Acta Scientiarum
http://periodicos.uem.br/ojs/acta
ISSN on-line: 1807-8664
Doi: 10.4025/actascitechnol.v41i1.37186

Application of Markov chain on daily rainfall data in Paraíba-Brazil from 1995-2015

Jader da Silva Jale[1*], Sílvio Fernando Alves Xavier Júnior[2], Érika Fialho Morais Xavier[2], Tatijana Stošić[1], Borko Stošić[1] and Tiago Alessandro Espínola Ferreira[1]

[1]Departamento de Estatística e Informática, Universidade Federal Rural de Pernambuco, Rua Manuel de Medeiros, s/n., 52171-900, Recife, Pernambuco, Brazil. [2]Departamento de Estatística, Universidade Estadual da Paraíba, Campina Grande, Paraíba, Brazil. *Author for correspondence. E-mail: jsj.brainstorm@gmail.com

ABSTRACT. This study analyzed the behavior of daily rainfall in the State of Paraíba using the data from five meteorological stations distributed across the mesoregions of this state. We used the three-state Markov Chain model, in which states are defined as dry, wet and rainy. We calculated transition probabilities among states, probabilities of equilibrium of states, and expected lengths of the defined states for all stations and seasons to investigate spatial/seasonal variability. Results showed that for the entire region and for all seasons, the probability of dry days is greater than the probability of rainy days; expected values of rainy spells are low, indicating that the rainfall regime in Paraíba is characterized by high rainfall intensity distributed over short rainy periods. The dry-dry transition probability presents the highest values for all seasons and stations, as well as the corresponding expected dry spell length, indicating that this region is subjected to prolonged dry periods. The transition probabilities that lead to dry condition are higher in the interior of the State, while probabilities that lead to rainy condition are higher in the coastal region as well as the probability of rainy days, which is greater in fall, during the rainy season.

Keywords: Markov chain; transition probability matrix; rainfall; Paraíba; Brazil.

Received on May 11, 2017.
Accepted on February 8, 2018.

Introduction

Depletion of water resources in different parts of the world is one of the renowned environmental problems of this century. Thus, an essential aspect of water resources planning is the analysis of long-term records of hydro–meteorological variables. As a primary input to the hydrological cycle, rainfall represents the potential availability of water resources of an area (Maruyama, Kawachi, & Singh, 2005). The knowledge about the daily rainfall distribution is fairly crucial for water use practices and future planning in agriculture (planting, irrigation and drainage), civil defense (risk of landslide, forest fires or floods), and hydrology (river flow estimation, sediment transport) (Cull, Hearn, & Smith, 1981; Ingram, Roncoli, & Kirshen, 2002; Seeger et al., 2004; Pereira, Trigo, Camara, Pereira, & Leite, 2005; Collischonn, Haas, Andreolli, & Tucci, 2005; Minder, Roe, & Montgomery, 2009; Guhathakurta, Sreejith, & Menon, 2011).

A variety of models has been proposed for simulating daily rainfall and its spatial pattern (Richardson, 1981; Wilks, 1998; Serinaldi, 2010; Asong, Khaliq, & Wheater, 2016; Müller & Haberlandt, 2016). The first order two state ('rain' and 'no rain') Markov model was applied on Tel Aviv daily rainfall data by Gabriel and Neuman (1962) and since then Markov chain models including multistate (Haan, Allen, & Street, 1976; Pegram, 2009) higher order (Lana & Burgueño, 1998; Deni, Jemain, & Ibrahim, 2009), hidden Markov model (Hughes, Guttorp, & Charles, 1999; Robertson, Kirshner, & Smyth, 2004) and non-homogeneous Markov model (Rajagopalan, Lall, & Tarboton, 1996) were widely used for modeling daily occurrence of precipitation.

The Northeast of Brazil has mostly a semi-arid climate. It is characterized by high seasonal and interannual rainfall variability, with extreme wet and dry episodes. Seasonal variation is associated with the migration and intensity of the Intertropical convergent zone (ITCZ) and South American monsoon system (Sams). During the rainy season which occurs in February-April, the Atlantic ITCZ reaches its southernmost position and directly overlies NE Brazil, merging with the Sams, which moves northward. For most of year, the climate is semiarid under the influence of the South Atlantic subtropical anticyclone (Robertson et al., 2004). Interannual variability of rainfall with severe droughts (excessive rainfall) occur when the southward seasonal migration of the ITCZ is inhibited (amplified) (Hastenrath & Heller, 1977). The interannual behavior of the Atlantic ITCZ is related to El Niño - Southern Oscillation (Enso) in the equatorial Indo-Pacific Ocean and the meridional sea surface temperature gradient (Mgrad) over the tropical Atlantic (see e.g. Lucena, Servain, & Gomes Filho, 2011; and references therein). During the positive-El Niño (negative La Niña) phase of Enso, there is a tendency for the inhibition (reinforcement) of the convective system over the west tropical Atlantic leading to less (more) rainfall over Northeast Brazil. During a negative phase of Mgrad, which is characterized by a negative SST (sea surface temperature) anomaly in the north tropical Atlantic and a positive SST anomaly in the south, there is a thermal gradient directed towards the Southern Hemisphere, the intertropical convergence zone (ITCZ) moves southward of its normal climatological position leading to increased rainfall over Northeast Brazil. This effect is reversed during a positive phase of Mgrad, the regular ITCZ moves northward and the rainfall rate is below the average in Northeast Brazil (Hastenrath, 1990; Wagner, 1996). The largest ITCZ displacements occur in Enso years in which preexisting Atlantic SST anomalies are such that they amplify the direct impact of Enso (Giannini, Saravanan, & Chang, 2004), which can lead to dramatic drought in the region with serious environmental, economic and social consequences (Marengo, Torres, & Alves, 2016). Rainfall spatial and temporal patterns in NEB were extensively studied, however most of the studies concentrate on interior dry region (Hastenrath, 1990; Uvo, Repelli, Zebiak, & Kushnir, 1998; Moscati & Gan, 2007; Hastenrath, 2012; Rao, Franchito, Santo, & Gan, 2016), while the knowledge of rainfall characteristics of eastern and southern part remain less complete (Chaves & Cavalcanti, 2001; Lyra, Oliveira-Júnior, & Zeri, 2014).

