.7675	22/02/1992	31/12/2014	1922.1
05	1358004	-13.0944	-58.1764	01/01/1997	31/12/1999	1523.3
06	254000	-2.4431	-54.7075	04/06/1995	30/11/2010	2375.9
07	1259001	-12.0603	-59.6503	01/01/1990	31/12/2001	1902.9
08	254003	-2.6419	-54.9439	01/01/1990	31/12/1998	1667.2
09	455001	-4.2772	-55.9931	01/01/1990	30/06/2014	2092.0
10	255002	-2.5644	-55.3742	01/01/2005	31/12/2014	2347.4
11	255001	-2.6508	-55.7206	01/01/1998	31/12/2014	1934.2
12	355001	-3.3500	-55.1167	01/03/2004	31/12/2014	2087.7
13	455004	-4.0894	-54.9028	01/01/1990	30/09/2003	1807.6
14	455002	-4.1750	-55.4269	01/03/2004	31/12/2014	2188.9
15	456002	-4.5500	-56.3000	01/03/2004	31/03/2014	2418.5
16	455003	-4.7547	-56.0794	01/01/1990	30/09/2003	1972.2
17	456001	-4.9469	-56.8822	01/01/1990	30/09/2003	1972.2
18	556000	-5.1542	-56.8556	01/03/2004	31/12/2012	2425.0
19	555000	-5.1825	-56.0578	01/01/1990	30/09/2003	2356.9
20	655003	-5.5000	-55.8333	01/03/2004	31/10/2012	2373.9
21	657000	-6.2358	-57.7756	01/03/2004	31/12/2014	2142.6
22	655004	-6.2575	-55.7733	01/03/2004	31/03/2013	1766.5
23	555002	-6.6714	-55.4958	01/03/2004	31/05/2014	2241.2

Continue...

Continued...

24	656003	-6.8000	-56.7333	01/03/2004	31/03/2013	2348.3
25	655002	-6.9678	-56.4728	01/03/2004	30/11/2012	1985.9
26	755000	-7.0606	-55.4078	01/03/2004	28/02/2013	2246.8
27	758000	-7.3389	-58.1550	01/03/2004	31/12/2014	2891.9
28	655001	-7.5067	-55.2614	01/01/1990	31/08/2002	2060.7
29	857000	-8.8703	-57.4164	01/03/2004	28/02/2013	2067.3
30	957001	-9.5664	-57.3947	01/09/1993	31/12/2013	1888.8
31	956001	-9.6425	-56.0183	01/10/1994	31/12/2013	2243.0
32	956002	-9.6939	-56.4742	01/11/1999	31/12/2014	2299.1
33	954001	-9.8172	-54.8858	01/04/1993	31/12/2014	2185.3
34	1058002	-9.8561	-58.2469	01/11/1994	31/08/2007	1920.2
35	956000	-9.8703	-56.1022	01/07/1998	30/11/2003	1898.3
36	958004	-9.9133	-58.5642	15/09/2004	31/12/2014	2085.3
37	1057001	-9.9414	-57.1319	01/09/1994	31/12/2014	2177.8
38	954002	-9.9758	-54.5653	01/01/2005	31/12/2014	1762.5
39	957002	-9.9761	-57.4736	01/07/2000	31/12/2014	2248.9
40	1055001	-10.1125	-55.5700	01/05/1995	31/12/2014	1940.9
41	1054002	-10.1500	-54.9186	13/12/2004	31/12/2014	2052.1
42	1055000	-10.2194	-54.9703	01/02/2004	31/10/2008	1673.1
43	1058003	-10.3286	-58.5003	01/01/1990	31/10/2004	1947.7
44	1056001	-10.3975	-56.4192	01/07/2005	31/12/2014	2119.9
45	1055004	-10.6044	-55.1033	01/08/2000	31/08/2012	1992.9
46	1058006	-10.6397	-58.0039	01/05/2001	31/01/2011	1951.6
47	1054000	-10.7461	-54.5461	01/05/1995	31/12/2014	1784.3
48	1055002	-10.7986	-55.4486	01/09/1993	31/12/2014	1734.5
49	1058004	-10.8342	-58.8033	03/07/2001	31/12/2014	2073.1
50	1055003	-10.9558	-55.5486	01/08/1994	31/12/2014	1990.7
51	1158004	-11.1392	-58.6150	01/10/2004	31/12/2014	1801.5
52	1157001	-11.2531	-57.5067	01/01/1997	31/08/2014	1934.3
53	1156002	-11.3047	-56.8250	01/08/2004	31/12/2014	2056.9
54	1158001	-11.3417	-58.3383	01/11/1991	31/12/2014	1785.5
55	1158002	-11.4081	-58.7186	01/01/1990	31/03/2006	1921.0
56	1156000	-11.4714	-56.4333	01/06/1992	31/12/2014	1845.2
57	1157000	-11.5358	-57.4172	01/03/1999	31/07/2011	1789.0
58	1156003	-11.6447	-56.1572	01/02/2008	31/12/2012	1967.4
59	1155000	-11.6531	-55.7017	01/08/2004	31/12/2014	1868.8
60	1156001	-11.6914	-56.4486	01/08/1992	31/12/2014	1843.1
61	1157002	-11.7650	-57.0419	04/12/1999	31/12/2014	1939.9
62	1158003	-11.7772	-58.0725	01/01/2008	31/08/2014	1832.8
63	1257000	-12.1169	-57.9992	01/03/1996	31/12/2014	1778.2
64	1255001	-12.6750	-55.7931	10/09/1996	31/12/2014	1667.9
65	1258001	-12.8675	-58.0703	01/10/2008	31/12/2014	1983.8
66	1256002	-12.9797	-56.1806	04/12/1999	31/12/2014	1795.7
67	1358007	-13.0267	-58.1881	01/03/2008	31/01/2014	1615.0
68	1357000	-13.0292	-57.0925	01/10/1996	31/12/2014	1642.3
69	1359000	-13.1831	-59.8769	07/10/1993	31/12/2014	2152.4
70	1356004	-13.4450	-56.7275	02/11/2004	31/12/2014	1623.0
71	1358002	-13.4667	-58.9750	01/11/1990	31/12/2014	2244.7
72	1355001	-13.5564	-55.3317	01/01/1992	31/10/2011	1721.6
73	1358001	-13.6414	-58.2892	19/04/1993	31/12/2014	1728.8
74	1357001	-13.6922	-57.8944	15/05/2000	30/09/2013	1760.6
75	1356002	-13.8156	-56.1222	01/08/2006	31/12/2014	1723.0
76	1358005	-13.9100	-58.8981	18/10/1999	31/12/2014	1929.8
77	1457003	-14.1847	-57.5069	01/01/1990	31/01/2008	1921.3
78	1455009	-14.2214	-55.5067	01/04/2006	31/03/2014	1671.9
79	1458002	-14.3842	-58.2344	01/09/2004	31/12/2014	1803.1
80	1454000	-14.4178	-54.0494	01/06/1994	31/08/2006	1742.4

2.2. Determination of rainy or dry days

Rainy or dry days are determined by applying the Markovian stochastic process, a widely used technique in the literature (Detzel and Mine, 2011; Dash, 2012; Stowasser, 2012;

Szyniszewska and Waylen, 2012; Baú *et al.*, 2013). The condition of rainy or dry state is associated with a probability of occurrence. The stochastic process adopted in the study was used to model rainfall occurrences by first-order MC (the probability of the rainfall state on the current day "*t*" depends only on the rainfall state of the previous day, *t-1*) and by two states (dry or rainy). Combination hypotheses for the determination of the probabilities of transition between states are carried out by a matrix of transition (MT) (Equation 1).

$$\text{MT} = \begin{vmatrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{vmatrix} \tag{1}$$

The transition probabilities between states follow Equations 2 and 3:

$$P_{00} = P[X_{t+1} = 0/X_t = 0] \tag{2}$$

$$P_{11} = P[X_{t+1} = 1/X_t = 1] \tag{3}$$

In the case of defining rainfall states on day "*t*", current days are tagged as "X_t", with "0" for dry days and "1" for rainy days. First order MCs consider combinations of dry (0) and rainy (1) states as follows:

P_{00} is the probability of not raining today because it did not rain yesterday;

P_{01} is the probability of not raining today because it rained yesterday;

P_{10} it is likely to rain today because it did not rain yesterday;

P_{11} it is likely to rain today because it rained yesterday.

