

STATISTICAL UNCERTAINTY IN DROUGHT FORECASTING USING MARKOV CHAINS AND THE STANDARD PRECIPITATION INDEX (SPI)

ESTÁCIO, Alyson Brayner Sousa – alysonbrayner@gmail.com
Universidade Federal do Ceará / UFC

SILVA, Samiria Maria Oliveira da – samiriamaria@hotmail.com
Universidade Federal do Ceará / UFC

SOUZA FILHO, Francisco de Assis – assissouzafilho@gmail.com
Universidade Federal do Ceará / UFC

Submetido em: 29/10/2020

Aceito para publicação em: 15/01/2021

Publicado em: 28/04/2021

DOI: <http://dx.doi.org/10.5380/abclima.v28i0.77590>

ABSTRACT: Droughts affect basic human activities, and food and industry production. An adequate drought forecasting is crucial to guarantee the survival of population and promote societal development. The Standard Precipitation Index (SPI) is recommended by the World Meteorological Organization (WMO) to monitor meteorological drought. Using drought classification based on SPI to build Markov chains is a common tool for drought forecasting. However, Markov chains building process produce uncertainties inherent to the transition probabilities estimation. These uncertainties are often ignored by practitioners. In this study the statistical uncertainties of using Markov chains for drought annual forecasting are assessed. As a case study, the dry region of the State of Ceará (Northeastern Brazil) is analyzed, considering the precipitation records from 1911 to 2019. In addition to 100-year database (1911-2011) for Markov chain modeling and 8-year data (2012-2019) for forecasting validation, four fictional database extensions were considered in order to assess the effect of database size in the uncertainty. A likelihood ratio is used to assess model performance. The uncertainties assessment showed that an apparent performant Markov chain model for drought class forecasting may not be more informative than the historic proportion of drought class. Considering these uncertainties is crucial for an adequate forecasting with Markov chains.

KEYWORDS: Stochastic process; Uncertainty; Drought index; Transition probability.

INCERTEZA ESTATÍSTICA NA PREVISÃO DE SECAS USANDO CADEIAS DE MARKOV E O STANDARD PRECIPITATION INDEX (SPI)

RESUMO: As secas afetam as atividades humanas básicas, além da produção de alimentos e da atividade industrial. Uma previsão adequada das secas é crucial para garantir a sobrevivência das populações e promover o desenvolvimento da sociedade local. O Standard Precipitation Index (SPI) é recomendado pela Organização Meteorológica Mundial (OMM) para o monitoramento meteorológico. Assim, a construção de cadeias de Markov a partir da classificação de secas com o SPI é uma estratégia comumente utilizada para a previsão de secas. No entanto, o processo de construção de cadeias de Markov produz incertezas inerentes à estimação das probabilidades de transição entre os estados. Essas incertezas são, na maioria das vezes, desprezadas em aplicações práticas. Nesse estudo foram analisadas as incertezas estatísticas associadas ao uso de cadeias de Markov para a previsão de secas na escala anual. Como estudo de caso, é analisado o território do Estado do Ceará na região semiárida, considerando os registros de precipitação de 1911 a 2019. Além de uma série de precipitação de 100 anos (1911-2011) usada para a construção do modelo de previsão com cadeias de Markov e de uma série de oito anos (2012-2019) reservada para a validação, quatro séries fictícias estendidas foram consideradas a fim de se avaliar o efeito do tamanho da série na incerteza da previsão. Uma razão de verossimilhança é utilizada como índice para se

avaliar a performance do modelo. A análise de incerteza mostrou que um modelo de cadeia de Markov para a previsão de secas pode ser menos eficiente que um modelo ingênuo que sempre prevê a proporção histórica das classes de seca. Considerar as incertezas estatísticas se mostrou crucial para uma previsão adequada ao se usar cadeias de Markov.

PALAVRAS-CHAVE: Processo estocástico; Incerteza; Índice de seca; Probabilidade de transição.

INCERTIDUMBRE ESTADÍSTICA EN EL PRONÓSTICO DE SEQUÍA UTILIZANDO CADENAS MARKOV Y EL STANDARD PRECIPITATION INDEX (SPI)

RESUMEN: Las sequías han afectado a las actividades humanas básicas, además de la producción de alimentos y la actividad industrial. El pronóstico adecuado de la sequía es crucial para garantizar la supervivencia de las poblaciones y promover el desarrollo de la sociedad local. O Standard Precipitation Index (SPI) es recomendado por la Organización Meteorológica Mundial (OMM) para el monitoreo meteorológico. Por lo tanto, la construcción de cadenas Markov a partir de la clasificación de sequías con SPI es una estrategia comúnmente utilizada para la predicción de sequías. Sin embargo, el proceso de construcción de la cadena Markov produce incertidumbres relevantes para estimar las probabilidades de transición entre estados. Estas incertidumbres se ignoran con mayor frecuencia en las aplicaciones prácticas. En este trabajo, se analizan las incertidumbres estadísticas asociadas con el uso de cadenas Markov para la predicción de sequías a escala anual. Como caso práctico, se analiza el territorio del Estado de Ceará en la región del semiárido, teniendo en cuenta los registros de lluvias de 1911 a 2019. Además de una serie de precipitaciones de 100 años (1911-2011) utilizada para la construcción del modelo de predicción de la cadena Markov y una serie de ocho años (2012-2019) reservada para la validación, se consideraron cuatro series ficticias extendidas para evaluar el efecto del tamaño de la serie en la incertidumbre prevista. Una razón de verosimilitud se utiliza como índice para evaluar el rendimiento del modelo. El análisis de incertidumbre mostró que un modelo de cadena Markov para la previsión de sequía puede ser menos eficiente que un modelo ingenuo que siempre predice la proporción histórica de las clases de sequía. Considerar las incertidumbres estadísticas resultó crucial para la predicción adecuada al usar cadenas Markov.

PALABRAS CLAVE: Proceso estocástico; Incertidumbre; Índices de sequía; Probabilidad de transición.

1. INTRODUCTION

Water resources availability is a primary determinant of any society development, since it is necessary for the maintenance of basic human activities, for food production, and for many industrial processes (ALLAN, 2005; FLÖRKE et al., 2013; ZHOU; DENG; WU, 2017). For this reason, populations living in regions with frequent dry periods need a better comprehension of the drought phenomenon in order to guarantee its survival and promote its development.

The American Meteorological Society divide the drought definition into four categories (i.e. meteorological, hydrological, agricultural, and socioeconomic), according to the drought effect in which we are interested (HEIM, 2002). The meteorological effects of the drought are the first ones to appear. Its consequences propagate along the hydro-ecological-social system, producing hydrological, agricultural, and socioeconomic drought. There is no promising technology in the present time that allow to avoid meteorological drought. One exception are the experiments in the second half of the 20th century that produced artificial rain by stimulating raindrop formation (BISWAS;

DENNIS, 1971; GABRIEL, 1967; HEILIG, 1994). However, that technology produced results limited to a local scale (KAMPE; WEICKMANN, 1951).

On the other hand, many strategies of water resources planning and management allow to prevent (or at least reduce) the effects of hydrological, agricultural, and socioeconomic droughts. Those strategies can range from developing hydraulic infrastructures (e.g. for water storage and transport) to promoting mechanisms of water demand reduction. However, even when these actions against the drought effects are well established in a community, the weather can still produce drier conditions with negative consequences to the society, which make the meteorological drought a critical issue. That is why modelling the meteorological droughts towards its predictability is crucial to assess the drought risk (KIM; JEHANZAIB, 2020; PONTES FILHO et al., 2019).

In order to quantify the intensity of drought in a certain region, many indexes have been designed during the last decades, including the Standardized Precipitation Index (SPI) (MCKEE; DOESKEN; KLEIST, 1993), the Palmer Drought Severity Index (PDSI) (PALMER, 1965), the Surface Water Supply Index (SWSI) (SHAFER; DEZMAN, 1982), and the Aggregated Drought Index (ADI) (KEYANTASH; DRACUP, 2004). Among these indexes, the SPI has been recommended by the World Meteorological Organization (WMO) to monitor meteorological drought conditions (PONTES FILHO et al., 2019). In Brazil, many studies have been used SPI to analyze extreme events (DOS SANTOS; CUNHA; RIBEIRO-NETO, 2019; GOIS et al., 2020; MACEDO; GUEDES; SOUSA, 2011; SIQUEIRA; NERY, 2015).