This study aimed to provide further insight into the pattern of rainfall distribution in the Brazilian northeast, specifically the State of Paraíba with a large part (about 80% of total territory) located in the so called ‘semiarid polygon’, making it extremely vulnerable to rainfall seasonal and interannual variability (Silva, Costa, Campos, & Dantas, 2009). We define three daily weather conditions (dry, wet and rainy) by using appropriate rainfall threshold and employ Markov chain to determine the probability of transitions between these states, equilibrium probability of each state and expected lengths of dry, wet and rainy spells. These results are then used to perform an analysis of temporal and spatial variability of rainfall.

Material and methods

Study area

Paraíba is a State in the Northeast of Brazil, located between the parallels 6° 2’ 24” S and 8° 18’ 36” S and meridians 34° 49’ 48” W and 38° 46’ 12” W, it borders the States of Pernambuco, Rio Grande do Norte and Ceará, limiting to the east with the Atlantic Ocean. The State is divided into four geographic mesoregions: Mata Paraibana (5.242 km^2), Agreste (12.914 km^2), Borborema (15.572 km^2) and Sertão (22.720 km^2). In most of its territory, Paraíba has a semi-arid climate, characterized by high seasonal and interannual rainfall variability and experiences extreme wet and dry conditions. Historically, Paraíba has a short rainy season, which usually occurs from April to June in Mata Paraibana, from March to May in Agreste and Borborema and from January to March, in Sertão (Soares, Paz, & Piccilli, 2016). The study area (State of Paraíba with division into four mesoregions) is shown in Figure 1.

Figure 1. The map of the State of Paraíba divided into mesoregions.

Data on daily rainfall in Paraíba for twenty years from January 1, 1995, to December 31, 2015, for five meteorological stations (João Pessoa, Areia, Campina Grande, Monteiro and Patos), which are distributed over all mesoregions were collected from Brazilian National Institute of Meteorology (Instituto Nacional de Meteorologia - Inmet) and are available at http://www.inmet.gov.br/. The daily rainfalls for the four (nominal) season periods of summer (Dec 21-Mar 20), fall (Mar 20-Jun 21), winter (Jun 21-Sep 22) and spring (Sep 22-Dec 21), were separated and analyzed to investigate spatial and seasonal variability. There are no missing observations in the entire daily rainfall data set recorded. All statistical analyses were run using the statistical software R. (R Core Team, 2016).

Markov chain

Markov chain models have been widely used for simulating discrete time series. Some applications include vegetation dynamics (Balzter, 2000), hydrological processes (Schoof & Pryor, 2008; Fu, Li, & Huang, 2012), wind speed (Shamshad, Bawadi, Hussin, Majid, & Sanusi, 2005) and urban growth (Le Gallo & Chasco, 2008). Markov chain models can vary in two properties: the number of states (different values that the variable can assume) and the order (number of previous values used to determine the state-to-state transition probabilities) (Schoof & Pryor, 2008).

Ratan and Venugopal (2013) report that the Indian Meteorological Department uses 2.5 mm as a threshold to define a rainy day. However, daily rainfall less than 5.0 mm is ineffective as this amount of rain would evaporate before entering the ground (Ali & Mubarak, 2017). Therefore, rainfall would only humidify the surface. This would be beneficial initially only for plantations of superficial or medium roots, e.g., onion and wheat, respectively (Ali & Mubarak, 2017). In this way, we use 2.5 and 5.0 mm cut-off point.

In this work, we used first order three-state Markov chain to study the behavior of rainfall occurrence in Paraíba. The states used were: dry (d), wet (w) and rainy (r). A day was considered dry (d) if rainfall occurrence on that day was at most 2.50 mm; wet (w) if rainfall occurrence was between 2.51 and 5.00 mm and, rainy (r) if rainfall was more than 5.00 mm. The probability of the rainfall process being in a particular state was calculated based on the first order Markov chain assumption that the current day's rainfall depends only on the preceding day's rainfall (Garg & Singh, 2010). The transition probability matrix is defined as $P_{ij} \equiv P(j|i)$, where $i, j \in S$ ($S = \{d, w, r\}$is the state space) and is listed in Table 1.

The transition probabilities are defined as follows: $P_{dd} = P(d|d)$is the probability of a dry day preceded by a dry day, $P_{dw} = P(w|d)$is the probability of a wet day preceded by a dry day, $P_{dr} = P(r|d)$ is the probability of a rainy day preceded by a dry day, and so on. The sum of probabilities of each row equals unity: $P_{dd} + P_{dw} + P_{dr} = 1$, $P_{wd} + P_{ww} + P_{wr} = 1$ and $P_{rd} + P_{rw} + P_{rr} = 1$.

Estimation for Markov Chains

The transition probability matrix is generally unknown and should be estimated through observations. In the literature, some methods for estimating the transition probabilities are: Maximum Likelihood Estimator (MLE), Bootstrap Method, Smoothed Estimators, Lagrange Multipliers (Zhang, Wang, & Zhang, 2014). Here, we used the MLE.

Let $n, m \in N$, with $n, m \geq 1$. Define $S = \{1, \ldots, m\}$. Consider $X_1^n = \{X_1, \ldots, X_n\}$ a sequence of random variables such that $P_{ij} = P(X_{t+1} = j \vee X_t = i)$ is independent from t, $\forall i, j \epsilon S$. Therefore, the sequence $X_1^n = \{X_1, \ldots, X_n\}$ is a Markov chain with state space S and transition probability $P_{ij} \forall i, j \epsilon S$, subjected to $\sum_{j=1}^{m} P_{ij} = 1$ and $0 \leq P_{ij} \leq 1$.