Calculation of probabilities is performed by counting the items in the historical records of the desired rainfall station, as described in Equations 4, 5, 6 and 7, where "*N*" represents the relation between the number of occurrences of combinations dry / rainy days of the historical series according to the rainfall station (j).

$$P_{00}(j) = \frac{N_{00}(j)}{N_{00}(j) + N_{10}(j)} \tag{4}$$

$$P_{01}(j) = \frac{N_{01}(j)}{N_{01}(j) + N_{11}(j)} = 1 - P_{11}(j) \tag{5}$$

$$P_{10}(j) = \frac{N_{10}(j)}{N_{00}(j) + N_{10}(j)} = 1 - P_{00}(j) \tag{6}$$

$$P_{11}(j) = \frac{N_{11}(j)}{N_{01}(j) + N_{11}(j)} \tag{7}$$

Where:

N_{00} – number of dry days with previous dry day;

N_{01} – number of dry days with previous rainy day;

N_{10} - number of rainy days with previous dry day;

N_{11} - number of rainy days with previous rainy day.

The values of daily rainfall as indicative of dry periods, also called minimum values, present a considerable variation among the authors found in the literature, such as, 0.3 mm (Baú *et al.*, 2013), 5.0 mm (Pizzato *et al.*, 2012; Viana *et al.*, 2002), 0.2 mm (Calgaro *et al.*, 2009),

0.1 mm (Dourado Neto *et al.*, 2005; Keller Filho *et al.*, 2006), 0.85 mm (Barron *et al.*, 2003) and 1 mm (Jeong *et al.*, 2013; Santos *et al.*, 2009), in which days with rainfall below these limits were classified as dry. Mehrotra and Sharma (2009) fined as a rainy day the one whose measured value reaches the threshold of 0.3 mm. Andrade Junior *et al.* (2001) and Viana *et al.* (2002) defined the day as dry, based on the occurrence of water deficit, that is, dry days are considered, those in which rainfall is less than the reference evapotranspiration. In the study by Vasconcellos *et al.* (2003), the authors defined the day as dry when the water storage in the soil, according to the water balance, is equal to or less than a certain critical value, conditioned by the atmospheric demand. However, the above rates depend on the study's aim, activity and type of environmental management under development. So that a day may be considered rainy, the criterion in current study is that minimum rainfall rate recorded in one day should be equal to or greater than 0.1 mm; therefore, if it is less than 0.1 mm, the day will be considered dry. This is the criterion used by INMET. This rate is equivalent to the smaller amount recorded by the pluviograph in conventional meteorological stations (CMS).

2.3. Sensitivity analysis

Sensitivity analysis was performed to determine the minimum period required for a given historical series of daily precipitation in order to estimate the probability of occurrence. In order to generate these probabilities, we verified the length of the historical series and their start and end date. In many studies applied to hydrological modeling there is a limitation in the available series, as well as in relation to the small period of observations and the low density of the data collection network. To analyze the behavior of pluviometric probabilities over time, a historical series of 30 years was selected, considering a failure limit of up to 1.8%. The four probabilities of monthly occurrences were generated for periods of 1, 2, 3, ..., 29 and 30 years. Thus, the station selected was station code 455001 (ID: 09), located in the state of the PA, in group 1 (G1) (Figure 1 and Table 1). The selected period was from 07/01/1984 to 06/30/2014 for 30 years and for the remaining periods (1 to 29 years) the final date was kept constant (day, month and year), varying only the start date and the year at each interval. To evaluate the behavior of the probabilities, a simple linear regression analysis was performed between the probabilities of occurrence generated for the 30-year series (keeping it as an independent variable in the regression analysis) and the respective probability of occurrence of each period of 1 to 29 years (dependent variables). Thus, for P00, P01, P10 and P11, the coefficient of determination (R^2), the adjusted determination coefficient (R^2a), the standard error, and the mean absolute and relative errors were obtained, considering p-value less than 0.05 at a significance level of 95%, with acceptance of the null hypothesis, that there is a correlation between the variables.

2.4. Cluster Analysis

The cluster analysis (CA) is a Multivariate Statistics technique that has the purpose of grouping individual items (such as objects, places or samples) into several groups, according to a classification criterion so that there is homogeneity within a group (or variables) and heterogeneity among the other groups formed based on their characteristics (Lyra *et al.*, 2014; Gonçalves *et al.*, 2016). The CA technique was used to determine the homogeneous pluviometric groups in the TRB, using all the values of the probability of occurrence of rainfall resulting from CM. The grouping criterion used was the method proposed by Ward Jr. (1963), which is a method of hierarchical data grouping that forms groups in such a way as to always achieve the smallest internal error between the vectors that make up each group and the average group vector. According to Johnson and Wichern (2007), this method is also called "minimal variance", because in each step of convergence of the method, the two clusters that have the smallest distance between them are combined to form a single group. For each phase, Ward's method uses Equation 8, which regulates the operation of the same and its convergence.

IPABH

$$E_{(G1G2)} = \sum_{V}^{p} \sum_{\substack{i=1 \\ i \in G1}}^{n} (x_{iv} - \bar{x}_v)^2 \tag{8}$$

Where, $E_{(G1G2)}$ is the mean rate for two clusters; $\bar{x}_v$ is the cluster mean for each variable "v".

Euclidean distance, the geometric distance between two objects i and i' was employed to measure similarity between clusters. Let x_{ij} be the observation of $i\text{-}th$ rainfall gauge station (i = 1, 2, ..., n), with reference to $j\text{-}th$ variable in each class (j = 1, 2, ..., p) studied. Standardized Euclidean distance (DE) between two stations i and i' is calculated by Equation. 9.

$$DE = \sum_{j=1}^{p}(x_{ij} - x_{i'j})^2 \tag{9}$$

Where, x_{ij} is the $j\text{-}th$ characteristic of the $i\text{-}th$ individual; x_{ij} is the j-th characteristic of the i-th individual.

Figure 2 shows the operational scheme of the methodology applied in this study, guiding the sequence of each step.
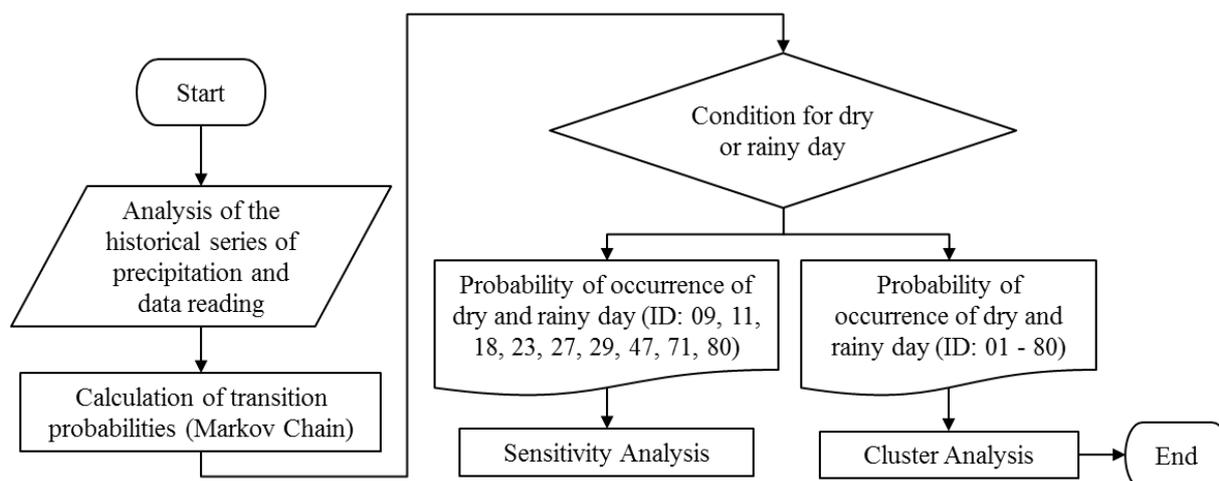
Figure 2. Methodology operational scheme.