An issue more challenging than classifying the historical meteorological conditions is forecasting it. An appropriate drought forecasting is central to water resources mid-term and short-term planning in dry regions. Many modelling approaches have been developed for drought forecasting in the last decade (FUNG et al., 2020).

Since the meteorological conditions are driven by the interaction between the atmosphere with the oceans and the land surface in a continuous process, the current meteorological state may keep information about the next meteorological conditions. According to that assumption, Markov chains could be used to evaluate the probability of drought in a region, using only the historical meteorological conditions (REZAEIANZADEH; STEIN; COX, 2016).

Markov chains driven by SPI states have been applied in the last years in order to analyze past drought (e.g. RAZIEI et al., 2007; SANTOS et al., 2019), allowing to assess the transition probability between the meteorological states, and also the return period and the residence time of different levels of droughts. In other studies the same tools were applied for drought forecasting (e.g. RAHMAT; JAYASURIYA; BHUIYAN, 2017; SIASAR; SHOJAEI, 2018). However, considering Markov chains for forecasting can lead to significant error, since a limited available data is often used to estimate the transition probability between the meteorological states (CANCELLIERE et al., 2007). The statistical uncertainties inherent to the transition probabilities estimation are often ignored by practitioners and its effect in forecasting is barely discussed in the literature.

This study's aim was analyzing the statistical uncertainties of using a Markov chain for drought forecasting, pointing to this framework limitations, in order to promote a better use of this tool towards drought predictability. As a

case study, a drought forecasting framework using the Markov chain with SPI states was applied to the dry region of the State of Ceará, in the Northeastern Brazil.

2. MATERIAL AND METHODS

2.1 STUDY AREA AND PRECIPITATION DATA

The State of Ceará is almost completely situated in the semi-arid region of Brazil, except by its coastal strip, where the tropical humid climate occurs. Because of the its frequent droughts, Ceará is included in the Polígono das Secas (drought polygon) as defined by Ministério da Fazenda (2010)(see Figure 1). This region has faced 10 droughts during the 20th century, which has shaped the local development (FRISCHKORN; SANTIAGO; DE ARAUJO, 2003).

Rainfall in Ceará concentrates in the first semester (mostly in the 4-month period between February and May), with almost no precipitation between June and November. High interannual rainfall variability is also characteristic of local climate. Mean annual rainfall range from 400 mm in the driest years to 1500 mm in the wettest years, with an average of 850 mm and a Coefficient of Variation of 0.3 (GÜNTNER; BRONSTERT, 2004; SOUZA FILHO, 2018).

As consequence of the dry and highly variable climate conditions, a policy of construction of artificial reservoirs to store water excess in the wet periods has been implemented in the State of Ceará for more than a century, in order create water reserves to enable local development. In spite of the well established reservoir network, with 18.6 bi m³ of total capacity (SOUZA FILHO, 2018), even the large and mid-sized reservoir are liable to dry out in droughts that last over 2 years (FRISCHKORN; SANTIAGO; DE ARAUJO, 2003). For this reason, an adequate drought forecasting is crucial for mid-term water resources management in the region.

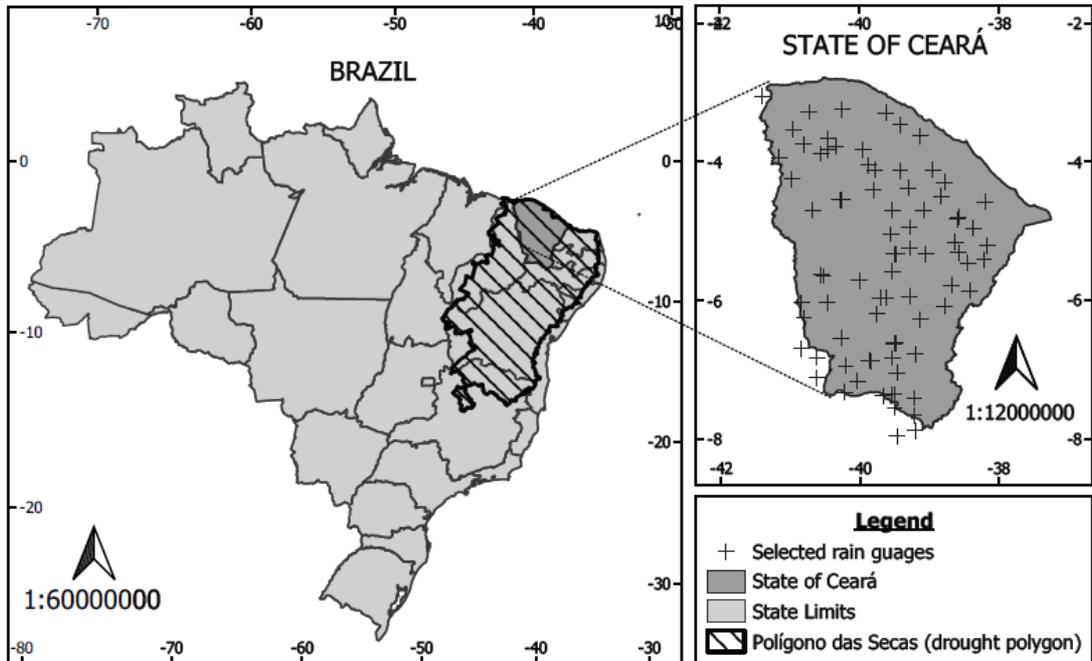


Figure 1 – Study Area and rain gauge network. Location of the State of Ceará and the Polígono das Secas in Brazil (on the left) and the selected 265 rain gauge distribution in the State of Ceará (on the top right).

An 815-rain-gauge network has been implemented and managed since 1910 by state and federal agencies (i.e. Departamento Nacional de Obras Contra as Secas; Superintendência do Desenvolvimento do Nordeste; and Fundação Cearense de Meteorologia e Recursos Hídricos), supporting local water resources management information. Precipitation data used in this study was obtained from 265 rain gauges (with more than 40% daily data available) from that network database, made available by Brazilian Water Agency (ANA) (see Figure 1).

Annual precipitation for the selected rain gauges was interpolated throughout State of Ceará using the Inverse Distance Weighting, obtaining annual precipitation in the average of the territory from 1911 to 2019, which was used for SPI calculation. The precipitation data is divided into two period: a 100-year period (1911-2011) devoted to Markov chain modelling, and an 8-year period (2012-2019), including a recent drought (2012-2017), reserved to model validation.

2.2 GENERAL FRAMEWORK

In order to analyze the Markov chain drought forecasting, considering the inherent statistical uncertainties, a four-step framework is proposed (see Figure 2). First, the SPI series for the annual precipitation in Ceará from 1911 to 2019 is calculated and classified in bands of drought magnitude. The drought state series from 1911 to 2011 is used for the traditional Markov chain modelling, with punctual estimation of the transition probabilities. Then, the uncertainty related to the transition probabilities estimations is explored using a Bayesian approach to determine confidence intervals to the transition probabilities.

Finally, that uncertainty is propagated to the drought forecasting, considering the period from 2012 to 2019.

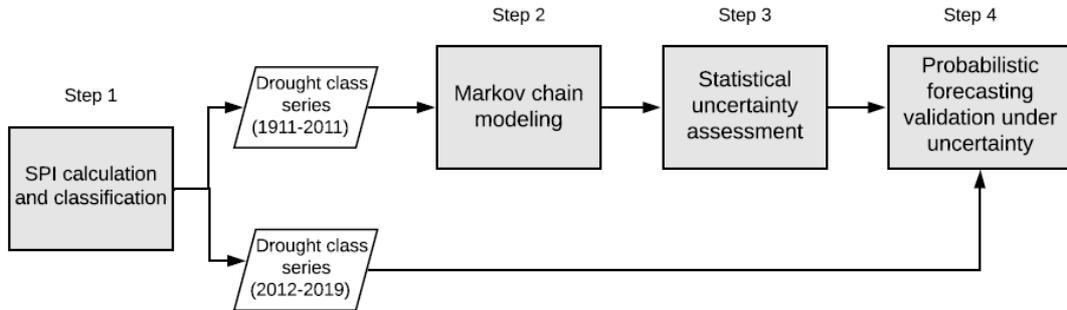


Figure 2 – Four-step framework for the analysis of the Markov chain drought forecasting in State of Ceará, considering the inherent statistical uncertainties.

The forecasting performance is evaluated in the modelling and validation periods with a likelihood ratio. In order to evaluate the effect of the modeling database size in the uncertainty, four fictional databases with series larger than the previous 100-year period were considered (i.e. 200-, 500-, 1000- and 2000-year database).