Taking a sample from the chain $x_1^n = \{x_1, \ldots, x_n\}$, we have the realization of the random variable $X_1^n = \{X_1, \ldots, X_n\}$ with probability $P(X_1^n = x_1^n) = P(X_1 = x_1) \prod_{t=2}^{n} P(X_t = x_t | X_{t-1} = x_{t-1}$. Let $\theta = \{P_{ij}, \forall i, j \in S\}$. Its likelihood function is given by $L(\theta) = P(x_1; \theta) \prod_{i=1}^{m} \prod_{j=1}^{m} {P_{ij}}^{n_{ij}}$ and $L(\theta) = logL(\theta) = logP(x_1; \theta) + \sum_{ij} n_{ij} \, logP_{ij}$, subjected to $\sum_j P_{ij} = 1$. We must choose a probability of any transition to express it in terms of others, say $P_{i1} = 1 - \sum_{j=2}^{n} P_{ij}$. Applying the derivative in $L(\theta)$ in relation to P_{ij}, we have $\partial L / \partial P_{ij} = n_{ij}/P_{ij} - n_{i1}/P_{i1} = 0 \rightarrow n_{ij}/\hat{P}_{ij} = n_{i1}/\hat{P}_{i1} \rightarrow n_{ij}/n_{i1} = \hat{P}_{ij}/\hat{P}_{i1}$. As it is true for all $j \neq 1$, then $\hat{P}_{ij} \propto n_{ij}$. Therefore, the probabilities $P_{ij}, i, j = d, w, r$ can be estimated from the corresponding observed absolute frequencies n_{ij} of days being in a particular state j preceded by a state i, as the maximum likelihood estimators (MLE) of $P_{ij}, (i, j = d, w, r)$ are given by $\hat{P}_{ij} = n_{ij}/\sum_{j=1}^{m} n_{ij}$. Craig and Sendi (2002) suggest Efron's bootstrap as a method to construct confidence intervals for transition probability matrix functions.

Table 1. Transition probability matrix for three-state Markov chain.

		Present Day (j)		
		dry (d)	wet (w)	rainy (r)
Previous Day (i)	dry (d)	P_{dd}	P_{dw}	P_{dr}
	wet (w)	P_{wd}	P_{ww}	P_{wr}
	rainy (r)	P_{rd}	P_{rw}	P_{rr}

Goodness-of-fit test

In order to model rainfall dynamics in Paraíba using Markov chain, we first tested the validity of the proposed three-state Markov chain approach: the null hypothesis H_0: Rainfall occurrences on consecutive days are independent, vs. alternative hypothesis H_1: Rainfall occurrences on consecutive days are not independent. This test is used to verify if the behavior of the rainfall can be explained by the model proposed: a Markov chain of three-states of order 1, i.e., rainfall occurrence on successive days is not independent (Garg & Singh, 2010). For three-state Markov chain, Wang and Martiz (1990) suggested a WS test statistics, given by Equation 1.

$$WS = \frac{A+B-1}{\sqrt{V(A+B-1)}} \underset{\rightarrow}{d} N(0,1) \quad (1)$$

where:
$A = P_{dd} + P_{ww} + P_{rr}, B = P_{rd}P_{dr} + P_{wr}P_{rw} + P_{dw}P_{wd} - P_{dd}P_{ww} - P_{dd}P_{rr} - P_{ww}P_{rr}$. The variance of $(A + B - 1)$ in (Equation 1) is given by $V(A + B - 1) = 2p_1p_2p_3\left(\frac{1}{n_dn_w} + \frac{1}{n_wn_r} + \frac{1}{n_rn_d}\right)$, n_d, n_w, n_r, are numbers of dry, wet, and rainy days and $p_{1,} p_2$ and p_3 are the stationary probabilities, which are calculated as $p_1 = [(1 + p) + (1 + s)\, p/q]^{-1}, p_2 = [r + ps/q]p_1$ and $p_3 = [p/q]p_1$ where $p = \left[P_{dr} + \frac{P_{wr}(1-P_{dd})}{P_{wd}}\right]\left(\frac{1}{1-P_{rr}}\right), q = 1 + \left[\frac{P_{wr}P_{rd}}{P_{wd}(1-P_{rr})}\right]; r = \left(\frac{P_{dw}}{1-P_{ww}}\right), s = \left(\frac{P_{rw}}{1-P_{ww}}\right)$.

According to the test procedures, critical region is $|\mathrm{WS}|_c \geq Z_\alpha$ at α level of significance i.e. the null hypothesis is rejected if $|WS| \geq Z_\alpha$, where Z_α is the 100 $(1 - \alpha)$ lower percentage point of a standard normal distribution (Garg & Singh, 2010).

Equilibrium probabilities

The equilibrium probabilities $\pi_{d,}\pi_w$ and π_r (of a dry, wet and rainy day, respectively) are found by solving the stationary matrix equation (Garg & Singh, 2010). This means that the probability of being in any state will not change over time. This indicates that the transition matrix will stabilize the values of its elements in the long term, i.e., there is a limit such that $lim_{n\to\infty} P_{ij}^{(n)} = \pi_j > 0$ where the π_j satisfy only the following stable state equations $\pi_j = \sum_{i=1}^{m} \pi_i \, P_{ij}$ to $j = 1, \ldots, m$. Then Equation 2:

$$(\pi_d\pi_w\pi_r) = (\pi_d\pi_w\pi_r)\begin{pmatrix} P_{dd} & P_{dw} & P_{dr} \\ P_{wd} & P_{ww} & P_{wr} \\ P_{rd} & P_{rw} & P_{rr} \end{pmatrix} (2)$$

so that the estimators of the long run equilibrium probabilities are given by solving the linear system of equations $\pi_d = \pi_dP_{dd} + \pi_wP_{wd} + \pi_rP_{rd}, \pi_w = \pi_dP_{dw} + \pi_wP_{ww} + \pi_rP_{rw}$ and $\pi_r = \pi_dP_{dr} + \pi_wP_{wr} + \pi_rP_{rr}$, together with the probability normalization condition $\pi_d + \pi_w + \pi_r = 1$.