3. RESULTS AND DISCUSSION

3.1. Sensitivity analysis
For the sensitivity analysis, the selected historical series was the one that presented values for R^2 and R^2_a equal to or greater than 0.8 in relation to the historical series of 30 years. From 3 years of data, considering the limit of failures in the registers equal to or lower than 1.8% for each analyzed period, it was possible to note that, with the exception of the linear coefficient (a), the angular coefficients determination (R^2) and adjusted determination (R^2_a) presented the same values (Table 2). This is due to the equation of the transition probabilities (Equation 2), since the behavior of P_{00} and P_{10} is complementary, as well as the behaviors of P_{01} and P_{11}. Therefore, these parameters tend to present the same slope of the regression line, although they cut the x-axis at different points.

Table 2. Linear regression analysis (linear coefficient (a), angular coefficient (b), determination (R^2) and adjusted determination (R^2_a)) between the probability of occurrence of rainfall data from Station 455001.

years	P_{00}				P_{01}				P_{10}				P_{11}			
	a	b	R^2	R^2_a	a	b	R^2	R^2_a	a	b	R^2	R^2_a	a	b	R^2	R^2_a
7	-0.05	1.02	0.97	0.97	-0.06	1.10	0.94	0.94	0.04	1.02	0.97	0.97	-0.05	1.10	0.94	0.94
6	-0.02	0.95	0.95	0.94	-0.06	1.11	0.93	0.93	0.06	0.95	0.95	0.94	-0.04	1.11	0.93	0.93
5	0.01	0.93	0.95	0.95	-0.05	1.08	0.90	0.89	0.06	0.93	0.95	0.95	-0.04	1.08	0.90	0.89
4	0.02	0.88	0.90	0.89	-0.02	1.00	0.88	0.87	0.10	0.88	0.90	0.89	0.02	1.00	0.88	0.87
3	0.11	0.72	**0.85**	**0.84**	-0.04	1.06	**0.94**	**0.93**	0.17	0.72	**0.85**	**0.84**	-0.02	1.06	**0.94**	**0.93**
2	0.12	0.69	0.77	0.75	0.00	1.01	0.96	0.95	0.19	0.69	0.77	0.75	-0.01	1.01	0.96	0.95
1	0.02	0.81	0.46	0.41	-0.05	1.14	0.85	0.84	0.17	0.81	0.46	0.41	-0.08	1.14	0.85	0.84

It was observed that R^2 and R^2_a presented values ranging from 0.8 to 0.9 in the historical series of 3, 4 and 5 years. From 6 years, the values of the coefficients are all above 0.9; therefore, the results of the other years were not shown in Table 2. Based on the analysis of the absolute and relative errors in each month for each probability of occurrence, it was verified that the highest values of these are found in the transition months (May and November) between the dry- and rainy period of the region where the rainfall season occurs.

As the period of the historical series available for generating the probabilities increases, the errors obtained in the estimates decrease. In Figure 3 (a, b, c and d), one can see the standard error and the mean absolute and relative errors for P_{00}, P_{01}, P_{10} and P_{11}, respectively. The level of significance was 95%, with a *p*-value of less than 0.05 in all analyses. The highest value was 0.0146 for the historical series of 1 year, indicating, therefore, strong statistical evidence of the relationship between the data. The range of inclination of the lines for 3 years of data, considering 95% confidence, presented the values 0.51 and 0.93 as lower and upper limit, respectively, in P_{00} and P_{10}. For P_{01} and P_{11}, the values were 0.86 and 1.26 as lower and upper limits, respectively. Since zero is not included in the confidence interval, it is possible to confirm the existence of a positive relationship between the analyzed data.

IPABH

Rev. Ambient. Água vol. 14 n. 3, e2284 - Taubaté 2019

(a)

(b)

(c)

(d)
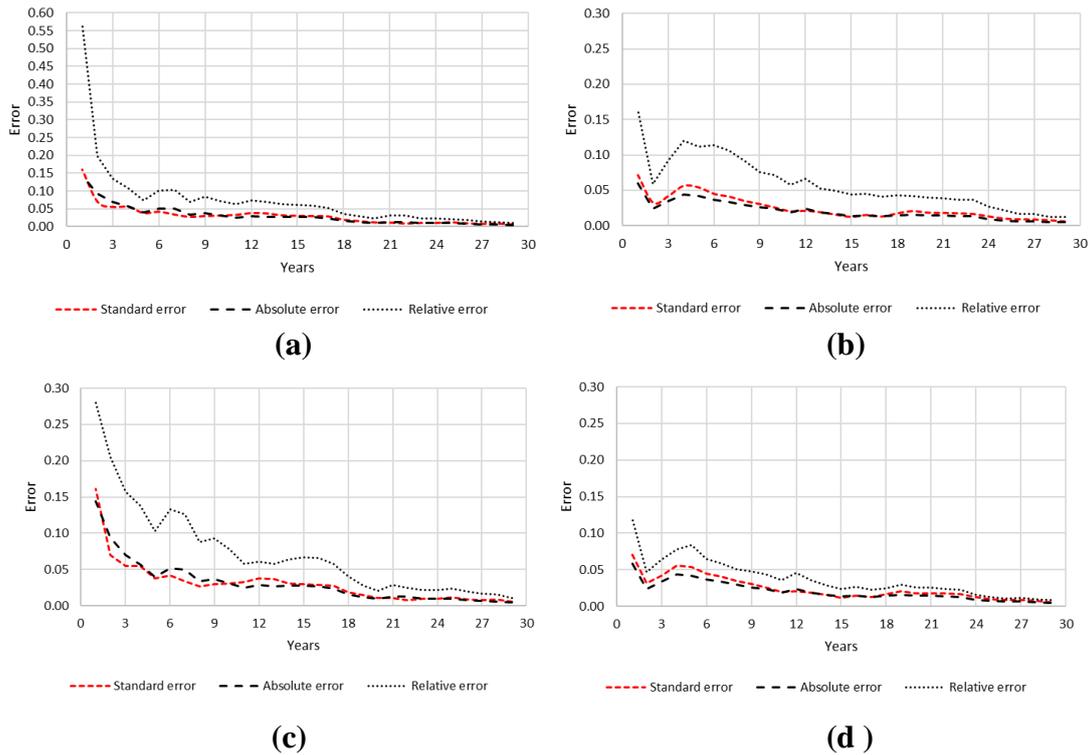
Figure 3 (a, b, c, d). Standard errors, absolute and relative sensitivity analysis of the probabilities of occurrence of rainfall data of 455001 Station for (a) P_{00}, (b) P_{01}, (c) P_{10} and (d) P_{11}.

In order to verify if this behavior could be applied in other stations, the analysis was performed this time for 8 distinct rain stations, with these codes: 758000 (ID: 27), 556000 (ID: 18), 555002 (ID: 23), 255001 (ID: 11), 857000 (ID: 29), 1454000 (ID: 80), 1358002 (ID: 71) and 1054000 (ID: 47). The starting date was the same as in Table 1, and from these the first 3 years of each historical series were separated, which were related to their maximum periods used in this study. In addition to the analyses performed at station code 455001 (ID: 09), the study also sought to verify possible interferences in the determination coefficients. The results appear in Figure 4.
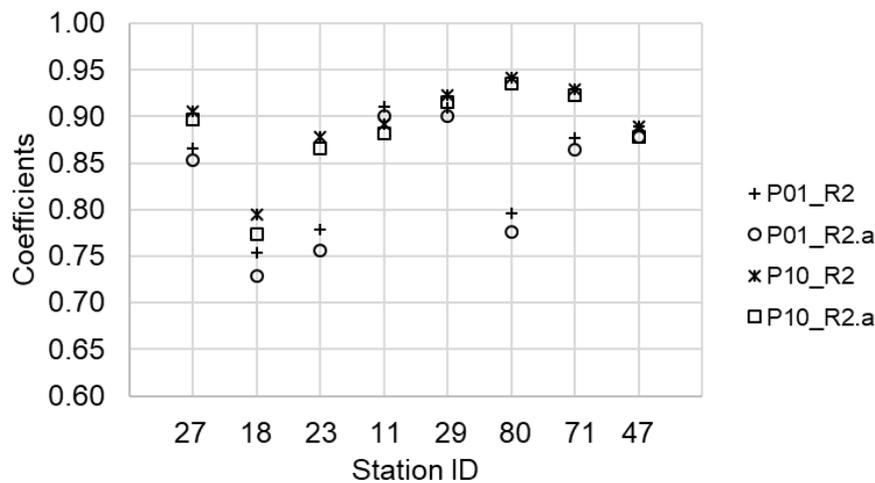
Figure 4. Coefficient of determination (R^2) and Adjusted Determination Coefficient (R^2_a) of the sensitivity analysis for the 8 rainy seasons.

