UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA HIDRÁULICA E AMBIENTAL
PROGRAMA DE PÓS-GRADUAÇÃO EM RECURSOS HÍDRICOS

LARISSA ZAIRA RAFAEL ROLIM

ANALYSIS OF STATIONARITY AND LOW FREQUENCY VARIABILITY OF THE STREAMFLOW TIME SERIES FROM THE NATIONAL INTERCONNECTED SYSTEM (NIS)

FORTALEZA
2019

LARISSA ZAIRA RAFAEL ROLIM

ANALYSIS OF STATIONARITY AND LOW FREQUENCY VARIABILITY OF THE STREAMFLOW TIME SERIES FROM THE NATIONAL INTERCONNECTED SYSTEM (NIS)

Dissertação apresentada ao Programa de Pós-Graduação em Recursos Hídricos e Saneamento Ambiental da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Mestra em Engenharia Civil. Área de concentração: Recursos Hídricos.

Orientador: Prof. Dr. Francisco de Assis de Souza Filho

FORTALEZA

2019

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

R653a Rolim, Larissa Zaira Rafael.
Analysis of stationarity and low frequency variability of the streamflow time series from the national interconnected system (NIS) / Larissa Zaira Rafael Rolim. – 2019.
98 f.

Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia de Transportes, Fortaleza, 2019.
Orientação: Prof. Dr. Francisco de Assis de Souza Filho.

1. Análise de Séries Temporais. 2. Estacionariedade. 3. Modelo de Markov Escondido. I. Título.
CDD 388

LARISSA ZAIRA RAFAEL ROLIM

ANALYSIS OF STATIONARITY AND LOW FREQUENCY VARIABILITY OF THE STREAMFLOW TIME SERIES FROM THE NATIONAL INTERCONNECTED SYSTEM (NIS)

Dissertation presented to the Postgraduate Program in Water Resources and Environmental Sanitation of the Federal University of Ceará, as a partial requirement to obtain a Master's degree in Civil Engineering. Area of concentration: Water Resources.

Approved in: 08/02/2019.

EXAMINATION BOARD

_____
Prof. Dr. Francisco de Assis de Souza Filho (Orientador)
Universidade Federal do Ceará (UFC)

_____
Profa. Dra. Ticiana Marinho Carvalho Studart
Universidade Federal do Ceará (UFC)

_____
Prof. Dr. Francisco das Chagas Vasconcelos Júnior
Fundação Cearense de Meteorologia e Recursos Hídricos

ACKNOWLEDGMENTS

To my family, especially to my parents Edinaldo Ferreira Rolim and Maria Aurilene Rafael Rolim for the support, patience and encouragement throughout this process.

To my companion, Konstantinos, for the support and encouragement throughout this master's degree.

To the Northeastern Drought Monitor project with the UFC and ANA / ASTEF, for the financial support with the maintenance of the aid grant.

To my teachers, especially to Prof. Dr. Francisco de Assis de Souza Filho for the excellent orientation and the time available.

To the colleagues of the Climate Risk Management (GRC) research group for help in this work and associated articles, especially Victor, Renan, Gabriela, Taís and Renata.

To the participants of the examining board Francisco das Chagas Vasconcelos Júnior and Ticiana Marinho Carvalho Studart, thank you for the availability and time spent on reading this work.

To the other professors of PÓS-DEHA for the knowledge obtained.