Distribution of events and expected length of dry, wet and rainy spells

Given a discretized series of days for the amount of rainfall each day, the distribution separating events follows a negative binomial distribution (Longley, 1953). Let $r > 0$ and $p \in (0,1)$. A random variable X has the negative binomial distribution $X \sim NB\,(r,p)$ if $P(X = k) = \{\Gamma(r + k - 1)/\Gamma(r)\Gamma(k)\}\, p^r(1 - p)^{k-1}$, $k = 1,2, \ldots$ The expected length of different spells is calculated as follows (Garg & Singh, 2010). The probability of dry spell of length d (a sequence of d consecutive dry days preceded and followed by wet or rainy days) is $p(d) = (P_{dd})^{D-1}(1 - P_{dd})$. The expected length of dry spell is given by $E(D) = 1/(1 - P_{dd})$. The expected length of dry spell is given by $E(D) = 1/(1 - P_{dd})$. Equivalently, the probability of a wet spell of length w (a sequence of w consecutive wet days preceded and followed by dry or rainy days) is $p(w) = (P_{ww})^{W-1}(1 - P_{ww})$ with the expected length of wet spell given by $E(W) = 1/(1 - P_{ww})$. The probability of a rainy spell of length r (a sequence of r consecutive rainy days preceded and followed by dry or wet days) is $p(r) = (P_{rr})^{R-1}(1 - P_{rr})$ with the expected length of rainy spell given by $E(R) = 1/(1 - P_{rr})$. Finally, the weather Cycle (WC) is defined as $E(WC) = E(D) + E(W) + E(R)$.

Results and discussion

Table 2 lists the geographic information about stations used in this study and descriptive statistics for 1995-2015 daily data.

It is seen in Table 2 that the coastal region of Paraíba (João Pessoa station) receives most rain (highest mean value), while Sertão (Pato station) is affected by the strongest wet episodes (highest maximum value). Table 3 presents the linear correlations and distances (km) in a straight line between the stations. When distance increases, linear correlation decreases among rainfall time series.

The hypothesis of dependence of rainfall occurrences on successive days is tested independently for the four seasons (fall, winter, spring, summer), and the estimated values of WS statistics (1) for the four seasons and their p-values are presented in Table 4.

The results of goodness-of-fit test (Table 4) show that rainfall occurrences on consecutive days in Paraíba for each of the four seasons are dependent ($p < 0.05$), satisfying a major property of the Markov chain model, except for Patos station for winter season ($p = 0.3276$). This occurs because there was no transition of the type$w \rightarrow w$ and $r \rightarrow w$in winter.

Table 5 lists the estimated equilibrium state probabilities, the expected length of different spells, the weather cycle, and the total number of days, for each season for five Paraíba's stations. It is seen in Table 5 that for the stations located in Mata Paraíbana (João Pessoa), Agreste (Areias and Campina Grande) and Borborema (Monteiro), the probability of rainy days is greatest in fall, during the rainy season, with the highest value for João Pessoa station (~35%), located in the coastal region. Patos station is located in Sertão with the rainy season in summer, and consequently the greater probability of rainy days (15%), although this probability is much smaller than for dry days (81%). For all seasons (including the rainy season), the probability of dry days (59-98%) is greater than the probability of rainy days (2-35%), indicating that the rainfall regime in Paraíba is characterized by high rainfall intensity over short rainy spells (the expected length of rainy spell during the rainy season is ~1.5 days). For all stations, except for Patos, the probability of dry days (87-96%) and the expected length of dry spells (13-42 days) is greater in spring, with more prolonged dry spells in Campina Grande and Monteiro (expected lengths of dry spells are 29.24 and 42.02, respectively). In Sertão (Patos station) with prolonged dry season, the probability of dry days (~98%) and the expected length of dry spells (~50 days) is greater in winter and spring. Weather cycle also shows seasonal and spatial variability: lowest for João Pessoa station in fall (6.07) and winter (5.97) and highest for Patos station in winter (51.79) and spring (52.45).

Table 2. Station locations and descriptive statistics (in mm) for daily rainfall.

Station	Long.	Lat.	Altitude	Median	Mean	Min.	Max.	Std.Dev.	C.V.
Areia	-35.68	-6.97	574.62	0.1	3.59	0	146	8.9026	2.48
C. Grande	-35.88	-7.22	547.56	0	2.183	0	110.1	6.6777	3.0587
J. Pessoa	-34.86	-7.1	7.43	0.2	5.11	0	186	13.56	2.6537
Monteiro	-37.06	-7.88	603.66	0	1.864	0	103.4	7.4048	3.9723
Patos	-37.26	-7.01	249.09	0	1.89	0	258.2	8.7564	4.6328

Table 3. Correlations between the daily rainfall series and distance between cities.

	Areia	C. Grande	J. Pessoa	Monteiro	Patos
Areia	1	0.6631 (35)	0.4475 (92)	0.1573 (182)	0.1461 (174)
C. Grande		1	0.4149 (113)	0.1877 (149)	0.1790 (154)
J. Pessoa			1	0.0915 (258)	0.1233 (265)
Monteiro				1	0.2687 (99)
Patos					1

Table 4. Estimated values of WS Test Statistics and the associated p-values.

