Sílvio Fernando Alves Xavier Júnior

Study of Trend Analysis and Sample Entropy of Precipitation in Paraíba, Brazil

Recife, PE

March 2016

Universidade Federal Rural de Pernambuco
Pró-Reitoria de Pesquisa e Pós-Graduação
Programa de Pós-Graduação em Biometria e Estatística Aplicada

Sílvio Fernando Alves Xavier Júnior

Study of Trend Analysis and Sample Entropy of Precipitation in Paraíba, Brazil

Recife, PE
March 2016.

Sílvio Fernando Alves Xavier Júnior

Study of Trend Analysis and Sample Entropy of Precipitation in Paraíba, Brazil

Thesis presented to the Programa de Pós-Graduação em Biometria e Estatística Aplicada, in compliance with legal requirements for obtaining the doctorate title.

Concentration field: Biometry and Applied Statistics

Advisor:
Prof. Dra. Tatijana Stosic

Co-advisor:
Prof. Dr. Carlos Antônio Costa dos Santos

Co-advisor:
Prof. PhD Vijay P. Singh

Recife, PE
March de 2016.

Ficha catalográfica

X3s Xavier Júnior, Sílvio Fernando Alves

Study of trend analysis and sample entropy of precipitation in Paraíba, Brazil / Sílvio Fernando Alves Xavier Júnior. – Recife, 2016.

84 f. : il.

Orientadora: Tatijana Stosic.

Tese (Programa de Pós Graduação em Biometria e Estatística Aplicada) – Universidade Federal Rural de Pernambuco, Departamento de Estatística e Informática, Recife, 2016.

Inclui referências e apêndice(s).

1. Rainfall 2. Trends 3. Sample Entropy 4. Paraíba I. Stosic, Tatijana, orientadora II. Título

CDD 574.018

Sílvio Fernando Alves Xavier Júnior

Study of Trend Analysis and Sample Entropy of Precipitation in Paraíba, Brazil

Thesis considered suitable to obtain doctorate title in Biometria e Estatística Aplicada by UFRPE, defended and unanimity approved on 22/03/2016 by the following exam committee.

Exam Committee:

Profa. Dra. Tatijana Stosic
Advisor
DEINFO - UFRPE

Prof. Dr. Carlos Antônio Costa dos Santos
Co-advisor
UACA - UFCG

Prof. Dr. Moacyr Cunha Filho
DEINFO - UFRPE

Prof. Dr. Tiago Almeida de Oliveira
DE - UEPB

Prof. Dr. Borko Stosic
DEINFO-UFRPE

Dedicated to my family and Érika, my lovely wife.

Acknowledgments

Firstly, I am so grateful to everything God, with his steadfast love, has done for me. Everything belongs to him.

I am also so grateful to my parents because they taught me everything, I love you so much! My niece, nephew, all my in-laws.

To my lovely wife because we are one and without you support and patient I couldn't finish it on time.

All Professors of my department at UEPB.

My advisor Prof. PhD Tatijana Stosic and Prof. PhD Carlos for the ideias and suggestions.

My advisor at Texas A&M University Prof. PhD Vijay P. Singh for all support and kindness. You are an example of Professor.

My partners from China, specially Xin Tong and Professor PhD Hongbo. I hope we can initiate a partnership as soon as possible.

My friends at Westminster Presbyterian Church in Bryan, specially Andrew and his wife Lauren! Thank you for everything.

All guys at Baptist Student Ministry...you do not know how you were amazing! Giggem Aggies!

All my friends at UFRPE, I missed you when I was in USA. You were so important!

My friend Jader who helped me a lot. Our partnership has just started and I think we can produce much more.

My friends Ruy de Deus and Fábio Jacques.

My friend Cilas and all my family at Igreja Presbiteriana em Bairro Novo...thank you for all prayers!

My friend Marco for his huge support!

To CAPES for my financial support in College Station, Texas.

"Now this is not the end. It is not even the beginning of the end. But it is, perhaps, the end of the beginning."

Winston Churchill

Abstract

The objective of this work was to present two different methodologies in order to obtain a better comprehension of rainfall phenomena and its consequences over a particular region which suffers from water scarcity. Firstly, semivariogram models were selected to estimate trends in monthly precipitation in Paraíba State-Brazil using ordinary kriging. The methodology involves the application of geostatistical interpolation of precipitation records of 51 years from 69 rainfall stations across the state. Analysis of semivariograms showed that anisotropy for specific months had a strong spatial dependence (Index of Spatial Dependence - IDE <25%). The trends were subjected to the following models: circular, spherical, pentaspherical, exponential, Gaussian, rational quadratic, K-Bessel and tetraspherical. The models with the best fit were selected by cross-validation and Error Comparison Index (ECI). Each data set month had a particular spatial dependence structure, which made it necessary to define specific models of semivariograms in order to enhance the adjustment of the experimental semivariogram. Besides, the standardized error prediction map and hot spot analysis were obtained with the aim of justifying the chosen models. Furthermore, one can see that a climate system is a complex nonlinear system. To describe the complexity characteristics of precipitation series in Paraíba, we propose the use of sample entropy, a kind of entropy-based algorithm, to measure the complexity of precipitation series. The Paraíba's four macro-regions: Mata, Agreste, Borborema, and Sertão were analyzed. Results of analysis show that complexities of monthly average precipitation have differences in the macro-regions. Sample entropy can reflect the dynamic change of precipitation series providing a new way to investigate the complexity of hydrological series. The complexity exhibits an areal variation of local water resources system which can influence the basis for utilizing and developing resources in dry areas.

Keywords: Rainfall, Trends, Paraíba, Sample Entropy.

Resumo

No presente trabalho duas metodologias diferentes foram investigadas a fim de obter uma melhor compreensão dos fenômenos de chuva e suas consequências sobre uma região particular que sofre escassez de recursos hídricos. Em primeiro lugar, foram selecionados modelos de semivariogramas para determinar as tendências de precipitação mensal no Estado da Paraíba usando krigagem ordinária. A metodologia envolve a aplicação de interpolação geoestatística de registros de precipitação de 51 anos de 69 estações pluviométricas em todo o Estado. As análises de semivariogramas mostraram que a ocorrência da anisotropia durante meses específicos teve uma forte dependência espacial (Índice de Dependência Espacial - IDE < 25 %). As tendências foram submetidas aos seguintes modelos: Gaussian, *rational quadratic*, circular, esférico, *pentaspherical*, exponencial, K-Bessel e *tetraspherical*. Os modelos com o melhor ajuste foram selecionados pela validação cruzada e Índice de Comparação de Erros (ICE). Cada conjunto de dados apresentou uma estrutura de dependência espacial particular, o que tornou necessário definir modelos específicos de semivariogramas, a fim de melhorar o ajuste do semivariograma experimental. Além disso, o mapa de previsão de erro padronizado e *Hot Spots* foram obtidos com o objetivo de justificar os modelos escolhidos. Um sistema climático é um sistema não-linear bastante complexo. Para descrever as características de complexidade das séries de precipitação na Paraíba, propomos o uso do *Sample Entropy*, um algoritmo baseado na entropia de Shannon, para medir a complexidade da série de precipitação. Quatro macro-regiões da Paraíba foram analisadas: Mata, Agreste, Borborema e Sertão. Os resultados das análises mostraram que as complexidades de precipitação mensal média têm diferenças nas macro-regiões. *Sample Entropy* pode refletir a mudança dinâmica da série de precipitação proporcionando uma nova maneira de investigar a complexidade das séries hidrológicas. A complexidade apresenta uma variação regional do sistema de recursos hídricos locais que podem influenciar a base para a utilização e desenvolvimento de recursos em áreas secas.

Palavras-chaves: Precipitação, tendências, Paraíba, entropia.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Analysis of long-term hydrometeorological variables plays a significant role in water resources planning and climate studies. In general, global warming intensifies the global hydrological cycle, thus increasing globally averaged precipitation, evaporation, and runoff (CLARK *et al.*, 1999). According to Milly *et al.* (2002), the amount, timing, and distribution of precipitation, evaporation, temperature, are connected with changes in the hydrologic cycle and point to changes in the availability of water as well as in the competition for water resources.

It has been noticeable that the exhaustion of water resources in different parts of the world is one of the preeminent environmental problems of this century. As reported by Clark *et al.* (1999), an important aspect of water resources planning is the analysis of long-term hydrometeorological variables. The advanced hydrological cycle is altering the spatiotemporal patterns of precipitation resulting in increased occurrences of precipitation extremes (EASTERLING *et al.*, 2000). In the outcome of the hydrological global warming and the resulting alteration of the hydrological cycle, there are growing concerns about the spatiotemporal patterns of precipitation regimes and the subsequent influences on the space-time variations of floods and droughts (ZHANG *et al.*, 2013).

Complex climatic systems have properties that cannot be entirely described by understanding the parts of the system. As stated by Gallagher *et al.* (1999), the properties of the system are different from the properties of the parts, and they depend on the entirety of the whole; the systemic properties disappear when the system breaks apart while the properties of the parts are retained.

On the report of Dong & Meng (2013), the long-term influence of various kinds of na-

tural factors (solar cycle, landform, geographical position) and human activities (afforestation and tree planting, hydraulic engineering construction) have led to the complexity in the characteristics of precipitation series. However, the complexity of water resources system is ignored when the optimal utilization of water resources issues are investigated, which makes it difficult to discover the internal information of water resources system and truly realize the optimal utilization of regional water resources after allocation.

An important factor is that rainfall represents the primary input to the hydrologic cycle, and can thus be recognized to represent the potential water resources availability of an area. The disorder (or uncertain) in the intensity and occurrence of rainfall in time is one of the primary constraints to water resources development and the water use practices (MARUYAMA *et al.*, 2005). Normally the potential availability of water resources is measured in terms of cumulative rainfall in the region of interest.

Rainfall is a periodical spatiotemporal phenomenon displaying significant spatial and temporal variability and rain gauge networks only collect point estimates. Therefore, providing an estimate of spatial rainfall distribution within an area from rain gauge data usually remains a barrier of interpolation (MIRÁS-AVALOS *et al.*, 2007).

Rainfall data is the essential input for hydrological systems, this type of data plays a fundamental role in understanding the hydrological cycle. It is essential an accurate estimation of precipitation. In developing countries, the availability of rainfall data is obstructed by the scarcity of precise, high-resolution precipitation. Since its inception, rainfall measurement principles have remained unchanged; non-recording and recording rain gauges are still the standard equipment for measuring ground-based precipitation notwithstanding that they only provide point measurements. According to Ochoa-Sánchez *et al.* (2014), rainfall amounts evaluated at different locations are usually extrapolated to obtain areal-averaged rainfall estimates.

As state in Young (2010), the rational use of water resources has become a major challenge for developing countries in recent decades. This concern seems from the indiscriminate use of the vital resource for the maintenance of modern society. The accelerated economic growth has led to the depletion of natural resources essential for survival and ecosystems. It is important to note that the responsible use and management of water resources has been considered a priority to improve the quality of life and, therefore, has set itself invaluable to free access.

Understanding of the variability of precipitation regime over tropical South America is crucial because of the strong dependence of water supply, hydro-energy, agriculture, and

transportation. Natural irrigation is the essential support of Brazilian agriculture. Hence, the importance of understanding the mechanism to produce water through precipitation variability in tropical South America has not only meteorological implications but also practical applications for the society in the region.

A probable scenario is drawn that due to climate change resulting from the current global warming the local rainfall pattern would be modified around the world. Nonetheless, it is important to establish efficient strategies for water resources development in conjunction with for meeting actual and future water demands, peculiarities of such a disorder over a particular region can be the predominant factors in making a decision on the precedences for development or demarcating the boundaries to formulate the necessity and availability of development.

Northern South America distinguishes by being an enormous and complex region where distinct weather systems act. According to Andreoli *et al.* (2012), Amazon region, which represents one of the most intense convective areas in the world, and northeast of Brazil, which is related to intense and prolonged droughts due to its semiarid climate, are inserted in northern South America. The cooperation between different atmospheric phenomena that appear in the whole region and local surface conditions (like vegetation, topography, and land use), generates a non-homogenous rainfall distribution that exhibits in a wide temporal and spatial range (SIERRA *et al.*, 2015).

As reported by Zhou & Lau (2001) and Takahashi (2004), precipitation variability in South America has been influenced by particular factors such as Atlantic Sea Surface Temperature (SST), El Niño-Southern Oscillation (ENSO), and Intertropical Convergence Zone (ITCZ). Grimm (2011) also considered that the ENSO impact on the characteristics of Brazil Northeast precipitation is related to the position of ITCZ and Atlantic and Pacific SST.

In general, it is known that several phenomena have a strong influence on the rainfall pattern within the Northeastern region of Brazil. One is the (ENSO), a global phenomenon that may cause severe droughts or excess of rainfall conditions depending on its intensity. Generally, the different phases of ENSO relate to years with below or above normal rainfall conditions. However, Northeast Brazil has experienced severe drought events that are not necessarily related to ENSO, but to the influence of different atmospheric systems that cause rainfall in this region (JUNIOR; ALMEIDA, 2012).

As in most parts of Northeast Brazil, Paraíba State has the following main agroecological zones: the *Mata* zone, the *Agreste* zone, and the *Sertão* zone (see Figure 1).

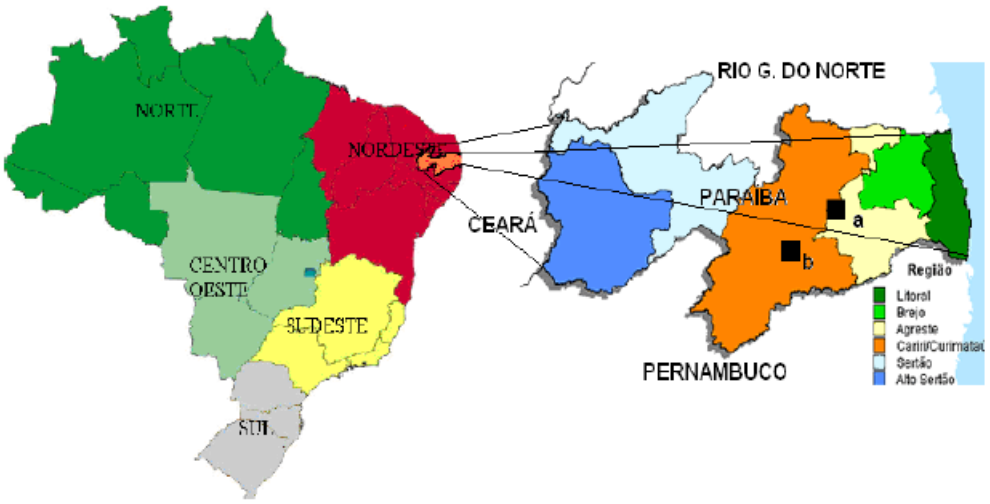


Figure 1: Homogenous climatic regions of Paraíba. Source:(ARAUJO, 2007)

In the *Mata* zone, there are two important factors for agriculture and water resources: higher precipitation (over 1,400 mm annually), and relatively fertile soils. This region has been, since colonization, widely dedicated to sugar cane production. The *Agreste* zone is located inland in the *Borborema* Highlands (between the *Mata* and *Sertão* zones), with an average annual precipitation around 700 mm, distributed irregularly but concentrated in the period from March to August (which is the season of least evapotranspiration), with mild night temperatures. The *Sertão* zone displays higher temperatures, and rains occur during the hottest months. In the *Agreste* zone, likewise in the *Sertão* zone, the dry season is long, continuing six to seven months and seven to eight months, correspondingly, with extreme droughts periods every 10 or 11 years (ARIAS; CABALLERO, 2015).

The hydrological series trend analysis is of paramount practical importance due to the effects of global climate change. Statistical procedures are used for detection of trends over time. Thus, a variety of statistical test methods have been used to detect trends in hydrometeorological and hydrological time series; these are classified as parametric and nonparametric tests (ZHANG *et al.*, 2006; CHEN *et al.*, 2007). Parametric tests are more robust but request that data be independent and normally distributed, which is barely true for hydrological time series data. For nonparametric tests, data have to be independent. However, outliers are better tolerated. The most common nonparametric tests for working with time series trends is the Mann-Kendall (MANN, 1945; KENDALL, 1975). The Mann-Kendall test is largely one used by researchers in studying hydrologic and hydrometeorological time series trends (YANMING *et al.*, 2012; WANG *et al.*, 2012; YANG *et al.*, 2012).

Geostatistics based on the theory of regionalized variables permits the analysis and interpretation of any spatially (temporally) referenced data (ISAH, 2009). It is increasingly preferred because it capitalizes on the spatial correlation between close observations to predict attribute values at unsampled locations (GOOVAERTS, 2000). Several studies Goovaerts (2000), Creutin & Obled (1982), Tabios & Salas (1985), Lebel *et al.* (1987) have demonstrated the estimation of precipitation by appropriate geostatistical tools permits more accurate results than other forms of interpolation. The possibility of quantifying uncertainty for an interpolated point or area is particularly useful, as it allows a more meaningful comparison with rainfall estimates generated by other means (e.g. radar, satellite, or numerical weather models). It also facilitates the investigation of the propagation of uncertainty in downstream models (e.g. hydrological or agricultural forecast models). However, interpretation of the kriging variance as an estimate of error depends on the data obeying the implicit statistical assumptions of kriging, but some caution may be needed.

The Shannon's entropy produces a measurement to evaluate rainfall variability. In recent years, the theory of entropy has been applied to a variety of hydrological variables. Rajagopal *et al.* (1987) suggested new perspectives for potential applications of entropy in water resources. Hydrological modeling using entropy and its application in water resources were presented by Singh (1989), Singh & Fiorentino (1992). Krstanovic & Singh (1992), they presented an investigation about information transfer between selected droughts or flood sequences using joint entropy and marginal entropy in long-term monthly rainfall series. Marginal entropy was used by Maruyama & Kawachi (1998) to investigate the characteristics of rainfall in Japan. Kawashi *et al.* (2001) presented isoentropy and isohyetal entropy maps to evaluate a degree of uncertainty of rainfall occurrence of annual rainfall pattern in Japan. Maruyama *et al.* (2005) evaluated the disorder by two different entropies, intensity entropy (IE) and apportionment entropy (AE) to represent the disorder in the intensity and occurrence of monthly rainfall in Japan. Silva *et al.* (2006) analyzed the complexity and predictability of rainfall related both to high temporal and spatial variability using entropy concepts. Mishra *et al.* (2009) presented an entropy investigation into the variability of precipitation. Using cluster analysis and entropy theory, Liu *et al.* (2013) analyzed the rainfall distribution in the Pearl River basin, China. Rajsekhar *et al.* (2015) proposed a multivariate drought index (MDI) to evaluate drought conditions and entropy maps at different time scales. Zhang *et al.* (2015) also applied multivariate drought index (MDI) associated with entropy concepts to understand the spatiotemporal patterns of precipitation regimes in the Huai River basin, China.

The following chapters show some applications to understand trends and complexity of rainfall in Paraíba. In chapter 2, the trends and their geostatistical behavior were investigated. In chapter 3, sample entropy was applied to verify the complexity and predictability of rainfall. Finally, in chapter 4 a general conclusion was presented to summarize what was verified and to project further applications with different techniques.

Chapter 2

Precipitation trend analysis by Mann-Kendall test: A case study of Paraíba, Brazil

2.1 Introduction

Recently, under a changing environment, some climatic time series have a random variable tendency. Due to the effects of possible global climate change on water engineering, conducting trend analysis using statistical methods is of paramount practical importance (ÖNÖZ; BAYAZIT, 2003).

The purpose of the trend test is to determine the increase or decrease of a series over a period of time (HELSEL; HIRSCH, 1992). Parametric and non-parametric statistical tests are used to decide whether there is a trend with statistical significance.

Minetti (1998) and Minetti *et al.* (2003) examined long-term trends in rainfall series in a variety of rain gauges around South America. They identified during the period from 1931 to 1999 that there was a steady decrease in annual rainfall for an area located west of the Andes and an increase to the east in the center of Argentina. A third region studied, further north in Argentina was identified a steady increase in rainfall until mid-1980 when high occurrences of El Niño caused declines in the indices in precipitation.

Silva *et al.* (2010) considered rainfall time series throughout Brazil during the period from 1961 to 2008; the authors found different patterns of increasing and decreasing rainfall trends. In annual terms, the range further east from Brazil's Southern region stands out as an area with significant positive trends from the center of Rio Grande do Sul to Paraná.

Pinheiro *et al.* (2013) analyzed the daily, monthly and annual time series of 18 rainfall stations in Southern of Brazil. To evaluate the occurrence of precipitation time series the Mann-Kendall test was used and showed a positive trend to annual extreme daily, total monthly and annual over time. They realized that there was an increase in rainfall in most of the analyzed stations with significant trends at 95 % in sixteen complete series analyzed. Several series have changes in the behavior of its average of the 1970s and 1980s.

Hence, the evaluation of historical trends or future projections on a regional or local scale is needed. So, in this chapter, an attempt is made to evaluate the trends in the monthly precipitation series in the Paraíba State, Brazil, and to find if there have been any significant changes in precipitation trends during 1962 to 2012. We use spatial clustering analysis to identify spatial explicit and statistically significant rainfall stations. Such an integrated spatiotemporal analysis facilitates a spatial quantification of precipitation variability.

2.2 Methodology

The Mann-Kendall test had been formulated by Mann (1945) as a non-parametric test for trend detection, and the statistical distribution of test had been given by Kendall (1975) for evaluating non-linear trend and the change point. The test has been extensively used to detect trends and spatial variation in hydroclimatic series that is, meteorological, hydrological and agrometeorological time series. Many authors [Tabari & Talaee (2011), Gocic & Trajkovic (2013), Dias *et al.* (2013), Sayemuzzaman & Jha (2014)] have used this test to evaluate the presence of significant climate trends in different parts of the world.

A factor that affects trend detection in a series is the presence of positive or negative autocorrelation (YUE *et al.*, 2003; NOVOTNY; STEFAN, 2007). A positive autocorrelation does not guarantee for a series of being detected as having trend. By contrast, for negative autocorrelation, this is reverse, where the trend is not detected. ρ_k, the autocorrelation coefficient of a discrete time series for lag-k, can be expressed as

$$\rho_k = \frac{\sum_{k=1}^{n-k}(x_t - \bar{x}_t)(x_{t+k} - \bar{x}_{t+k})}{\left[\sum_{t=1}^{n-k}(x_t - \bar{x}_t)^2 \times \sum_{n-k}^{t=1}(x_{t+k} - \bar{x}_{t+k})^2\right]^{\frac{1}{2}}} \tag{2.1}$$

where, $\bar{x}_t$ and $Var(\bar{x}_t)$ are the sample mean and sample variance of the first $(n-k)$ terms, respectively, and $\bar{x}_{t+k}$ and $Var(\bar{x}_{t+k})$ are the sample mean and the sample variation of the

last $(n-k)$ terms, respectively. Besides, the hypothesis of serial independence is tested by the lag-1 autocorrelation coefficient as $H_0 : \rho_1 = 0 \times H_1 : |\rho_1| > 0$ using the following statistic

$$t = |\rho| \sqrt{\frac{n-2}{1-\rho^2}} \tag{2.2}$$

where, the statistic of the test has a Student's t-distribution with $(n-2)$ degrees of freedom. If occurs that $|t| > t_{\frac{\alpha}{2}}$, so the null hypothesis is rejected at the chosen significance level α (CUNDERLIK; BURN, 2004).