R programming language was used for the SPI calculation and Markov chain modeling and application.

2.3 STANDARDIZED PRECIPITATION INDEX

Standardized Precipitation Index (SPI) relates precipitation events to its probability. SPI calculation consists of a variable transformation from the aggregated monthly precipitation into a normally distributed variable (i.e. the SPI itself). Different time intervals (i.e. 3, 6, 12, 24, or 48 months) can be considered to aggregate precipitation for SPI calculation. In that scheme, a SPI value is calculated for each month by considering the aggregate precipitation of the previous months, as in a moving average scheme. Each aggregation time interval imply in a different analysis purpose (MATHBOUT et al., 2018; MERABTI et al., 2018). In this study, it was selected the SPI-12 (i.e. with a time interval of 12 months for precipitation aggregation) that represents long-term precipitation patterns and which is related to water scarcity (WMO, 2012).

As in the original application of SPI (i.e. MCKEE; DOESKEN; KLEIST, 1993), aggregated precipitation was assumed Gamma distributed.

A Gamma distributed variable x has the following density function:

$$g(x) = \frac{\beta^\alpha}{\int_0^\infty \tau^{\alpha-1} e^{-\tau} d\tau} x^{\alpha-1} e^{-\beta x} \tag{1}$$

Where α is the shape parameter, β is the rate parameter.

The parameters α and β were estimated by $\hat{\alpha}$ and $\hat{\beta}$, using the maximum likelihood method, as following:

$$\hat{\alpha} = \frac{1}{4A} \left(1 + \sqrt{1 + \frac{4A}{3}} \right) \tag{2}$$

$$\hat{\beta} = \frac{\hat{\alpha}}{\bar{x}} \tag{3}$$

Where \bar{x} is the average observation of x and A is calculated as following:

$$A = \ln(\bar{x}) - \bar{x} \tag{4}$$

Since Gamma distribution does not admit $x = 0$, a correction of the accumulated probability function is usually applied for the SPI calculation (e.g. SIQUEIRA; NERY, 2015). Nevertheless, as only the 12 months accumulated precipitations were considered in this study, $x = 0$ is completely unlikely, even in a semi-arid region. Then, that correction was irrelevant for the current study.

In view of the Markov chain application, using the monthly SPI-12 would artificially introduce dependence between consecutive series values, since a 12-month aggregated precipitation in the series differs from the next month aggregated precipitation for only one out of twelve months (as in a moving average scheme). For that reason, only the last value of SPI-12 of each year was considered, i.e. the sum of the 12-month precipitations for each year. This resulting series is here called the annual SPI.

Although the annual SPI was calculated for the entire precipitation series (1911-2019), Gamma parameters were fitted to the precipitation of the modelling period (1911-2011), which contains the data that is assumed to be known at the model building. The precipitation data was fitted to a gamma distribution using the package `fitdistrplus` in R (DELIGNETTE-MULLER; DUTANG, 2015).

In order to analyze the dry ($SPI < 0$) and wet periods ($SPI > 0$), it was calculated the basic statistics (i.e. mean and coefficient of variation) of the duration and severity for both dry and wet periods. According to Mishra and Singh (2010), the duration of a dry period is the interval time that the drought index keeps negative, while the severity of this dry period is the cumulative value of the index during that interval time. Similar parameters were calculated for the wet periods in order to compare them to dry ones.

For annual SPI classification, four drought magnitude classes were considered (see Table 1). The limits for that classification were inspired in those defined by Mckee, Doesken and Kleist (1993). Nevertheless, in the current study, the original severe and extreme drought classes were merged into a single one class (i.e. SED), in order to improve the statistical significance of these drought classes observations. The resulting four classes were used to define the states in Markov chain scheme. These states were represented by the variable X , which changes in the series as the SPI changes.

Table 1- Drought classification according to the SPI value.

Original drought class	SPI value	This work drought class
Extreme Drought	≤ -2.00	Severe or Extreme Drought (SED)
Severe Drought	-1.50 to -1.99	
Moderate Drought	-1.00 to -1.49	Moderate Drought (MD)
Mild Drought	0.00 to -0.99	Mild Drought (mD)
No class	> 0	No drought (ND)

Source: adapted from Mckee, Doesken and Kleist (1993)

2.4 MARKOV CHAIN FOR DROUGHT FORECASTING

A Markov chain is a sequence of random variables whose probability of assuming a specific value does not depend on the other antecedents, when we fix the value of the immediately previous antecedent (BLITZSTEIN; HWANG, 2019). Mathematically, the sequence of random variables $X_0, X_1, X_2, X_3, \dots$ is a Markov chain if for all $n \geq 0$:

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = P(X_{n+1} = j | X_n = i) \quad (5)$$

Where $P(A|B)$ is the conditional probability of observing the event A if it is known that event B occurs. $P(X_{n+1} = j | X_n = i)$ is called the transition probability from the state i to state j . For a homogenous Markov chain, the transition probabilities are assumed to be independent of n .

The ensemble of transition probabilities $p_{i,j} = P(X_{n+1} = j | X_n = i)$ of a Markov chain is summarized in its transition matrix $TM = [p_{i,j}]_{m \times m}$ with m lines and m columns, where m is the number of values that the variable X_n may assume.

In the Markov chain scheme for drought forecasting, X is defined as the annual drought magnitude, varying from one to four, according to the drought class (Table 1).

When considering an annual-step Markov chain scheme for drought forecasting, it is assumed that the drought class in a certain year keeps information about the probability of observing another drought class the next year. Since drought is a phenomenon that develops slowly in the time (CANCELLIERE et al., 2007), annual-step Markov chain hypothesis is reasonable.

Additionally, the transition probabilities were assumed not to change during the analyzed period. In other words, the Markov chain is assumed to be homogeneous, which allow to consider the entire period data-set (except the data reserved for validation) to estimate unique transition probabilities.

The transition probabilities can be punctually estimated (i.e. with a single value for each transition) from the number of transitions observed in the database, as shown in the equation 2 (MISHRA; SINGH; DESAI, 2009):

$$p_{i,j} = P(X_{n+1} = j | X_n = i) \approx \frac{n_{i,j}}{\sum_{k=1}^4 n_{i,k}} \quad (6)$$

Where $n_{i,j}$ is the number of transitions observed from state i to state j (i.e. from drought class i to drought class j), with i and $j \in \{1,2,3,4\}$ according to the value of X , and $\sum_{k=1}^4 n_{i,k}$ is the number of observations of state i , independently of the next state.

As consequence of $p_{i,j}$ definition, $\sum_{j=1}^4 p_{i,j}$ (i.e. the sum of the elements of a generic line i of the transition matrix) is equal to one.

Given a transition matrix $TM = [p_{i,j}]_{m \times m}$, a probabilistic forecasting of the drought class (s_{n+1}) can be obtained by multiplying the probabilistic estimation of the precedent year drought class (s_n) by TM , as in the equation:

$$s_{n+1} = s_n \times TM \quad (7)$$

Where s_n and s_{n+1} are 1×4 matrixes with elements between zero and one, and summing one, which represent the probabilistic estimation of observing each

one of the four drought classes in the years n and $n+1$, respectively. TM is the 4×4 transition matrix.

When the drought class of a year n is known to be equal to i , the state s_n is defined as $s_n = [k_1 k_2 k_3 k_4]$, with $k_j = 1$ if $i = j$, and $k_j = 0$ if $i \neq j$. When the drought class is not known, the elements of matrix s_n are the probabilities estimated for each class. In that case, s_n can be estimated with s_{n-1} by multiplying it by TM (i.e. $s_n = s_{n-1} \times TM$). Thus, a m -step probabilistic forecasting with Markov chain is obtained by:

$$s_{n+m} = s_n \times TM^m \tag{8}$$

When TM is multiplied by itself many times it tends to converge to a matrix with identical lines, called steady state of Markov chain (BLITZSTEIN; HWANG, 2019). The convergence of TM^m means that the elements of the Markov chain keep information about the next states only until a certain number k of steps. After that, the Markov chain forecasting for the element $k+n$ in the chain is independent of element n .

In this study, the punctual estimations of the transition probabilities were calculated by counting the transitions observed in the 100-year modelling period, in order to obtain TM (as in equation 2). Then, considering the real states observed in the modeling and validation periods, the probabilistic forecast of drought class was performed for different forecasting horizons (i.e. 1-, 2-, 3-, 4- and 5-step scheme).