RESUMO

Uma abordagem comum na modelagem de séries temporais hidrológicas é a suposição de que a série é estacionária, embora essa hipótese tenha sido questionada devido aos impactos das mudanças climáticas e usos do solo motivados por ações antropogênicas. Desta forma, o presente trabalho pretende fazer um diagnóstico das séries temporais do Sistema Interligado Nacional (SIN) no Brasil, analisando a estacionariedade, tendência e modo de variação. Em seguida, foi realizada uma análise apenas da baixa frequência (16-32 anos), decomposta pela transformada de ondeleta, utilizando o índice padronizado de escoamento (SRI), o *changepoint* e o Modelo de Markov Escondido para identificar os diferentes estados das vazões do reservatório de Sobradinho, a fim de entender os impactos da baixa frequência na série temporal completa e desenvolver um modelo de previsão. Foi utilizado um modelo autorregressivo e um HMM para prever o próximo estado da série temporal de baixa frequência. Primeiro, a estacionariedade foi analisada por testes estatísticos e testes de raiz unitária. Para avaliar a presença de tendência, foram utilizados os testes de Mann-Kendall e inclinação de Sen. Posteriormente, as séries temporais foram decompostas pela Decomposição Completa em Modos Empíricos por Conjunto com Ruído Adaptativo e Transformada de Ondeleta. Os resultados entre os testes estatísticos e de raiz unitária classificaram mais da metade das séries como não estacionárias. Foi avaliado que os testes de raiz unitária são mais indicados ao modelar séries com a presença de uma tendência. A análise com o *changepoint* mostrou um melhor resultado na identificação de eventos extremos de baixa frequência em relação ao SRI. Foi avaliada a função de distribuição cumulativa utilizando a séries temporal original e foi observado que as distribuições são diferentes para os anos classificaram os anos secos/úmidos, assim, a proposta de identificar os estados e prever o próximo estado foi justificada. Um modelo para previsão foi desenvolvido ajustando um modelo autorregressivo para a série temporal de baixa frequência, mas apresentou bons resultados apenas para previsão de 1 ano à frente. No modelo HMM, os resultados mostraram uma redução do período úmido entre os anos 2010-2016 e uma probabilidade crescente de um período normal foi verificada. A curto prazo, é possível identificar o próximo estado mais provável para o reservatório de Sobradinho usando o HMM, e este foi uma boa indicação em relação à série observada. Em conclusão, a maior parte das estações analisadas não são estacionária, embora ainda sejam modeladas desta forma, e a previsão HMM para as vazões do reservatório de Sobradinho provou ser uma ferramenta importante para auxiliar na previsão do próximo estado da variabilidade de baixa frequência, proporcionando um mecanismo para o gerenciamento e operação de reservatórios,

especialmente para o deste estudo, devido à sua importância econômica para o setor hidrelétrico.

Palavras-Chave: Análise de Séries Temporais. Estacionariedade. Modelo de Markov Escondido.

ABSTRACT

A common approach in the hydrological time series modeling is the assumption that the series is stationary, although this belief has been questioned due to the impacts of climatic changes and land use change motived by anthropogenic actions. Thus, the present work intends to make a diagnosis of streamflow time series of the National Interconnected System (NIS) in Brazil analysing the stationary, trend and mode of variation. Then, the analysis is performed for the low frequency only (16-32 years), decomposed by the wavelet transform, and use the Standard Runoff Index (SRI), changepoint and a Hidden Markov Model (HMM) to identify the different states of the Sobradinho's reservoir inflow to understand the impacts of this frequency in the time series and develop a forecast model. It was used the Autoregressive model and a HMM to predict the next state in the low frequency time series. First, the stationarity was analyzed by statistical and unit root tests. In order to evaluate the presence of trend was used the Mann-Kendall and Sen's slope tests. Then, the time series were decomposed by Complete Ensemble Empirical Mode Decomposition with Adaptive Noise and Wavelet Transform. Results between the statistical and unit root tests classified more than a half of the series as nonstationary. It was evaluated that the unit root tests are more indicated when modeling series with the presence of a trend. The analysis with the changepoint showed a better result in the identification of extreme events of the low frequency in relation to SRI. It was evaluated the cumulative distribution function using the original time series and was observed that the they have different distributions for years classified as dry/wet years, thus, the proposal to identify the states and forecast the next state was justified. A forecasting model was developed adjusting an autoregressive model to the low frequency time series, however it presented good results for 1-year ahead forecast only. Regarding the HMM model, the results shown a reduction of the wet period between the years 2010-2016 and an increasing probability of a normal period was verified. For a short-term, it is possible to identify the next most probable state for the Sobradinho's reservoir and the next state was a good indication compared to the historical time series using the HMM. In conclusion, most of the station gage analyzed are not stationary, although they are still model as such and the HMM forecast for Sobradinho's reservoir inflow proved to be an important tool to assist in predicting the next state of the low frequency variability, thus, providing an mechanism to managing and operating reservoirs, especially for the one in this study due to its economic importance for the hydroelectric sector.

Keywords: Time series analysis. Stationarity. Hidden Markov Model.