The computational method for the Mann Kendall test considers the time series for n data points R_i and R_j as two subsets of data where $i = 1, 2, \ldots, n-1$ and $j = i+1, i+2, \ldots, n$. The data values are measured as an ordered time series. Thus, each data value is compared with all subsequent data values. If a data value from a later time period is higher than a data value from an earlier time period, statistic S is incremented by 1 (DRAPELA; DRAPELOVA, 2011). Further, if the data value from a later time period is lower than a data value from a previous time period, statistic S is decremented by 1 (SHAHID, 2011). The net result of all such increments yields the final value of S.

Given a time series $X = x_1, \ldots, x_n$ that is ranked from $R_i = R_1, \ldots, R_n$ the statistic S is calculated as

$$S = \sum_{i<j} sgn(R_j - R_i). \tag{2.3}$$

Each data point x_i is taken as a reference point which is compared with the rest of the data points x_j such as

$$Z_i = sgn(x_j - x_i) = \begin{cases} +1, & \text{if} \quad x_j > x_i \\ 0, & \text{if} \quad x_j = x_i \\ -1, & \text{if} \quad x_j < x_i \end{cases} \tag{2.4}$$

(MANN, 1945); (KENDALL; STUART, 1967) reported that when $n \geq 8$ statistic S is approximately normally distributed with

$$E(S) = 0. \tag{2.5}$$

The variance is given as

$$Var(S) = \frac{1}{18}[n(n-1)(2n+5) - \sum_{i=1}^{m} t_i(i)(i-1)(2i+5)] \tag{2.6}$$

Wherein t_i is considered as the number of ties up of sample i, and the summation is over

all ties. The standardized normal test statistic is calculated as

$$Z(n) = \frac{S}{[Var(S)]^{1/2}} \tag{2.7}$$

Test statistics Z_c is computed as

$$Z_c = \begin{cases} \dfrac{S-1}{\sqrt{Var(S)}}, & \text{if} \quad S > 0 \\ 0, & \text{if} \quad S = 0 \\ \dfrac{S+1}{\sqrt{Var(S)}}, & \text{if} \quad S < 0 \end{cases} \tag{2.8}$$

Z_c has a standard normal distribution. A positive value of Z_c implies an ascendant value of trend and vice versa. A significant level α is also utilized for testing either an ascendant or descendant monotone trend, that is, a two-tailed test. If $Z_c > Z_{\frac{\alpha}{2}}$, where α states the significant level, so the trend is considered as significant.

2.2.1 Modified Mann-Kendall test

A peculiar problem in detecting and interpreting trends in hydrological data is associated with the confusing effect of serial dependence. Pre-whitening is used for detecting a trend in a time series in the presence of autocorrelation (CUNDERLIK; BURN, 2004). However, pre-whitening is stated to reduce the rate of detection of significant trend in the Mann-Kendall test. Therefore, the Modified Mann-Kendall test is used to minimize the rate of detection of significant trend in the Mann-Kendall test (RAO *et al.*, 2003). In this work, the autocorrelation between the ranks of observations ρ_k was estimated after deducting an estimate of a non-parametric trend.

Significant values of ρ_k have only been used for calculating the variance correction factor n/n_s^*, as the variance of S is underestimated for the positively autocorrelated data:

$$\frac{n}{n_s^*} = 1 + \frac{2}{n(n-1)(n-2)} \times \sum_{k=1}^{n-1}(n-k)(n-k-1)(n-k-2)\rho_k \tag{2.9}$$

where n represents the actual number of observations, n_s^* is expressed as an effective number of observations to account for the autocorrelation in the data, and p_k is considered as the autocorrelation function for the ranks of observations (MONDAL *et al.*, 2012).

For calculating the corrected variance we assume as Rao *et al.* (2003)

$$V^* = V(S) \times \frac{n}{n_s^*} \tag{2.10}$$

wherein $Var(S)$ is obtained from Equation (2.6).

2.3 Geostatistical Techniques

Geostatistics has been defined by Matheron (1963) as "the application of probabilistic methods to regionalized variables" which designates any function displayed in real space. Different from conventional statistics, whatever the complexity and the irregularity of the real phenomenon, geostatistics searches to exhibit a structure of spatial correlation. Geostatistical methods use semivariograms as a core tool to characterize the spatial dependence in the property of interest (LY *et al.*, 2011). Geostatistics uses the concept of random functions to build a model for physical reality, bringing up these two apparently contradictory characteristics random and structured. As reported by Gruijter & Marsman (1985), Heuvelink *et al.* (1997), its basic apparatus is variogram analysis which contains the study of the variogram function of specific physical variable value or of a water quality parameter under study.

2.3.1 Kriging

Kriging is the term used by geostatisticians for a family of generalized least-squares regression methods that use available data in a specific search neighborhood to estimate the values at unsampled locations (ISAAKS; SRIVASTAVA, 1989; GOOVAERTS, 1997). As stated in Berke (1999), it is based on a spatial linear model for the data which specify a parametric spatial mean function and spatial dependence structure. As stated by Isaaks & Srivastava (1989), kriging uses a variogram model to characterise spatial correlation. A variogram describes in terms of variance how spatial variability changes as a function of distance and direction.

Kriging uses statistical models and allows a variety of map outputs, including predictions, prediction standard errors, probability, etc. Today, a number of variants of kriging are in general use, such as *Simple Kriging, Ordinary Kriging, Universal Kriging, Block Kriging, Co-Kriging*, and *Disjunctive Kriging*. Among the various forms of kriging, *Ordinary Kriging* has been used widely as a reliable estimation method (YAMAMOTO, 2000). In short, *Ordinary Kriging* is the basic form of kriging. It has been widely used with rainfall data. The prediction by *Ordinary Kriging* is a linear combination of measured values.

2.3.2 Semivariogram

Semivariogram is a convenient tool for the analysis of spatial dependence structure (CRASSIE, 1993). If the spatial dependence exists, its degree is quantified by comparing the models to the experimental semivariogram. Using Eq. 2.11 to compute experimental semivariogram from the data under study is the only certain way to describe how semivariance changes with distance, determine which semivariogram model should be used. By changing h, both in distance and direction, a set of the sample (or experimental) semivariograrns for the data is obtained (BURROUGH, 1986)

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2 \tag{2.11}$$

where $\gamma(h)$ is the semivariance as well as $N(h)$ is the number of $Z(x_i)$ and $Z(x_i + h)$, separated by an h vector. A variety of theoretical models can be utilised to adjust from the experimental semivariogram to the theoretical semivariogram. Notwithstanding, Johnston *et al.* (2001) showed 11 theoretical models.

i) Spherical

$$\gamma(h) = (\tau^2 + \sigma^2) \left[\frac{3}{2} \left(\frac{h}{\phi} \right) - \frac{1}{2} \left(\frac{h}{\phi} \right)^3 \right]$$

ii) Exponential

$$\gamma(h) = (\tau^2 + \sigma^2) \left[1 - \exp \left(\frac{-h}{\phi} \right) \right]$$

iii) Gaussian

$$\gamma(h) = (\tau^2 + \sigma^2) \left[1 - \exp \left(-3 \left(\frac{-h}{\phi} \right) \right)^2 \right]$$

iv) Linear

$$\gamma(h) = \tau^2 + [h(\sigma^2/\phi)]$$

v) Circular

$$\gamma(h) = \tau^2 + \left\{ (2\sigma^2/\pi) \left[(h/\phi) \left[1 - \left(\frac{h}{\phi} \right)^2 \right]^{0,5} + \arcsin(h/\phi) \right] \right\}$$

vi) Tetraspherical

$$\gamma(h) = \tau^2 + \left(\frac{2\sigma^2}{\phi}\right)\left[\arcsin\left(\frac{h}{\phi}\right) + \left(\frac{h}{\phi}\right)\left(1 - \left(\frac{h}{\phi}\right)^2\right)^{0,5} + \frac{2h}{3\phi}\left(1 - \left(\frac{h}{\phi}\right)^2\right)^{1,5}\right]$$

vii) Pentaspherical

$$\gamma(h) = \tau^2 + \left\{\sigma^2\left[\left(\frac{15h}{8\phi}\right) - \frac{5}{4}\left(\frac{h}{\phi}\right)^3 - \left(\frac{3}{8}\right)\left(\frac{h}{\phi}\right)^3\right]\right\}$$

viii) Rational Quadratic

$$\gamma(h) = \tau^2\left\{\frac{\left[19\left(\frac{h}{\phi}\right)^2\right]}{\left[1 + 19\left(\frac{h}{\phi}\right)^2\right]}\right\}$$

ix) Hole Effect

$$\gamma(h) = \tau^2 + \left\{\sigma^2\frac{\left[1 - \sin\left(\frac{2\pi h}{\phi}\right)\right]}{\sin\left(\frac{2\pi h}{\phi}\right)}\right\}$$

x) K-Bessel

$$\gamma(h) = \tau^2 + \sigma^2\left[1 - \frac{\left(\frac{\Omega_{\theta_k} h}{\phi}\right)^{\theta_k}}{2^{\theta_{K-1}}\Gamma\left(\theta_k\right)} K_{\theta_k}\left(\frac{\Omega_{\theta_k} h}{\phi}\right)\right]$$

where Ω is the numeric value of $\gamma(h) = 0.95\tau^2$ for any θ_k. K_ϕ is the modified Bessel function of second and order ϕ (ABRAMOWITZ; STEGUN, 1964).

xi) J-Bessel

$$\gamma(h) = \tau + \left\{\sigma^2\left[1 - \left[\frac{2^{\tau^2+\sigma^2}\Gamma\left(\tau^2 + \sigma^2 + 1\right)}{\Omega_{\left(\frac{\tau^2 \times h}{\phi}\right)}}\right] J_{(\tau^2+\sigma^2)}\right]\left(\frac{\Omega_{(\tau^2+\sigma^2)\times h}}{\phi}\right)\right\}$$

$J_{(\tau^2+\sigma^2)}$ is the J-Bessel function (ABRAMOWITZ; STEGUN, 1964).

Applying the algorithm of weighted least squares (WLS), these models were adjusted to the experimental semivariogram, and the subsequent model parameters were defined: **nugget effect** σ^2, **sill** $\tau^2 + \sigma^2$, and **range** ϕ. In order to verify the existence of spatial dependence, the spatial dependence index (SDI), proposed by Cambardella *et al.* (1994), was applied, which is the ratio representing the percentage of data variability explained by spatial dependence. The SDI is estimated as follows: $SDI = [\tau^2/(\sigma^2+\tau^2)] \times 100$, being classified as strong (SDI$\leq$ 25%), medium ($25 < SDI < 75\%$), and low ($SDI \geq 75\%$).

2.3.3 Cross-Validation

Goovaerts (1997) argues that cross-validation allows comparing the impact of interpolators among the real estimated values, in which the model with more accurate prediction is chosen. Cross-validation allows determination of models that provide the best prediction (JOHNSTON *et al.*, 2001). The cross-validation technique was used to select the semivariogram model (WEBSTER; OLIVER, 2007). Faraco *et al.* (2008) recognized the cross-validation criterion as the most adequate for choosing the best semivariogram adjustment. The semivariogram models were tested for each parameter data set. The quality of prediction performances was assessed by cross-validation. Cross-validation was conducted to assess the accuracy of Ordinary Kriging through some statistical measurements as follows:

Mean Prediction Errors (ME)

$$ME = \sum_{i=1}^{n} \frac{\left[\hat{Z}_{(}t_i) - Z_{(}t_i)\right]}{n} \tag{2.12}$$

Mean Standardized Prediction Errors (MS)

$$MS = \frac{\sum_{i=1}^{n} \frac{\left[\hat{Z}_{(}t_i) - Z_{(}t_i)\right]}{\hat{\sigma}_{t_i}}}{n} \tag{2.13}$$

Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{\left(\left[\hat{Z}_{(}t_i) - Z_{(}t_i)\right]\right)^2}{n}} \tag{2.14}$$

Average Standardized Error (ASE)

$$ASE = \sqrt{\frac{\sum_{i=1}^{n} \hat{\sigma}_{t_i}}{n}} \tag{2.15}$$

Root Mean Squared Standardized Error (RMSSE)

$$RMSSE = \sqrt{\frac{\sum_{i=1}^{n} \frac{\left[\hat{Z}_{(}t_i) - Z_{(}t_i)\right]}{\hat{\sigma}_{t_i}}}{n}} \tag{2.16}$$

According to Johnston *et al.* (2001), to evaluate a model that provides accurate predictions, the Standardized Prediction Errors (MS) should be close to 0, the Square Root of Standardized Mean Error (RMSSE) should be close to 1. The Average Standard Error

(ASE) and the Root Mean Square Error (RMSE) and should be as small as possible. In order to verify the best choice among J different fitted models and their MS values close to 0 and RMSSE values close to 1, Santos *et al.* (2012) suggested Error Comparison Index (ECI), which can be calculated as follows:

$$ICE_i = A_i + B_i \tag{2.17}$$

where,

$$A_i = \begin{cases} \dfrac{ABS(MS)_i}{MAX(ABS(MS))}, & \text{when} \quad MAX(ABS(MS)) > 0 \\ 1, & \text{when} \quad MAX(ABS(MS)) = 0 \end{cases} \tag{2.18}$$

$$B_i = \begin{cases} \dfrac{ABS(RMSS-1)_i}{MAX(ABS(RMSS-1))}, & \text{when} \quad MAX(ABS(RMSS-1)) > 0 \\ 1, & \text{when} \quad MAX(ABS(RMSS-1)) = 0 \end{cases} \tag{2.19}$$

The best-fitted model among J different models is one which presents the lowest ECI value.

2.4 Data preprossessing

2.4.1 Study area

The study was carried out in Paraíba State, located in Brazilian Northeast, as shown in Figure 2. The Brazil Northeast has 1.5×10^6 km^2 of area ranging between $1 - 18°S$ and $35 - 47°W$, the region is influenced by different meteorological systems with distinct characteristics (FILHO *et al.*, 2014). As stated in Liebmann *et al.* (2011), Kovadio *et al.* (2012), in Northeast extreme events are related to deficit precipitation (semiarid region) or excess precipitation (capitals or coast regions). Nevertheless, the amount of precipitation in Paraíba is related to various meteorological systems: Intertropical Convergence Zone (ITCZ), High-Level Cyclone Vortex (HLCV), Jet Streams. According to Macedo *et al.* (2010), and it is one of the Brazilian states which presents a marked water scarcity, as it has a semi-arid climate in most of its territory.

2.4.2 Data

The data required for calculation of trends was obtained for a total of 69 precipitation stations having a common period from 1962 to 2012. The data were collected from *Instituto Nacional de Meteorologia* (INMET) and acquired by *Unidade Acadêmica de Ciências Atmosféricas* (UACA-UFCG). Rainfall gauges present a homogeneous spatial distribution around the state and include all micro-regions of Paraíba state. The spatial locations of stations are shown in Figure 2 and their geographical locations are shown in Table 1.

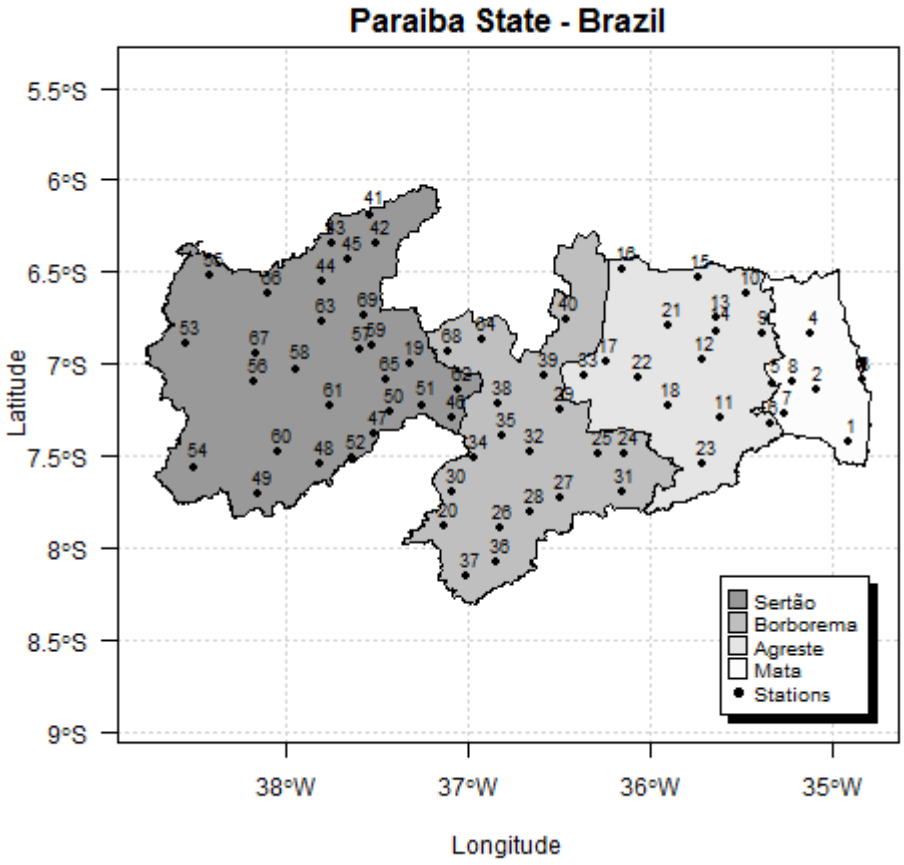


Figure 2: Spatial distribution of 69 precipitation stations used in the study.

Table 1: Geographical coordinates of precipitation stations used in the study.

Number	Code	Name	Longitude	Latitude
1	3940819	Alhandra	-34,9106	-7,4256
2	3849384	Cruz do Espírito Santo	-35,0911	-7,1408
3	3940226	João Pessoa/DFAARA	-34,8333	-7,0833
4	3839679	Mamanguape	-35,1214	-6,8356
5	3849232	Caldas Brandão	-35,3244	-7,1025
6	3849636	Itabaiana	-35,3375	-7,3250
7	3849545	Pilar	-35,2608	-7,2675
8	3849254	Sapé	-35,2233	-7,0925
9	3839727	Araçagi	-35,3878	-6,8333
10	3839208	Caiçara	-35,4681	-6,6147
11	3848579	Ingá	-35,6119	-7,2925
12	3838962	Areia	-35,7178	-6,9756
13	3838575	Bananeiras	-35,6342	-6,7514
14	3838675	Serraria	-35,6386	-6,8192
15	3838055	Araruna	-35,7397	-6,5314
16	3827973	Cuité	-36,1492	-6,4850
17	3837953	Olivedos	-36,2436	-6,9886
18	3848428	Campina Grande/EMBRAPA	-35,9042	-7,2256
19	3845045	Patos	-37,3131	-7,0008
20	3882792	Monteiro	-37,1269	-7,8850
21	3838526	Casserengue/Salgado, St.	-35,8944	-6,7931
22	3847188	Pocinhos	-36,0592	-7,0778
23	3858065	Aroeiras	-35,7111	-7,5458
24	3847979	Boqueirão/Boqueirão, Aç.	-36,1358	-7,4908
25	3857044	Cabaceiras	-36,2869	-7,4922
26	3856828	Camalaú	-36,8256	-7,8900
27	3856498	Caraúbas	-36,4903	-7,7253
28	3856667	Congo	-36,6586	-7,8022
29	3847505	Gurjão	-36,4892	-7,2478
30	3855383	Prata	-37,0842	-7,6950
31	3857471	Riacho de Santo Antônio	-36,1561	-7,6942
32	3846969	Serra Branca	-36,6600	-7,4819
33	3847128	Soledade	-36,3619	-7,0608

Continued on next page

Table 1 – (continued)

Number	Code	Name	Longitude	Latitude
34	3856008	Sumé/Bananeiras, Fz.	-36,9631	-7,5069
35	3846739	São José dos Cordeiros	-36,8058	-7,3908
36	3866128	São João do Tigre	-36,8472	-8,0800
37	3865397	São Sebastião do Umbuzeiro	-37,0097	-8,1517
38	3846434	Taperoá	-36,8281	-7,2164
39	3846185	Juazeirinho	-36,5800	-7,0683
40	3837507	Pedra Lavrada	-36,4644	-6,7553
41	3824396	Belém do Brejo do Cruz	-37,5356	-6,1864
42	3825701	Brejo do Cruz	-37,4997	-6,3483
43	3824751	Catolé do Rocha	-37,7467	-6,3439
44	3834137	Jericó	-37,8000	-6,5500
45	3824992	Riacho dos Cavalos/Jenipapeiro dos Carreiros	-37,6531	-6,4353
46	3845583	Desterro	-37,0881	-7,2903
47	3845703	Imaculada	-37,5094	-7,3822
48	3854036	Jurú	-37,8067	-7,5478
49	3853467	Manaíra	-38,1525	-7,7069
50	3845514	Mãe dŧÁgua	-37,4253	-7,2572
51	3845448	Teixeira	-37,2497	-7,2217
52	3854072	Água Branca	-37,6367	-7,5119
53	3832789	Cajazeiras	-38,5444	-6,8942
54	3852197	Conceição	-38,5019	-7,5600
55	3833018	Uiraúna	-38,4092	-6,5231
56	3843166	Aguiar	-38,1733	-7,0933
57	3834877	Condado	-37,5947	-6,9231
58	3844008	Coremas/Coremas, Aç.	-37,9428	-7,0250
59	3834894	Malta	-37,5197	-6,9033
60	3843992	Nova Olinda	-38,0425	-7,4819
61	3844448	Olho d‘Água	-37,7506	-7,2278
62	3845289	Passagem	-37,0475	-7,1364
63	3834538	Pombal	-37,8006	-6,7719
64	3836715	Santa Luzia	-36,9181	-6,8681
65	3845113	Santa Terezinha	-37,4450	-7,0842
66	3833285	São Francisco	-38,0947	-6,6178
67	3833869	São José da Lagoa Tapada	-38,1619	-6,9422
68	3835882	São Mamede	-37,1036	-6,9306

Continued on next page

Table 1 – (continued)

Number	Code	Name	Longitude	Latitude
69	3834389	Vista Serrana/Desterro da Malta	-37,5683	-6,7386

2.5 Results and Discussion

In this section, we present trend analysis results for monthly precipitation. In order to choose the best-fitted model that would predict the trend of precipitation in Paraíba, we used cross validation. The cross validation that examines the validity of fitted models and parameters of semivariograms for precipitation parameters are given in Table 2 and Table 3. Data analysis and Mann-Kendal test were done using Team (2014), and spatial trends, experimental semivariograms and geostatistical analysis were done using *ArcGis* (ESRI., 2014).