Rev. Ambient. Água vol. 14 n. 3, e2284 - Taubaté 2019

IPABHi

Analyzing the results of both tests, the length of the historical series presented a greater influence on the results of the probabilities of occurrences when compared with the obtained Ps (simulated precipitation) if only the start and end dates are changed. This can be explained by MT itself, in that the larger the amount of information available to generate the Ps, the closer to 1 the coefficient of determination is presented (Table 2 and Figure 4).

In a historical series of 3 years, the limit of 1.8% corresponds to approximately 19 days of failure, being distributed throughout the 12 months, so that they do not present consecutively or concentrated in a single month. Otherwise, the lower the total available period, the greater the interference in the estimate of Ps in the month in question. Another possible interference is the occurrence of ENSO phenomena, since these will be related to the frequency and distribution of dry- and rainy days throughout the year, since the occurrence of the previous day is the information used to calculate the Ns of each probability. The occurrence of large-scale natural events (eg volcanic eruptions and forest fires) and low-frequency atmospheric-ocean phenomena (El Niño and La Niña) are pointed out by Salas *et al.* (2012) as factors that affect the statistical balance of hydrological series. In the study by Baú *et al.* (2013) in the Paraná Hydrographic Basin III, the analysis of the probability of occurrence results showed that the daily precipitation behavior maintained a pattern of quantity and occurrence simultaneously with the appearance of El Niño and La Niña phenomena. These possible interferences justify the reduction in the values found for the analyzed coefficients (R^2 and R^2_a), where the lowest coefficient obtained corresponds to 0.73 for R^2_a of P_{01} in station code 556000 (ID: 18), which, although less than the others, can still be considered as a good correlation between the data. The behavior of the *p*-value remained above 0.05 in all analyses. Thus, from 3 years of data, the probabilities of occurrence tend to present behavior statistically similar to the probabilities of the larger historical series, thus allowing the use of data with historical series from 3 years. It is worth mentioning that in the study are historical series ranging from 3 to 25 years, so it is not necessary to exclude any rainfall in this analysis.

3.2. Probability of rainfall occurrence and Cluster Analysis

The four probabilities of occurrence (P_{00}, P_{01}, P_{10} and P_{11}) were determined for each month of the historical series of the rainfall gauge stations of the TRB, calculated according to the number of dry and rainy days (Ns). Figures 5 and 6 show the boxplots of probabilities P_{10} and P_{01}, respectively, for the 80 stations studied. For each Boxplot (or box diagram), the vertical bar indicates the minimum and maximum value of the sample, the values being discrepant or outliers (if any) represented by circumferences. The horizontal lines of the gray box represent from bottom to top, respectively, the 1st quartile, the 2nd quartile or the median and the 3rd quartile. It may be observed that probability P_{01} presents less dispersion and asymmetry than probability P_{10}. The presence of outliers, especially in the transition months (May and October), may be explained by the individual behavior of each station along the TRB. Mean probability rates in the basin denote that the dry period occurs in the months between May and September and the rainy season from October to April. A similar result was obtained by Collischonn *et al.* (2008), who reported that the region had two well-defined seasons, a dry season from May to September and a rainy season from October to April, with annual rainfall rates varying between 1,600 mm and 2,700 mm. Figure 5, representing P_{10}, the month of February showed the lowest data variability. However, the lowest rainfall probabilities occurred in July, a fact confirmed by Figure 6 for P_{01}. Once more, dry days were a great probability in July. As a rule, May is considered the transition month between the dry and rainy periods in the TRB. Change is more pronounced in June because, depending on the rainy season and its location, rainfall probability approaches zero, confirming July and August as the driest months in the TRB.
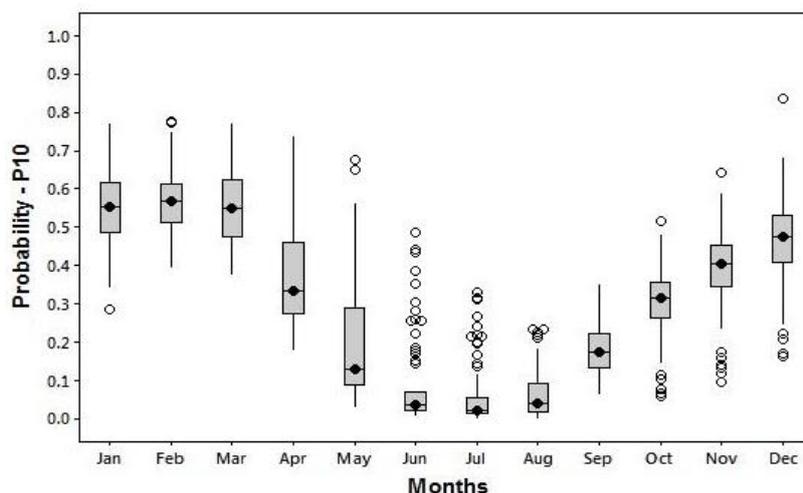
Figure 5. Boxplot of the probability of the sequence of rainy present day with previous dry (P_{10}) for the 80 rainy seasons of the TRB.
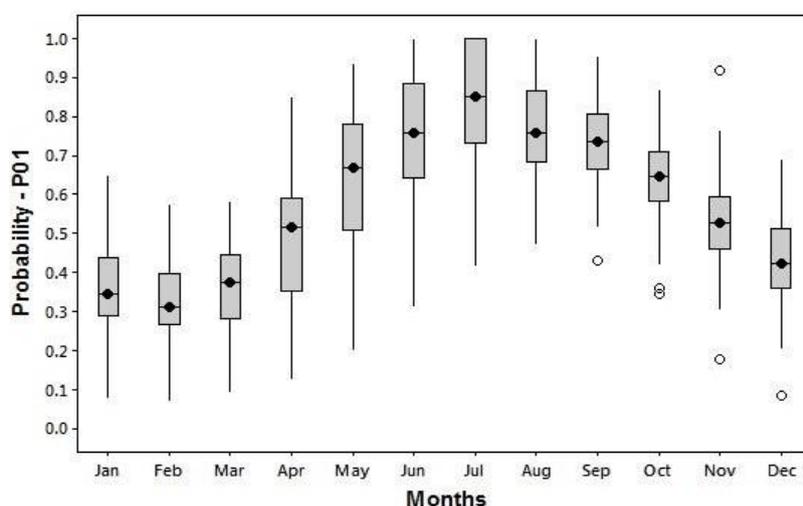
Figure 6. Boxplot of the probability of the current dry day sequence with previous rainy (P_{01}) for the 80 rainy seasons of the TRB.

Figure 7 shows the dendrogram with cluster formation and the results of the sensitivity analysis concerning the selection of Euclidean distance. Rainfall gauge stations, clustered according to similarity, are represented on the x-axis. The y-axis represents the measure of similarity for binding distances. The data used for the grouping were all 48 values of Ps obtained for each of the 80 rainfall stations. In the sensitivity analysis, the distance 3.5 was selected (red line) because it presented a better distribution of the formed clusters, resulting in 8 homogeneous ones. The stations of each cluster are identified in Table 3, while Figure 8 represents their distribution in the TRB.