2.5 LIKELIHOOD RATIO

In order to evaluate the forecasting performance (in the n -step schemes) both in the 100-year period for modeling and in the 8-year period of validation, the concept of likelihood was used. The likelihood measure the probability of obtaining a result that was actually observed, assuming that the analyzed probabilistic model is valid.

Let $so_n = [k_1 k_2 k_3 k_4]$ (a matrix 1×4 with $k_j \in \{0,1\}$) indicate the observed state in the year n and so_{n+1} (similar to so_n) indicate the observed state in the year $n+1$. The probability of obtaining so_{n+1} if the Markov chain is valid can be calculated as:

$$p_{n+1} = so_n \times TM \times (so_{n+1})^T \tag{9}$$

Where p_{n+1} is the likelihood value in the year $n+1$ for the Markov chain forecasting model; so_n and so_{n+1} are 1×4 matrixes with elements zero or one, indicating the observed drought class in the years n and $n+1$, respectively; TM is the 4×4 transition matrix.

A wise way to evaluate the forecasting performance of a probabilistic model is comparing its likelihood to the likelihood of a simpler and evident probabilistic model, for example a model that represent the average pattern of the inferred variable. In this study, the simpler model is the one that always expects the historic proportion of the drought classes. That simpler model is from now on called climatology model, since it represents the average climatological pattern. The probability of obtaining the so_{n+1} considering this simpler model is calculated as:

$$pc_{n+1} = HP \times (so_{n+1})^T \tag{10}$$

Where pc_{n+1} is the likelihood value in the year $n+1$ for a model that always expects the historic proportion of the drought classes; so_{n+1} is 1×4 matrix with elements zero or one, indicating the observed drought class in the year $n+1$; and HP is a 1×4 matrix with the historic proportion of the drought classes.

In order to compare the Markov chain forecasting performance with the climatology model, the normalized likelihood ratio (equation 7) was used. It compares the likelihood of both models in an ensemble of years (SOUZA FILHO; LALL, 2003). The effect of the number of years is removed by the exponent $1/N$.

$$LR = \left(\frac{\prod_{n=1}^N p_n}{\prod_{n=1}^N pc_n} \right)^{\frac{1}{N}} \tag{11}$$

Where $\prod_{n=1}^N p_n$ and $\prod_{n=1}^N pc_n$ are the likelihood values for the ensemble of $N+1$ analyzed years, considering Markov chain and the climatology model, respectively; LR is the likelihood ratio.

The likelihood ratio varies from zero to positive infinite. It is superior to one if the analyzed probabilistic model is more performant than the climatology model.

2.6 MARKOV CHAIN UNCERTAINTIES

Using Markov chain for forecasting can produce inaccuracies inherent to the way the transition probabilities are estimated. The historical series used to estimate those probabilities can be quite informative depending on the data availability, but it is crucial to verify the statistical uncertainty related to those estimations and its consequences on forecasting performance.

Thus, in addition to the punctual estimation, the transition probabilities were estimated in this study by confidence intervals, considering that the number of the observed transitions follows the Binomial distribution. In fact, if the transition probability from state i to state j is $p_{i,j}$, the number $X_{i,j}$ of observations of class j after the N_i observations of class i follows $Bi(N_i, p_{i,j})$, with the probability mass function (BLITZSTEIN; HWANG, 2019):

$$P(X_{i,j} = x) = \binom{N_i}{x} p_{i,j}^x (1 - p_{i,j})^{N_i - x} \tag{12}$$

With $x \in \{0, 1, \dots, N_i\}$

If $x_{i,j}$ transitions were observed, the best punctual estimation for $p_{i,j}$ is $x_{i,j}/N_i$, where N_i is the number of observations of class i . However, if we are interested in finding an interval to estimate $p_{i,j}$ with confidence level of $CL\%$, we should find the values $(p_{i,j})_{min}$ and the $(p_{i,j})_{max}$ that make true the equality :

$$P((p_{i,j})_{min} < p_{i,j} < (p_{i,j})_{max} | X_{i,j} = x_{i,j}) = CL\% \tag{13}$$

According to Bayes's rule (BLITZSTEIN; HWANG, 2019) the conditional probability above is equal to:

$$\frac{P(X_{i,j}=x_{i,j} | (p_{i,j})_{min} < p_{i,j} < (p_{i,j})_{max}) \times P((p_{i,j})_{min} < p_{i,j} < (p_{i,j})_{max})}{P(X_{i,j}=x_{i,j})} \tag{14}$$

Where:

$$P(X_{i,j} = x_{i,j} | (p_{i,j})_{min} < p_{i,j} < (p_{i,j})_{max}) = \frac{\int_{(p_{i,j})_{min}}^{(p_{i,j})_{max}} P(X_{i,j}=x_{i,j}|p_{i,j}=k)dk}{(p_{i,j})_{max}-(p_{i,j})_{min}} \quad (15)$$

$$P((p_{i,j})_{min} < p_{i,j} < (p_{i,j})_{max}) = (p_{i,j})_{max} - (p_{i,j})_{min} \quad (16)$$

$$P(X_{i,j} = x_{i,j}) = \frac{1}{N_{i+1}} \quad (17)$$

When a symmetric confidence interval was possible, it was admitted:

$$(p_{i,j})_{min} = \frac{x_{i,j}}{N_i} - \Delta p_{i,j} \text{ and } (p_{i,j})_{max} = \frac{x_{i,j}}{N_i} + \Delta p_{i,j} \quad (18)$$

However, when $x_{i,j}/N_i$ was too close to the extreme values (i.e. zero or one), there was not enough margin to consider a symmetric interval with confidence level $CL\%$. In those cases, one of the interval limits was fixed. In other words, $(p_{i,j})_{min}$ was fixed to zero if $x_{i,j}/N_i$ was too close to zero or $(p_{i,j})_{max}$ was fixed to one if $x_{i,j}/N_i$ was too close to one.

Then, in order to determine $(p_{i,j})_{min}$ and $(p_{i,j})_{max}$ values that make equality 13 true, a univariate research algorithm based in the gold section procedure combined with the parabolic interpolation (BRENT, 1973) was applied to find $\Delta p_{i,j}$ for the symmetric intervals, or directly $(p_{i,j})_{min}$ or $(p_{i,j})_{max}$ for the asymmetric one. Thus, the confidence intervals for each transition probability $p_{i,j}$ in TM was estimated, quantifying its statistical uncertainty.

In view of that uncertainty, it would be more reasonable to consider that the transition matrix used for drought forecasting has transition probabilities between the calculated limits rather than consider values fixed to the observed proportions $x_{i,j}/N_i$. In that approach, an infinite number of transition matrixes are possible, respecting the transition probabilities limits and the sum of one for the probabilities in the same initial state (i.e. same line). Each one of those likely transition matrixes has its own performance in prediction, that can be assessed by the likelihood ratio.

In order to evaluate the effects of transition probabilities uncertainties in forecasting performance, optimization was carried out to find the transition matrixes (respecting the transition probabilities constraints) that maximize and minimize the likelihood ratio. In view of the complexity involved to that constrained optimization problem with a non-linear objective function, an exhaustive search (with a step of 0.01 for the probabilities) was conducted, allowing to determine the transition matrixes that maximize and minimize the likelihood ratio, determining an uncertainty band for it.

In addition, assuming that no transition has null probability (i.e. all transition even rare are possible to happen) a residual probability was introduced when the transition probabilities interval included the value zero. Because of the search method the residual probability was 0.01.

In the traditional Markov chain scheme for forecasting without uncertainties considerations, the chosen transition matrix is that one that maximizes the forecasting performance during the modelling period, which correspond to the matrix with transition probabilities equal to the observed transition proportions (i.e. $p_{i,j} = x_{i,j}/N_i$). The same transition matrix may not

produce the best performance in other periods, which requires a validation of the Markov chain scheme for drought prediction with a period different from the one used for modelling. Thus, the uncertainty band of the likelihood ratio was evaluated both in the modelling and the validation period.

Since the uncertainty in the transition probabilities estimation is a function of the total number of observations (N_i) in each initial state, four fictional scenarios with 200, 500, 1000 and 2000 years of data set were considered, in order to evaluate the effect of the database size in the uncertainty band of the likelihood ratio. For those scenarios, the observed proportions $x_{i,j}/N_i$ are assumed to be equal to the proportions observed in the 100-year modeling period, but varying the number of observations, in a way as if the proportions were corroborated by a larger number of observations. That effect was evaluated in modelling and validation forecasting performance.