According to Owolawi & Afullo (2007), Sana *et al.* (2014), the RMS statistic is widely used to select a semivariogram model or an implementation method. Along with RMS, the ASE statistic is used to validate statistical models. Therefore, the knowledge of obtaining the lowest ASE value and closest statistic RMS benefits in choosing the best model. On the verification by RMSS, MS, ASE and RMS statistics may generate a small distraction, we used the Error Comparison Index (ECI) suggested by Santos *et al.* (2012) because it helped compare the statistics used in cross validation process and chose the best geostatistical model by the lowest ECI value.

One can note from Table 2 that the model selection which presents the best semivariogram fit may include all errors of prognosis in an integrated manner. Nonetheless, in January the circular model showed the lower ECI value, that is, 0.0559; even showed a moderate SDI classification as reported by Cambardella *et al.* (1994). For February, the model which presented the lowest ECI (0.6012) was the pentaspherical. February is a month of transition, that is, starts a new precipitation cycle (which impacts directly the trend), the end of rainy season in Paraíba hinterland region and the beginning of rainy season in the coastal region.

In March, the Gaussian model presented the lowest ECI value (0.1528). The tetraspherical model showed the lowest ECI value (0.5380) in April; Gomes *et al.* (2014), which also studied precipitation trends in Paraíba, suggested Mátern model. In May, the lowest

ECI (0.2343) was showed by K-Bessel model and SCI approximately zero which featured a strong spatial dependence as stated in (CAMBARDELLA *et al.*, 1994).

In June, the rational quadratic showed the lowest ECI value (0.3872) and a moderate spatial dependence. In July, the pentaspherical model showed the lowest ECI value (0.8411) but very close to the spherical model which showed an ECI value of (0.8500). Ly *et al.* (2011) conducted a study with daily data in a 30-year period by comparing the ratio of influence of precipitation in catchment areas of Belgium, and realized a significant occurrence of the pentaspherical model. Similar to March, the gaussian model showed the lowest ECI value (0.6077) in August.

Nevertheless, similar to April the tetraspherical model showed the lowest value of ECI (0.8071) in September; this value is not too far from the spherical model ECI value. As stated in Cambardella *et al.* (1994) the model showed a moderate spatial dependence. In October, the exponential model showed the lowest value of ECI (0.6617). In contrast, Viola *et al.* (2014) studied the distribution and the erosive potential of rains (October to April) and the exponential model not suited. Thus, in November the spherical model showed the lowest ECI value (0.8031) was as stated in Viola *et al.* (2014) who highlighted the influence of the spherical model in this period.

Similar to July, in December the rational quadratic model was the one with the lowest value of ECI (0.7735).

Table 2: Values of MS, RMSE, MS, RMSSE, ASE and spatial dependency index (SDI) and error comparison index (ECI) for trends of precipitation between January and December.

Month	Model	ME	RMSE	MS	RMSSE	ASE	SDI	ECI
January	Gaussian	-0.0069	0.935	-0.004	0.996	0.947	51.29	0.224
	Exponential	0.0141	0.953	0.011	0.993	0.970	26.95	0.516
	Spherical	0.0034	0.941	0.003	0.9991	0.950	41.29	0.162
	Circular	-0.002	0.936	-0.0011	0.999	0.946	42.78	**0.058**
	Rational Quadratic	0.0116	0.953	0.006	0.981	0.980	39.95	0.498
	Tetraspherical	0.0048	0.942	0.004	0.997	0.954	40.16	0.199
February	Spherical	-0.0148	0.957	-0.009	1.039	0.912	43.24	0.800
	Circular	-0.0282	0.958	-0.021	1.047	0.907	47.69	1.130
	Hole Effect	-0.0231	1.035	-0.020	1.060	0.970	45.19	1.332
	K-Bessel	-0.0137	0.947	-0.009	1.033	0.905	0	0.706
	Tetraspherical	-0.0184	0.955	-0.014	1.031	0.917	39.07	0.759
	Pentaspherical	-0.0146	0.953	-0.0114	1.025	0.920	34.90	**0.601**
March	**Gaussian**	-0.0013	1.2012	0.0037	0.999	1.199	53.49	**0.152**
	Exponential	-0.0382	1.0708	-0.0171	0.963	1.118	14.24	1.396
	Spherical	-0.0075	1.1362	-0.0003	0.977	1.165	38.21	0.461
	Quadratic	-0.0415	1.0156	-0.0181	0.949	1.094	19.21	1.718
	Tetraspherical	-0.0076	1.1172	0.0001	0.978	1.147	33.33	0.428
	Pentaspherical	-0.0075	1.1032	0.0008	0.980	1.133	29.03	0.417
April	Gaussiano	-0.0600	1.2102	-0.0417	0.990	1.228	48.03	1.045
	Spherical	-0.0491	1.2082	-0.0344	0.992	1.223	27.40	0.857
	Circular	0.0637	1.2059	-0.0441	0.986	1.228	38.24	1.139
	Hole Effect	-0.0333	1.1836	-0.0251	0.906	1.286	23.16	1.566
	Rational Quadratic	-0.0406	1.1738	-0.0272	0.946	1.236	0.09	1.182
	Tetraspherical	-0.0327	1.2097	-0.0227	0.997	1.21	16.80	**0.538**

Continued on next page

Table 2 – (continued)

Month	Model	ME	RMSE	MS	RMSSE	ASE	SDI	ECI
May	Gaussian	-0.04311	1.0121	-0.0322	1.034	0.980	34.35	1.282
	Spherical	-0.03757	1.0025	-0.0258	1.037	0.973	9.88	1.159
	Circular	-0.0535	1.0176	-0.0393	1.051	0.969	18.39	1.687
	K-Bessel	-0.0159	0.9927	-0.0086	1.001	0.986	0	**0.234**
	J-Bessel	-0.01634	0.9995	-0.0124	1.006	1.005	0	0.404
	Tetraspherical	-0.0417	1.0229	-0.0308	1.044	0.979	2.72	1.373
June	Gaussian	0.0066	2.0526	0.0035	1.120	1.798	28.54	1.154
	Exponential	-0.0019	2.0570	-0.0011	1.068	1.913	0	0.613
	J-Bessel	-0.0471	2.1392	-0.0232	1.066	2.001	54.15	1.550
	Rational Quadratic	-0.0082	2.0497	-0.004	1.024	1.9784	26.36	**0.387**
	K-Bessel	0.0025	2.0516	0.0014	1.1007	1.831	30.17	0.897
	Pentaspherical	0.0106	2.0539	0.0054	1.055	1.921	22.30	0.689
July	Spherical	-0.02610	0.9658	-0.0214	0.983	0.988	52.34	0.850
	Circular	-0.0349	0.9628	-0.0298	0.988	0.978	52.33	1.015
	Hole Effect	-0.0393	0.9750	-0.0345	0.972	1.007	72.59	1.366
	Pentaspherical	-0.0186	0.9664	-0.0144	0.968	1.004	53.08	**0.841**
	Rational Quadratic	-0.0376	0.9804	-0.0339	0.926	1.058	68.80	1.982
	J-Bessel	-0.0263	0.9586	-0.0220	0.962	1.004	64.77	1.147
August	**Gaussian**	-0.0021	1.3396	-0.0003	1.047	1.275	60	**0.607**
	Exponential	0.0112	1.2936	0.0087	1.043	1.236	31.59	1.217
	Spherical	0.0062	1.3228	0.0055	1.055	1.250	50.11	1.113
	Circular	0.0029	1.3328	0.0031	1.059	1.254	52.20	0.970
	J-Bessel	-0.0123	1.3363	-0.0084	1.082	1.232	57.42	1.669
	Rational Quadratic	0.0168	1.2897	0.0126	1.069	1.205	30.26	1.842
September	J-Bessel	0.0043	1.1225	0.0037	1.054	1.065	51.11	1.073
	Rational Quadratic	0.0133	1.1327	0.0104	1.076	1.053	36.05	1.993
	Tetraspherical	-0.0032	1.1208	-0.0028	1.040	1.078	39.34	**0.807**
	Gaussian	-0.0045	1.1363	-0.0042	1.05	1.071	49.78	1.161
	Exponential	0.0094	1.1416	0.0080	1.076	1.060	26.92	1.772
	Spherical	0.0016	1.1274	0.0012	1.061	1.061	39.73	0.916

Continued on next page

Table 2 – (continued)

Month	Model	ME	RMSE	MS	RMSSE	ASE	SDI	ECI
October	Gaussian	-0.0241	1.4932	-0.0143	1.071	1.391	46.52	1.412
	Exponential	-0.0269	1.5077	-0.0141	0.991	1.524	67.31	**0.661**
	Spherical	-0.0283	1.5061	-0.0158	1.051	1.437	53.15	1.234
	K-Bessel	-0.0175	1.4874	-0.0095	1.065	1.394	47.43	1.146
	J-Bessel	-0.0168	1.4950	-0.0089	1.059	1.409	55.54	1.053
	Pentaspherical	0.0385	1.1594	0.0249	1.085	1.452	25.70	1.999
November	Exponential	-0.0054	1.0956	-0.0040	1.041	1.047	51.55	1.117
	Spherical	0.0004	1.0790	0.0024	1.032	1.043	62.67	**0.803**
	Circular	-0.0049	1.0825	-0.0023	1.036	1.041	63.86	0.879
	K-Bessel	-0.0135	1.0870	-0.0111	1.032	1.049	70.77	1.592
	J-Bessel	-0.0120	1.0731	-0.0087	1.051	1.019	67.38	1.728
	Rational Quadratic	-0.0126	1.1140	-0.0107	1.055	1.048	54.67	1.961
December	Gaussian	0.0288	1.1308	0.0195	1.062	1.044	48.04	1.326
	Exponential	0.0303	1.1396	0.0199	1.041	1.076	17.19	1.150
	Spherical	0.0384	1.1408	0.0248	1.077	1.0393	28.58	1.660
	Circular	0.0311	1.1372	0.0206	1.074	1.0392	36.77	1.478
	Rational Quadratic	0.0254	1.1339	0.018	1.007	1.1106	55.33	**0.773**
	Hole Effect	0.0349	1.1146	0.0216	1.108	0.996	41.18	1.827

Table 3: Variographic parameters of experimental semivariogram for rainfall trends in Paraíba.

Month	Variographic parameters	
January	$\hat{\tau}^2$	0.5613
	σ^2	0.7509
	$\hat{\tau}^2 + \hat{\sigma}^2$	1.3125
	$\hat{\phi}$	1.3949

Month	Variographic parameters	
February	$\hat{\tau}^2$	0.31365
	σ^2	0.58504
	$\hat{\tau}^2 + \hat{\sigma}^2$	0.89869
	$\hat{\phi}$	0.72139

Month	Variographic parameters	
March	$\hat{\tau}^2$	1.17701
	σ^2	1.02311
	$\hat{\tau}^2 + \hat{\sigma}^2$	2.20012
	$\hat{\phi}$	1.9136

Month	Variographic parameters	
April	$\hat{\tau}^2$	0.27912
	σ^2	1.38172
	$\hat{\tau}^2 + \hat{\sigma}^2$	1.66084
	$\hat{\phi}$	0.57556

Month	Variographic parameters	
May	$\hat{\tau}^2$	0
	σ^2	1.22825
	$\hat{\tau}^2 + \hat{\sigma}^2$	1.22825
	$\hat{\phi}$	0.5755

Month	Variographic parameters	
June	$\hat{\tau}^2$	1.02348
	σ^2	2.85857
	$\hat{\tau}^2 + \hat{\sigma}^2$	3.88205
	$\hat{\phi}$	0.55513

Month	Variographic parameters	
July	$\hat{\tau}^2$	0.59547
	σ^2	0.52618
	$\hat{\tau}^2 + \hat{\sigma}^2$	1.12165
	$\hat{\phi}$	1.07346

Month	Variographic parameters	
August	$\hat{\tau}^2$	1.41874
	σ^2	0.94568
	$\hat{\tau}^2 + \hat{\sigma}^2$	2.36442
	$\hat{\phi}$	2.07576

Month	Variographic parameters	
September	$\hat{\tau}^2$	0.79313
	σ^2	1.22278
	$\hat{\tau}^2 + \hat{\sigma}^2$	2.01591
	$\hat{\phi}$	2.07576

Month	Variographic parameters	
October	$\hat{\tau}^2$	1.93543
	σ^2	0.93967
	$\hat{\tau}^2 + \hat{\sigma}^2$	2.8751
	$\hat{\phi}$	1.30947

Month	Variographic parameters	
November	$\hat{\tau}^2$	0.81326
	σ^2	0.48429
	$\hat{\tau}^2 + \hat{\sigma}^2$	1.29755
	$\hat{\phi}$	1.4994

Month	Variographic parameters	
December	$\hat{\tau}^2$	0.75012
	σ^2	0.60538
	$\hat{\tau}^2 + \hat{\sigma}^2$	1.3555
	$\hat{\phi}$	1.55466

The variograms for the ordinary kriging are presented in Table 3. Parameters were obtained by using measurement errors to estimate **nugget**, global variance to estimate the **sill** and the mean distance to the nearest neighbor to estimate the **range**. Using the parameter estimation, we can obtain the semivariogram's equation as follows:

January

$$\hat{\gamma}(h) = 0.5613 + \left\{ \frac{1.5018}{\pi} \times \left[\frac{h}{1.39495} \times \left[1 - \left(\frac{h}{1.39495} \right)^2 \right]^{2.5} + \arcsin\left(\frac{h}{1.39495} \right) \right] \right\}$$

February

$$\hat{\gamma}(h) = 0.31365 + \left\{ 0.58504 \times \left[\frac{15h}{5.77112} - \frac{5}{4} \times \left(\frac{h}{0.72139} \right)^3 - \frac{3}{8} \times \left(\frac{h}{0.72319} \right)^5 \right] \right\}$$

March

$$\hat{\gamma}(h) = 2.20012 \left[1 - \exp\left(-3 \left(\frac{h}{1.9136} \right)^2 \right) \right]$$

April

$$\hat{\gamma}(h) = 0.27912 + \left(\frac{0.55824}{0.57556} \right) \times \left[\arcsin\left(\frac{h}{0.57556} \right) + \left(\frac{h}{0.57556} \right) \left(1 - \left(\frac{h}{0.57556} \right)^2 \right)^{0.5} + \frac{2h}{1.72668} \left(1 - \left(\frac{h}{0.57556} \right)^2 \right)^{1.5} \right]$$

May

$$\hat{\gamma}(h) = 1.22825 \left[\frac{1 - (\Omega_{\theta_k} h)^{\Omega_{\theta_k}}}{2^{\Omega_k^{-1}} \Gamma(\theta_k)} K_{\theta_k} \left(\frac{\Omega_{\theta_k}}{0.5755} \right) \right].$$

Where, Ω_{θ_k} is a numerical value so that $\gamma(h) - 0.95(\tau^2 + \sigma^2)$ for any Ω_k. $\Gamma(\theta_k)$ is the gamma function, and K_{θ_k} is the modified Bessel function of second order θ_k.

June

$$\hat{\gamma}(h) = 1.02348 \times \left\{ \left[19 \times \left(\frac{h}{(1.02348)^2} \right)^2 \right] / \left[1 + 19 \left(\frac{h}{1.02348} \right)^2 \right] \right\}$$

July

$$\hat{\gamma}(h) = 0.59547 + \left\{ (0.52618) \times \left[\left(\frac{15h}{8.58768} \right) - \frac{5}{4} \times \left(\frac{h}{1.07346} \right)^3 - \frac{3}{8} \left(\frac{h}{1.07346} \right)^2 \right] \right\}$$

August

$$\hat{\gamma}(h) = 2.36442 \left[1 - \exp\left(-3\left(\frac{h}{2.07576}\right)^2\right)\right]$$

September

$$\hat{\gamma}(h) = 0.79313 + \left(\frac{2.44556}{2.07576}\right) \times \left[\arcsin\left(\frac{h}{2.07576}\right) + \left(\frac{h}{2.07576}\right)\left(1 - \left(\frac{h}{2.07576}\right)^2\right)^{0,5} + \frac{2h}{6.22728}\left(1 - \left(\frac{h}{2.07576}\right)^2\right)^{1.5}\right]$$

October

$$\hat{\gamma}(h) = 2.8751 \left[1 - \exp\left(\frac{-h}{1.30947}\right)\right]$$

November

$$\hat{\gamma}(h) = 1.29755 \left[\frac{3}{2}\left(\frac{h}{1.4994}\right) - \frac{1}{2}\left(\frac{h}{1.4994}\right)^3\right]$$

December

$$\hat{\gamma}(h) = 0.75012 \times \left\{\left[19 \times \left(\frac{h}{(1.55466)^2}\right)^2\right] / \left[1 + 19\left(\frac{h}{1.55466}\right)^2\right]\right\}$$

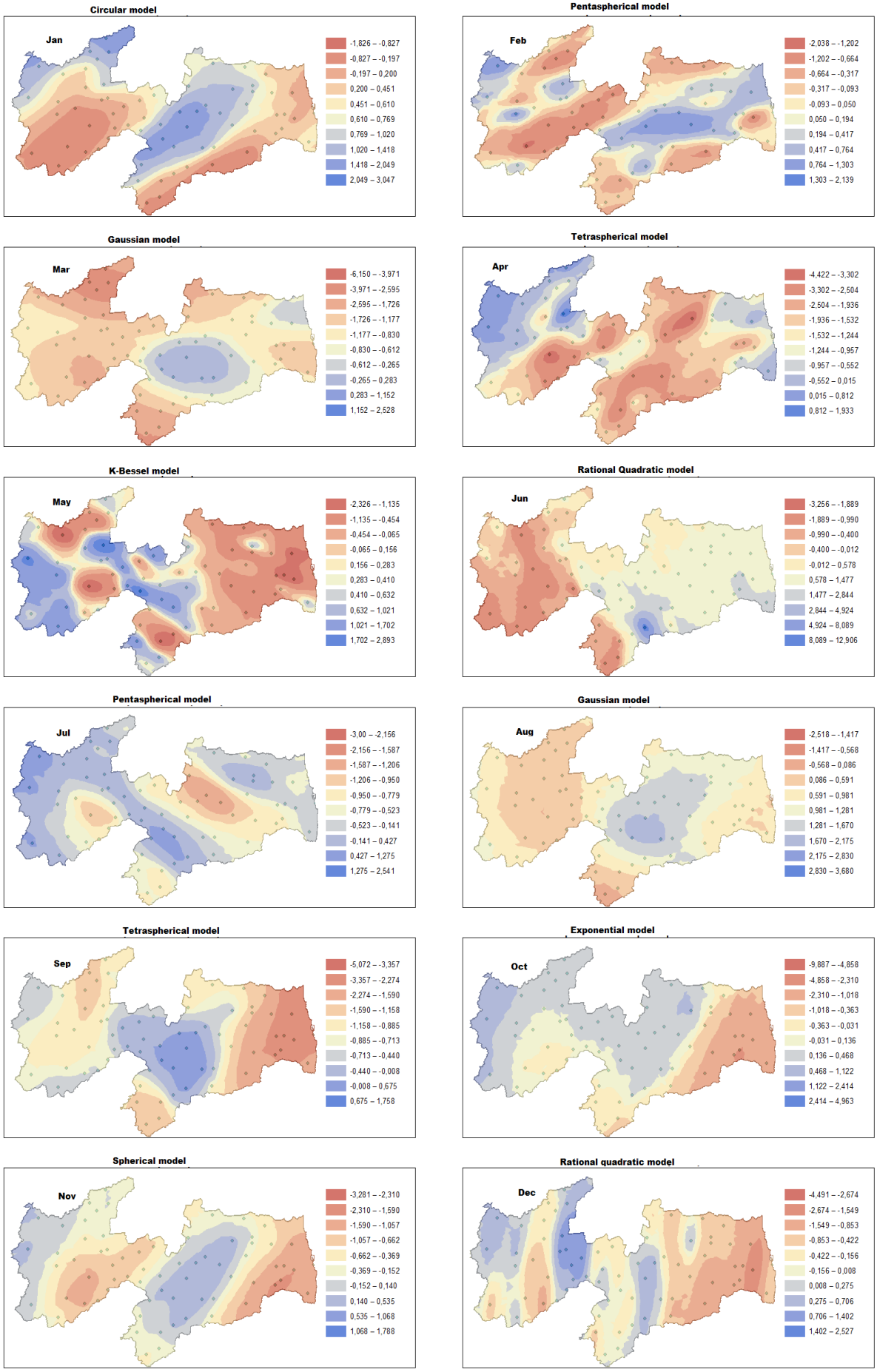


Figure 3: Spatial variation of monthly rainfall trends (mm/month) for Paraíba from 1962 to 2012.

The successful performance of a method depends on numerous factors, in particular, temporal and spatial resolutions of data, and parameters of the model, such as the semivariogram in the case of kriging. Figure 3 on the top left shows the trend interpolation map for January adjusted according to ECI results shown in Table 2. One can see a high concentration of increasing trends in the north of *Sertão* region and in the central part of the state with a slight displacement to *Planalto da Borborema* and *Agreste* region. Carvalho *et al.* (2004) emphasized the influence of South Atlantic Convergence Zone (SCAZ) over the maximum values of precipitation in Northeast of Brazil. The intraseasonal anomalies in precipitation are related to anomalies of Pacific Ocean. In February, different from January, there is an increase of significant trends in the North of *Sertão* region and an intensified displacement between *Borborema* and *Agreste* and until the North of *Mata* region. These results are in agreement with (CARVALHO *et al.*, 2004).

In March, it is conceivable that there is a huge concentration of trends between *Borborema* and *Agreste* and not so significant in the North of *Mata*. This is related to the end of rainy season between the months of January and February. In April, the interpolation map shows increasing trends in the North and West of *Sertão* as well as appearance of trends in the North and South of *Mata* with great influence over the coast. In the study of Rodrigues *et al.* (2011) it was verified, regarding seasonal rainfall distribution, that increasing trends take place during March, April and May (autumn in South Hemisphere) caused by the influence of *El Niño* and *La Niña* phenomena.