After the use of the CA technique, the probability of rain occurrence was based on 8 rainy seasons, one station from each homogeneous cluster, as well as on their location in the TRB so that they could be located in different regions. The selected stations were: 555002 (ID: 23), 255001 (ID: 11), 556000 (ID: 18), 758000 (ID: 27), 1054000 (ID: 47), 857000 (ID: 29), 1358002 (ID: 71) and 1454000 (ID: 80) (Table 1). Only the probability of rainfall after a dry day (P10) and rainfall after a wet day (P11) was plotted. These two probabilities have been chosen because they are sufficient to define the Markovian process used in the model. Probabilities rates of P00 and P01 are not presented in these graphs because they are complementary to the probability rates of rain occurrence.
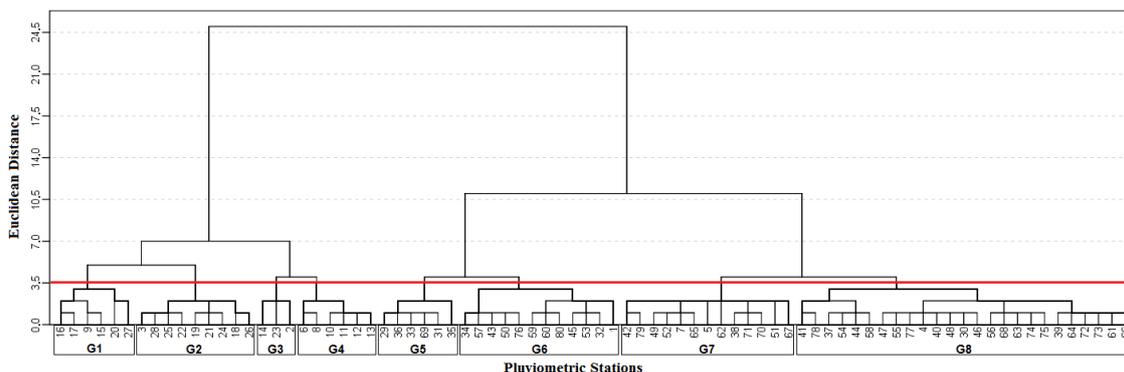
Figure 7. Dendrogram - Stations grouped according to their similarity, based on Euclidean distance.

Table 3. Identification of stations of the 8 homogeneous rainfall clusters (G_1, G_2, G_3, G_4, G_5, G_6, G_7 and G_8).

Cluster	Identification
1	16; 17; 9; 15; 20; 27.
2	3; 28; 25; 22; 19; 21; 24; 18; 26.
3	14; 23; 2.
4	6; 8; 10; 11; 12; 13.
5	29; 36; 33; 69; 31; 35.
6	34; 57; 43; 50; 76; 59; 60; 80; 45; 53; 32; 1.
7	42; 79; 49; 52; 7; 65; 5; 62; 38; 71; 70; 51; 67.
8	41; 78; 37; 54; 44; 58; 47; 55; 77; 4; 40; 48; 30; 46; 56; 68; 63; 74; 75; 39; 64; 72; 73; 61; 66.
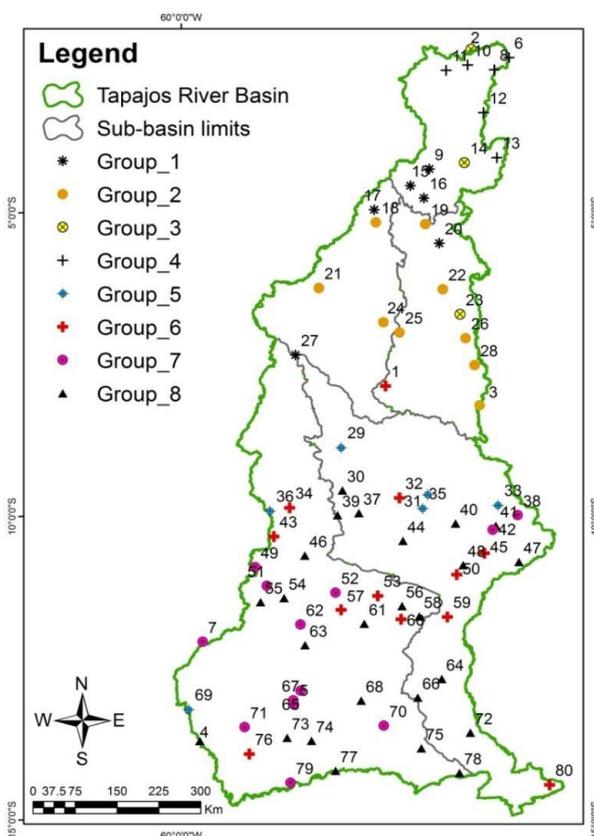
Figure 8. Spatial distribution of homogeneous rainfall clusters (G_1, G_2, G_3, G_4, G_5, G_6, G_7 and G_8) according to similarity.

IPABH

Analyzing the distribution of clusters in the TRB area, it can be observed that clusters 1, 2, 3 and 4 are located in the northern region of the basin, while clusters 5, 6, 7 and 8 are distributed in the central and southern areas. It may be seen that the stations inserted in the same cluster, although presenting different rates for the occurrence of rains, have dry and rainy periods divided in a similar way in relation to the months. This fact can be verified in Figure 9 (a, b, c, d, e, f, g, h), in which the rates of clustered stations in the clusters were plotted, presenting in common the months of June, July and August as the driest; except in cluster 4, in which the smallest occurrences of rainy days occur in the months of August, September and October. For cluster 3, although the rainfall stations show relatively different values of $P10$, the dry period is the same (June to November). Figures 10 and 11 show the results for one station from each cluster.

a) G_1

b) G_2

c) G_3

d) G_4

e) G_5

f) G_6
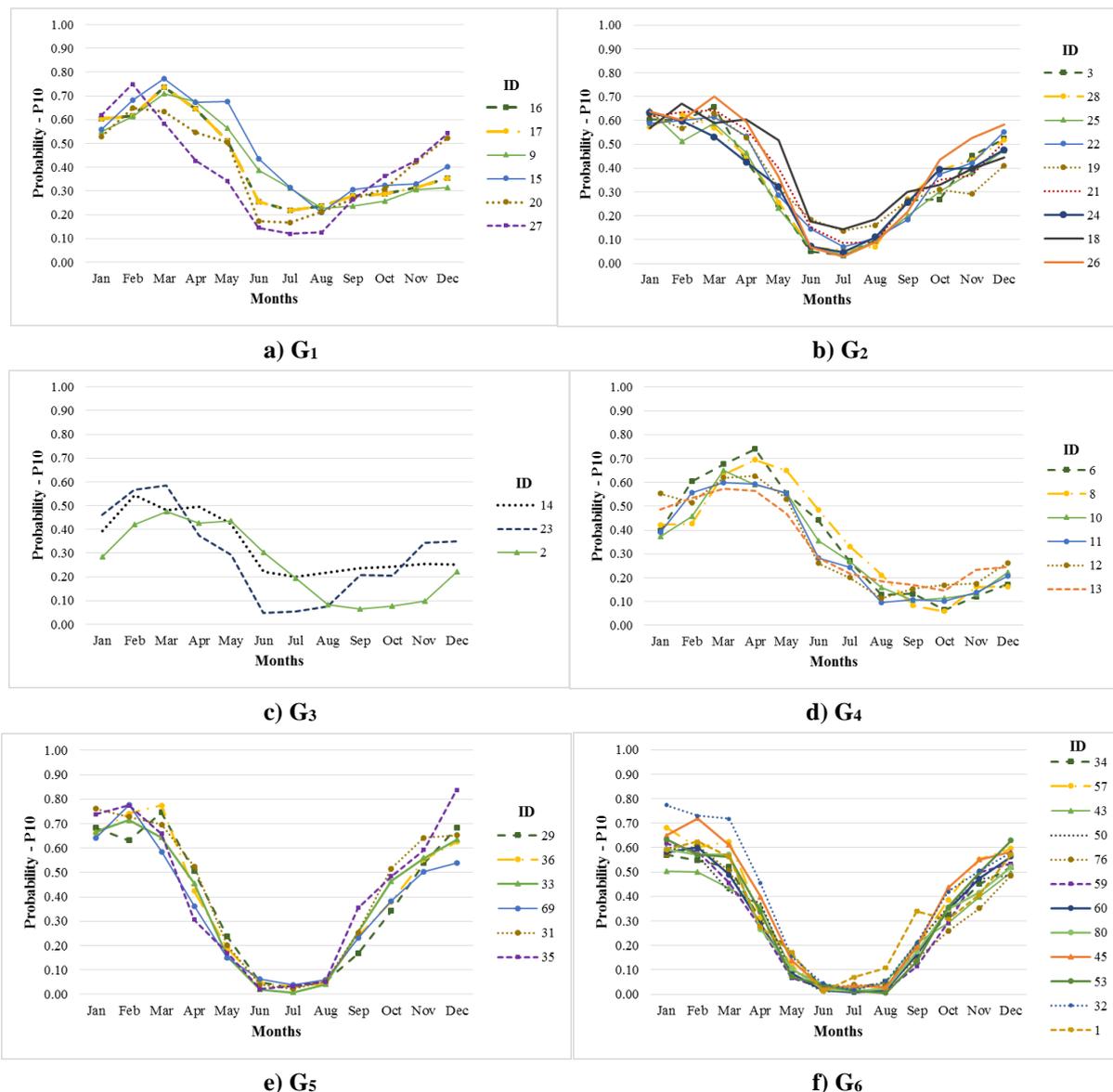
Figure 9 (a, b, c, d, e, f, g, h). Probabilities of transition - rainfall after a dry day (P_{10}) - rainfall gauge stations: clusters G_1, G_2, G_3, G_4, G_5, G_6, G_7 and G_8. Continue.