LIST OF FIGURES

Figure 1 – Systemic Approach for time series modeling constructed by Box and Jenkins (a) and Salas (b) 16
Figure 2 – National Interconnected System 25
Figure 3 – Interregional transfers of the National Interconnected System 26
Figure 4 – Spatial distribution of the Base Posts 27
Figure 5 – Methodology used in this article 28
Figure 6 – Results of statistical tests for stationarity analysis 38
Figure 7 – Results of unit root tests for stationary analysis 41
Figure 8 – Results of trend analysis 42
Figure 9 – Percentage of explained variance for the analyzed stations using wavelet transform. (a) Explained variance for the 2-4 years period; (b) Explained variance for the 4-8 years period; (c) Explained variance for the 8-16 years period; (d) Explained variance for the 16-32 years period and (e) Explained variance for the residue (higher than 32 years) 43
Figure 10 – (a) Years with significance in the 2-4 years frequency, (b) Years with significance in the 4-8 years frequency, (c) Years with significance in the 8-16 years frequency, (d) Years with significance in the 16-32 years frequency. The figure is organized by region from South (Top), then Southeast, Northeast and North (Bottom) 46
Figure 11 – Percentage of explained variance using CEEMDAN decomposition. (a) Explained variance for IMF1 (2-4 years); (b) Explained variance for the IMF2 (5-13 years); (c) Explained variance for IMF3 (13-20 years); (d) Explained variance for IMF4 (16-25 years) and (e) Explained variance for IMF5 (higher than 25 years) 48
Figure 12 – Location of the Sobradinho Reservoir 57
Figure 13 – Methodology applied in the article 58
Figure 14 – Morlet wavelet 59
Figure 15 – Example of a basic Hidden Markov Model 64
Figure 16 – Average wavelet power 66
Figure 17 – (a) Standardized Time Series; (b) Low frequency time series (16-32 years) 66
Figure 18 – SRI 12 of the streamflow time series 67
Figure 19 – Changepoint detection using the penalty value of 1 68
Figure 20 – Changepoint detection using the penalty value of 0.1 68
Figure 21 – Cumulative distribution function of the streamflow time series using the three-state classification based on the changepoint detection 70
Figure 22 – Cumulative distribution function of the streamflow time series using the two-state classification based on the changepoint detection 70
Figure 23 – Regime posterior probabilities of a two-state model 72
Figure 24 – Most probable path given by the Viterbi algorithm for the two-state model 72
Figure 25 – Cumulative distribution function of the streamflow time series using the two-state HMM 73
Figure 26 – Regime Posterior Probabilities of a three-state HMM 74
Figure 27 – Most probable path given by the Viterbi algorithm for the three-state model 74
Figure 28 – Cumulative distribution function of the streamflow time series using the three-state HMM 75
Figure 29 – State variation of the streamflow series (left) and the low frequency series (right) 75

Figure 30 – Regime posterior probabilities of a multivariate three-state HMM 77
Figure 31 – Most probable path given by the Viterbi algorithm for a multivariate three-state model 78
Figure 32 – Cumulative distribution function of the streamflow time series using a multivariate three-state HMM 78
Figure 33 – Regime posterior probabilities of a multivariate three-state HMM using streamflow and PDO 80
Figure 34 – Most probable path given by the Viterbi algorithm for a multivariate three-state model using streamflow and PDO 80
Figure 35 – Comparison between Viterbi results of univariate three-state model and the multivariate three-state model using PDO 81
Figure 36 – Regime posterior probabilities of a multivariate three-state HMM using streamflow and AMO 82
Figure 37 – Most probable path given by the Viterbi algorithm for a multivariate three-state model using streamflow and AMO 83
Figure 38 – Forecast of streamflow time series using an AR model of order 1. (a) Next state forecast, (b) the next three states and (c) the next five states 83
Figure 39 – Probability of the next state of the univariate HMM model 85

LIST OF TABLES

Table 1 – P-value results of the statistical tests applied to the Base Posts. The results that rejected the null hypothesis (p-value>0,05) are highlighted, thus, defining if the series was classified as stationary/nonstationary. 36
Table 2 – Results of the unit-root tests applied to the Base Posts time series 39
Table 3 – Summary of the results from the stationarity and trend tests 51
Table 4 – SPI values 61
Table 5 – Changepoint segmentation with penalty value of 1 68
Table 6 – Changepoint segmentation with penalty value of 0.1 69
Table 7 – Performance of HMM with different number of states K 71
Table 8 – Statistical properties of the low frequency series 76
Table 9 – Performance of multivariate HMM with different number of states K 76
Table 10 – Performance of multivariate HMM (streamflow and PDO) with different number of states K 79
Table 11 – Performance of multivariate HMM (streamflow and AMO) with different number of states K 81
Table 12 – Results of the prediction using an AR model of order 1 84
Table 13 – Probability of the next state of the univariate HMM model 85