Station	Fall	Winter	Spring	Summer
Areia	11.54 (p < 0.001)	36.69 (p < 0.001)	117.76 (p < 0.001)	123.98 (p < 0.001)
C. Grande	43.24 (p < 0.001)	9.3 (p < 0.001)	2.99 (p = 0.0014)	7.8 (p < 0.001)
João Pessoa	14.97 (p < 0.001)	6.7 (p < 0.001)	6.17 (p < 0.001)	4.26 (p < 0.001)
Monteiro	53.59 (p < 0.001)	156.17 (p < 0.001)	188.05 (p < 0.001)	20.93 (p < 0.001)
Patos	11.66 (p < 0.001)	0.45 (p = 0.3276)*	2.27 (p = 0.0117)	5.26 (p < 0.001)

These results imply qualitative agreement with findings of Robertson et al. (2004) who applied hidden Markov model (HMM) to describe daily rainfall occurrence at 10 gauge stations in the State of Ceará, northeast Brazil, during the February-April wet season 1975-2002. They identified four 'hidden' rainfall states, where two of the states are found to correspond to wet or dry conditions at all stations, with similar relative frequencies. They also found that the wet state tends to be more prevalent during March (fall), however the dry state is more prevalent at the beginning of the FMA (February-March-April) rainy season.

Table 6 presents the fit for the duration of days according to the occurrence of states. The dispersion parameter refers to the important property of the negative binomial distribution of the over-dispersion, since its variance exceeds the mean. Note that João Pessoa presented the highest mean duration and standard error for the three events d, w and r. This indicates that there is a greater variability in the duration of these events/states around their average duration. We highlight the Areia station, with behavior closer to João Pessoa in relation to the other stations. This suggests the influence of the climatology of the Borborema Plateau, which is hot and humid.

Figure 2 illustrates the interpolated maps of transition probabilities estimated among states d, w, r for winter season. In winter, the dry–dry transition presented the highest probability values above 65% (a), while dry–wet (b) and rainy–wet (h) transition presented the lowest probability, below 10%. The dry-rainy (c), wet-rainy (f) and rainy-rainy (i) transition probability increased towards the coastline, while dry-dry (a), wet-dry (d) and rainy–dry (g) transition probability increased in the Coastal–Sertão direction. Sertão region (Patos station) showed the highest transition probabilities that lead to dry condition (Figure 2 a, d, g) and lowest transition probabilities that lead to wet and rainy conditions (Figure 2 b, c, e, f, h, i).

For all seasons and for all stations, there was the highest probability of the transition from a dry day to another dry day, reflecting the climate characteristics in the Northeast of Brazil: the rainfall is concentrated in short time periods with prolonged dry spells (Lucena et al., 2011). The transition probabilities from wet and rainy conditions to dry condition are higher in the Sertão region, while transition probabilities that lead to rainy conditions are higher in the coastal region.

Table 5. Equilibrium state probabilities, expected length of different spells, weather cycle, and total number of days for Paraíba's rainfall stations for different seasons. The observed lengths of different spells and weather cycles are given in parentheses.

Station	Season	π_d	π_w	π_r	Expected Length of Season's Spell				Total
					Dry (d)	Wet (w)	Rainy (r)	Weather Cycle	
Areia	Fall	0.63	0.1	0.27	3.55 (4)	1.09 (1)	1.7 (2)	6.34 (6)	1840
	Winter	0.65	0.09	0.26	3.89 (4)	1.19 (1)	1.73 (2)	6.81 (7)	1879
	Spring	0.87	0.09	0.24	13.09 (13)	2.19 (2)	1.3 (1)	16.58 (17)	1800
	Summer	0.76	0.1	0.14	6.08 (6)	1.83 (2)	1.48 (1)	9.39 (9)	1780
Campina Grande	Fall	0.72	0.1	0.18	4.61 (5)	1.29 (1)	1.5 (2)	7.4 (7)	1840
	Winter	0.76	0.09	0.15	5.41 (5)	1.18 (1)	1.39 (1)	7.9 (8)	1879
	Spring	0.96	0.02	0.02	29.24 (29)	1.15 (1)	1.12 (1)	31.51 (32)	1800
	Summer	0.87	0.04	0.09	10.13 (10)	1.11 (1)	1.33 (1)	12.57 (13)	1780
João Pessoa	Fall	0.57	0.08	0.35	2.94 (3)	1.12 (1)	2.01 (2)	6.07 (6)	1840
	Winter	0.59	0.11	0.3	3.02 (3)	1.14 (1)	1.81 (2)	5.97 (6)	1879
	Spring	0.9	0.05	0.05	13.53 (14)	1.16 (1)	1.26 (1)	15.95 (16)	1800
	Summer	0.79	0.07	0.14	6.6 (7)	1.12 (1)	1.43 (1)	9.15 (9)	1780
Monteiro	Fall	0.81	0.05	0.14	7.04 (7)	1.22 (1)	1.54 (2)	9.8 (10)	1840
	Winter	0.9	0.03	0.07	17.27 (17)	1.33 (1)	2.1 (2)	20.7 (21)	1879
	Spring	0.96	0.02	0.02	42.02 (42)	2 (2)	1.47 (1)	45.49 (45)	1800
	Summer	0.86	0.03	0.11	10.11 (10)	1.12 (1)	1.51 (2)	12.74 (13)	1780
Patos	Fall	0.83	0.04	0.13	8.25 (8)	1.12 (1)	1.48 (1)	10.85 (11)	1840
	Winter	0.98	0.01	0.01	49.75 (50)	1 (1)	1.04 (1)	51.79 (52)	1879
	Spring	0.97	0.01	0.02	50 (50)	1.08 (1)	1.37 (1)	52.45 (52)	1800
	Summer	0.81	0.04	0.15	7.16 (7)	1.06 (1)	1.55 (2)	9.77 (10)	1780

Table 6. The fit of the day's duration for the events d, w and r to the negative binomial distribution.