3. RESULTS AND DISCUSSION

3.1 STANDARDIZED PRECIPITATION INDEX IN MODELING PERIOD (1911-2011)

Annual SPI in Ceará territory between 1911 and 2011 ranged from -2.1 (in 1919) to 3.1 (in 1985) (see Figure 3). Statistical analysis in R showed that dry periods ($SPI < 0$) had average duration (2.5 year) slightly longer than wet periods (2.1 years). On the other hand, the coefficient of variation of dry periods duration ($CV=0.9$) was significantly larger than for wet periods ($CV=0.4$).

Similarly, severities of dry and wet periods were close (1.8 in both cases), while the coefficient of variation of dry periods severity ($CV=1.0$) was larger than that one of wet periods severity ($CV=0.7$).

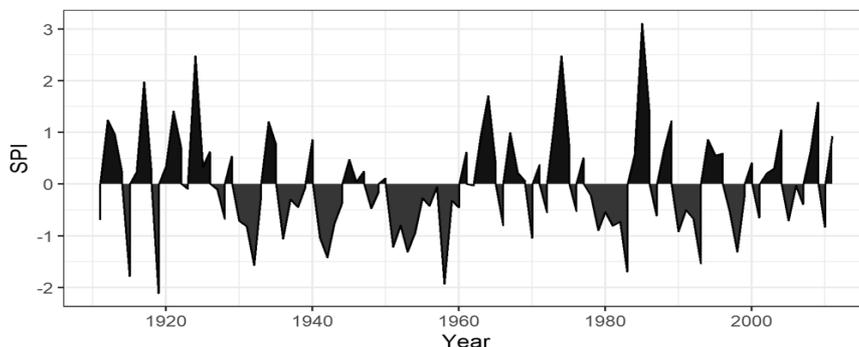


Figure 3 - SPI series in Ceará territory between 1911 and 2011.

Freitas and Billib (1997) argue that there is no definitive periodicity in the incidence of drought in Ceará. In fact, a larger dispersion of dry periods parameters, with larger coefficients of variation, showed that drought forecasting is a challenging task in the territory, since drought severity and duration are more irregular than in wet periods. It worth reminding that the weather conditions in Ceará are mainly related to the Intertropical Convergence Zone, that may control drought occurrence. Effectively, previous investigations showed that the annual variability of precipitation in the region has the sea surface temperature in the Atlantic and Pacific as a primary factor (KWON et al., 2012; SUN et al., 2005; WARD; FOLLAND, 1991).

Using the drought classification, the SPI series was transformed in a drought class series (a series of the variable X), which was the income to design the Markov chain model for drought forecast.

3.2 MARKOV CHAIN FOR DROUGHT FORECASTING

The transition matrix obtained for the four drought classes (i.e. severe or extreme drought = SED, moderate drought = MD, mild drought = mD, and no drought = ND) was:

	<i>SED</i>	<i>MD</i>	<i>mD</i>	<i>ND</i>
<i>SED</i>	0.00	0.00	0.33	0.67
<i>MD</i>	0.00	0.14	0.71	0.14
<i>mD</i>	0.10	0.05	0.44	0.41
<i>ND</i>	0.04	0.09	0.33	0.54

According to the estimated transition probabilities, before a year with no drought (ND) or with a mild drought (mD) is more likely to keep the same state the next year. On the other hand, before a year with extreme or severe drought (SED) the climatic system is more likely to finish the drought (ND), and before a year with moderate drought (MD) the system tends to regress to a mild drought (mD). This result indicates that more intensive droughts in Ceará, reaching the SED state, tend to finish quickly before the peak, while less intensive droughts are more persistent.

Siasar and Shojaei (2018) found similar outcomes in Saravan (in Iran), which might indicate a standard in drought pattern for semi-arid regions. For Santos et al. (2019) study, considering four rain gauges in São Francisco basin (also in Brazilian Polígono das Secas), the transition probabilities between the drought classes (based on the SPI-12) indicated a consistent tendency of keeping the current state, with maximum probabilities located in the diagonal of the transition matrix. However, in that study, monthly SPI-12 was considered for Markov chain modelling, which explains the slower transition between the drought classes if compared to the new results.

The stationary state of Markov chain (see Table 2) was obtained after four steps with a precision of 0.01 for the transition probabilities, which indicates that the Markov chain scheme can keep information about the next states at maximum for the following three years. Above that, the drought forecasting with the Markov chain indicated probabilities similar to considering the historic portion of time in each drought class, regardless the antecedent state, as in the climatology model.

Table 2- Stationary state of Markov chain after four steps, considering a precision of 0.01 for the transition probabilities.

Drought class	Probability
Severe to Extreme Drought (SED)	0.06
Moderate Drought (MD)	0.07
Mild Drought (mD)	0.41
No drought (ND)	0.46

Source: Authors (2021)

When the calibration period (1911-2011) was considered, the likelihood ratio indicated that in the first steps of forecasting the Markov chain scheme is

more informative (likelihood ratio > 1) than the climatology model. The likelihood ratio was 1.05 in the one-year drought forecasting, dropping to 1.01 in the two-year forecasting (see Figure 4). Above three steps the index converged to one, which means that for longer forecasting horizons, with the convergence of the transition matrix, the Markov chain is as performant as the simpler historic proportion.

Even with likelihood ratio superior to one, the model performance can be considered modest if compared to the 2.89 likelihood ratio found by Souza Filho and Lall (2003) for a probabilistic model for inflow forecasting in Ceará. However, performance of the Markov chain model is still satisfactory, considering it uses only the annual precipitation data.

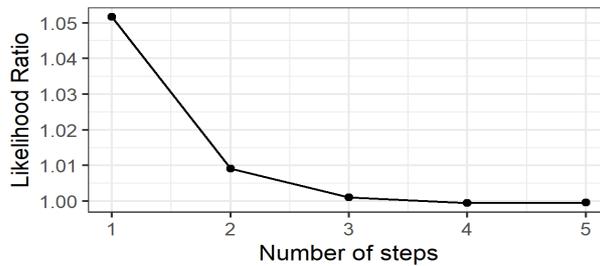


Figure 4 – Likelihood ratio versus the number of steps of forecast horizon in the 100-year calibration period. The likelihood ratio converges to the unit above 3 steps.

3.3 MARKOV CHAIN UNCERTAINTIES

In the 100-year calibration period (1911-2011), severe or extreme droughts were observed only six times, which produce a relevant uncertainty on estimating the transition probabilities from this limited data. This issue is inherent to extreme events assessment, since more these events are relevant, more they are rare, and difficult to predict (GHIL et al., 2011).

Similarly, only seven moderate droughts were observed in the calibration period. Because of those limitations in the data, no transition from SED to SED or to MD and no transition from MD to SED were observed (see detected transitions in Table 3).

Table 3 – Number of transitions observed between the drought classes from 1911 to 2011.

Drought class in the previous year	Number of observations $x_{i,j}$ (1911-2011)				Total (N_i)
	Drought class in the next year				
	SED	MD	mD	ND	
SED	0	0	2	4	6
MD	0	1	5	1	7
mD	4	2	18	17	41
ND	2	4	15	25	46

Source: Authors (2021)

Not detecting some transitions in the observed data is a common issue in Markov chain modelling (SANTOS et al., 2019; SIASAR; SHOJAEI, 2018). However, it does not mean that those transitions has null probability, as the transition matrix may indicate. It actually means that this probability is small

enough not to be detected in the few observations (i.e. six or seven) of the previous drought class (i.e. SED or MD).

Even when the transitions were detected in calibration period, there is a statistical uncertainty related to the sample size (i.e. the total number of observations in each drought class). For instance, the estimations for the new state probabilities after a year with no drought (with 46 observations) may be more reliable than estimations for the states after a severe or extreme drought, with only six observations.

In order to standardize the reliability of the transition probabilities, confidence intervals for the transition probabilities were estimated with 95% of confidence level (see Table 4). Intervals centered in the average value of the observed transition proportions were possible in six out of six-teen transitions. The other ten transitions, with observed proportion close to the extreme values (i.e. zero and one), needed asymmetric intervals, which are inferiorly limited by zero or superiorly limited by one, in order to compute 95% of confidence level.

Table 4 – Confidence intervals for the transition probabilities between the drought classes from 1911 to 2011 estimated with 95% confidence level.