LIST OF ABBREVIATIONS AND ACRONYMS

AR	Autoregressive Model
MA	Moving Averages Model
HMM	Hidden Markov Model
NOS	National System Operator
CEEMDAN	Complete Ensemble Empirical Mode Decomposition with Adaptive Noise
NIS	National interconnected system
ADF	Augmented Dickey-Fuller
PP	Phillips-Perron
BP	Base Posts
MME	Ministry of Mines and Energy
ERC	Energy Research Company
MPEO	Monthly Program of the Energy Operation
EMD	Empirical Mode Decomposition
IMF	Intrinsic Mode Function
EEMD	Ensemble Empirical Mode Decomposition
DWT	Discrete Wavelet Transform
CWT	Continuous Wavelet Transform
PDF	Probability Density Functions
KPSS	Kwiatkowski–Phillips–Schmidt–Shin
SRI	Standard Runoff Index
SPI	Standardized Precipitation Index
SPEI	Standardized Precipitation Evapotranspiration Index
SSI	Standardized Soil Moisture Index
PELT	Pruned Exact Linear Time
NAO	North Atlantic Oscillation
ENSO	El Niño-Southern Oscillation
PDO	Pacific Decadal Oscillation
AMO	Atlantic Multidecadal Oscillation
GEV	Generalized Extreme Value
BS	Binary Segmentation
SN	Segmentation Neighborhood
OP	Optimal Partitioning
CDF	Cumulative Distribution Function
BIC	Bayesian Information Criterion
AIC	Akaike Information Criterion

SUMMARY

1 INTRODUCTION 15
2 OBJECTIVE 19
2.1 MAIN OBJECTIVE 19
2.2 SPECIFIC OBJECTIVES 19
3 STATIONARITY ANALYSIS AND PRINCIPAL MODES OF INTERANNUAL AND DECADAL VARIABILITY OF THE HISTORICAL HYDROLOGICAL TIME SERIES OF THE NATIONAL SYSTEM OPERATOR (NSO) 20
3.1 ABSTRACT 20
3.2 INTRODUCTION 21
3.3 STUDIED AREA: NATIONAL INTERCONNECTED SYSTEM (SIN) 24
3.4 METHODS 27
3.4.1 Structural Analysis of Time Series 28
3.4.1.1 Complete EEMD With Adaptative Noise 28
3.4.1.2 Wavelet Analysis 29
3.4.2 Stationarity Assessment 31
3.4.2.1 Augmented Dickey-Fuller (ADF) 31
3.4.2.2 KPSS 32
3.4.2.3 Phillips-Perron 32
3.4.2.4 Dickey Fuller-GLS (ERS) 32
3.4.2.5 Mann-Kendall 33
3.4.2.6 Wilcoxon test 33
3.4.2.7 Student's t-test 34
3.4.2.8 Cox-Stuart test 34
3.4.3 Trend Analysis 35
3.4.3.1 Sen's Slope 35
3.5 RESULTS AND DISCUSSION 35
3.5.1 Statistical tests for stationarity analysis 35
3.5.2 Unit-root tests for stationarity analysis 38
3.5.3 Mann-Kendall and Sen's Slope 42
3.5.4 Wavelet Analysis and Complete EEMD With Adaptative Noise 43
3.6 ANALYSIS OF THE RESULTS 50
3.7 CONCLUSION 51
4 PLURIANNUAL STREAMFLOW FORECASTING USING HIDDEN MARKOV MODEL 53
4.1 ABSTRACT 53
4.2 INTRODUCTION 53
4.2 DATA AND METHODS 55
4.2.1 Wavelet transforms 58
4.2.2 Standard Runoff index 60
4.2.3 Changepoint detection 61
4.2.4 Autoregressive model 62
4.2.5 Hidden Markov Models 63
4.3 RESULTS 66
4.3.1 Time series decomposition 66
4.3.2 State detection 67

4.3.3 Low frequency forecast model *83*
4.4 CONCLUSION **85**
5 CONCLUSIONS **88**
REFERENCES **91**

1 INTRODUCTION

Time series analysis has been a central topic in many knowledge areas, especially in stochastic hydrology (Salas 1980; Garfias-Soliz et al., 2010; Sang, 2013). A time series is a set of measures or numerical values of any time-ordered variable. It can be either discrete (e.g. number of floods in a given period) or continuous (e.g. hydraulic conductivity in space or time flow), where the first can be obtained by a sampling interval of a continuous series. Among the main objectives of the time series analysis are forecasting long- or short-term future values; prediction of the frequency of extreme events; generate synthetic series; describe the behavior by verifying the existence of seasonal trends or variations; investigation of operating system rules in reservoirs (Morettin and Toloi, 2006).