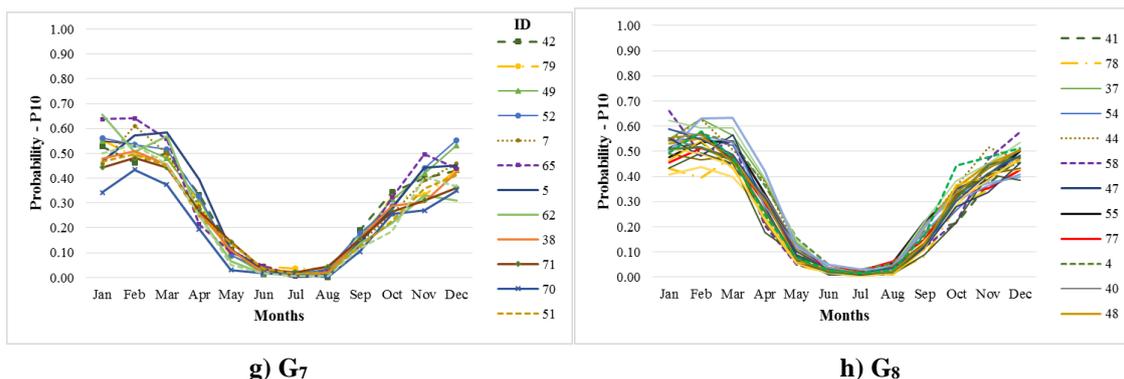**g) G₇**   **h) G₈**

Figure 9 (a, b, c, d, e, f, g, h). Continued.

Figures 10 and 11 demonstrate that transition probabilities provide information on the dry or rainy periods of each season. It is possible to predict the magnitude of each period over each rainfall season. This may be noticed when comparing the probability rates in stations subjected to different climatic factors, for instance, stations 11 (code: 255001) and 80 (code: 1454000), located respectively north and south of TRB, where different biomes predominate, such as the Amazon biome at the mouth and the savannah biome at the headwaters (Mancuzzo *et al.*, 2011). This fact provides particular aspects with regard to climate and rainfall frequency. Climatic factors correspond to the static geographical features of the landscape, such as latitude, altitude, relief and vegetation (Mendonça and Danni-Oliveira, 2007).
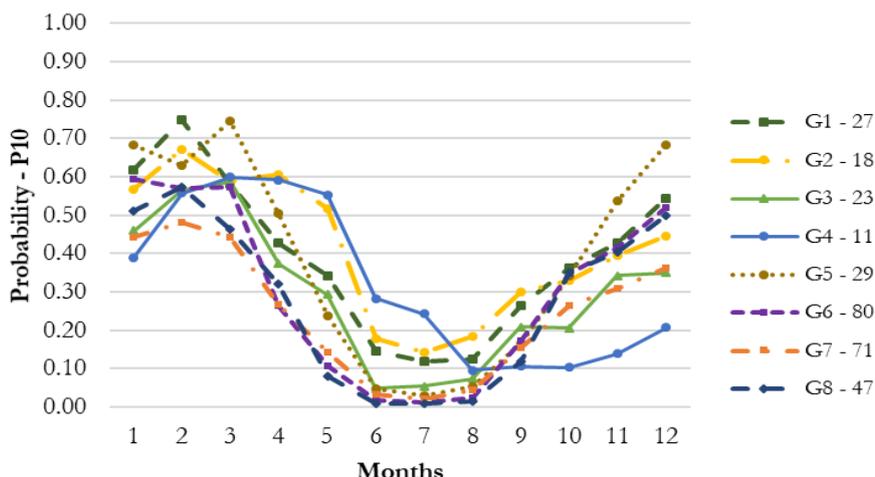

Figure 10. Transition Probabilities - Rain after a dry day (P_{10}) in 8 seasons.
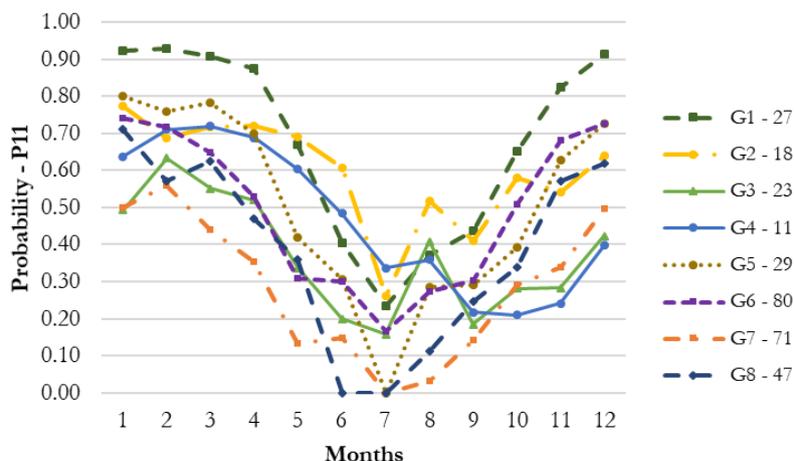

Figure 11. Transition Probabilities - Rain after a rainy day (P_{11}) in 8 seasons.

When the rates of P_{10} of stations 11 and 80 are compared, one may observe that the probability of rain after a dry day presented a rainy and dry period during a good part of the month. It was greater in the region where pluviometric station 255001 was installed. Similar results were reported by Mancuzzo *et al.* (2011) for the state of Mato Grosso (MT) where rainfall rates were higher in areas characterized by the Amazon biome and lowest for the Savannah and Pantanal biomes. According to Ziegler *et al.* (2004), the vegetation cover may influence the rainfall percentage of a given region, since, due to soil cover, the recharge of the surface and underground aquifers tend to increase or decrease due to the direct interference of the flow component. In their study conducted in Cáceres (MT), a municipality located in the south of the state of Mato Grosso, Pizzato *et al.* (2012) registered that rainfall behavior in this region differed from that in a study by Moreira *et al.* (2010) in Nova Maringá (MT) in the north of the same state. The authors concluded that, in the northern region, the drought period occurs earlier when compared to the Pantanal Matogrossense region.

In the case of the study by Moraes *et al.* (2005), the authors observed that December characterizes the beginning of the rainy season in most localities of the state of Pará (PA). However, in a small area south of the state, the beginning of the rainy season may occur in the month of October. However, in a wide range that goes from southwest to south-east, including center-south, the beginning of the rainy season occurs in November. This result does not differ from that found by Menezes *et al.* (2015) when they divided the state of PA into three homogeneous rainfall regions where the biggest rainfall rate (mm) occurred in the south-north region of the state. The occurrence of rainfall increases in most of the state of Pará in December also. In the case of annual rainfall distribution in the state of Pará, and taking into consideration the occurrence of El Niño and La Niña, results of a study by Gonçalves *et al.* (2016) showed that both events triggered the highest rainfall indexes, mainly in the northeast of the state, followed by the south region with the lowest rainfall rate.

In the case of station 255001, the period with the highest rainfall probability comprises the period between January and June, and the period with the lowest rainfall probability comprises the period between July and December, contrary to the result of Moraes *et al.* (2005). However, in the case of station 1454000, the period with the highest rainfall probability lies between October and March. Lowest rainfall probability is more pronounced between April and September. In most stations, especially those located between the headwaters and the center of the basin, probabilities P_{10} indicated a low transition probability in the driest period, i.e., rates close to zero. This transition increased significantly in stations located at the mouth. This may be explained by the type of climate of the region (*"Am"*), characterized by a brief dry season and intense rains during the rest of the year. Figure 11 also reveals that P_{11} rates were higher when compared to those of P_{10}. According to Baú *et al.* (2013), these results tend to confirm the hypothesis of persistence of the preservation of rainfall data from the previous day in the generation of probability of rainfall.