Drought class in the previous year	Confidence interval for transition probabilities							
	Drought class in the next year							
	SED		MD		mD		ND	
	min	max	min	max	min	max	min	max
SED	0.00	0.34	0.00	0.35	0.01	0.66	0.34	0.99
MD	0.00	0.31	0.00	0.47	0.40	1.00	0.00	0.47
mD	0.00	0.20	0.00	0.14	0.29	0.59	0.27	0.56
ND	0.00	0.13	0.00	0.18	0.19	0.46	0.40	0.68

Source: Authors (2021)

Although a same confidence level was adopted (i.e. 95%), the length of the confidence interval (see Table 5) varied according to the total number of observations in the previous year drought class (N_i) and according to the observed proportion ($x_{i,j}/N_i$) estimating the transition probability. In fact, smaller samples imply in greater uncertainty, which explains why the length of the confidence interval was larger for transitions with SED and MD as initial states (i.e. lines 1 and 2). In addition, considering a same initial state (i.e. same line), the uncertainty is bigger for the observed proportions closer to 0.5 (see Table 4 e Table 5), since when the populational proportion ($p_{i,j}$) is extreme (i.e. close to 0 or close to 1) the sample proportion ($x_{i,j}/N_i$) tends to be less dispersed (and less uncertain) than when it is close to 0.5 (BLITZSTEIN; HWANG, 2019).

Table 5 – Length of the confidence interval for the transition probabilities between the drought classes from 1911 to 2011 estimated with 95% confidence level.

Drought class in the previous year	Length of the confidence interval for trans. probability			
	Drought class in the next year			
	SED	MD	mD	ND
SED	0.35	0.35	0.65	0.65
MD	0.31	0.47	0.60	0.47
mD	0.20	0.14	0.29	0.29
ND	0.13	0.18	0.26	0.28

Source: Authors (2021)

The search for the transition matrixes that minimize and maximize the forecasting performance in the modelling period showed that the likelihood ratio can vary from 0.75 to 1.05. The inferior value of likelihood ratio below one indicates that, considering the transition probabilities uncertainty, there is no guarantee that the Markov chain scheme for drought forecasting is more performant than the climatology model.

The assessment of different scenarios of database size showed a reduction of likelihood ratio range with the number of year (see Figure 5), when the 1-step forecasting scheme is considered.

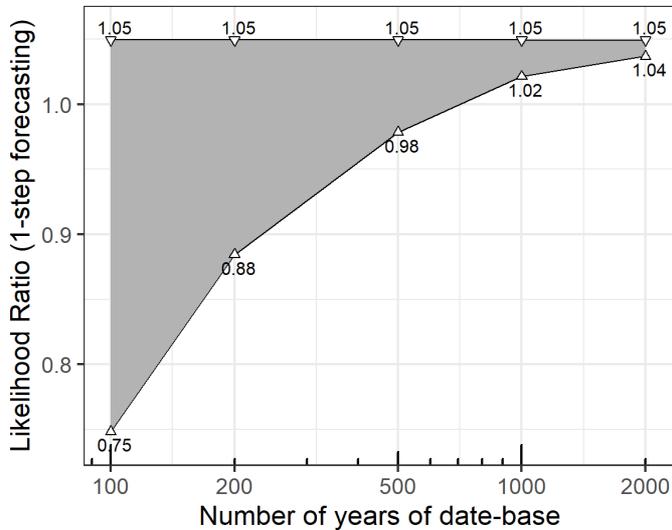


Figure 5 – Uncertainty band (95% confidence level) for the likelihood ratio (in the modelling period) in 1-step Markov chain forecasting, varying with the data-base size (in logarithmical scale).

The uncertainty band for the likelihood ratio is superiorly limited by 1.05 for all database sizes. In fact, the maximum likelihood ratio is obtained when transition probabilities considered in *TM* are equal (or as closest as possible) to the proportion of transitions observed in analyzed period. Thus, for the modeling period (1911-2011), the maximum likelihood ratio is obtained with the punctual estimations of the transition probabilities

3.4 PROBABILISTIC DROUGHT FORECASTING (2012-2019)

Using the period from 2012 to 2019 for validation allowed to analyze the model performance in a real drought forecasting. The drought class probabilities for each year, estimated from the previous year class (i.e. in the 1-step scheme) generally followed the observed SPI (see Figure 6)

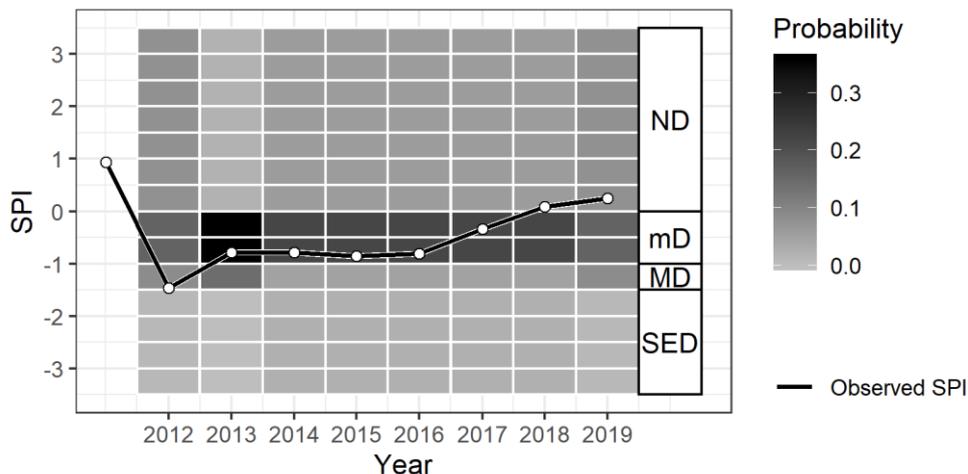


Figure 6 – Probabilistic drought class forecasting (in the 1-step scheme) and observed SPI during the validation period (2012-2019) The forecasting probabilities were represented by the colors. The drought classes were divided into intervals of 0.5 SPI, in order not to contaminate the graphic representation of forecasting probabilities by the drought classes length.

Forecasting performance for the 8-year validation period showed a likelihood ratio superior to one in the 1-step and 2-step schemes, with values of 1.15 and 1.01 respectively. The likelihood ratio converges to values close to one above three steps. (see Figure 7).

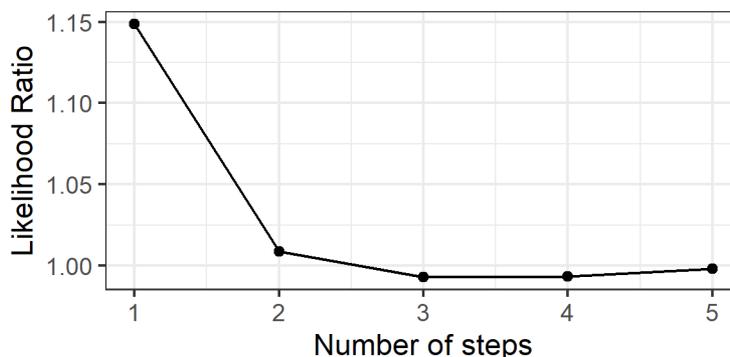


Figure 7 – Likelihood ratio versus the number of steps of forecast horizon in the 8-year validation period. The likelihood ratio converges to one above 4 steps. The likelihood ratio axe is plotted in logarithm scale.

Even in that not-calibrated period, including a 6-year drought (2012-2017), the Markov chain scheme is more informative than the climatology model, with performance even superior to what was observed in calibration for the first step. On the other hand, the performance in validation deteriorates with the number of steps fast than in calibration.

In spite of the apparent good performance in validation, considering the uncertainty of the transition probabilities within the estimated confidence intervals, the punctual estimation of the probabilities may hide other likely transition matrixes with different likelihood ratios.

The maximum likelihood ratio in validation is 1.53. The matrix that maximizes the likelihood ratio is that one with transition probabilities as close as possible to the transition proportions observed in the period. Since the validation period does not include years with extreme or severe droughts (which are quite rare), the transition probabilities with initial state in extreme or severe droughts (i.e SED) are irrelevant (i.e. first line is irrelevant).

Similarly, the transition matrix that minimize the likelihood ratio is that one with transition probabilities as different as possible from the observed proportions. The minimum likelihood ratio is 0.63 for the one-step forecasting, pointing again that there is no guarantee of good forecasting performance.

Considering the scenarios of different database sizes, as the database becomes larger, the confidence interval for the likelihood ratio becomes smaller (see Figure 8).