There are two main approaches when it comes to investigating natural phenomena, the deterministic approach, and the stochastic approach. The first is based on cause-and-effect relationships in which their variables are free from random variations, while the second will not necessarily be linked to those cause and effect relationships, being thought of as random variables that have probability distributions (Yevjevich, 1987; Salas et al., 1980). A time series is described as a random or stochastic process if it contains a stochastic component. Most hydrological time series are composed of a stochastic component superimposed on a deterministic component.

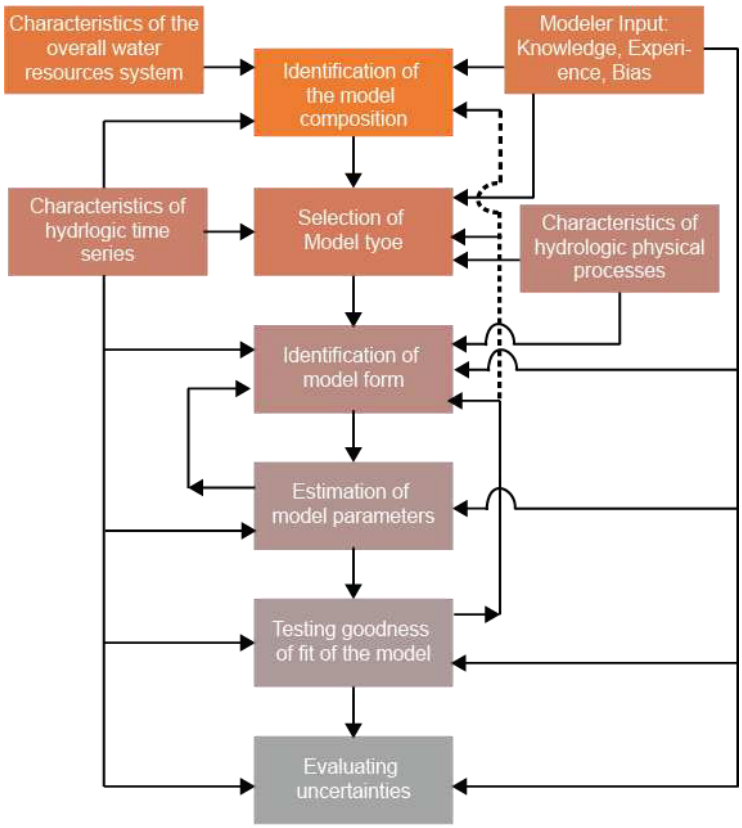
The deterministic component can be classified as periodic, trend components or a combination of both. The trend component, in a time series, is the result of the gradual changes of nature or anthropogenic in the hydrological environment, which produce the time series. The trend analysis and its potential impacts are an important component in the planning and management of water resources, especially under temporal changes in hydrologic regimes. Therefore, a careful analysis of the flow regime and the identification of temporal changes in the hydrologic cycle is a significant topic in water resources studies (Alves et al., 2013; Sethi et al., 2015; Bayazit, 2015).

The periodic component is present in most continuous monthly series, where a cyclic pattern is observed. The stochastic component represents an irregular variation, in comparison to the others, but continuous during the time series. This can be caused by sample observation errors or by random fluctuations in natural physical processes (Morettin and Toloi, 2006).

In the analysis of a time series model, the best procedure is to identify a model that fits a certain data sets properly, which may be a stochastic or deterministic model. Thus, Box and Jenkins (1970) constructed a model of identification, estimation and diagnosis to check the stages in the elaboration of the model (Hipel, McLeod, 1994). Salas (1980) also proposed a systematic approach to hydrologic time series modeling composed of six main phases, illustrated in Figure 1.