In May, there is a strong increase of trends in the West of *Sertão* and in the South of *Borborema*, also noticeable is a sensible increase of trends in the North of *Agreste* region and in the South of *Mata* region. Rodrigues *et al.* (2011) related this displacement of trends to the occurrence of *El Niño* and *La Niña*. In June, one realizes that there is a strong increase of the trends in the region over the Southern *Borborema* region. An increase of the trend can be visualized in the South of *Mata* region. Rodrigues *et al.* (2011), regarding the seasonal distribution of precipitation, observed a negative trend intensity during the months of June, July and August (winter in South Hemisphere), not the periods of extreme events.

In July, there was a concentration of an increasing trend almost in the entire region further west of *Sertão* and a large displacement in the southernmost of *Borborema*. Another displacement of trends can be visualized from the North of *Agreste* through North and South of *Mata* addressing the coast. In August, there was a huge concentration of the increase of trends in South Central *Borborema* region with a sensible movement to

Agreste region.

In September, there was an increase of trends in the Northwest and South of *Sertão* region and a substantial concentration in the South of *Borborema* region with a slight movement to *Sertão* and *Agreste*. In October, for the first time, there was a predominance of increasing trends in the unbroken range of Paraíba, that is, much of *Sertão*, *Borborema* and North of *Agreste* region. This predominance was also described by Gomes *et al.* (2014) which detailed the displacement of precipitation associated with various meteorological system activities, such as: ICTZ, cold fronts, etc. As stated in Uvo *et al.* (1998), when the ITCZ does not reach its most southern position (close to the northeast) droughts appears in the Northeast Brazil, but when the ITCZ stays longer in the south, heavy rains occur in this region. The anomalies in the ITCZ shifts are mainly produced by variations in the SST interhemispheric gradient in the Atlantic ocean.

In November, it was clear that the significant values for the increasing trends were found in the Northwest of *Sertão* and in the range between the Southern region of *Borborema* and in the North of *Agreste* region. In the month of December, the last but not least, unlike the months of October and November, there was not a continuation of the increase of trend in any region of Paraíba. However, significant increases were observed in the northwest and northeast of *Sertão*, central and South *Borborema* region and north of *Agreste*. The mixed existence of positive and negative trends, along with the differences in the results referring to the particular observed time interval, does not allow to draw conclusions of an ordinary tendency for the investigated large-scale area. Illustrated analysis describes the observational evidences of potential dynamics in cumulative, monthly precipitation, and they have been applied at the point scale to each selected gauged site.

Figure 4 shows the spatial distribution of standardized monthly trend errors from January to December. January and February noticed a higher concentration in the regions where trends showed higher significant values. This is in agreement with the proper choice of the model.

In March, the errors were mostly well distributed in the central region evidencing uniformity in *Sertão*, *Borborema* and *Agreste*. In April, one can see that errors were essentially well distributed in the region ranging from *Borborema* until *Mata*, as well as the South of *Agreste* and *Sertão*.

It is interesting to note that in May, the errors are almost homogeneously distributed throughout Paraíba, highlighting *Borborema* region. Similarly in June, it was found that the errors are almost evenly distributed in *Borborema* and *Mata* regions, the highest

concentration in the regions with the largest increases in trends.

In July, the errors are well distributed in the central region of Paraíba (much of *Sertão* and *Borborema* regions), there is a certain homogeneity in the *Mata* region. Equivalent to July, in August the errors are well distributed in central region (particularly *Borborema* region), there is also a certain homogeneity in the *Sertão* and *Mata* regions. In September, one can see almost the same characteristics as presented in the previous month.

The distribution of standard monthly trend errors for the month of October showed that the errors are mostly well distributed over much of *Sertão* region, *Borborema* and North of *Agreste* region. In November, it is observed that the errors are distributed largely in the Northwest of *Sertão* region, and in the range between *Borborema's* Southern region to North of *Agreste* region. In December, the errors are mostly well distributed largely in regions where there has been an increasing trend as well as in Southern *Agreste* region and almost the entire region of *Mata* and Coast of Paraíba.

Somehow the density can tell where clusters in our data exist, but not if clusters are statistically significant. Figure 5 shows hot spots analysis for rainfall trends in Paraíba. According to (ESRI) (2014), this tool identifies statistically significant spatial clusters of high values (hot spots) and low values (cold spots). It automatically aggregates incident data (e.g precipitation, deforestation and so on), identifies an appropriate scale of analysis, and corrects for both multiple testing and spatial dependence. This tool interrogates data in order to determine settings that will produce optimal hot spot analysis results. Hot spot analysis uses vectors (not rasters) to identify the locations of statistically significant hot spots and cold spots in data. A high Z-score and a small P value for a feature (in our study rainfall trends) indicate a significant hot spot. A low negative Z-score and small P value indicate a significant cold spot. The higher (or lower) the Z-score, the more intense the clustering. A Z-score near zero means no spatial clustering.

In January, hot spots were found in the North of *Agreste*, specifically for *Casserenge, Araruna, Areia, Caiçara, Bananeiras, Araçagi* precipitation stations (see Tabel 1 and Figure 2 for more detail), all values showed 99% confidence. There was no evidence of cold spots, that is, it means no decreasing trends for this month in the period studied. In February, the hot spots were between *Borborema* and *Agreste*, that is, there was a displacement as compared with the previous month. *Serra Branca* and *Areia* showed 95% confidence. On the other hand, *Imaculada, Jurú* and *Água Branca* showed cold spots a level of 95% confidence, this fact is related to the low tendency of rainfall in the *Sertão*.

In March, hot spots are concentrated between *Agreste* and *Borborema*, especially *São*

José dos Cordeiros, Serra Branca, Gurjão, Olivedos, Soledade which showed a level of 99% confidence. The cold spots moved from the South to the North of *Sertão*, especially for *Belém do Brejo da Cruz, Catolé do Rocha, Jericó, Riacho dos Cavalos* which showed a level 99% of confidence. The following month there was a reversal of hot and cold spots. Northern and West of *Sertão* showed the hot spots values, highlighting *Brejo da Cruz, Cajazeiras, Uiraúna* precipitation stations which showed a level of 99% confidence, this was confirmed by trend analysis. The precipitation station of *Água Branca* showed the high level of cold spots at a level of 99% confidence; this is related to the decreasing of rainfall trends in the region.

In May, the hot spots moved from *Sertão* to the center of the state; this is influenced by changes in the rainfall regime. *Taperoá, Brejo da Cruz, São Mamede* precipitation stations showed a level of 95% confidence. The totality of cold spots values is concentrated in the *Mata* region. *Caldas Brandão, Cruz do Espírito Santo, Mamanguape, Sapé* precipitation stations exhibited a level of 99% confidence. In June the hot spots values are concentrated in the *Borborema* region, especially *Caraúbas, Serra Branca, São José dos Cordeiros* which showed a level of 99% confidence. The cold spots are located in the center of the *Sertão* region and only *Olho D'Água* precipitation station showed a level of 95% confidence.

In July, the cold spots values are located in the South of *Sertão* and in the center of *Borborema* region, especially *Olivedo* and *Taperoá* precipitation stations. On the other hand, hot spots values are located in the Northwest of *Sertão*, especially *Cajazeiras* and *Uiraúna* precipitation stations and at *Serra Branca* precipitation station in the center of the *Borborema* region. In August we can observe that hot spot values are located at the borderline between *Sertão* and *Borborema* regions, this is related to rainy periods in *Borborema*; especially *Cabaceiras, Gurjão, Serra Branca* and *São José dos Cordeiros* where the precipitation stations showed the highest trend values. The cold spot values are located in the South of *Borborema* region and also concentrated in the center of *Sertão*, especially *Condado, Coremas, Olho D'água* and *Pombal* precipitation stations.

In September, the hot spots values are still in the same region of previous month but more precipitation stations showed statistically significant trends; emphasis for *Gurjão, Serra Branca, São José dos Cordeiros, Taperoá.* The cold spots moved from *Borborema* to *Mata* region, this is related to Spring season which is characterized by nonoccurrence of rainy periods specifically in the *Mata* region. In October, the hot spots values barely showed statistically significant trends, except in *Cuité* which is located in Borborema's North and *Pocinhos* which is also located in the Borborema region. The cold spots values,

as observed in the previous month, are located in the *Mata* region too, especially *Alhandra, Cruz do Espírito Santo, Caldas Brandão* and *Itabaiana*.

In November, the hot spots values scattered through *Borborema* region and also appeared in the southwest of *Sertão* region, especially *Olivedos, Serra Branca* and *Soledade* which showed a level of 99% confidence. The majority of cold spots are located in the *Mata* region, especially *Cruz do Espírito Santo, Caldas Brandão, Itabaiana, Pilar* but a modest concentration is also located in the south of *Sertão* region. In December, the hot spots values are concentrated in the *Sertão* region, this month represents the beginning of rainy season, especially *Patos, Malta* and *Vila Serrana/Desterro da Mata*. The cold spots showed a behaviour similar to the previous month but a displacement is noticeable through *Borborema* region.

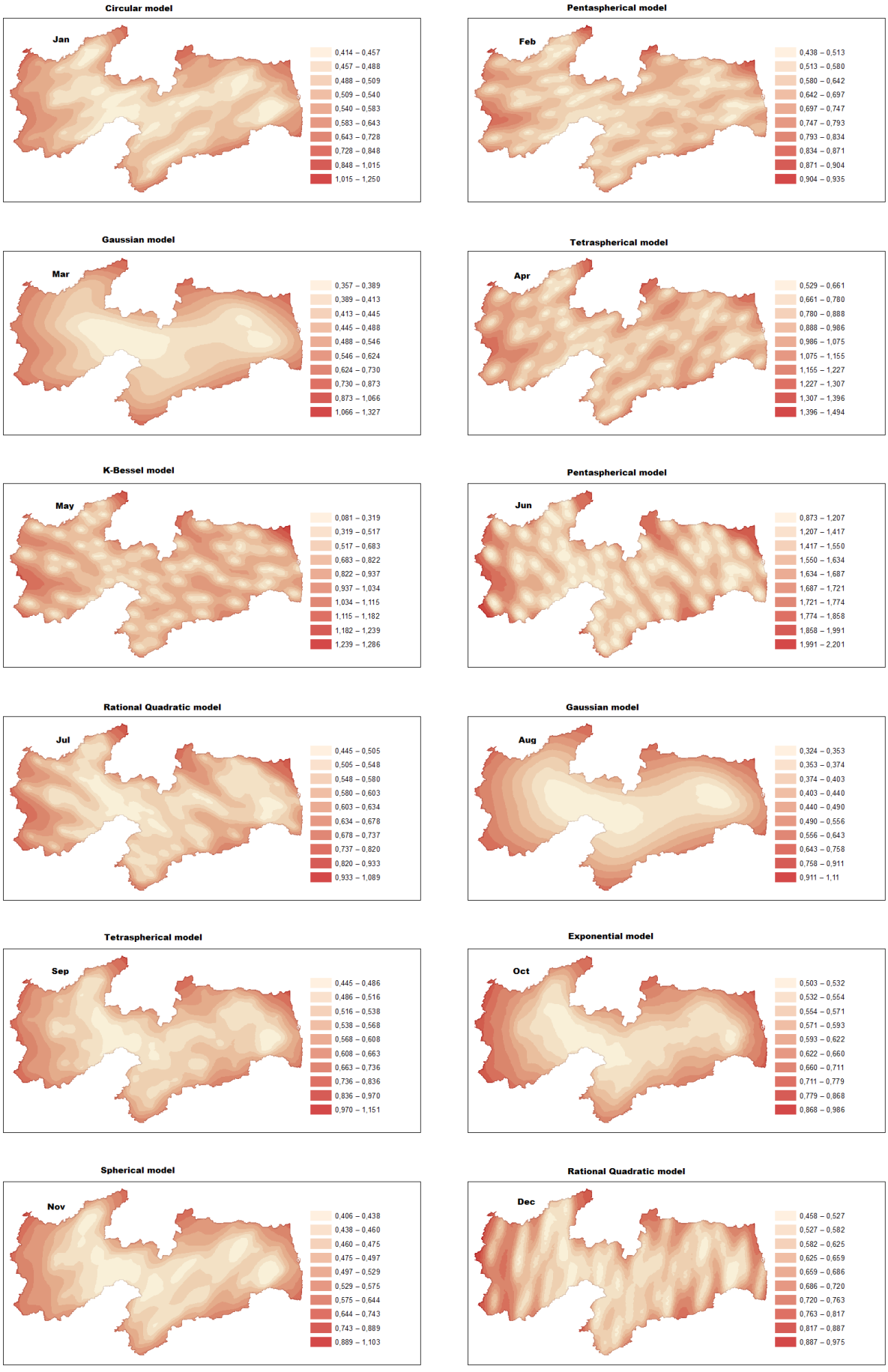


Figure 4: Spatial variation of monthly rainfall trends prediction errors for Paraíba from 1962 to 2012.

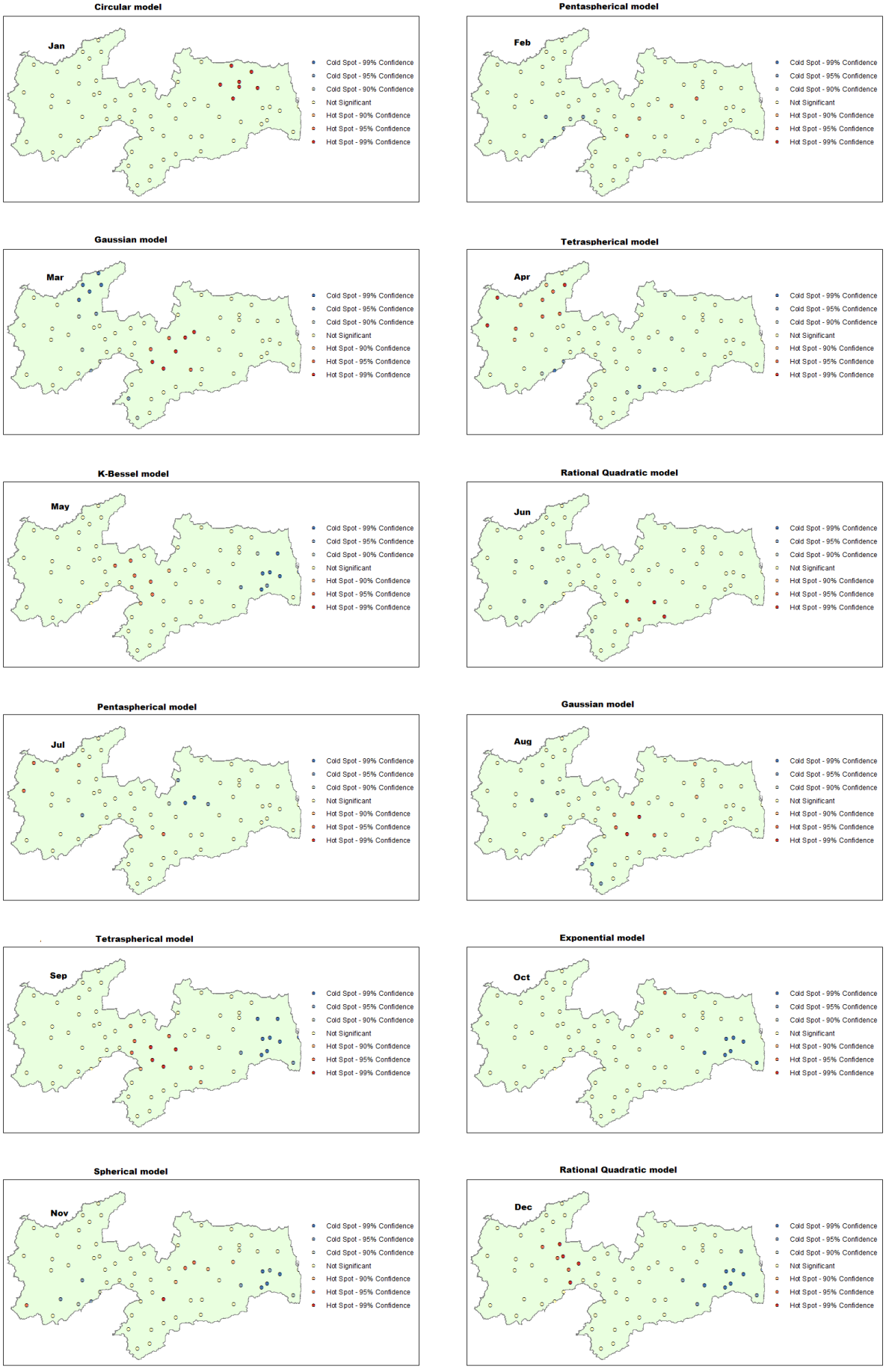


Figure 5: Hot Spots analysis for monthly rainfall trends for Paraíba from 1962 to 2012.

2.6 Conclusions

In this chapter, we analyze rainfall trends over the entire Paraíba State where significant changes in rainfall trends have occurred during all months. It is noticeable that rainfall trends vary across the state for some reasons (e.g. topography, forest cover, ITCZ, distance from coast). The impact of possible changes in rainfall trend intensity adds complexity to the implications of our results. Average rainfall trend levels may vary slightly, but if rain falls simply at one time in the seasonal calendar or too early or too late in the agricultural cycle, and then everything which is related may suffer.

For a better understanding of trends, monthly rainfall trend maps have been generated by following a methodology that includes geostatistical techniques. Geostatistical algorithms were applied to estimate rainfall trend patterns from data recorded at 69 rainfall stations distributed across the state. The hot spots analysis over all rainfall stations reveals increasing (decreasing) trends that, associated with public policies, facilitates an overview of scarcity of resources to understand the rainfall phenomenon and the low rainfall level during the drought period which affects most parts of the State.

It is worthwhile emphasizing that the trend results presented in this study were not sufficient to approve climatic change in Paraíba. Future studies are needed to address the issue of trend attribution and to attempt to establish a linkage between climatic change and observed hydrologic trends.

Chapter 3

An Application of Sample Entropy to Precipitation in Paraíba State, Brazil

3.1 Introduction

Silva *et al.* (2003) addressed the issue of spatial and temporal rainfall variability since the uncertainty and error over time constitute a significant problem in climatic studies. In tropical regions, specifically in the Brazilian Northeast, this variability increases as there are activities of various meteorological systems, such as ITCZ, the frontal systems, the east breezes and cyclonic vortices.

As a result of the occurrence of different types of water demand during different months and different places, it is necessary to study the variability of precipitation based on a concept (idea) of a general study. To this end, an entropy-based approach is deemed to be an appropriate approach to evaluate the disorder of the space-time variability of precipitation (MISHRA *et al.*, 2009).

There are many nonlinear dynamical methods utilized for measuring and understanding complex climatic systems and their associated variables. Notwithstanding, the procedures often need very long data sets that can be difficult or impossible to obtain (RICHMAN; MOORMAN, 2000). The authors drew a parallel between approximate entropy and sample entropy. The first has the capability to discern changing complexity from a relatively small amount of data that holds in a variety of contexts and is characterized to be a powerful tool for analyzing short and noisy time series. However, it is heavily dependent on the record length and consistently lower than expected for short records.

Quantifying the amount of regularity for a time series is an important task for understanding the system's behavior. One of the most prominent regularity measurements for a time series is the sample entropy ($SampEn$)[Richman & Moorman (2000)]; which when compared with the approximate entropy, is an unbiased estimator of the conditional probability that two similar sequences of m consecutive data points (m is the embedding dimension) will stand similar when one more consecutive point is included (COSTA *et al.*, 2003).

The $SampEn$ characterizes the complexity narrowly on a time scale defined by the sampling procedure which is used to acquire the time series under evaluation. In this chapter, we used $SampEn$ to evaluate the complexity or irregularity of monthly rainfall series over 69 precipitation stations in Paraíba during 1962 to 2012 (see Figure 2 and Table 1 for more details).

3.2 Methodology

The $SampEn$ computational algorithm was first published by Richman & Moorman (2000) and has been since applied by (LAKE *et al.*, 2002; SCHUANGCHENG *et al.*, 2006). $SampEn$ is defined as the natural logarithm of conditional probability of two similar sequences to m points, according to tolerance level r they remain similar to $m+1$ points (RICHMAN; MOORMAN, 2000). In the following, we provide a brief description of the calculation, as applied to the measurement of complexity for a monthly precipitation series.

Given a sequence U_N, consisting of N monthly precipitation observation data, $u(1), u(2), \ldots, u(N)$,

(i) create m-vectors $x(1), x(2), \ldots, x(N-m+1)$ in an embedding space R^m using the method of delay and $x(i)$ is defined as follows:

$$x(i) = \{u(i), u(i+1), \ldots, u(i+m-1)\}(1 \leq i \leq N-m+1) \tag{3.1}$$

wherein m represents the embedding dimension.

(ii) Designate the Euclidean distance between $x(i)$ and $x(j)$, i.e.,

$$d[x(i), x(j)] = max_{k=1,2,\ldots,m}(|u(i+k-1) - u(j+k-1)|) \tag{3.2}$$

(iii) Designate the criterion of similarity r, $r \in \mathbb{R}_+^*$. For given $x_m(i)$, count the number of j ($j = 1 \sim N - m, j \neq i$) so that $d[x(i), x(j)] < r$ is denoted as $B_i^m(r)$. It follows that, for $i = 1 \sim N - m$,

$$B^m(r) = \frac{\sum_{i=1}^{N-m} B_i^m(r)}{(N-m)} \tag{3.3}$$

(iv) Correspondingly, given $x_{m+1}(i)$, count the number of j ($j = 1 \sim N - m, j \neq i$) such that $d[x(i), x(j)] < r$ is denoted as $A_i^m(r)$. It follows that for $i = 1 \sim N - m$,

$$A^m(r) = \frac{\sum_{i=1}^{N-m} A_i^m(r)}{(N-m)} \tag{3.4}$$

(v) The *SampEn* is defined as

$$SampEn(m, r) = \lim_{N \to \infty} \left\{ -ln\left[\frac{A^m(r)}{B^m(r)}\right] \right\} \tag{3.5}$$

Indeed, N, the number of data points, is finite and the result obtained by (i)-(v) is an estimate of approximate entropy (ApEn) when the data length is N, that can be written as

$$SampEn(m, r, N) = -ln\left[\frac{A^m(r)}{B^m(r)}\right] \tag{3.6}$$

$SampEn(m, r, N)$ is the negative natural log of the conditional probability that two sequences similar within a tolerance r for m points remain similar at the next point, where N is the total number of points and self matches are not included (LAKE *et al.*, 2002). The threshold factor or filter r is an important parameter. In principle, with an infinite amount of data, it should approach zero. With a finite amount of data, or with measurement noise, the r value tipically varies between 10% and 20% of the time series standard deviation (PINCUS, 1991).