State	Parameters	Areia	C. Grande	J. Pessoa	Monteiro	Patos
d	Dispersion	0.40 (0.07)	0.46 (0.07)	0.42 (0.07)	0.69 (0.11)	0.62 (0.10)
	Average	22.76 (5.4)	13.31 (2.6)	25.82 (6.2)	8.22 (1.28)	7.06 (1.14)
w	Dispersion	0.02 (0.09)	0.01 (0.006)	0.01 (0.006)	0.02 (0.01)	0.005 (0.003)
	Average	10.35 (11.44)	6.37 (7.9)	11.78 (17.62)	2.75 (2.23)	2.38 (4.16)
r	Dispersion	0.04 (0.01)	0.01 (0.007)	0.04 (0.01)	0.03 (0.01)	0.02 (0.009)
	Average	18.85 (14.04)	10.23 (10.26)	21.62 (16.03)	6.58 (4.68)	5.91 (4.8)

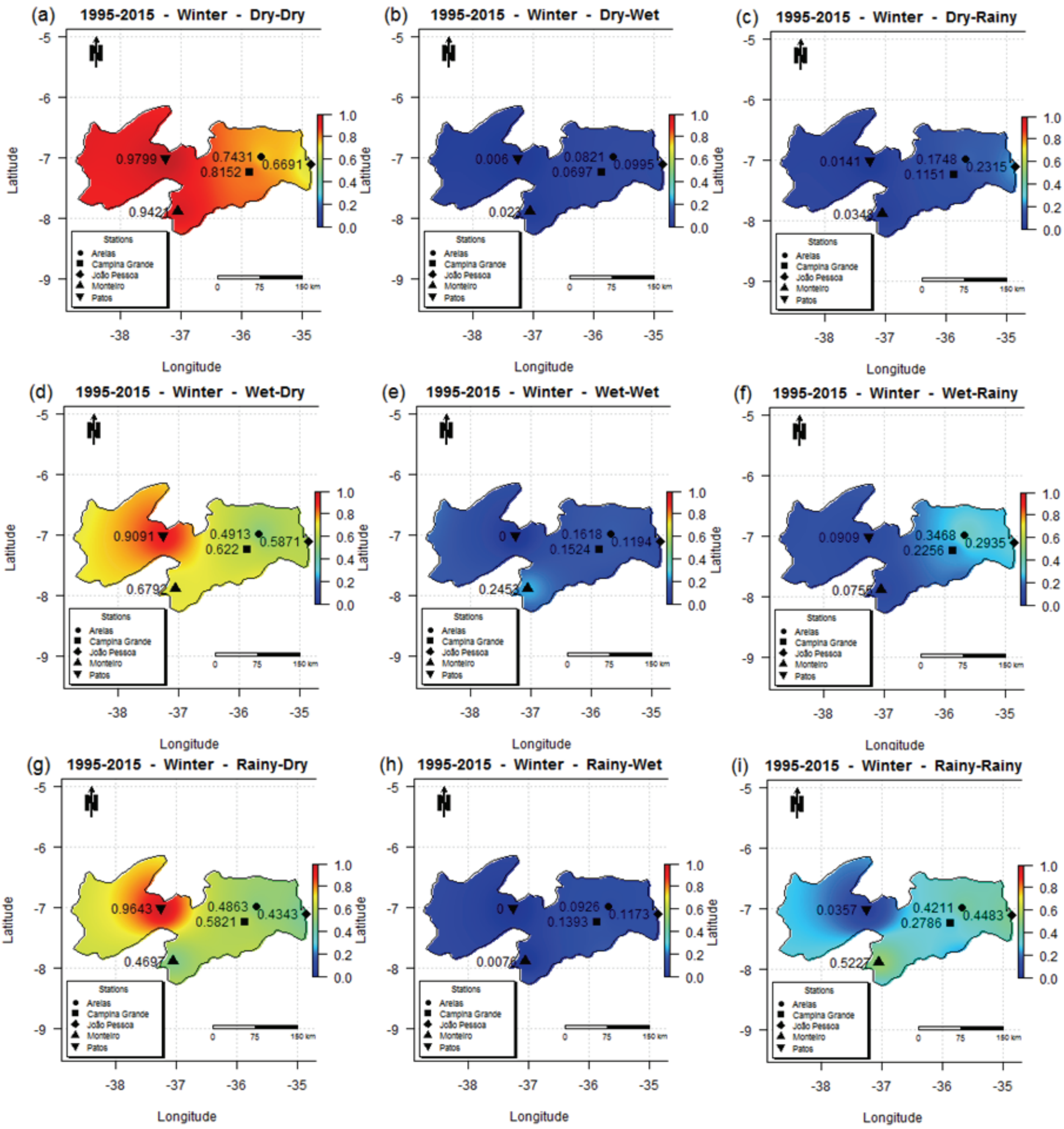

Figure 2. Interpolated maps of transition probabilities for winter season in the period 1995-2015.

Conclusion

The results of *WS* goodness-to-fit test indicate that Markov chain is an appropriate model for rainfall dynamics in Paraíba. For all stations/seasons, the probability of dry days is greater than the probability of rainy days (regime characterized by high rainfall intensity distributed over short rainy spells). The estimated transition probabilities among states show that the dry-dry transition presents the highest values for all seasons/stations, as well as the expected dry spell length (severe droughts). The transition probabilities that lead to rainy condition are higher in the coastal region, while the transition probabilities that lead to dry condition are higher in Sertão.

Acknowledgements

We acknowledge the support of the Brazilian agency *Coordenação de Aperfeiçoamento de Pessoal de Nível Superio*r (Capes) and anonymous reviewers for the constructive criticisms and suggestions that contributed to the improvement of this work.