4. CONCLUSION

The Tapajos Basin region has two well-defined seasons, a dry season from May to September and a rainy season from October to April, with May and October characterized as transition periods. Factors of the probability transition matrix show variability in time and also the influence of the geographical position of the rainfall gauge stations on the determination of dry and rainy periods in specific localities of the Tapajos River Basin. Further, 8 regions with rainfall probability were identified by the clustering technique. This identification shows the difference in specific behavior of each rainfall station within the Tapajos River Basin. The insertion of the probability of occurrence analysis for different rainfall volumes according to annual variation may be recommended. For agricultural activity, the definition of these regions

Rev. Ambient. Água vol. 14 n. 3, e2284 - Taubaté 2019

IPABHi

assists in the identification of sites with similar dry and rainy periods, optimizing their planning. Thus, in the months with the greatest number of dry days, being June, July and August in a good part of the basin, and the months of August, September and October for cluster 4, it is necessary to use more irrigation. In defining the best sowing period, the farmer should take into account crop cycles (short or long cycle) in order to avoid crop losses due to the increase in rainy days in the region of interest.

5. ACKNOWLEDGEMENTS

6. REFERENCES

ALMEIDA, A. Q.; RIBEIRO, A.; PAIVA, Y. G.; RASCON, N. J. L.; LIMA, E. P. Geoestatística no estudo de modelagem temporal da precipitação. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 15, n. 4, p. 354–358, 2011.

ANDRADE JÚNIOR, A. S. de; FRIZZONE, J. A.; BASTOS, E. A.; CARDOSO, M. J.; RODRIGUES, B. H. N. Estratégias ótimas de irrigação para a cultura da melancia. **Pesquisa Agropecuária Brasileira**, v. 36, p. 301-305, 2001.

BARRON, J.; ROCKSTROM, J.; GICHUKI, F.; HATIBU, N. Dry spell analysis and maize yields for two semi arid locations in East Africa. **Agricultural and forest Meteorology**, Boston, v. 117, n. 1-2, p. 23 – 37, 2003.

BAÚ, A. L.; AZEVEDO, C. A. V. de; BRESOLIN, A. de A. Modelagem da precipitação pluvial diária intra-anual da Bacia Hidrográfica Paraná III associada aos eventos ENOS. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 17, n. 8, p. 883–891, 2013.

BERTONI, J. C.; TUCCI, C. E. M. Precipitação. In: TUCCI, C. E. M. **Hidrologia**: Ciência e aplicação. Porto Alegre: UFRGS, 2007. p. 177-241.

BREINL, K.; TURKINGTON, T.; STOWASSER, M. Stochastic generation of multi-site daily precipitation for applications in risk management. **Journal of Hydrology**, v. 498, p. 23-35, 2013. https://doi.org/10.1016/j.jhydrol.2013.06.015

CABRERA, J. L. B.; ROMERO, E. A.; SUCH, V. Z.; GARCÍA, C. G.; PORRÚA, F. E. Cluster analysis for validated climatology stations using precipitation in Mexico. **Atmosfera**, v. 25, n. 4, p. 339-354, 2012.

CALGARO, M.; ROBAINA A. D.; PEITER, M. X.; BERNARDON, T. Variação espaço-temporal dos parâmetros para a modelagem estocástica da precipitação pluvial diária no Rio Grande do Sul. **Engenharia Agrícola**, v. 29, n. 2, p.196-206, 2009.

CARVALHO, A. L. de; SOUZA, J. L. de; LYRA, G. B.; WANDERLEY, H. S. Aplicação da Cadeia de Markov para Dias Secos e Chuvosos. **Revista Brasileira de Meteorologia**, v. 32, n. 2, 207-214, 2017. http://dx.doi.org/10.1590/0102-77863220001

COLLISCHONN, B.; COLLISCHONN, W.; TUCCI, C. E. M. Daily hydrological modeling in the Amazon basin using TRMM rainfall estimates. **Journal of Hydrology**, v. 360, n. 1-4, p. 207–216, 2008. https://doi.org/10.1016/j.jhydrol.2008.07.032

CONAB. **Séries históricas relativas às safras 1976/77 a 2018/2019 de área plantada, produtividade e produção**. 2019. Available at: https://www.conab.gov.br/info-agro/safras/serie-historica-das-safras?start=20 Access: 16 Fev. 2019.

DAMÉ, R. DE C. F.; TEIXEIRA, C. F. A.; LORENSI, R. P. Simulação de precipitação com duração horária mediante o uso do modelo Bartlett-Lewis do pulso retangular modificado. **Revista Brasileira de Agrociência**, v. 13, n. 1, p. 13-18, 2007. http://dx.doi.org/10.18539/cast.v13i1.1304

DASH, P. R. A Markov Chain modelling of daily precipitation occurrences of Odisha. **International Journal of Advanced Computer and Mathematical Sciences**, v. 3, n. 4, p. 482-486, 2012.

DETZEL, D. H. M.; MINE, M. R. M. Generation of Daily Synthetic Precipitation Series: Analyses and Application in La Plata River Basin. **The Open Hydrology Journal**, v. 5, 2011.

DOURADO NETO, D.; ASSIS, J. P.; TIMM, L. C.; MANFRON, P. A.; SPAROVEK, G.; MARTIN, T. N. Ajuste de modelos de distribuição de probabilidade a séries históricas de precipitação pluvial diária em Piracicaba-SP. **Revista Brasileira de Agrometeorologia**, v. 13, n. 2, p. 273-283, 2005.

FERNANDES, H. C.; HAMAKAWA, P. J.; LANÇAS, K. P. Metodologia e cálculo dos dias trabalháveis com máquinas florestais na região de Botucatu, SP. **Engenharia Agrícola**, v. 22, n. 1, p. 68-74, 2002.

GARCIA, R. G.; DALLACORT, R.; KRAUSE, W.; SERIGATTO, E. M.; FARIA JÚNIOR, C. A. Calendário agrícola para a cultura do milho em Sinop (MT). **Pesquisa Agropecuária Tropical**, v. 43, n. 2, p. 218-222, 2013. http://dx.doi.org/10.1590/S1983-40632013000200014

GONÇALVES, M. F.; BLANCO, C. J. C.; SANTOS, V. C. dos.; OLIVEIRA, L. L. dos S.; PESSOA, F. C. L. P. Identification of Rainfall Homogenous Regions taking into account El Niño and La Niña and Rainfall Decrease in the state of Pará, Brazilian Amazon. **Acta Scientiarum**, v. 38, n. 2, p. 209-216, 2016. https://dx.doi.org/10.4025/actascitechnol.v28i2.26534

JEONG, D. I.; ST-HILAIRE, A.; OUARDA, T. B. M. J.; GACHON, P. Projection of future daily precipitation series and extreme events by using a multi-site statistical downscaling model over the great Montréal area, Québec, Canada. **Hydrology Research,** v. 44, n. 1, 2013. https://doi.org/10.2166/nh.2012.183

JOHNSON, R. A.; WICHERN, D. W. **Applied Multivariate Statistical Analysis**. 6th ed. Upper Saddle River: Pearson Prentice Hall, 2007.

KELLER FILHO, T.; ZULLO JUNIOR, J.; LIMA, P. R. S. R. Análise da transição entre dias secos e chuvosos por meio da Cadeia de Markov de terceira ordem. **Pesquisa Agropecuária Brasileira**, v. 41, n. 9, p. 1341-1349, 2006. http://dx.doi.org/10.1590/S0100-204X2006000900001

KOTTEK, M.; GRIESER, J.; BECK, C.; RUDOLF, B.; RUBEL, F. World map of the Köppen-Geiger climate classification updated. **Meteorologische Zeitschrift,** v. 15, n. 3, p. 259–263, 2006.