Because of the hypothesis of keeping the observed transition proportions in modelling data for all fictional database sizes, the transition probabilities are implicitly supposed to be equal to the observed proportions. Thus, the uncertain band tends to converge to 1.15, that is the likelihood ratio obtained in the validation period with the punctual estimation of transition probabilities.

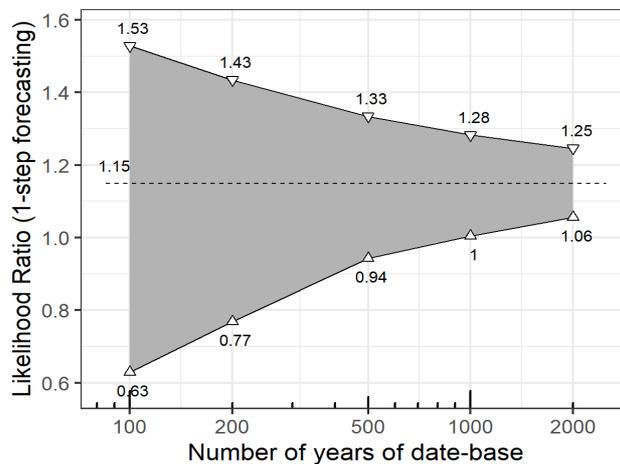


Figure 8 - Uncertainty band (95% confidence level) for the validation likelihood ratio in 1-step Markov chain forecasting, varying with the data-base size (in logarithmical scale).

Only with a 1000-year data set it would be possible to assure (with 95% confidence level) likelihood ratio superior to one. On the other hand, for a 500-year data set the likelihood ratio interval (from 0.94 to 1.33) is already mostly superior to one. Since real database of measured precipitation with more than 500 years are rare, precipitation reconstructed by paleoclimatic analysis would be necessary for a less uncertain Markov chain drought class forecasting. Pimentel et al. (2011) for instance estimated last 10k years precipitation in Ceará using sea surface temperature as proxy. In addition to the data availability issue, using large precipitation data asks for a review of stationarity hypothesis.

4. CONCLUSIONS

In this study, SPI proved to be more than an index for drought statistical analysis. Indeed, it provided a fair indication of drought magnitude that make possible a drought class forecasting within a Markov chain scheme.

The Markov chain model for drought class forecasting in the State of Ceará provided initially a performance superior to a simpler model that always expects the historic proportion of the drought classes (i.e. the climatology model). The superiority of performance was observed mainly when the drought class was estimated from the previous year conditions (i.e. in the 1-step forecasting scheme) and deteriorates for longer horizons. The likelihood ratio for the 1-step forecasting scheme was 1.05 in the modelling period (1911-2011) and 1.15 in the validation period (2012-2019), that includes a 6-year drought (2012-2017).

The analysis of the statistical uncertainties of the transition probabilities showed that the likelihood ratio in the 1-step forecasting scheme was actually in the interval 0.75-1.05 in modelling and in the interval 0.63-1.53 in validation, without guarantee of a performance better than the climatology model. Then, the statistical uncertainties assessment was able to show that an apparent performant Markov chain model for drought class forecasting may not be more informative than the historic proportion of drought class.

The analysis of the effect of database size in the uncertainty showed that a 1000-year database would be necessary to guarantee likelihood superior to one in the case study. Even if such large databases may be made available by paleoclimatic studies, the hypothesis of stationarity of precipitation must be verified, as the pattern of drought classes transitions may change along the centuries.

Improvement of the likelihood optimization method is recommended as future work, since it is necessary to produce more precise results (i.e. without need of discretization for the transition probabilities), and necessary to allow likelihood uncertainty assessment for horizons longer than one step.

The Markov chain continue to be an important tool for analyzing the transitions of states conducted by the natural random processes and it can be valuable to forecast the next states, including the drought magnitude. However, considering the analysis of the statistical uncertainty, inherent to the transition probabilities estimation, is crucial for an adequate forecasting.

ACKNOWLEDGEMENTS

The research was partially supported by grants from the Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brasil (CNPq) (NEXUS Project, Nº 441457/2017-7), Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (FUNCAP), and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES).

REFERENCES

- ALLAN, J. A. Water in the environment/ socio-economic development discourse: Sustainability, changing management paradigms and policy responses in a global system. *Government and Opposition*, [S. l.], v. 40, n. 2, p. 181–199, 2005. DOI: 10.1111/j.1477-7053.2005.00149.x.
- BISWAS, K. R.; DENNIS, A. S. Formation of a Rain Shower by Salt Seeding. *Journal of Applied Meteorology*, 1971.
- BLITZSTEIN, J. K.; HWANG, J. *Introduction to Probability*. 2nd. ed. [s.l.] : Chapman & Hall/CRC Texts in Statistical Science, 2019.
- BRENT, R. *Algorithms for Minimization without Derivatives*. NJ: Prentice-Hall: Englewood Cliffs, 1973.
- CANCELLIERE, A.; MAURO, G. Di; BONACCORSO, B.; ROSSI, G. Drought forecasting using the standardized precipitation index. *Water Resources Management*, [S. l.], v. 21, n. 5, p. 801–819, 2007. DOI: 10.1007/s11269-006-9062-y.
- DELIGNETTE-MULLER, Marie Laure; DUTANG, Christophe. fitdistrplus: An R Package for Fitting Distributions. *Journal of Statistical Software*, [S. l.], v. 64, n. 4, p. 1–34, 2015.
- DOS SANTOS, Sergio Rodrigo Quadros; CUNHA, Ana Paula Martins do Amaral; RIBEIRO-NETO, Germano Gondim. Avaliação De Dados De Precipitação Para O Monitoramento Do Padrão Espaço-Temporal Da Seca No Nordeste Do Brasil. *Revista Brasileira de Climatologia*, [S. l.], v. 25, p. 80–100, 2019. DOI: 10.5380/abclima.v25i0.62018.
- FLÖRKE, Martina; KYNAST, Ellen; BÄRLUND, Ilona; EISNER, Stephanie; WIMMER, Florian; ALCAMO, Joseph. Domestic and industrial water uses of the past 60 years as a mirror of socio-economic development: A global simulation study. *Global Environmental Change*, [S. l.], v. 23, n. 1, p. 144–156, 2013. DOI: 10.1016/j.gloenvcha.2012.10.018.
- FREITAS, M. A. S.; BILLIB, M. H. A. Drought prediction and characteristic analysis in semiarid Ceará, northeast Brazil. In: SUSTAINABILITY OF WATER RESOURCES UNDER INCREASING UNCERTAINTY 1997, Rabat. *Anais [...]*. Rabat: IAHS, 1997. p. 105–112.
- FRISCHKORN, H.; SANTIAGO, M. F.; DE ARAUJO, J. C. Water Resources of Ceará and Piauí. In: T. GAISER, M. KROL, H. FRISCHKORN, J. C. De Araújo (org.). *Global Change and Regional Impacts*. Berlin: Springer-V, 2003. p. 87–94.
- FUNG, K. F.; HUANG, Y. F.; KOO, C. H.; SOH, Y. W. Drought forecasting: A review of modelling approaches 2007–2017. *Journal of Water and Climate Change*, [S. l.], v. 11, n. 3, p. 771–799, 2020. DOI: 10.2166/wcc.2019.236.
- GABRIEL, KR. The Israeli artificial rainfall stimulation experiment. Statistical evaluation for the period 1961–1965. *Proceedings of the Fifth Berkeley Symposium*, [S. l.], 1967.
- GHIL, M. et al. Extreme events: Dynamics, statistics and prediction. *Nonlinear Processes in Geophysics*, [S. l.], v. 18, n. 3, p. 295–350, 2011. DOI: 10.5194/npg-18-295-2011.

GOIS, Givanildo De; OLIVEIRA-JÚNIOR, José Francisco De; PAIVA, Roberta Fernanda da Paz de Souza; FREITAS, Welington Kiffer; TERASSI, Paulo Miguel de Bodas; SOBRAL, Bruno Serafini. Variabilidade pluviométrica, indicadores de seca e a aplicação do índice SPI para a região do médio Vale Paraíba do Sul no Estado do Rio de Janeiro. *Revista Brasileira de Climatologia*, [S. l.], v. 27, p. 122–157, 2020.

GÜNTNER, A.; BRONSTERT, A. Representation of landscape variability and lateral redistribution processes for large-scale hydrological modelling in semi-arid areas. *Journal of Hydrology*, [S. l.], v. 297, n. 1–4, p. 136–161, 2004. DOI: 10.1016/j.jhydrol.2004.04.008.

HEILIG, Morton L. *United States Patent Office* 1994.