References

Ali, M., & Mubarak, S. (2017). Effective rainfall calculation methods for field crops: an overview, analysis and new formulation. *Asian Research Journal of Agriculture, 7*(1), 1-12. doi: 10.9734/ARJA/2017/36812

Asong, Z. E., Khaliq, M. N., & Wheater, H. S. (2016). Multisite multivariate modeling of daily precipitation and temperature in the Canadian Prairie Provinces using generalized linear models. *Climate Dynamics, 47*(9-10), 2901-2921. doi: 10.1007/s00382-016-3004-z

Balzter, H. (2000). Markov chain models for vegetation dynamics. *Ecological Modelling, 126*(2), 139-154. doi: 10.1016/S0304-3800(00)00262-3

Chaves, R. R., & Cavalcanti, I. F. A. (2001). Atmospheric circulation features associated with rainfall variability over southern Northeast Brazil. *Monthly Weather Review, 129*(10), 2614-2626. doi: 10.1175/1520-0493(2001)129<2614:ACFAWR>2.0.CO;2

Collischonn, W., Haas, R., Andreolli, I., & Tucci, C. E. M. (2005). Forecasting River Uruguay flow using rainfall forecasts from a regional weather-prediction model. *Journal of Hydrology, 305*(1), 87-98. doi: 10.1016/j.jhydrol.2004.08.028

Craig, B. A., & Sendi, P. P. (2002). Estimation of the transition matrix of a discrete□time Markov chain. *Health Economics, 11*(1), 33-42. doi: 10.1002/hec.654

Cull, P. O., Hearn, A. B., & Smith, R. C. G. (1981). Irrigation scheduling of cotton in a climate with uncertain rainfall. *Irrigation Science, 2*(3), 127-140. doi: 10.1007/BF00257975.

Deni, S. M., Jemain, A. A., & Ibrahim, K. (2009). Fitting optimum order of Markov chain models for daily rainfall occurrences in Peninsular Malaysia. *Theoretical and Applied Climatology, 97*(1-2), 109-121. doi: 10.1007/s00704-008-0051-3

Fu, D. Z., Li, Y. P., & Huang, G. H. (2012). A fuzzy-Markov-chain-based analysis method for reservoir operation. *Stochastic Environmental Research and Risk Assessment, 26*(3), 375-391. doi: 10.1007/s00477-011-0497-1

Gabriel, K. R., & Neumann, J. (1962). A Markov chain model for daily rainfall occurrence at Tel Aviv. *Quarterly Journal of the Royal Meteorological Society, 88*(375), 90-95. doi: 10.1002/qj.49708837511

Garg, V. K., & Singh, J. B. (2010). Three-state Markov chain approach on the behaviour of rainfall. *New York Science Journal, 3*(12), 76-81. doi: 10.7537/marsnys031210.14

Giannini, A., Saravanan, R., & Chang, P. (2004). The preconditioning role of tropical Atlantic variability in the development of the Enso teleconnection: Implications for the prediction of Nordeste rainfall. *Climate Dynamics, 22*(8), 839-855. doi: 10.1007/s00382-004-0420-2

Guhathakurta, P., Sreejith, O. P., & Menon, P. A. (2011). Impact of climate change on extreme rainfall events and flood risk in India. *Journal of Earth System Science, 120*(3), 359-373. doi: 10.1007/s12040-011-0082-5

Haan, C. T., Allen, D. M., & Street, J. O. (1976). A Markov chain model of daily rainfall. *Water Resources Research, 12*(3), 443-449. doi: 10.1029/WR012i003p00443

Hastenrath, S. (1990). Prediction of Northeast Brazil rainfall anomalies. *Journal of Climate, 3*(8), 893-904. doi: 10.1175/1520-0442(1990)003<0893:PONBRA>2.0.CO;2

Hastenrath, S. (2012). Exploring the climate problems of Brazil's Nordeste: a review. *Climatic Change, 112*(2), 243-251. doi: 10.1007/s10584-011-0227-1

Hastenrath, S., & Heller, L. (1977). Dynamics of climatic hazards in northeast Brazil. *Quarterly Journal of the Royal Meteorological Society, 103*(435), 77-92. doi: 10.1002/qj.49710343505

Hughes, J. P., Guttorp, P., & Charles, S. P. (1999). A non□homogeneous hidden Markov model for precipitation occurrence. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 48*(1), 15-30. doi: 10.1111/1467-9876.00136

Ingram, K. T., Roncoli, M. C., & Kirshen, P. H. (2002). Opportunities and constraints for farmers of west Africa to use seasonal precipitation forecasts with Burkina Faso as a case study. *Agricultural Systems, 74*(3), 331-349. doi: 10.1016/S0308-521X(02)00044-6

Lana, X., & Burgueño, A. (1998). Daily dry-wet behaviour in Catalonia (NE Spain) from the viewpoint of Markov chains. *International Journal of Climatology, 18*(7), 793-815. doi: 10.1002/(SICI)1097-0088(19980615)18:7<793::AID-JOC262>3.0.CO;2-M

Le Gallo, J., & Chasco, C. (2008). Spatial analysis of urban growth in Spain, 1900-2001. *Empirical Economics, 34*(1), 59-80. doi: 10.1007/s00181-007-0150-5

Longley, R. W. (1953). The length of dry and wet periods. *Quarterly Journal of the Royal Meteorological Society, 79*(342), 520-527. doi: 10.1002/qj.49707934208

Lucena, D. B., Servain, J., & Gomes Filho, M. F. (2011). Rainfall response in Northeast Brazil from ocean climate variability during the second half of the twentieth century. *Journal of Climate, 24*(23), 6174-6184. doi: 10.1175/2011JCLI4194.1

Lyra, G. B., Oliveira□Júnior, J. F., & Zeri, M. (2014). Cluster analysis applied to the spatial and temporal variability of monthly rainfall in Alagoas state, Northeast of Brazil. *International Journal of Climatology, 34*(13), 3546-3558. doi: 10.1002/joc.3926

Marengo, J. A., Torres, R. R., & Alves, L. M. (2016). Drought in Northeast Brazil—past, present, and future. *Theoretical and Applied Climatology, 129*(3-4), 1189-1200. doi: 10.1007/s00704-016-1840-8