IPABHi

LENNARTSSON, L.; BAXEVANI, A.; CHEN, D. Modelling precipitation in Sweden using multiple step markov chains and a composite model. **Journal of Hydrology**, v. 363, p. 42-59, 2008. http://dx.doi.org/10.1127/0941-2948/2006/0130

LYRA, G. B.; OLIVEIRA-JÚNIOR, J. F.; ZERI, M. Cluster analysis applied to the spatial and temporal variability of monthly rainfall in Alagoas state, Northeast of Brazil. **International Journal of Climatology**, v. 34, n. 3, p. 3546-3558, 2014. https://doi.org/10.1002/joc.3926

MANCUZZO, F. F. N.; MELO, D. C. R.; ROCHA, H. M. Distribuição espaço-temporal e sazonalidade das chuvas no estado do Mato Grosso. **Revista Brasileira de Recursos Hídricos**, v. 16, n. 4, p. 157-167, 2011. http://dx.doi.org/10.21168/rbrh.v16n4.p157-167

MARCELINO, A. S.; ARAÚJO, L. E.; ANDRADE, E. C. A.; ALVES, A. S. Avaliação temporal da climatologia do litoral norte da Paraíba. **Revista Brasileira de Geografia Física**, v. 5, n. 3, p. 467-472, 2012.

MARENGO, J. A.; SCHAEFFER, R.; ZEE, D.; PINTO, H. S. **Mudanças climáticas e eventos extremos no Brasil**. Available at: http://www.fbds.org.br/cop15/FBDS_MudancasClimaticas.pdf. Access: Mar. 2017.

MARTIN, T. N.; STORCK, L.; DOURADO NETO, D. Simulação estocástica da radiação fotossinteticamente ativa da temperatura do ar por diferentes métodos. **Pesquisa Agropecuária Brasileira**, v. 42, n. 9, p. 1211-1219, 2007.

MEHROTRA, R.; SHARMA, A. Assessing rainfall availability over the Sydney region in a future climate using stochastic downscaling. In: WORLD IMACS / MODSIM CONGRESS, 18. 2009, Cairns. **Anais[...]** Cairns: MODSIM, 2009.

MENDONÇA, F.; DANNI-OLIVEIRA, I. M. **Climatologia**: Noções básicas e climas do Brasil. São Paulo: Oficina de Texto, 2007.

MENEZES, F. P.; FERNANDES, L. L.; ROCHA, E. J. P. da. O uso da estatística para regionalização da precipitação no Estado do Pará, Brasil. **Revista Brasileira de Climatologia**, v. 16, 2015. http://dx.doi.org/10.5380/abclima.v16i0.40023

MORAES, B. C. de; COSTA, J. M. N. da; COSTA, A. C. L. da.; COSTA, M. H. Variação espacial e temporal da precipitação no Estado do Pará. **Revista Acta Amazônica**, v. 35, n. 2, p. 207 – 214, 2005.

MOREIRA, P. S. P.; DALLACORT, R.; MAGALHÃES, R. A.; INOUE, M. H.; STIELER, M. C.; SILVA, D. J. da; MARTINS, J. A. Distribuição e probabilidade de ocorrência de chuvas no município de Nova Maringá-MT. **Revista de Ciências Agro-Ambientais**, v. 8, n. 1, p. 9- 20, 2010.

PIZZATO, J. A.; DALLACORT, R.; TIEPPO, R. C.; MODOLO, A. J.; CREMON, C.; MOREIRA, P. S. P. Distribuição e probabilidade de ocorrência de precipitação em Cáceres (MT). **Pesquisa Agropecuária Tropical**, v. 42, n. 2, p. 137-142, 2012. http://dx.doi.org/10.1590/S1983-40632012000200006

RAZIEI, T.; BORDI, I.; PEREIRA, L. S. Regional drought modes in Iran using the SPI: the effect of time scale and spatial resolution. **Water Resources Management**, v.27, n.6, p.151–159, 2012. https://doi.org/10.1007/s11269-012-0120-3

SALAS, J. D.; RAJAGOPALAN, B.; SAITO, L.; BROWN, C. Special Section on Climate Change and Water Resources: Climate Nonstationarity and Water Resources Management. **Journal of Water Resources Planning and Management**, v. 138, n. 5, p. 385–388, 2012.

SANTOS, C. A. C.; BRITO, J. I. B.; RAMANA RAO, T. V.; MENEZES, H. E. A. Tendências dos índices de precipitação no Estado do Ceará. **Revista Brasileira de Meteorologia**, v. 24, n. 1, p. 39-47, 2009.

SANTOS, C. A. dos; SERRÃO, E. A. O.; GONÇALVES, L. de J. M.; WANZELER, R. T. S.; LIMA, A. M. de. Zoneamento da Distribuição da Precipitação Pluviométrica na Bacia Hidrográfica do Rio Tapajós. **Enciclopédia Biosfera**, v. 10, n. 18; p. 3092-3106, 2014.

SELVARAJ, R. S.; SELVI, S.T. Stochastic modeling of daily rainfall at Aduthurai. **International Journal of Advanced Computer and Mathematical Sciences**, v. 1, n. 1, p.52-57, 2010.

SEMENOV, M. A. Simulation of extreme weather events by a stochastic weather generator. **Climate Research**, v. 35, n. 3, p. 203-212, 2008. http://dx.doi.org/10.3354/cr00731

SHARIF, M.; BURN, D. H.; WEY, K. M. Daily and Hourly Weather Data Generation using a K-Nearest Neighbour Approach. In: CANADIAN HYDROTECHNICAL CONFERENCE, 2007, Winnipeg. **Proceedings[...]** Winnipeg: CHC, 2007. p. 1-10.

STOWASSER, M. Modelling rain risk: a multi-order Markov chain model approach. **The Journal of Risk Finance**, v. 13, n. 1, p. 45 – 60, 2012. https://doi.org/10.1108/15265941211191930

SUKLA, M. K.; MANGARA, A. K. J.; SAHOO, L. N. Markov Chain Modeling of Daily Rainfall Occurrence in the Mahanadi Delta of India. **Journal of Applied Mathematics and Statistical Sciences,** v. 1, n. 1, p. 21-30, 2016.

SZYNISZEWSKA, A. M.; WAYLEN, P. R. Determining the daily rainfall characteristics from the monthly rainfall totals in central and northeastern Thailand. **Applied Geography**, v. 35, n. 1-2, p. 377-393, 2012. https://doi.org/10.1016/j.apgeog.2012.09.001

VASCONCELLOS, S. L. B.; ANDRÉ, R. G. B.; PERECIN, D. Probabilidade de ocorrência de dias secos para a região de Jaboticabal-SP. **Revista Brasileira de Agrometeorologia**, v. 11, n. 2, p. 321-325, 2003.

VELA, R. H. N.; DALLACORT, R.; NIED, A. H. Distribuição descendial, mensal e totais de precipitação na região de Tangará da Serra - MT. *In:* CONGRESSO BRASILEIRO DE ENGENHARIA AGRÍCOLA, 36., 2007, Bonito. **Anais[...]** Bonito: SBEA, 2007. p. 1-4.

VIANA, T. V. A.; AZEVEDO, B. M.; BOMFIM, G. V.; ANDRADE JUNIOR, A. S. Probabilidade de ocorrência de períodos secos e chuvosos, em Pentecoste, CE. **Irriga (Botucatu)**, v. 7, n. 03, p. 226-229, 2002. https://doi.org/10.15809/irriga.2002v7n3p226-229

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

Rev. Ambient. Água vol. 14 n. 3, e2284 - Taubaté 2019

IPABH↑

YANG, T.; SHAO, Q.; HAO, Z.; CHEN, X.; ZHANG, Z.; XU, C.; SUN, L. Regional frequency analysis and spatio-temporal pattern characterization of rainfall extremes in the Pearl River Basin, China. **Journal of Hydrology**, v. 380, n. 3-4, p. 386-405, 2010. https://doi.org/10.1016/j.jhydrol.2009.11.013

YOO, C.; LEE, J.; RO, Y. Markov Chain Decomposition of Monthly Rainfall into Daily Rainfall: Evaluation of Climate Change Impact. **Advances in Meteorology**, v. 2016, 10 p., 2016. http://dx.doi.org/10.1155/2016/7957490

ZIEGLER, A. D.; GIAMBELLUCA, T. W.; TRAN, L. T.; VANA, T. T.; NULLET, M. A.; FOX, J.; VIEN, T. D.; PINTHONG, J.; MAXWELL, J. F.;EVETT, S. Hydrological consequences of landscape fragmentation in mountainous northern Vietnam: evidence of accelerated overland flow generation. **Journal of Hydrology**, n. 287, p. 124-146, 2004. https://doi.org/10.1016/j.jhydrol.2003.09.027

IPABH