HEIM R R, Jr. A review of twentieth-century drought indices used in the United States. *Bulletin of the American Meteorological Society*, [S. l.], n. August, 2002.

KAMPE, H. J.; WEICKMANN, H. K. The effectiveness of natural and artificial aerosols as freezing nuclei. *Journal of Meteorology*, [S. l.], v. 8, p. 283–288, 1951.

KEYANTASH, John A.; DRACUP, John A. An aggregate drought index: Assessing drought severity based on fluctuations in the hydrologic cycle and surface water storage. *Water Resources Research*, [S. l.], v. 40, n. 9, p. 1–14, 2004. DOI: 10.1029/2003WR002610.

KIM, Tae Woong; JEHANZAIB, Muhammad. Drought risk analysis, forecasting and assessment under climate change. *Water (Switzerland)*, [S. l.], v. 12, n. 7, p. 1–7, 2020. DOI: 10.3390/W12071862.

KWON, Hyun Han; DE ASSIS DE SOUZA FILHO, Francisco; BLOCK, Paul; SUN, Liqiang; LALL, Upmanu; REIS, Dirceu S. Uncertainty assessment of hydrologic and climate forecast models in Northeastern Brazil. *Hydrological Processes*, [S. l.], v. 26, n. 25, p. 3875–3885, 2012. DOI: 10.1002/hyp.8433.

MACEDO, Maria José Herculano; GUEDES, Roni Valter de Souza; SOUSA, Francisco de Assis Salviano. Monitoramento e intensidade das secas e chuvas na cidade de Campina Grande/PB. *Revista Brasileira de Climatologia*, [S. l.], v. 8, p. 105–117, 2011.

MATHBOUT, Shifa; LOPEZ-BUSTINS, Joan A.; MARTIN-VIDE, Javier; BECH, Joan; RODRIGO, Fernando S. Spatial and temporal analysis of drought variability at several time scales in Syria during 1961–2012. *Atmospheric Research*, [S. l.], v. 200, n. May 2017, p. 153–168, 2018. DOI: 10.1016/j.atmosres.2017.09.016.

MCKEE, Thomas B.; DOESKEN, Nolan J.; KLEIST, John. The relationship of drought frequency and duration to time scales. In: EIGHTH CONFERENCE ON APPLIED CLIMATOLOGY 1993, Anaheim, California. *Anais [...]*. Anaheim, California p. 17–22.

MERABTI, Abdelaaziz; MARTINS, Diogo S.; MEDDI, Mohamed; PEREIRA, Luis S. Spatial and Time Variability of Drought Based on SPI and RDI with Various Time Scales. *Water Resources Management*, [S. l.], v. 32, n. 3, p. 1087–1100, 2018. DOI: 10.1007/s11269-017-1856-6.

MINISTÉRIO DA FAZENDA. *Manual de Preenchimento da Declaração do Imposto sobre a Propriedade Territorial Rural*. Receita Federal. [s.l.: s.n.] 2010.

MISHRA, A. K.; SINGH, V. P.; DESAI, V. R. Drought characterization: A probabilistic approach. *Stochastic Environmental Research and Risk Assessment*, [S. l.], v. 23, n. 1, p. 41–55, 2009. DOI: 10.1007/s00477-007-0194-2.

MISHRA, Ashok K.; SINGH, Vijay P. A review of drought concepts. *Journal of Hydrology*, [S. l.], v. 391, n. 1–2, p. 202–216, 2010. DOI: 10.1016/j.jhydrol.2010.07.012.

PALMER, W. C. *Meteorological drought*. [s.l.] : U.S. Dept. of Commerce Weather Bureau Research Paper 45, 1965.

PIMENTEL, Felipe Viana; COSTA, Alexandre Araújo; ARAGÃO, Dias Tyhago; RIOS, Francisco Franklin Sousa; LIMA, Vanessa Araújo. A influência das paleotemperaturas da superfície do mar na precipitação sobre o Nordeste brasileiro durante o Holoceno. In: XIII ABEQUA CONGRESS - THE SOUTH AMERICAN QUATERNARY: CHALLENGES AND PERSPECTIVES 2011, *Anais [...]*. [s.l.: s.n.] p. 1–3.

PONTES FILHO, João Dehon; PORTELA, Maria Manuela; STUDART, Ticiania Marinho de Carvalho; SOUZA FILHO, Francisco de Assis. A Continuous drought probability monitoring system, CDPMS, based on copulas. *Water (Switzerland)*, [S. l.], v. 11, n. 9, p. 1–18, 2019. DOI: 10.3390/w11091925.

RAHMAT, Siti Nazahiyah; JAYASURIYA, Niranjali; BHUIYAN, Muhammed A. Short-term droughts forecast using Markov chain model in Victoria, Australia. *Theoretical and Applied Climatology*, [S. l.], v. 129, n. 1–2, p. 445–457, 2017. DOI: 10.1007/s00704-016-1785-y.

RAZIEI, T.; ARASTEH, Daneshkar; AKHTARI, R.; SAGHAFIAN, B. Investigation of Meteorological Droughts in the Sistan and Balouchestan Province, Using the Standardized Precipitation Index and Markov Chain Model. *Iran-Water Resources Research*, [S. l.], v. 3, n. 1, p. 86–76, 2007.

REZAEIANZADEH, Mehdi; STEIN, Alfred; COX, Jonathan Peter. Drought Forecasting using Markov Chain Model and Artificial Neural Networks. *Water Resources Management*, [S. l.], v. 30, n. 7, p. 2245–2259, 2016. DOI: 10.1007/s11269-016-1283-0.

SANTOS, Esdras Adriano Barbosa Dos; STOSIC, Tatijana; BARRETO, Ikaro Daniel de Carvalho; CAMPOS, Laélia; SILVA, Antonio Samuel Alves Da. Application of Markov chains to Standardized Precipitation Index (SPI) in São Francisco River Basin. *Ambiente e Agua - An Interdisciplinary Journal of Applied Science*, [S. l.], v. 14, n. 3, p. 1, 2019. DOI: 10.4136/ambi-agua.2311.

SHAFER, B. A.; DEZMAN, L. E. *Development of a Surface Water Supply Index (SWSI) to assess the severity of drought conditions in snowpack runoff areas (Colorado)*. *Proceedings: Eastern Snow Conference, 39th annual meeting*, 1982.

SIASAR, Hadi; SHOJAEI, Saeed. Drought forecast using standardized precipitation index and Markov chain in Iran. [S. l.], n. December, p. 1–10, 2018. DOI: 10.20944/preprints201812.0276.v1.

SIQUEIRA, Beatriz; NERY, Jonas Teixeira. Aplicação E Análise Do Índice Padronizado De Precipitação No Circuito Das Águas Paulista. *Revista Brasileira*

de *Climatologia*, [S. l.], v. 16, p. 93–107, 2015. DOI: 10.5380/abclima.v16i0.40331.

SOUZA FILHO, F. A. de. *Projeto Ceará 2050: Diagnóstico dos Recursos Hídricos no Ceará*. Fortaleza: [s.n], 2018.

SOUZA FILHO, F. A.; LALL, U. Seasonal to interannual ensemble streamflow forecasts for Ceara , Brazil: Applications of a multivariate , semiparametric algorithm. [S. l.], v. 39, n. 11, p. 1–13, 2003. DOI: 10.1029/2002WR001373.

SUN, LIQIANG; MONCUNILL, DAVID FERRAN; LI, HUILAN; MOURA, ANTONIO DIVINO; SOUZA FILHO, FRANCISCO DE ASSIS. Climate Downscaling over Nordeste , Brazil , Using the NCEP RSM97. *Journal of Climate*, [S. l.], v. 18, p. 551–567, 2005.

WARD, M. Neil; FOLLAND, Chris K. Prediction of seasonal rainfall in the north nordeste of Brazil using eigenvectors of sea-surface temperature. *International Journal of Climatology*, [S. l.], v. 11, n. 7, p. 711–743, 1991. DOI: 10.1002/joc.3370110703.

WMO. *Standardized precipitation index user guide*. Geneva. 2012 Disponível em: http://www.wamis.org/agm/pubs/SPI/WMO_1090_EN.pdf.

ZHOU, Qing; DENG, Xiangzheng; WU, Feng. Impacts of water scarcity on socio-economic development: A case study of Gaotai County, China. *Physics and Chemistry of the Earth*, [S. l.], v. 101, p. 204–213, 2017. DOI: 10.1016/j.pce.2017.03.009.