Both methods (ApEn and *SampEn*) are used to quantify the complexity of nonlinear time series. These methods were widely used in physiological processes (TANG *et al.*, 2004) and climate series (SCHUANGCHENG *et al.*, 2006). According to Richman & Moorman (2000), large *SampEn* values signify a more complex (less often) time series. The algorithm used to compute *SampEn* can be visualized in Appendix A.

3.2.1 Results and Discussion

In the previous section $SampEn(m, r, N)$ was described as an algorithm to sample entropy, wherein N is the length of time series. The two input parameters, m and r, are the keys of $SampEn$. Schuangcheng *et al.* (2006) studied daily temperature series and showed spatial patterns with variations of the input parameter r, and noticed that a gradual decrease in $SampEn$ occurred with an increase of filter r. On the other hand, Dong & Meng (2013) calculated the filter $r = kSD$, $k = 0.10 \sim 0.25$ step by 0.01. Here, $m = 2, r = kSD$, and the k values were 0.05, 0.10, 0.15, 0.20.

Figure 6 shows the values of filter r for all precipitation stations. The $SampEn$ values of all stations range from 0.7274 to 4.5326 ($N = 612, m = 2, r = 0.05$). When the value of filter is shifted to 0.10, 0.15 and 0.20, the $SampEn$ values of stations range from 0.9319 to 2.8039, from 0.7784 to 2.2616, from 0.7139 to 1.9131, respectively. The higher values of $SampEn$ are concentrated at the first 20 precipitation stations which are located in Mata and Borborema regions (see Figure 2 for more detail). These higher values are related to orography, sea breezes and differences in vegetation.

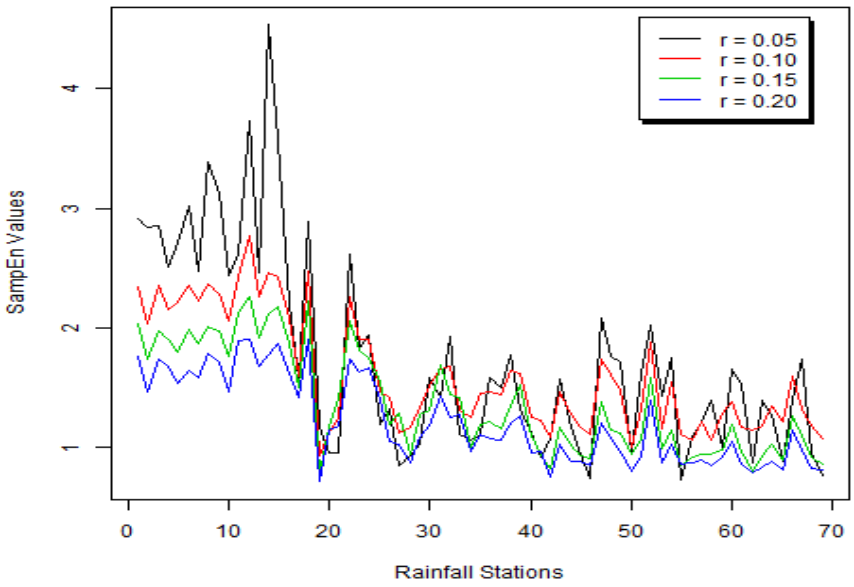


Figure 6: $SampEn$ values for filter $r = 0.05, 0.10, 0.15, 0.20$ for all 69 precipitation stations in Paraíba, Brazil.

What causes the spatial pattern of monthly precipitation series in the study area? According to Schuangcheng *et al.* (2006), the theory of nonlinear dynamics, a time series must be nonstationary if it is formed from the interaction of several nonlinear processes.

In a complex system, the areas interacted by two or more subsystems always show high nonlinearity.

SampEn of a system reflects the degree of its isolation from the surroundings. Greater signal regularity may indicate increased system isolation (PINCUS, 1994). If the hypothesis were true, *SampEn* could be a measure of the pairing between subsystems or system and external environment or factors which influence precipitation (air masses, convergence zones). The spatial patterns of *SampEn* indicate that the complexity of monthly precipitation series is high enough in areas influenced by two or more precipitation regime types than those controlled by only one precipitation regime type. In order to gain a better understanding of the effects of precipitation around Paraíba State, we analyzed *SampEn* for each specific region and its precipitation associated stations.

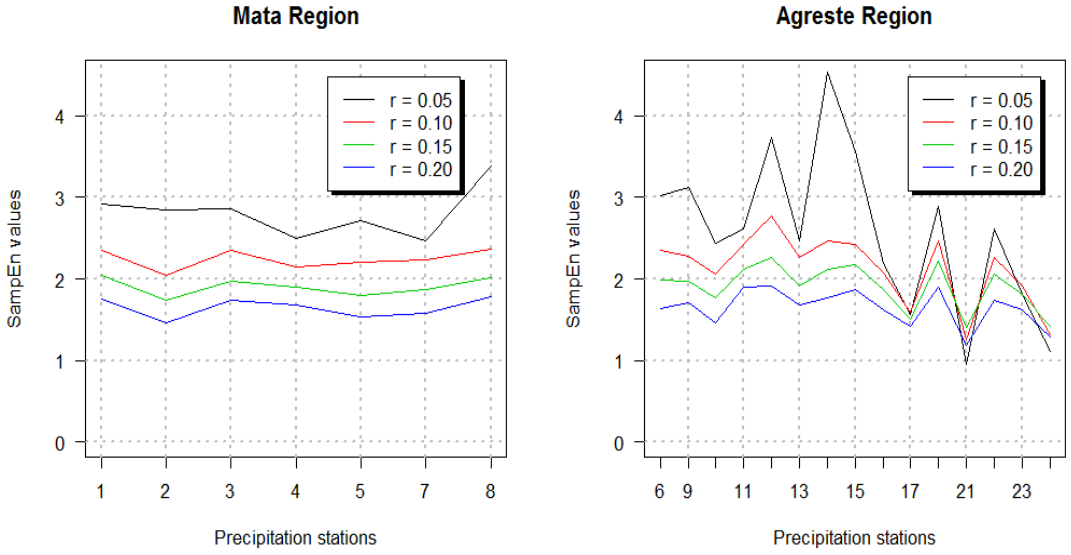
Figure 7 shows the *SampEn* values of each specific macro-region of Paraíba. On the top left is presented Mata region, this region has six precipitation stations, and the filter $r = 0.05$ has the highest *SampEn* values of the region with a maximum of 3.3911 for station 8 (Sapé). For filter $r = 0.10$, the *SampEn* values increase from 2.0374 to 2.3599 (Sapé). For filters $r = 0.15$ and $r = 0.20$ the values vary from 1.7352 to 2.0378 and 1.4642 to 1.7896, respectively. When *SampEn* values are particularly close (r=0.15 and r=0.20), they reflect that Mata region has a system of abundant precipitation, especially in the months from March to July, when winter is regular which influences the regularity of the *SampEn* values.

In Agreste region, we notice a higher range of the *SampEn* values compared with Mata region. When filter $r = 0.05$ the *SampEn* values are from 0.9561 from 4.5326 (Serraria), mean 2.5792 and standard deviation 0.9748. In general, even a noticeable increasing filter $r = 0.05$ is not able to explain the influence of *SampEn* in precipitation in the region. The filter $r = 0.10$ presents the *SampEn* values with a maximum of 2.8039 (Areias), minimum of 1.2512, mean of 2.1315, and standard deviation of 0.4417. The majority of stations have the *SampEn* values lower than the minimum values of subsequent filters r. The filters $r = 0.15$ and $r = 0.20$ have the following *SampEn* values, maximum 2.2616/1.9131, minimum 1.4106/1.1796, mean 1.9077/1.6458, standard deviation 0.2805/0.22670. The rivers in this area are almost always temporary, because they reduce their flow or dry up in times of drought. A striking factor that determines this condition is the rains that start to decline making it dry, this fact is associated with *SampEn* values because the higher the level, the stronger the complexity.

On the bottom left is Borborema region which has the same climatic conditions as

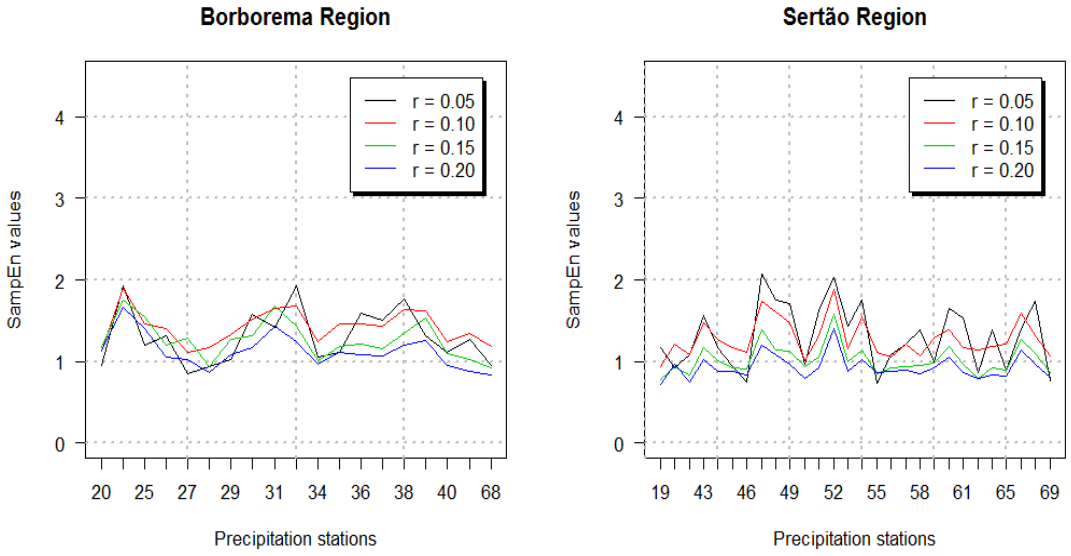
Mata region, i.e., the climate is hot dump with autumn/winter rains, but as one progresses inward the region the amount of precipitation falls whose average is around 800-1000 mm. Sousa *et al.* (2012) conducted a study using 77 precipitation station daily data for a period of 10 years. The marginal entropy values were analyzed. According to them, the rainy season has high entropy values in the middle of Borborema region. Indeed, when filter $r = 0.05$ the $SampEn$ values had a maximum of 1.9537 (Boqueirão/Boqueirão Aç.), minimum of 0.8425, mean of 1.3078 and standard deviation of 0.3381.

When filter r takes on subsequent values (0.10, 0.15 and 0.20), the $SampEn$ values have a maximum of 1.9052/1.7520/1.6654, minimum of 1.1174/0.9176/0.8327, mean of 1.4208/1.2657/1.1273 and standard deviation of 0.2144/0.2384/0.2102. These values imply that $SampEn$ shows rainfall complexity of region which is associated with prolonged dry periods. The Sertão region is the highest territory of Paraíba macro-regions. For filter $r = 0.15$ and $r = 0.20$ all the $SampEn$ values were closer, with a maximum of 1.5869/1.3957, minimum of 0.7784/0.7139, mean of 1.0188/0.9279, standard deviation of 0.1825/0.1471. When filter $r = 0.10$ the $SampEn$ values had a maximum of 1.8817, minimum of 0.9319, mean of 1.2774 and standard deviation of 0.2320. Filter $r = 0.05$ showed high $SampEn$ values but all of them were low in comparison to other macro-regions. According to Sousa *et al.* (2012), the highest values of marginal entropy were found in the entire Sertão region. In fact, the complexity associated with rainy/dry periods is the main factor explained by the $SampEn$ values.



(a)

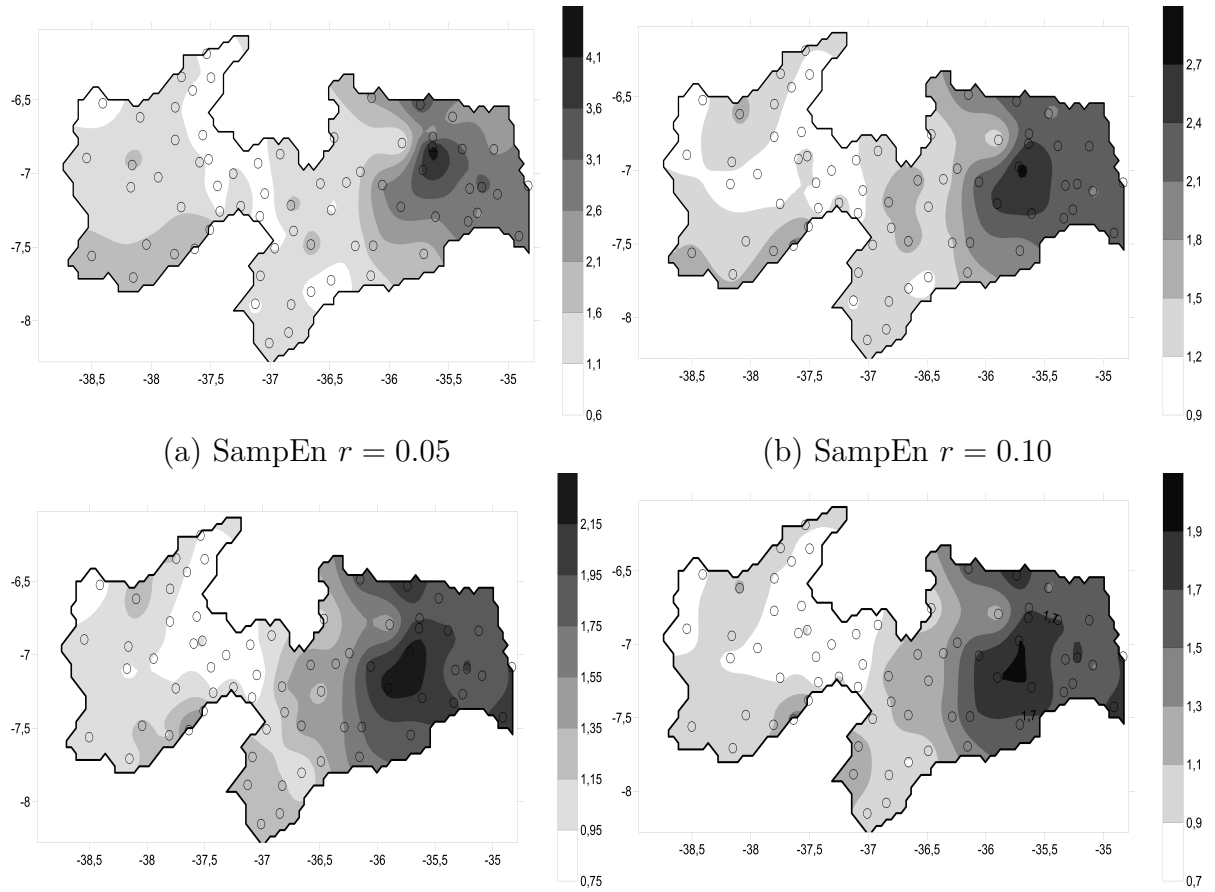
(b)



(c)

(d)

Figure 7: Comparison of *SampEn* among 4-regions in Paraíba.



(a) SampEn $r = 0.05$

(b) SampEn $r = 0.10$

(c) SampEn $r = 0.15$

(d) SampEn $r = 0.20$

Figure 8: Spatial view of *SampEn* for 69 precipitation stations in Paraíba.

From Figure 8 we can notice that when filter r=0.05, the *SampEn* values show a higher complexity in Mata region and almost all Agreste region. The lowest *SampEn* values are concentrated in Sertão region. In fact, (a) does not describe the real prediction of precipitation in Paraíba. (b) Filter $r = 0.10$ indicates higher complexity in Mata region and at least in more than 70% of Borborema region which leads to a lower prediction of precipitation. (c) and (d) both exhibit a similar behaviour in regions with higher *SampEn* values, this fact confirms the lower predictability of precipitation. The lower values of *SampEn* have more emphasis when $r = 0.20$, which is associated with regions that suffer with great dry periods.

3.2.2 Temporal trends of SampEn

In order to detect the *SampEn* differences of temporal patterns in different climate (region) divisions, we use *sliding windows*. This *sliding windows* approach computes the relevant test statistic that is capable of detecting serial dependence for the first window of a specified length, and then rolls the sample one point eliminating the first observation

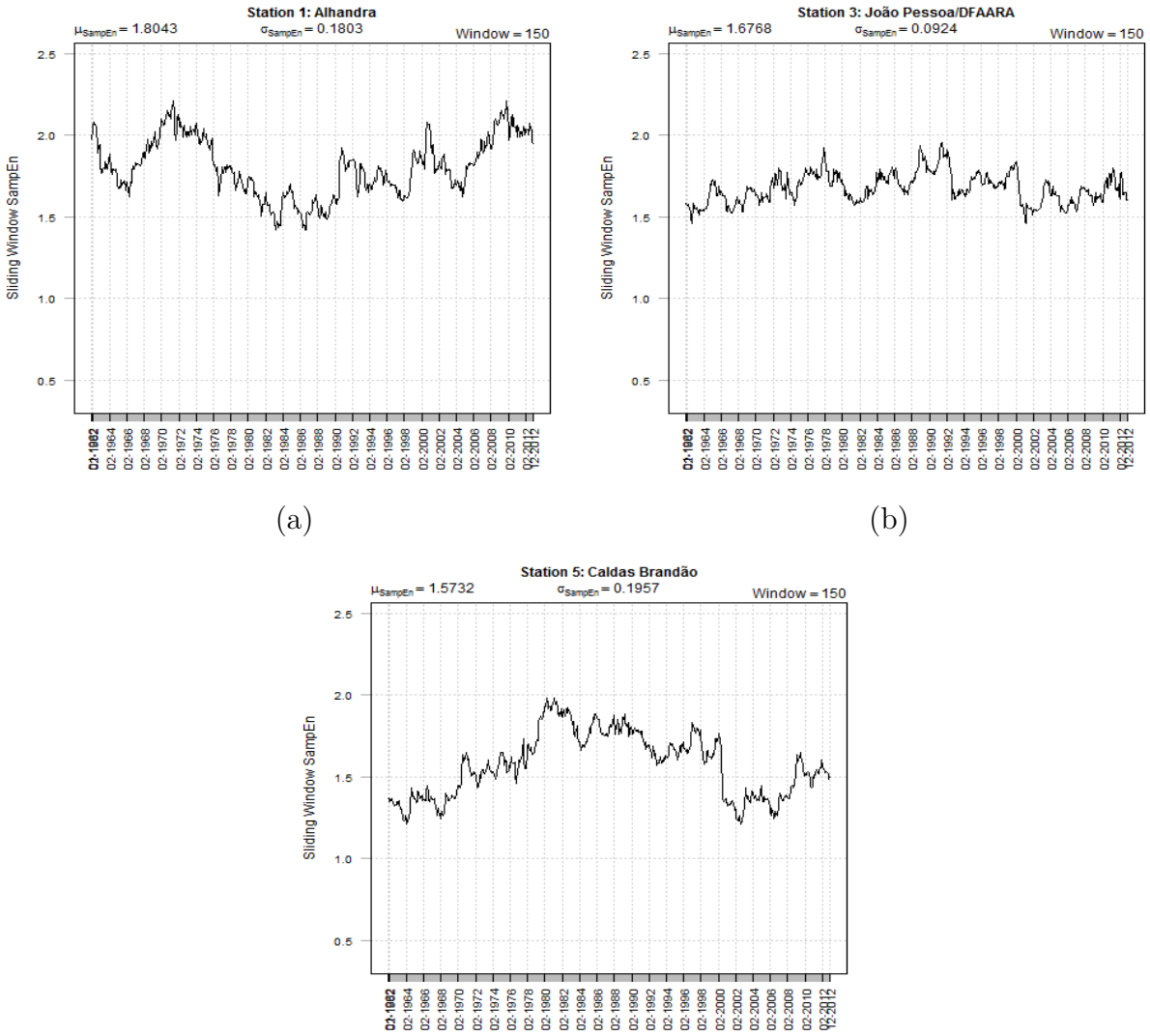
and including the next one for reestimation of the test statistic. This process continues until the last observation is used (LIM *et al.*, 2006). In this study, in a fixed-length *sliding windows* of 150 monthly length observations, the first window starts from month 1 (January 1962) and ends in month 150 (June 1974); the second window constitutes observations running from month 2 (February 1962) through month 151 (July 1974), and so on. Proportionate stratification was the method utilized to choose precipitation stations, 22 of 69 precipitation stations were randomly selected. All *SampEn rolling windows* figures are available in Appendix B.

Figure 9 shows *SampEn* for *sliding windows* of 150 months (Mata region) with a step of one month between two windows. In (a) the *SampEn* values started decreasing until 1966 but the complexity increased between 1967 and 1972 which provided less stability and less predictability. In (b) the higher complexity is concentrated between 1962 and 1978, 1982 and 1998 and from 2000, these periods showed lower precipitation predictability. In (c) the higher complexity is concentrated until 1980 which is associated with less stability and predictability. On the other hand, from 1988 more predictability and stability are shown.

The large oscillation of the amount of rain in Paraíba is due to the impact of meteorological and climate phenomena active in the region. The Mata region is the one that has the highest concentration of precipitation in Paraíba which influences the higher values of *SampEn*. According to Beserra *et al.* (2011), El Niño and La Niña influence enough in the rainy season and this variability may be influenced by the surface temperature gradient of the southern sea or inter-hemispheric (GRADM).

As stated in Sousa *et al.* (2012), the east coast of Paraíba, more specifically regions surrounding the Agreste and the Mata region's, also has the highest precipitation values, exceeding 450 mm. That is, at those locations where entropy is high, precipitation reached its highest values. Therefore, the uncertainty of precipitation is lower in the less rainy periods and locations.

Figure 10 shows *SampEn sliding windows* values for 5 precipitation stations in Agreste region. (a), (b) and (d) presented a similar behavior, that is, first they showed an increase until 1986 in (a), until 1994 in (b) and until 1985 in (d). These factors were influenced by *El Niño, La Niña*, which influenced the high complexity and affected the predictability. However, (c) and (e) showed increase and decrease through the period.



(a) (b)

(c)

Figure 9: *SampEn* sliding windows for Mata region of Paraíba State during 1962-2012.