Maruyama, T., Kawachi, T., & Singh, V. P. (2005). Entropy-based assessment and clustering of potential water resources availability. *Journal of hydrology, 309*(1), 104-113. doi: 10.1016/j.jhydrol.2004.11.020

Minder, J. R., Roe, G. H., & Montgomery, D. R. (2009). Spatial patterns of rainfall and shallow landslide susceptibility. *Water Resources Research, 45*(4), 1-11. doi: 10.1029/2008WR007027

Moscati, M. C. L., & Gan, M. A. (2007). Rainfall variability in the rainy season of semiarid zone of Northeast Brazil (NEB) and its relation to wind regime. *International Journal of Climatology, 27*(4), 493-512. doi: 10.1002/joc.1408

Müller, H., & Haberlandt, U. (2016). Temporal rainfall disaggregation using a multiplicative cascade model for spatial application in urban hydrology. *Journal of Hydrology, 556*, 847-864. doi: 10.1016/j.jhydrol.2016.01.031

Pegram, G. G. (2009). A nested multisite daily rainfall stochastic generation model. *Journal of Hydrology, 371*(1), 142-153. doi: 10.1016/j.jhydrol.2009.03.025

Pereira, M. G., Trigo, R. M., Camara, C. C., Pereira, J. M., & Leite, S. M. (2005). Synoptic patterns associated with large summer forest fires in Portugal. *Agricultural and Forest Meteorology, 129*(1), 11-25. doi: 10.1016/j.agrformet.2004.12.007

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, AT: R Foundation for Statistical Computing.

Rajagopalan, B., Lall, U., & Tarboton, D. G. (1996). Nonhomogeneous Markov model for daily precipitation. *Journal of Hydrologic Engineering, 1*(1), 33-40. doi: 10.1061/(ASCE)1084-0699(1996)1:1(33)

Rao, V. B., Franchito, S. H., Santo, C. M., & Gan, M. A. (2016). An update on the rainfall characteristics of Brazil: seasonal variations and trends in 1979-2011. *International Journal of Climatology, 36*(1), 291-302. doi: 10.1002/joc.4345

Ratan, R., & Venugopal, V. (2013). Wet and dry spell characteristics of global tropical rainfall. *Water Resources Research, 49*(6), 3830-3841. doi: 10.1002/wrcr.20275

Richardson, C. W. (1981). Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research, 17*(1), 182-190. doi: 10.1029/WR017i001p00182

Robertson, A. W., Kirshner, S., & Smyth, P. (2004). Downscaling of daily rainfall occurrence over northeast Brazil using a hidden Markov model. *Journal of climate, 17*(22), 4407-4424. doi: 10.1175/JCLI-3216.1

Schoof, J. T., & Pryor, S. C. (2008). On the proper order of Markov chain model for daily precipitation occurrence in the contiguous United States. *Journal of Applied Meteorology and Climatology, 47*(9), 2477-2486. doi: 10.1175/2008JAMC1840.1

Seeger, M., Errea, M. P., Begueria, S., Arnáez, J., Martı, C., & Garcıa-Ruiz, J. M. (2004). Catchment soil moisture and rainfall characteristics as determinant factors for discharge/suspended sediment hysteretic loops in a small headwater catchment in the Spanish Pyrenees. *Journal of Hydrology, 288*(3), 299-311. doi: 10.1016/j.jhydrol.2003.10.012

Serinaldi, F. (2010). Multifractality, imperfect scaling and hydrological properties of rainfall time series simulated by continuous universal multifractal and discrete random cascade models. *Nonlinear Processes in Geophysics, 17*(6), 697-714. doi: 10.5194/npg-17-697-2010

Shamshad, A., Bawadi, M. A., Hussin, W. W., Majid, T. A., & Sanusi, S. A. M. (2005). First and second order Markov chain models for synthetic generation of wind speed time series. *Energy, 30*(5), 693-708. doi: 10.1016/j.energy.2004.05.026

Silva, L. L., Costa, R. F., Campos, J. H. D. C., & Dantas, R. T. (2009). Influência das precipitações na produtividade agrícola no Estado da Paraíba. *Revista Brasileira de Engenharia Agrícola e Ambiental, 13*(4), 454-461. doi: 10.1590/S1415-43662009000400013

Soares, A. S. D., Paz, A. R., & Piccilli, D. G. A. (2016). Avaliação das estimativas de chuva do satélite TRMM no Estado da Paraíba. *Revista Brasileira de Recursos Hídricos, 21*(2), 288-299. doi: 10.21168/rbrh.v21n2.p288-299

Uvo, C. B., Repelli, C. A., Zebiak, S. E., & Kushnir, Y. (1998). The relationships between tropical Pacific and Atlantic SST and northeast Brazil monthly precipitation. *Journal of Climate, 11*(4), 551-562. doi: 10.1175/1520-0442(1998)011<0551:TRBTPA>2.0.CO;2

Wagner, R. G. (1996). Mechanisms controlling variability of the interhemispheric sea surface temperature gradient in the tropical Atlantic. *Journal of Climate, 9*(9), 2010-2019. doi: 10.1175/1520-0442(1996)009<2010:MCVOTI>2.0.CO;2

Wang, D. Q., & Martiz, J. S. (1990). Note on testing a three state Markov Chain for independence. *Journal Statistics Computation and Simulation, 37*(1-2), 61-68. doi: 10.1080/00949659008811294

Wilks, D. S. (1998). Multisite generalization of a daily stochastic precipitation generation model. *Journal of Hydrology, 210*(1), 178-191. doi: 10.1016/S0022-1694(98)00186-3

Zhang, S., Wang, H., & Zhang, X. (2014). Estimation of channel state transition probabilities based on Markov chains in cognitive radio. *Journal of Communications, 9*(6), 468-474. doi: 10.12720/jcm.9.6.468-474.