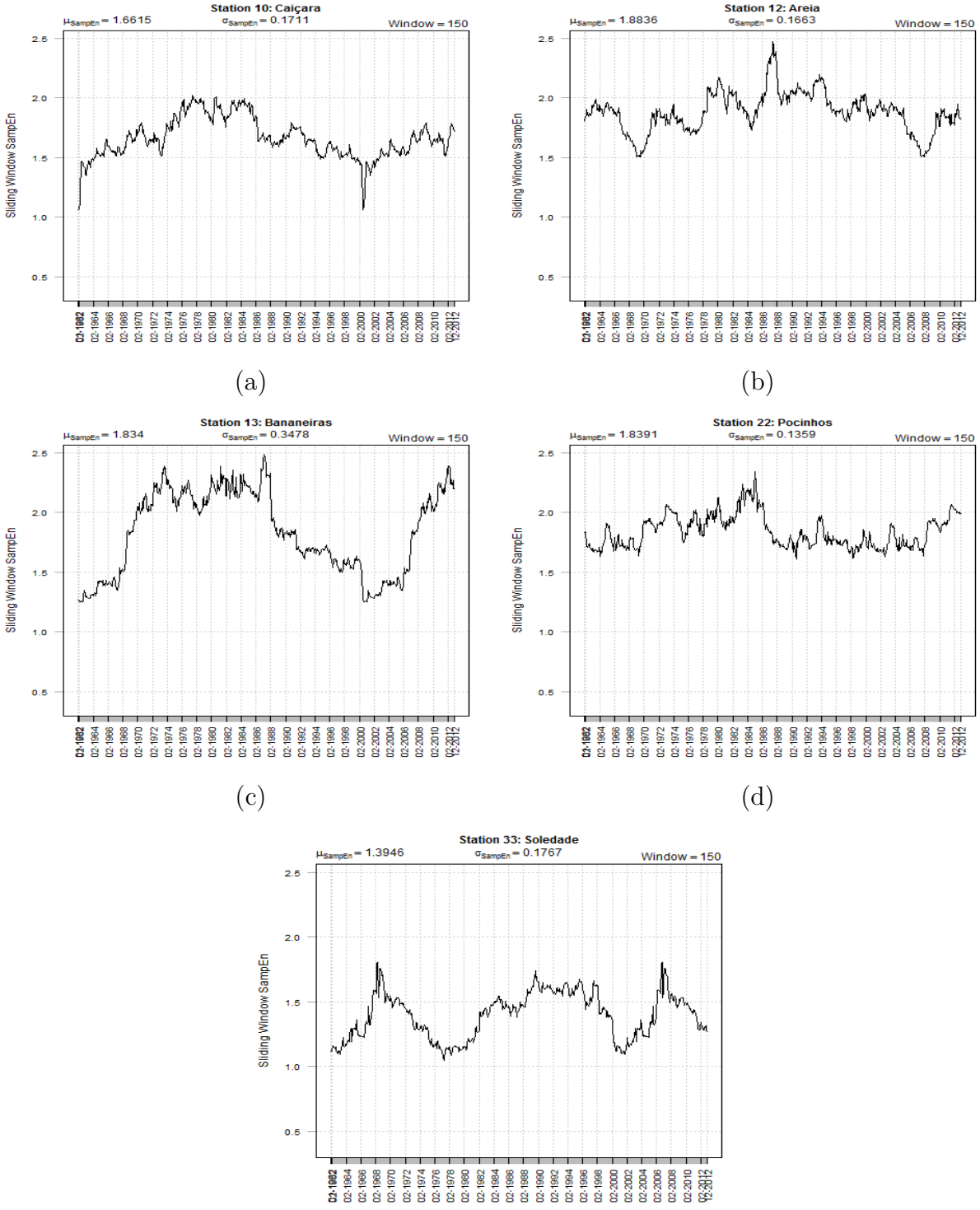


Figure 10: SampEn sliding windows for Agreste region of Paraíba State during 1962-2012.

According to Nóbrega *et al.* (2014), it can be seen that when these areas of the Atlantic sea surfaces are warmer than the climatological average, there is a reduction of such extreme levels of precipitation in Borborema region. It indicates that when the waters in the northern part of the Atlantic (TNAI) are colder, the ITCZ is shifted to the south of its dominant position, thereby influencing the precipitation regime positively in the northeast of Brazil. However, the rates of extreme precipitation events in Borborema and Agreste are more correlated with the SST (Sea Surface Temperature) in the Atlantic.

Figure 11 shows *SampEn sliding windows* for Borborema region. (a) shows an extreme decrease of *SampEn* values from 1976 to 1990. (b) and also, (d) present a decrease of *SampEn* values from 1978 to 2000 for (b), and from 1970 to 2002 for (d). (c) presents a continuous increase of *SampEn* values, except for the periods after 1997. (e) presents an increasing period, until 1984, which has the highest *SampEn* value of the region. There is another increasing period from 1986 until 2000 where there is a small decay but another increasing period starts after 2002. (f) presents a peculiar characteristic because it starts with a huge decrease of *SampEn* value and reaches the lowest value of the region but after 1982 an increase of *Sampen* values help to understand better the rainfall behavior of this station.

Sertão region features moderately low water availability. As stated by Sousa *et al.* (2012), the entropy of the annual period remained higher in municipalities with regard to the coast in comparison with Sertão locations. On the other hand, the Sertão region, which showed relatively low rainfall, exhibited the lowest entropy values, which provided high stability and high predictability of rainfall.

Figure 12 exposes four precipitation stations for Sertão region. (a), (b), (c) and (d) present decreasing *SampEn sliding windows* values, not so similar as seen before. In (a) the decrease is until 1990 when it reaches the lowest value of the region. (b) and (c) presented an alternation between increasing and decreasing periods. However, even (d) which presented the highest *SampEn* value (1.8213) concentrated for the most part of the period of decreasing *SampEn sliding windows* values, especially from 1970 until 2000. Figure 13, as well as the anterior, shows another four precipitation stations for Sertão region. (a) shows increasing *SampEn sliding windows* values, especially between 1962 and 1998. On the other hand, (b), (c) and (d) are characterized by decreasing values associated with low complexity and high prediction.

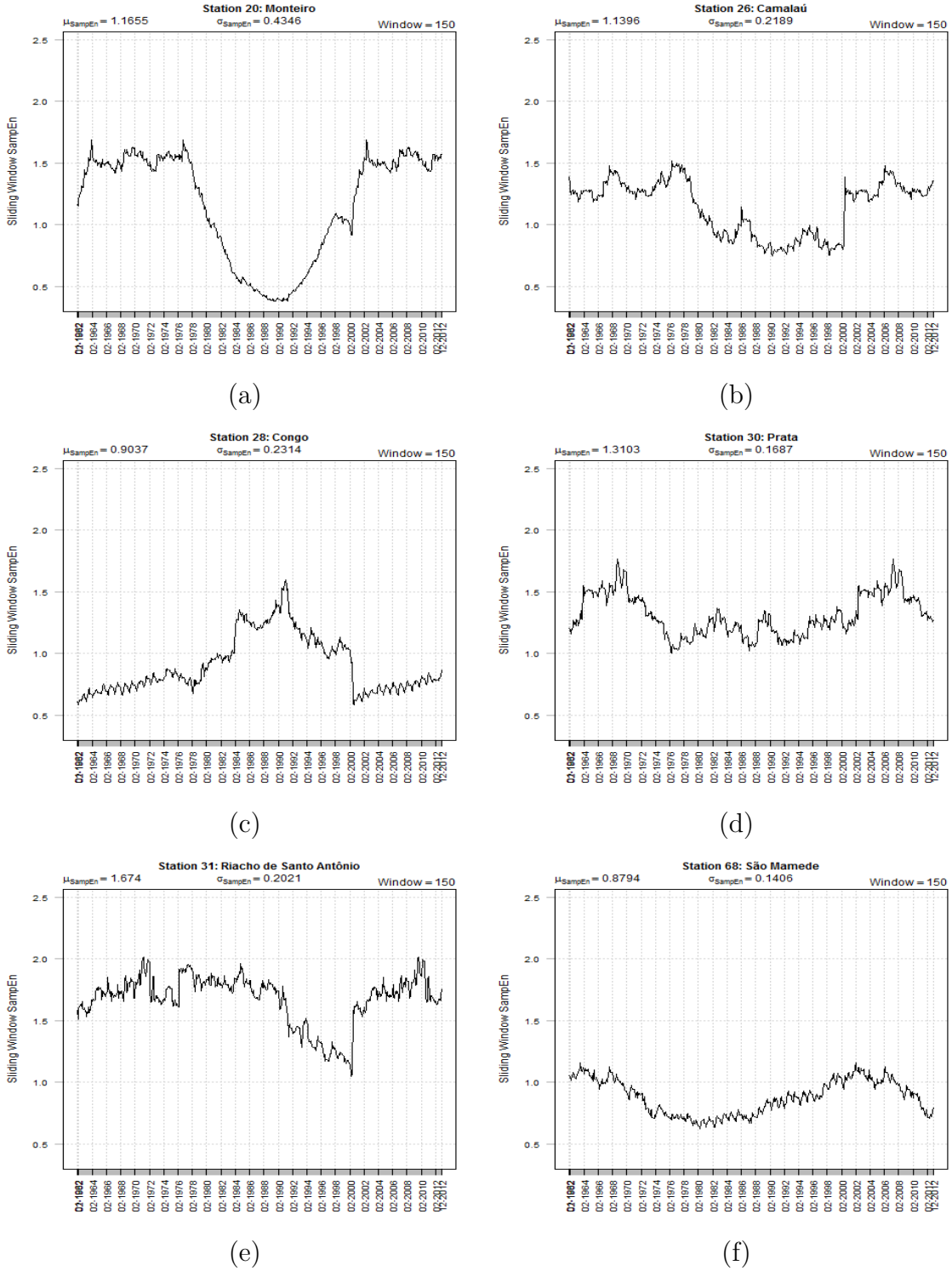


Figure 11: *SampEn* sliding windows for Borborema region of Paraíba State during 1962-2012.

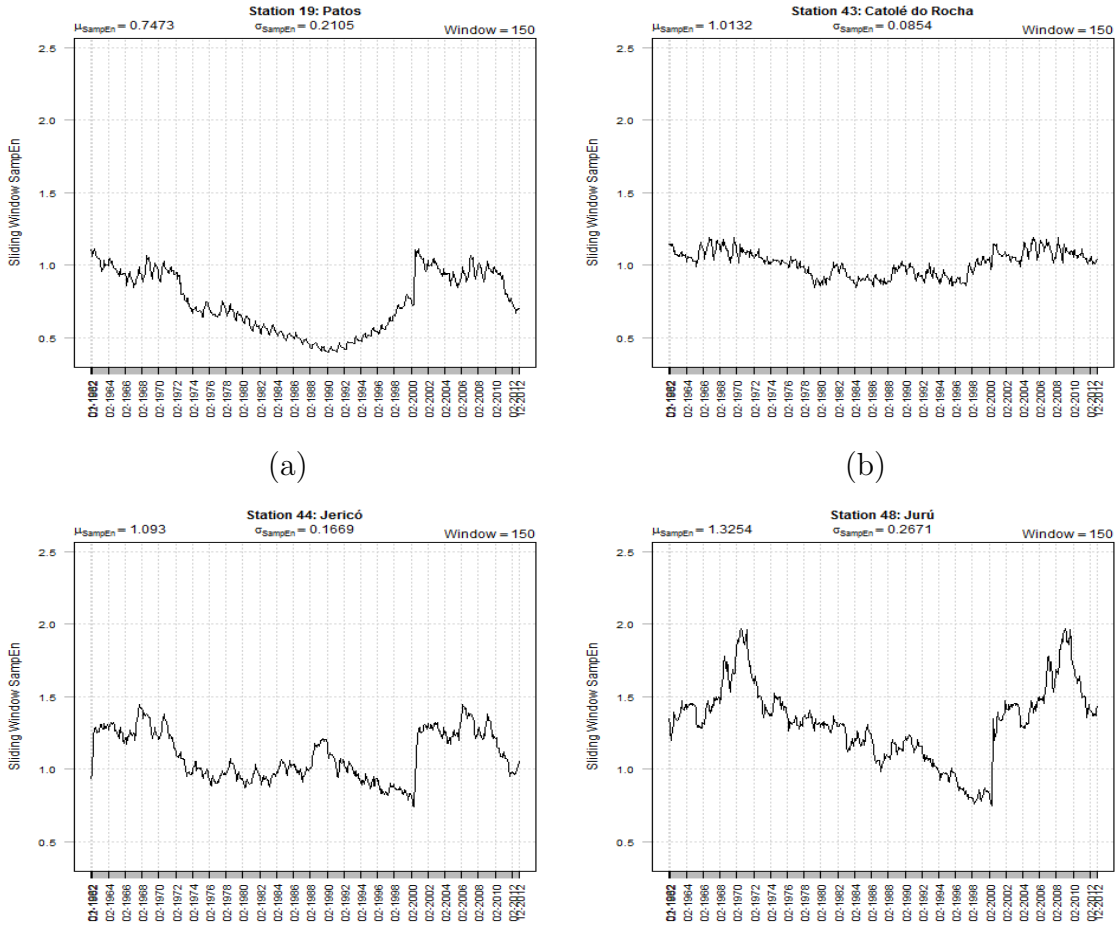


Figure 12: *SampEn* sliding windows for Sertão region of Paraíba State during 1962-2012.

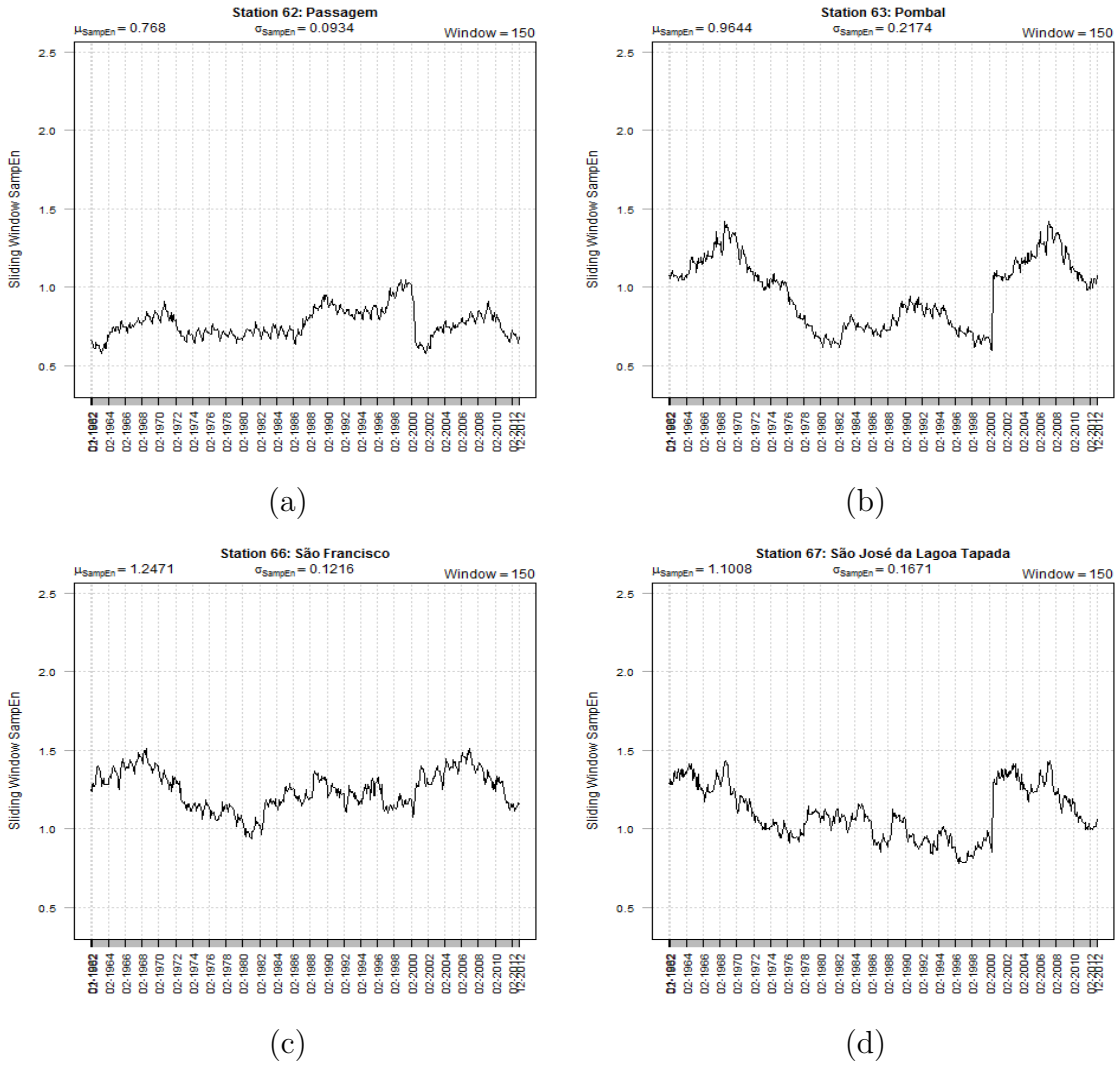


Figure 13: *SampEn* sliding windows for Sertão region of Paraíba State during 1962-2012.

3.3 Conclusions

Analysis of the complexity characteristics of precipitation is crucial to improve the efficiency of regional precipitation use and water resources planning. *SampEn* can model with less data the nonlinearity of time series. This method is introduced for measuring the complexity of monthly precipitation series of different regions of Paraíba. By calculating *SampEn* of random data using the same parameters ($N = 42229, m = 2, r = 0.05, 0.10, 0.15, 0.20$), the maximal *SampEn* values tend to approach 4.53 and minimal *SampEn* values tend to approach 0.71 for a regular series. Results of *SampEn* values of all stations indicate that precipitation series present significant complexity or irregularity.

Geostatistical analysis and spatial patterns of *SampEn* values of monthly precipitation series show significant regional differences in the study area. For filter $r = 0.15$ and $r = 0.20$ the high *SampEn* values occur in Mata region, almost all of the Borborema region. Over the micro regions of Alto-Sertão and Cariri occurrence of extreme rain and dry events are likely, hence the high variability of precipitation accumulated deviations influence the low values of *SampEn*, high prediction, and low complexity.

One cannot ignore the future trends of climate systems because of past and current climate change. Not only Paraíba, but the entire northeast region of Brazil has gone through the worst drought in 40 years. Our studies also recommend that *SampEn* is a potential tool for climate prediction and the length of the data may influence the *SampEn* values. In a nonstationary system, there is an inverse relationship between the values of *SampEn* and filter r, i.e., *SampEn* regularly decreases with a sensible increase of filter r.

In conclusion, this work testifies that entropy measurements have the potential to be an important alternative nonlinear way for analyzing precipitation series or even climate series.

Chapter 4

General Conclusions

The Paraíba state is dominated by scarcity of rainfall and water resources in the most parts of its territory; emphasis for hinterland and Sertão regions. Trends Analysis and Sample Entropy were evaluated and compared with focus on spatiotemporal variations of precipitation regimes based on monthly precipitation data covering the period of 1962-2012. In this work, the following conclusions are drawn:

1) Although the trends values presented statistically significant results practically over all four regions of Paraíba through Mann-Kendall's test evaluation; maybe it is necessary an application of Mann U Whitney test to evaluate an existence of homogeneity of the data over a particular micro-region;

2) It was noticed that transition of precipitation changes is evident in both space and time scale. Thus, it is crucial an investigation if theses changes are connected to disorders and predictability. To this purpose, the coupling of entropy with monthly rainfall data enables a relative assessment of the potential availability of water resources at specific local considering both the aggregate and the temporal variability of rainfall.

3) Sample Entropy was applied to investigate the disorder/predictability and behavior of monthly rainfall in Paraiba, but a different approach is necessary for a better knowledge about the disorder/predictability; for instance, Cross-Entropy.

4) It is ongoing an investigation using Disorder Indices (DI) such as marginal entropy (ME) and apportionment entropy (AE), which exhibit variations at different time scales, and also multivariate disorder index (MDI) to represent disorder properties of precipitation process.

5) Results of this work are of significance for planning and management of water resources and agricultural irrigation in regions characterized by long drought periods. Besides, this work can help to enhance human a better understanding of regional responses of hydrological cycle and its variations.

References

ABRAMOWITZ, M.; STEGUN, I. A. *Handbook of mathematical functions: with formulas, graphs,and mathematical tables.* [S.l.]: Courier Corporation, 1964.

ANDREOLI, R. V. *et al.* Seasonal anomalous rainfall in the central and eastern Amazon and associated anomalous oceanic and atmospheric patterns. *International Journal of Climatology*, n. 32, p. 1193–1205, 2012.

ARAUJO, R. A. F. *Morbidades em função de varáveis meteorológicas em Campina Grande, PB.* Dissertação (Mestrado) — UACA - Departamento de Meteorologia - UFCG, Campina Grande, Julho 2007.

ARIAS, D.; CABALLERO, J. *PARAÍBA STATE, BRAZIL - AGRICULTURAL SECTOR RISK ASSESSMENT.* [S.l.], November 2015. Accessed on December, 15. Disponível em: <http://www-wds.worldbank.org/external/default/WDSContentServer-/WDSP/IB/2015/11/20/090224b0831edcc3/1_0/Rendered/PDF/Risk0assessment-.pdf>.

BERKE, O. I. Estimation and Prediction in the Spatial Linear Model. *Water, Air and Soil Polution*, v. 110, p. 215–223, 1999.

BESERRA, L. D.; SERVAIN, J.; FILHO, M. F. G. Rainfall Response in Northeast Brazil from Ocean Climate Variability during the Second Half of the Twentieth Century. *Journal of Climate*, n. 24, p. 6174–6184, 2011.

BURROUGH, P. A. *Principles of geographical information systems for land resources assessment.* [S.l.]: Taylor & Francis, 1986.

CAMBARDELLA, C. A. *et al.* Field-scale variability of soil proprieties in central Iowa soils. *Soil Science Society America Journal.*, v. 58, p. 1240–1248, 1994.

CARVALHO, L. M. V.; JONES, C.; LIEBMANN, B. The South Atlantic Convergence Zone: Intensity, form, persistence, and relationship with intraseasonal to interanual activity and extreme rainfall. *Journal of Climate*, v. 17, p. 88–108, 2004.

CHEN, H. *et al.* Historical temporal trends of hydro-climatic variables and runoff response to climate variability and their relevance in water resource management in the Hanjiang basin. *Journal of hydrology*, v. 344, n. 3, p. 171–184, 2007.

CLARK, P. U.; ALLEY, R. B.; POLLARD, D. Northern Hemisphere ice-sheet influences on global climate change. *Science*, v. 286, n. 5442, p. 1104–1111, 1999.

COSTA, M. *et al.* Multiscale entropy analysis of human gait dynamics. *Physica A*, n. 330, p. 56–60, 2003.

CRASSIE, N. A. C. *Statistics for Spatial Data, revised edition.* New York: John Wiley & Sons, 1993.

CREUTIN, J. C.; OBLED, C. Objective Analysis and Mapping Techniques for Rainfall Fields: An Objective Comparison. *Water Resources Research*, v. 18, n. 2, p. 413–431, 1982.

CUNDERLIK, J. M.; BURN, D. H. Linkages between regional trends in monthly maximum flows and selected climatic variables. *Journal of Hydrologic Engineering*, v. 9, n. 4, p. 246–256, 2004.

DIAS, M. A. S. *et al.* Changes in extreme daily rainfall for São Paulo, Brazil. *Climatic Change*, v. 116, n. 3-4, p. 705–722, 2013.

DONG, L.; MENG, L. Application of Sample Entropy on Measuring Precipitation Series Complexity in Jiansanjiang Branch Bureau of China. *Nature Environment and Pollution Technology*, v. 12, n. 2, p. 249–254, 2013.

DRAPELA, K.; DRAPELOVA, I. Application of mann-kendall test and the sen's slope estimates for trend detection in deposition data from Bílý Kríz (Beskydy Mts., the Czech Republic) 1997-2010. *Beskdy Mendel University in Brno*, v. 4, n. 2, p. 133–146, 2011.

EASTERLING, D. R. *et al.* Climate extremes: observations, modeling, and impacts. *Science*, American Association for the Advancement of Science, v. 289, n. 5487, p. 2068–2074, 2000.

ESRI. *ArcGIS. Version 10.2.1.* 2014. Redlands, CA.

(ESRI), E. S. R. I. *ArcGIS Desktop Help 10.2 Geostatistical Analyst.* Redlands, CA., 2014. Disponível em: <\url http://resources.arcgis.com/en/help/main/10.2/index.ht>.

FARACO, M. A. *et al.* Selection criteria of spatial variability models used in thematical maps of soil physical attributes and soybean yield. *Revista Brasileira de Ciência do Solo*, v. 32, p. 463–476, 2008.

FILHO, W.; LUCIO, P.; SPYRIDES, M. Precipitation Extreme Analysis over the Brazilian Northeast via Logistic Regression. *Atmospheric and Climate Sciences*, v. 4, n. 1, p. 53–59, 2014.

GALLAGHER, R.; APPENZELLER, T.; NORMILE, D. Beyond reductionism. *Science*, v. 284, n. 5411, p. 79–109, 1999.

GOCIC, M.; TRAJKOVIC, S. Analysis of changes in meteorological variables using Mann-Kendall and Sen's slope estimator statistical tests in Serbia. *Global and Planetary Change*, v. 100, p. 172–182, 2013.

GOMES, O. M.; MOREIRA, G. R.; OLINDA R. A.; SANTOS, C. A. C. Trend surface analysis applied to rainfall data from the state of Paraíba. *Rev. Bras. Biom.*, v. 32, n. 3, p. 412–429, 2014.

GOOVAERTS, P. *Geostatistics for natural resources evaluation.* New York: Oxford University Press, 1997.

GOOVAERTS, P. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of hydrology*, n. 228, p. 113–129, 2000.

GRIMM, A. M. Interannual climate variability in south america: impacts on seasonal precipitation, extreme events, and possible effects of climate change. *Stochastic Environmental Research and Risk Assessment*, v. 25, n. 4, p. 537–554, 2011.

GRUIJTER, J. J.; MARSMAN, B. A. *Transect sampling for reliable information on mapping on units.* Wageningen, Netherlands.:: Pudoc, 1985.

HELSEL, D. R.; HIRSCH, R. M. *Statistical methods in water resources.* [S.l.]: Elsevier, 1992.

HEUVELINK, G. B. M.; MUSTERS, P.; PEBESMA, E. J. *Spatio-temporal kriging of soil water content.* Dordrecht, Netherlands.: Kluwer Academic Publishing., 1997.

ISAAKS, E. H.; SRIVASTAVA, R. M. *An introduction to applied geostatistics.* New York: Oxford University, 1989.

ISAH, A. Spatio-Temporal Modeling of Nonstationary Processes. *Journal of Science, Education and Technology*, v. 2, n. 1, p. 372–377, 2009.

JOHNSTON, K. *et al. Using ArcGIS geostatistical analyst.* [S.l.]: Esri Redlands, 2001.

JUNIOR, J. C. B.; ALMEIDA, H. A. The influence of El Niño phenomen over the rainfall regime in sousa microregion, PB. *Revista Geonorte*, v. 2, n. 4, p. 904–915, 2012.

KAWASHI, T.; MARUYAMA, T.; SINGH, V. Rainfall Entropy For Delineation Of Water Resources In Japan. *Journal of hydrology*, v. 246, p. 36, 2001.

KENDALL, M. *Rank Correlation Methods. Griffin & Co, London.* [S.l.], 1975.

KENDALL, M. A.; STUART, A. *The advanced theory of statistics.* [S.l.]: Charles Grifin, 1967.

KOVADIO, Y. K. *et al.* Heavy Rainfall Episodes in the Eastern Northeast Brazil Linked to Large-Scale Ocean-Atmosphere Conditions in the Tropical Atlantic. *Advances in Meteorology*, v. 2012, p. 1–16, 2012.

KRSTANOVIC, P.; SINGH, V. Transfer of information in monthly rainfall series of San Jose, California. In: *Entropy and Energy Dissipation in Water Resources.* [S.l.]: Springer, 1992. p. 155–173.

LAKE, D. E.; RICHMAN, J. S.; GRIFFIN, M. P. Sample entropy analysis of neonatal heart rate variability. *American Journal of Physiology-Regulatory Integrative and Comparative Physiology*, n. 283, p. 789–797, 2002.

LEBEL, T. *et al.* On the Accuracy of Areal Rainfall Estimation: A Case Study. *Water Resources Research*, v. 23, p. 2123–2134, 1987.

LIEBMANN, B. *et al.* Mechanisms associated with large daily rainfall events in Northeast Brazil. *Journal of Climate*, v. 24, n. 2, p. 376–396, 2011.

LIM, K. P.; BROOKS, R. D.; HINICH, M. Testing the assertion that emerging Asian stock markets are becoming more efficient. *SSRN Working Paper Series*, 2006. Disponível em: <http://ssrn.com/abstract=906515>.

LIU, B. *et al.* Entropy-based assessment and zoning of rainfall distribution. *Journal of hydrology*, v. 490, p. 32–40, 2013.

LY, S.; CHARLES, C.; DEGRÉ, A. Geostatistical interpolation of daily rainfall at catchment scale: the use of several variogram models in the ourthe and ambleve catchments, Belgium. *Hydrology and Earth System Sciences*, v. 15, n. 7, p. 2259–2274, 2011.

MACEDO, M. J. H. *et al.* Analysis of the standardized precipitation index for the Paraíba state, Brazil. *Revista Ambiente & Água*, v. 5, n. 1, p. 204–212, 2010.

MANN, H. B. Nonparametric tests against trend. *Econometrica: Journal of the Econometric Society*, p. 245–259, 1945.

MARUYAMA, T.; KAWACHI, T. Evaluation of rainfall characteristics using entropy. *Journal of Rainwater Catchment Systems*, v. 4, n. 1, p. 7–10, 1998.

MARUYAMA, T.; KAWACHI, T.; SINGH, V. P. Entropy-based assessment and clustering of potential water resources availability. *Journal of hydrology*, v. 309, n. 1, p. 104–113, 2005.

MATHERON, G. Principles of Geo-Statistics. *Economic Geology*, v. 58, p. 1245–1266, 1963.

MILLY, P. *et al.* Increasing risk of great floods in a changing climate. *Nature*, v. 415, n. 6871, p. 514–517, 2002.

MINETTI, J. L. Trends and jumps in the annual precipitation in South America, south of the 15 degrees S., 11,. *Atmósfera*, v. 11, p. 205–221, 1998.

MINETTI, J. L. *et al.* Non-linear trends and low frequency oscillations in annual pre-cipitation over Argentina and Chile, 1931-1999. *Atmósfera*, v. 16, n. 2, p. 119–135, 2003.

MIRÁS-AVALOS, J. M. *et al.* Mapping monthly rainfall data in Galicia (NW Spain) using inverse distances and geostatistical methods. *Advances in Geosciences*, n. 10, p. 51–57, 2007.

MISHRA, A. K.; ÖZGER, M.; SINGH, V. P. An entropy-based investigation into the variability of precipitation. *Journal of hydrology*, v. 370, n. 1, p. 139–154, 2009.

MONDAL, A.; KUNDU, S.; MUKHOPADHYAY, A. Rainfall trend analysis by Mann-Kendall test: A case study of north-eastern part of Cuttack district, Orissa. *International Journal of Geology, Earth and Environmental Sciences*, v. 2, n. 1, p. 70–78, 2012.

NÓBREGA, J. N. d. *et al.* Extreme precipitation events in the mesoregions of Paraíba and its relationship with the tropical oceans SST. *Revista Brasileira de Meteorologia*, v. 29, n. 2, p. 197–208, 2014.

NOVOTNY, E. V.; STEFAN, H. G. Stream flow in Minnesota: indicator of climate change. *Journal of Hydrology*, v. 334, p. 319–333, 2007.

OCHOA-SÁNCHEZ, A. *et al.* Evaluation of TRMM 3B42 precipitation estimates and WRF retrospective precipitation simulation over the Pacific-Andean region of Ecuador and Peru. *Hydrology and Earth System Sciences & Discussions*, v. 11, p. 411–449, 2014.

ÖNÖZ, B.; BAYAZIT, M. The power of statistical tests for trend detection. *Turkish Journal of Engineering and Environmental Sciences*, v. 27, n. 4, p. 247–251, 2003.

OWOLAWI, P. A.; AFULLO, T. J. Rainfall rate modeling and worst month statistics for millimetric line-of-sight radio links in South Africa. *Radio Science*, v. 42, n. 6, p. 1–11, 2007.

PINCUS, S. M. Approximate entropy as a measure of system Complexity. *Proceedings of the National Academy of Sciences*, n. 88, p. 2297–2301., 1991.

PINCUS, S. M. Greater signal regularity may indicate increased system isolation. *Math*, n. 122, p. 161–181, 1994.

PINHEIRO, A.; GRACIANO, R. L. G. A.; SEVERO, D. L. A.-s. Tendência das séries temporais de precipitação da região sul do Brasil. *Revista Brasileira de Meteorologia*, v. 28, p. 281 – 290, 09 2013.

RAJAGOPAL, A.; TEITLER, S.; SINGH, V. P. Some new perspectives on maximum entropy techniques in water resources research. In: *Hydrologic Frequency Modeling*. [S.l.]: Springer, 1987. p. 347–365.

RAJSEKHAR, D.; SINGH, V. P.; MISHRA, A. K. Multivariate drought index: An information theory based approach for integrated drought assessment. *Journal of hydrology*, v. 526, p. 164–182, 2015.

RAO, A. R.; HAMED, K. H.; CHEN, H. L. *Nonstationarities in Hydrologic and Environmental Time Series.* [S.l.]: Kluwer Academic Publishers, 2003.

RICHMAN, S. J.; MOORMAN, J. R. Physiological time series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology*, v. 278, n. 6, p. H2039–H2049, 2000.

RODRIGUES, R. R. *et al.* The Impacts of Inter El Niño Variability on the Tropical Atlantic and Northeast Brazil Climate. *Journal of Climate*, v. 24, p. 3402–3422, 2011.

SANA, R. S. *et al.* Spatial variability of physical-chemical attributes of soil and its effects on cotton yield. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v. 18, n. 10, p. 994–1002, 2014.

SANTOS, D. *et al.* Spatial variability of physical attributes of a distroferric Red Latosol after soybean crop. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v. 16, n. 8, p. 843–848, 2012.

SAYEMUZZAMAN, M.; JHA, M. K. Seasonal and annual precipitation time series trend analysis in North Carolina, United States. *Atmospheric Research*, n. 137, p. 183–194, 2014.

SCHUANGCHENG, L. *et al.* Mesaruement of climate Complexity using sample Entropy International. *Journal of Climatology*, v. 26, n. 15, p. 2131–2139, 2006.

SHAHID, S. Trends in extreme rainfall events of Bangladesh. *Theoretical and applied climatology*, v. 104, n. 3-4, p. 489–499, 2011.

SIERRA, J. P.; ARIAS, P. A.; VIEIRA, S. C. Precipitation over Northern South America and Its Seasonal Variability as Simulated by the CMIP5 Models. *Advances in Meteorology*, v. 2015, p. 1–23, 2015.

SILVA, F. D. S. *et al.* Tendência de alterações climáticas da precipitação observadas no Brasil de 1961 a 2008 utilizando dados gradeados. In: *XII Congresso Brasileiro de Meteorologia, Anais... Belém*. [S.l.: s.n.], 2010.

SILVA, M. *et al.* Complexity and predictability of daily precipitation in a semi-arid region: an application to Ceara, Brazil. *Nonlinear Processes in Geophysics*, v. 13, n. 6, p. 651–659, 2006.

SILVA, V. P. R. *et al.* Análises da precipitação pluvial no Estado da Paraíba com base na teoria da entropia. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v. 7, n. 2, p. 269–274, 2003.

SINGH, V.; FIORENTINO, M. A historical perspective of entropy applications in water resources. In: *Entropy and energy dissipation in water resources*. [S.l.]: Springer, 1992. p. 21–61.

SINGH, V. P. Hydrologic modelling using entropy. *J. Inst. Eng. Civil Engng Div*, v. 70, p. 55–60, 1989.

SOUSA, E. P. *et al.* The Entropy Theory for Analysing Rainfall in Paraíba State. *Revista Brasileira Geografica Física*, v. 5, n. 2, p. 386–399, 2012.

TABARI, H.; TALAEE, P. H. Recent trends of mean maximum and minimum air temperature in the western half of Iran. *Meteorology and Atmospheric Physics*, v. 11, p. 121–131, 2011.

TABIOS, G. Q.; SALAS, J. D. A Comparative Analysis of Techniques for Spatial Interpolation of Precipitation. *Water Resources Bulletin*, v. 21, n. 3, p. 365–380, 1985.

TAKAHASHI, K. The atmospheric circulation associated with extreme rainfall events in Piura, Peru, during the 1997–1998 and 2002 El Niño events. In: *Annales Geophysicae*. [S.l.: s.n.], 2004. v. 22, n. 11, p. 3917–3926.

TANG, X. J. *et al.* Complexity measurements of electroencephalograph recordings using sample entropy algorithm in patients with temporal lobe epilepsy. *Acta Biochimica et Biophysica Sinica*, v. 20, n. 5, p. 382–392, 2004.

TEAM, R. C. *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013*. [S.l.], 2014.

UVO, C. B. *et al.* The relationships between tropical Pacific and Atlantic SST and Northeast Brazil monthly precipitation. *Journal of Climate*, v. 11, p. 551–562, 1998.

VIOLA, M. R. *et al.* Distribution and erosive potential of rainfall in the State of Tocantins, Brazil. *Pesquisa Agropecuária Brasileira*, v. 49, n. 2, p. 125–135, 2014.

WANG, Q. X. *et al.* Change trends of temperature and precipitation in the Loess Plateau Region of China, 1961–2010. *Global and Planetary Change*, v. 92, p. 138–147, 2012.

WEBSTER, R.; OLIVER, M. A. *Geostatistics for environmental scientists.* [S.l.]: Chichester: J. Wiley, 2007.

YAMAMOTO, J. K. An alternative measure of the reliability of ordinary kriging estimates. *Mathematical Geology*, v. 32, n. 4, p. 489–509, 2000.

YANG, Z.; YAN, Y.; LIU, Q. The relationship of streamflow-precipitation-temperature in the Yellow River Basin of China during 1961-2000. *Procedia Environmental Sciences*, v. 13, p. 2336–2345, 2012.

YANMING, Z.; JUN, W.; XINHUA, W. Study on the change trend of precipitation and temperature in Kunming city based on Mann-Kendall analysis. In: *Future Computer, Communication, Control and Automation*. [S.l.]: Springer, 2012. p. 505–513.

YOUNG, M. C. F. *Considerations on the implementation of the new water resources policy in the State of Rio de Janeiro*. 2010. `http://www.ecoeco.org.br/conteudo/publicacoes/encontros/v_en/Mesa2/10.pdf`. Acessed 25 August 2015.

YUE, S.; PILON, P.; PHINNEY, B. Canadian streamflow trend detection: impacts of serial and crosscorrelation. *Hydrological Sciences Journal*, v. 48, n. 1, p. 51–63, 2003.

ZHANG, Q. *et al.* Spatio-temporal relations between temperature and precipitation regimes: Implications for temperature-induced changes in the hydrological cycle. *Global and Planetary Change*, v. 111, p. 57–76, 2013.

ZHANG, Q. *et al.* Observed trends of annual maximum water level and streamflow during past 130 years in the Yangtze River basin, China. *Journal of hydrology*, v. 324, n. 1, p. 255–265, 2006.

ZHANG, Q. *et al.* Entropy-based spatiotemporal patterns of precipitation regimes in the Huai River basin, China. *International Journal of Climatology*, p. n/a–n/a, 2015. ISSN 1097-0088. Disponível em: <http://dx.doi.org/10.1002/joc.4498>.

ZHOU, J.; LAU, K. M. Principal modes of interannual and decadal variability of summer rainfall over South America. *International Journal of Climatology*, v. 21, n. 13, p. 1623–1644, 2001.

Appendix A

Algorithm 1: SampEn Algorithm.

Input: Parameters m, r, N

Output: $SampEn$ value of the time series

// Carry out $N-m+1$ vectors of length m

for i $\leftarrow 1$ **to** $N-m+1$ **do**

  $x(i) \leftarrow \{x(i), x(i+1), \cdots, x(i+m-1)\}$;

end

// Define the distance d_m between $x(i)$ and $x(j)$ vectors $\forall i \neq j$

for k $\leftarrow 0$ **to** $m-1$ **do**

  $d_m\big[x(i), x(j)\big] \leftarrow max\Big\{|x(i+k) - x(j+k)|\Big\}$;

end

for i $\leftarrow 1$ **to** $N-m$ **do**

  $B_i^m(r) \leftarrow \dfrac{B_i}{(N-m-1)}$;

  $A_i^m(r) \leftarrow \dfrac{A_i}{(N-m)}$;

end

/* Where B_i is the number of vectors $x(j)$ similars to the vectors $x(i)$, $\forall i \neq j$, such that $d_m[x(i), x(j)] < r$ */

/* Furthermore, A_i is the number of vectors $x(j)$ of size $(m+1)$ similars to the vectors $x(i)$, $\forall i \neq j$, within the r distance */

// Compute $B^m(r)$ and $A^m(r)$

$B^m(r) \leftarrow \dfrac{1}{N-m}\left(\sum_{i=1}^{N-m} B_i^m(r)\right)$;

$A^m(r) \leftarrow \dfrac{1}{N-m}\left(\sum_{i=1}^{N-m} A_i^m(r)\right)$;

/* Finally, compute the $SampEn$ statistics */

$SampEn(m, r, N) \leftarrow -Ln\left(\dfrac{A^m(r)}{B^m(r)}\right)$;

return $SampEn(m, r, N)$;

Appendix B

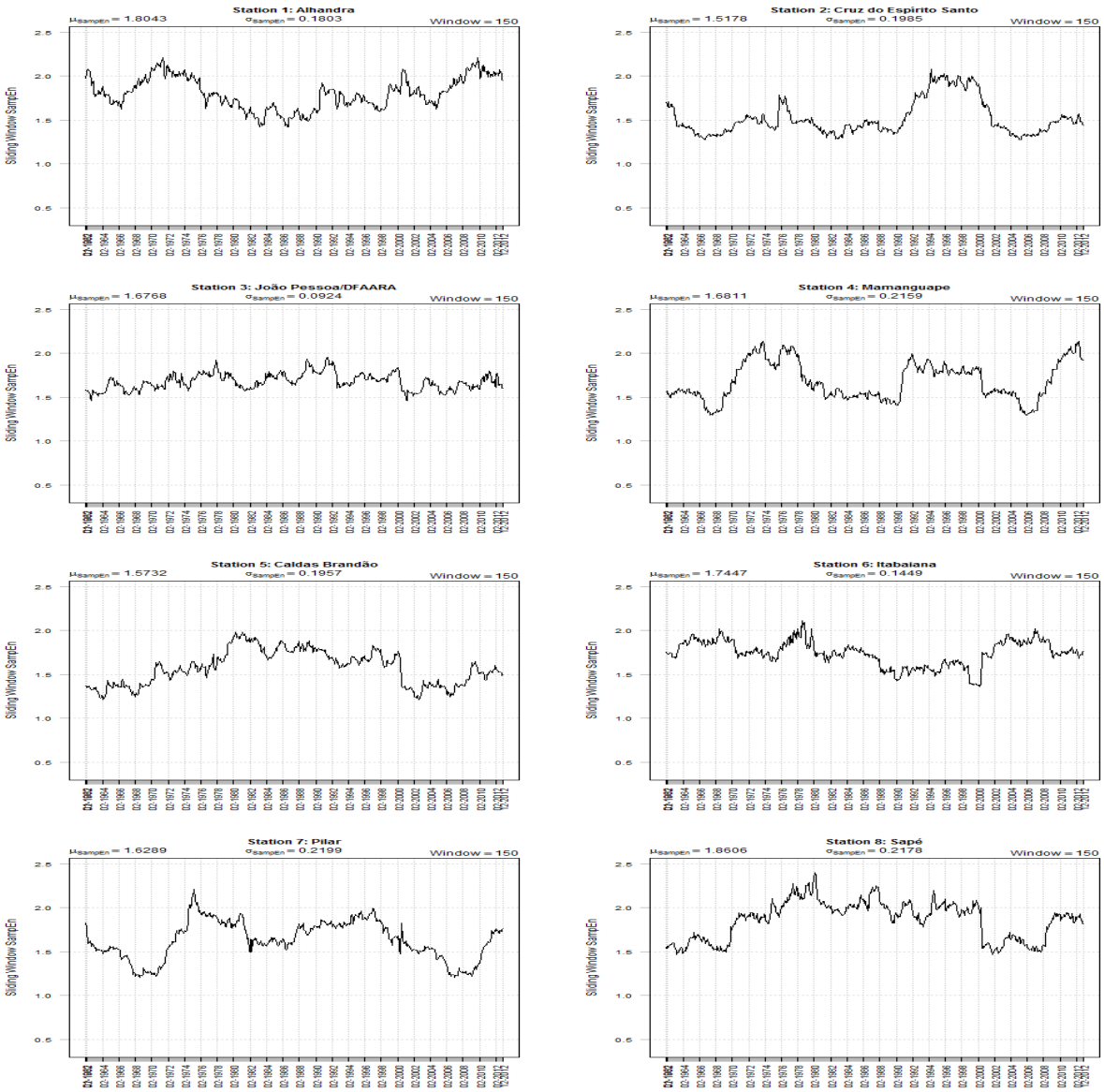


Figure 14: *SampEn* sliding windows from station 1 to 8.

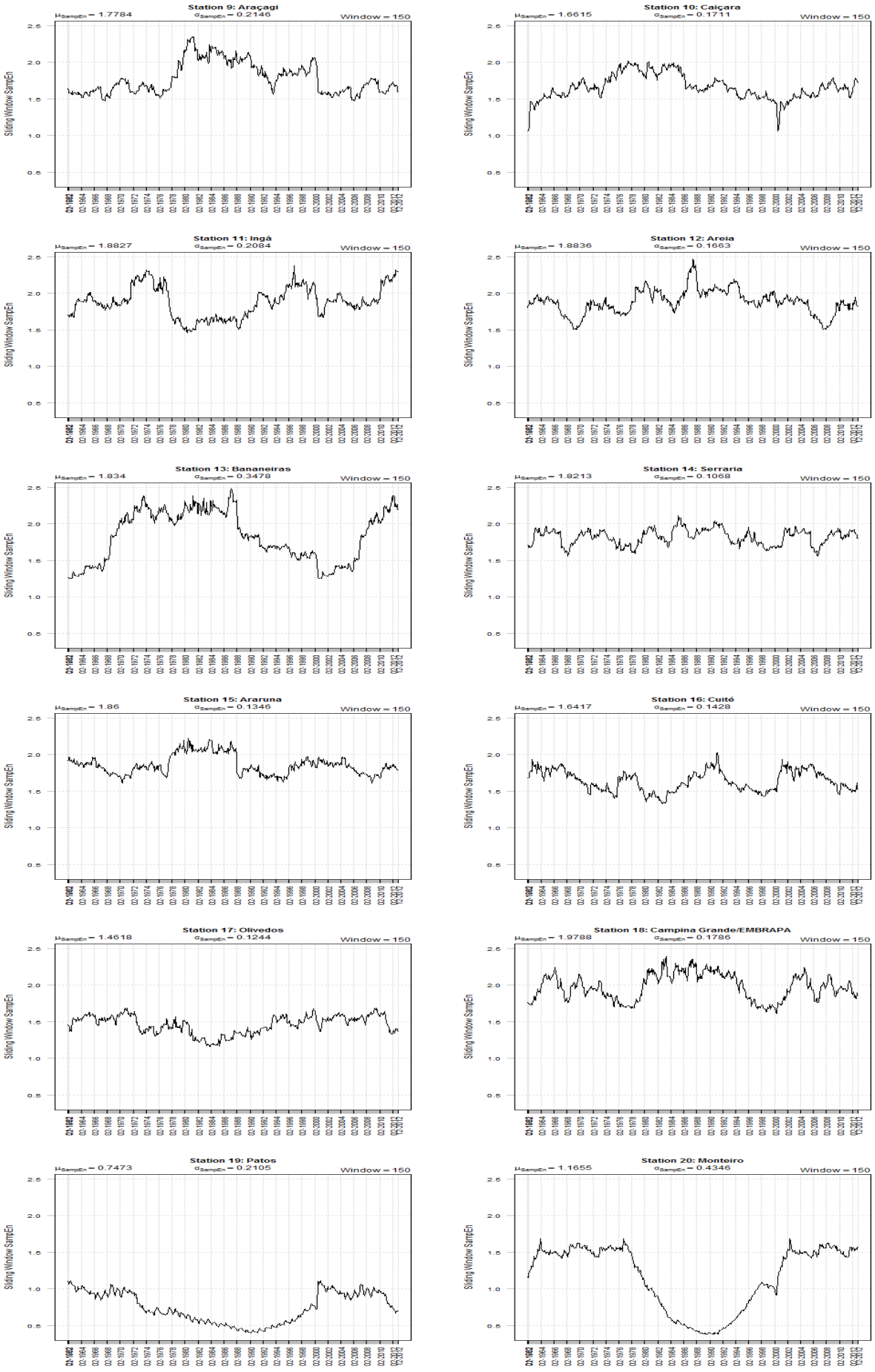


Figure 15: *SampEn* sliding windows from station 9 to 20.

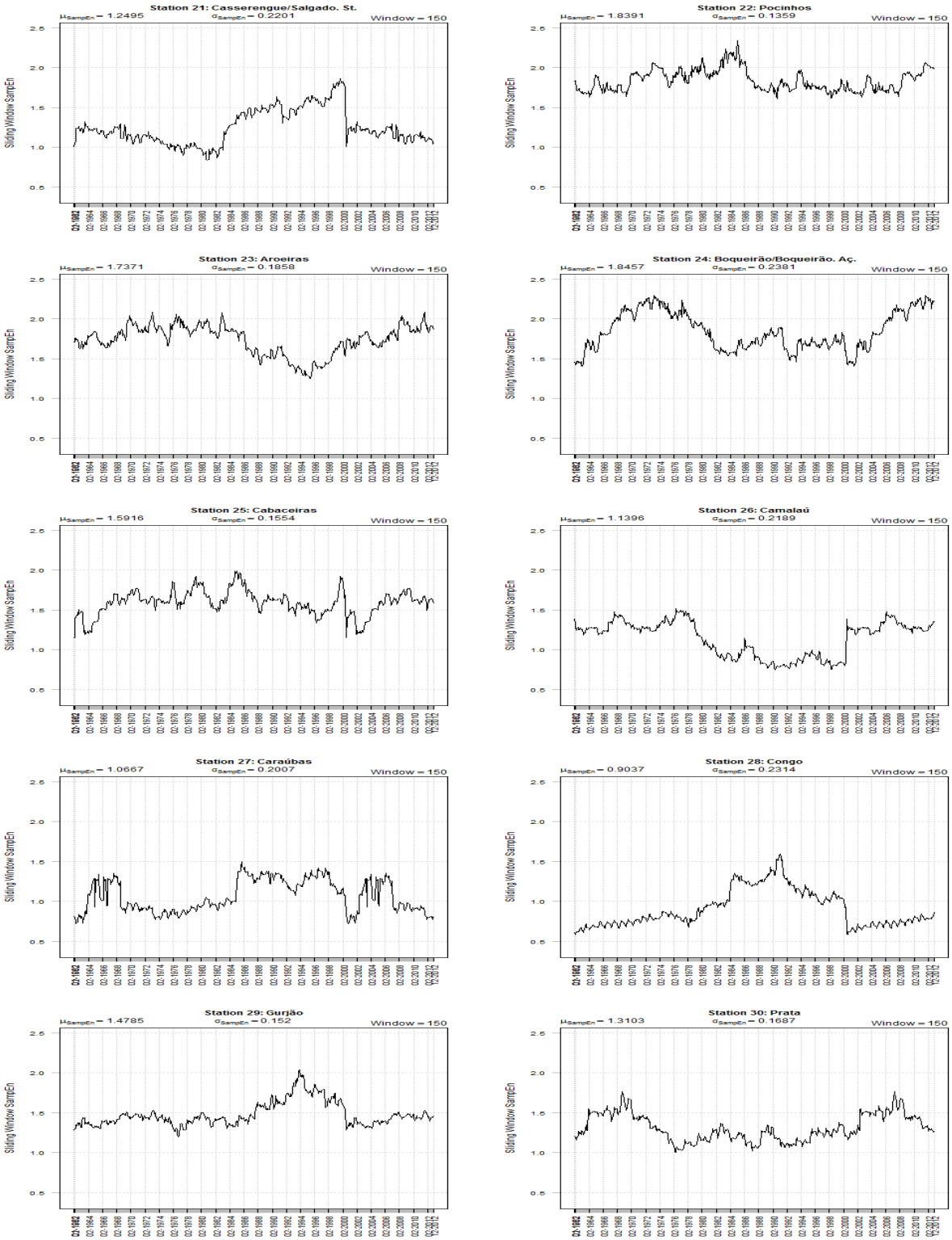


Figure 16: *SampEn* sliding windows from station 21 to 30.

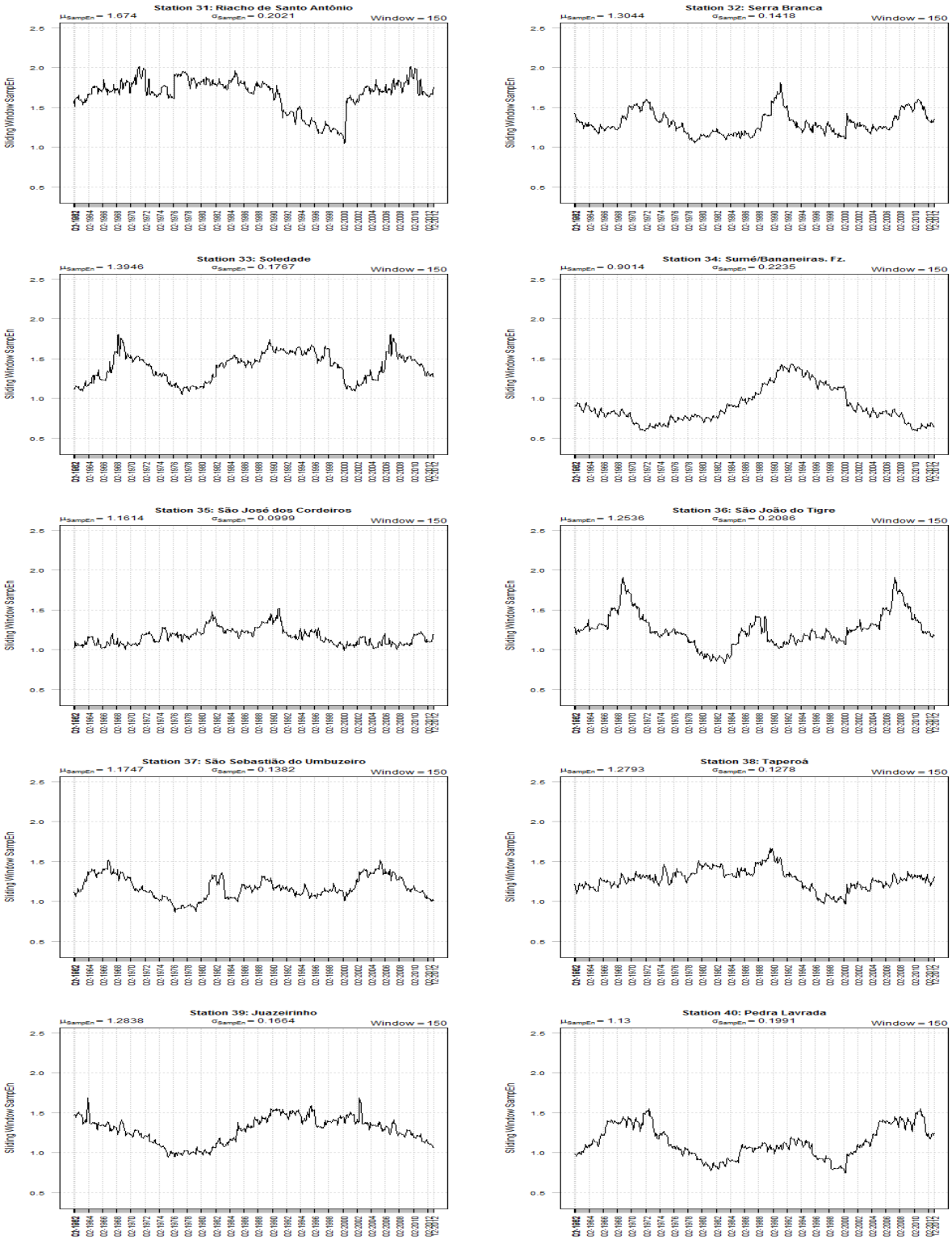


Figure 17: *SampEn* sliding windows from station 31 to 40.

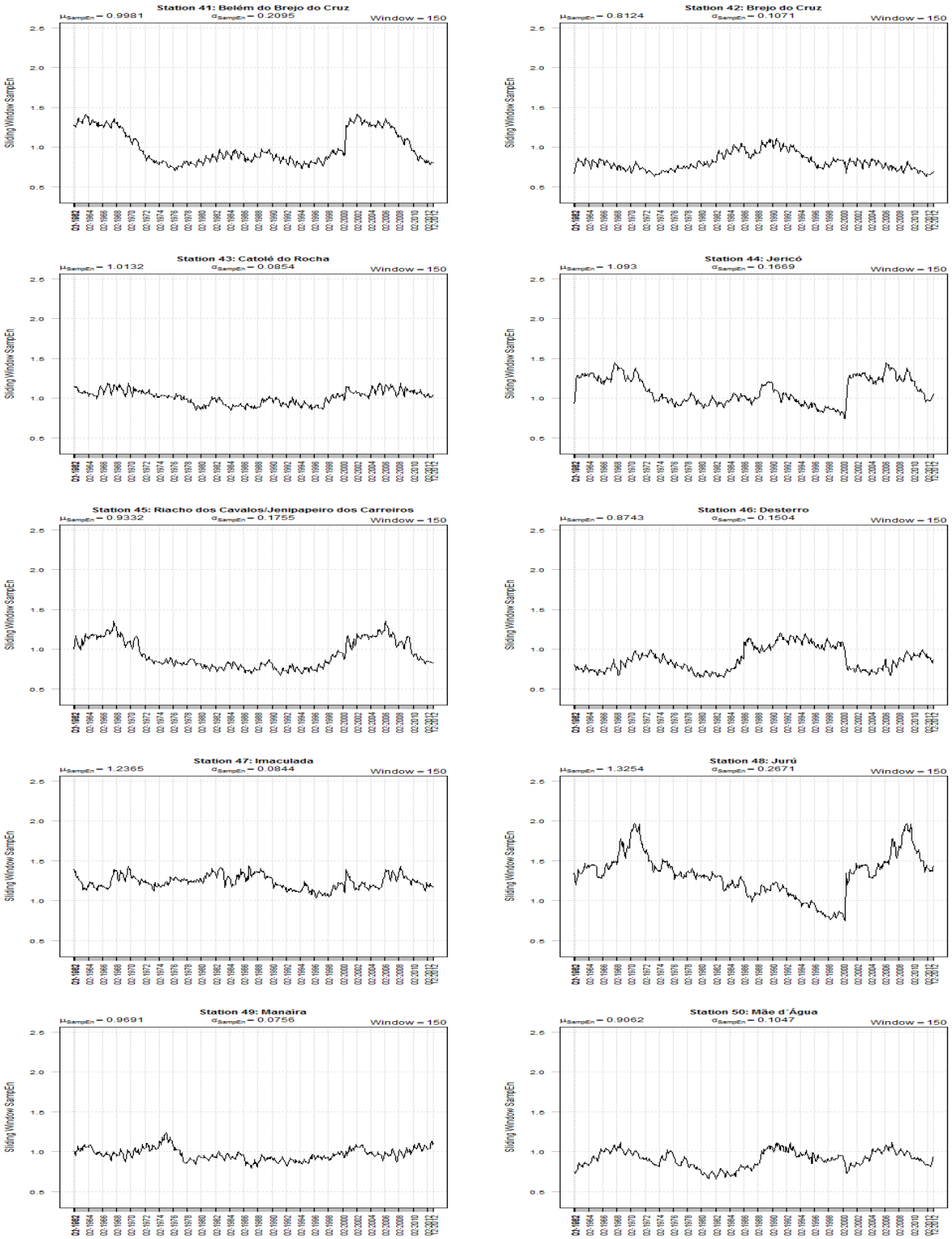


Figure 18: *SampEn* sliding windows from station 41 to 50.

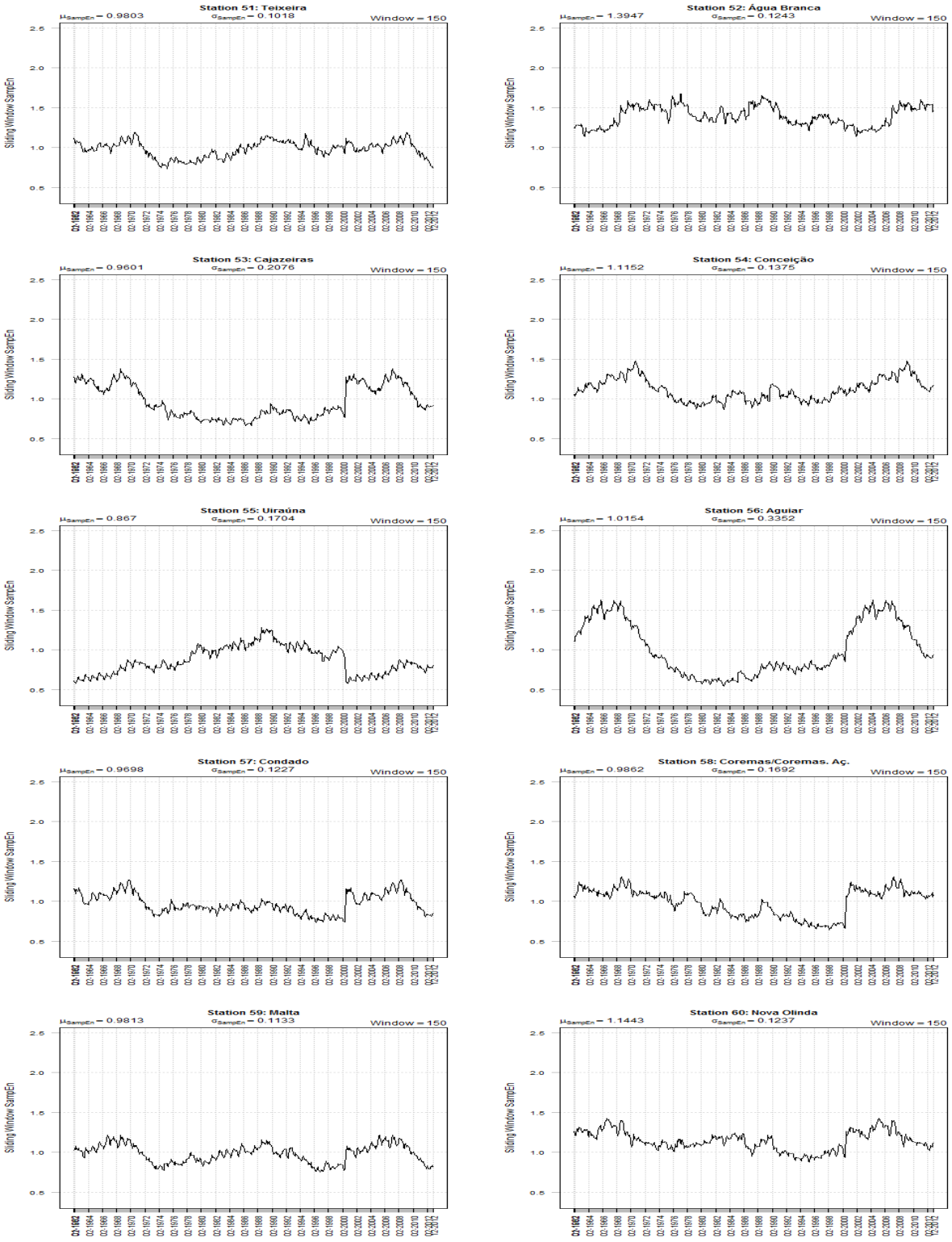


Figure 19: *SampEn* sliding windows from station 51 to 60.

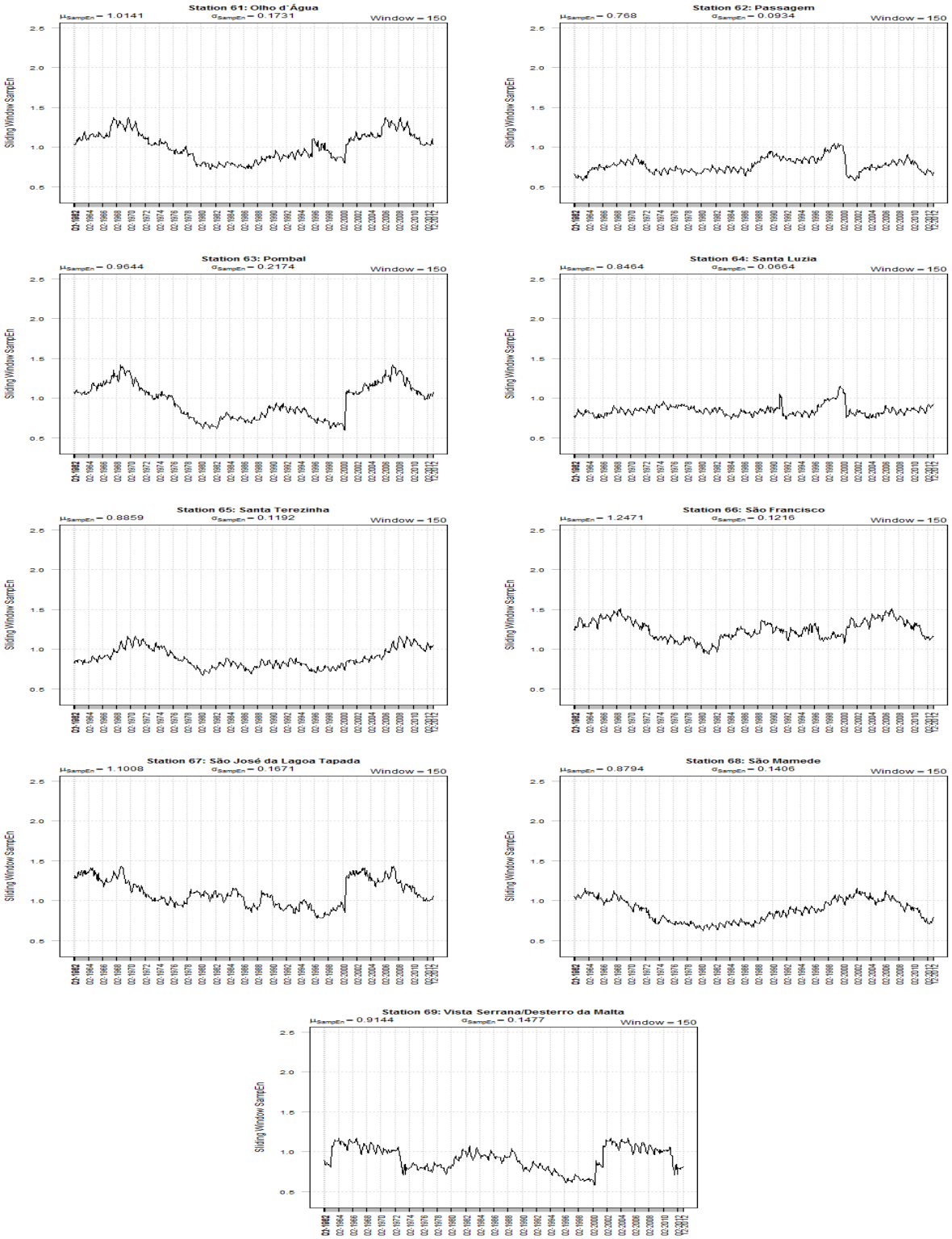


Figure 20: *SampEn* sliding windows from station 61 to 69.