Figure 1 – Systemic Approach for time series modeling constructed by Box and Jenkins (a) and Salas (b)

(a)



(b)

Source: Adapted from Salas (1980)

There are two main approaches in time series analysis, the parametric models and non-parametric. In the parametric models, it is made prior assumptions about the nature of the data and, generally, the parameters are fixed early in the process. For the nonparametric models,

the structure of the model not defined a priori, and normally the parameters are determined by the data themselves (Morettin and Toloi, 2006; Sivakumar, 2017). Some examples of parametric models are autoregressive models (AR) or moving averages (MA), Markov Chain Process, and others. Whereas, among the non-parametric models, are Nearest Neighbors resampling, Kernel density estimator, and others.

In many cases, the hydrological problems studied do not require detailed discussion about the physical process, but only a representation of these time series processes. Stochastic models can be used to represent, in a simplified way, these hydrological time series. Stochastic modeling places greater emphasis on the statistical characteristics of hydrological processes Thus, a stochastic process is a set of random variables, which are defined in a probability space.

One of the main assumptions made when evaluating a time series is stationary, hence its statistical properties do not change over time, contemplating an equilibrium around a constant mean (Morettin and Toloi, 2006; Brockwell and Davis, 2016). Mathematically the stationarity properties can be described as:

$$E(X_t) = \mu \quad \text{(Eq. 01)}$$

$$Var(X_t) = \sigma^2 \quad \text{(Eq. 02)}$$

$$Cov(X_t, X_{t+L}) = \lambda_L \quad \text{(Eq. 03)}$$

A stochastic process is purely stationary when the distribution of X_t is not time dependent and all simultaneous distributions of the random process variables depend only on the mutual interval time.

The stationarity hypothesis is still used in many design and planning situations, although many authors assert that stationarity is dead (Milly et al., 2008) and do not have to be used in time series modelling. It is speculated that the responsibility of the high nonstationary of the hydrological series is related to the climatic changes and anthropogenic actions (Milly et al., 2008; Detzel et al., 2016; Serinaldi and Kilsby, 2015, Chen et al., 2017). The stationarity assumption in streamflow modeling may lead to the underestimation or overestimation of the hydrologic risk which can generate poor planning and functioning of the water resources systems. Many studies have been developed for the investigation of the stationarity of the Brazilian hydrological series (Müller et al., 1998; Batista et al. 2009; Detzel et al. 2011).

Among the methods used in time series analysis which operate over multi-temporal scales and show nonstationary characteristics, it is the wavelet transform which has been widely applied in various geographic basins and regions worldwide to expose the complex

characteristics of hydrologic processes (Torrence and Compo, 1998; Labat, 2005; Silveira et al., 2011; Alves et al., 2013; Nourani, 2013; Sang, 2013). This multi-temporal analysis can capture the different time variabilities of the time series (interannual and interdecadal scale). Furthermore, the better understanding of the different frequencies presented in time series can improve water system modeling (Hoek and Bos, 2007). With the increasing length of historical hydrological time series, the presence of low frequency structures of climate and hydrologic time series has become an important feature in hydrological analysis (Kwon et al., 2007).

The inspection of low frequency oscillations in hydrological process has driven many researches in this field of study, especially models coupled with the climate indexes which are known for modulating the variability of hydrological variables (Kwon et al., 2008; Grimm and Sabóia, 2015). The changes in the hydrological variability can greatly affect the operation of reservoirs, and the prediction of these shifts in the mean can assist in the decision-making process and collaborate in the performance of water systems based on future forecasts.

Thus, the analysis of historical time series in a country like Brazil, where the spatial-temporal variability is very large and the its hydrological cycle is highly variable, the identification of these different behaviour as well as the influence of large-scale climatic forcing on the low frequency variability of the streamflow's can improve water systems planning and management.

The dissertation is organized on the following five parts: the mains and specific objectives, article one entitled as "Stationarity analysis and principal modes of interannual and decadal variability of the historical hydrological time series of the national system operator (NSO)", article two entitled as: "Pluriannual streamflow forecasting using Hidden Markov Model", the conclusions obtained in those studies and the reference used.

2 OBJECTIVE

2.1 Main objective

To analyze the time series of inflows identifying the presence of stationarity, trend and verifying the modes of variation present in these series in order to analyse the low frequency and develop a forecast model using the Hidden Markov Model (HMM).

2.2 Specific objectives

- Evaluate stationarity comparing different statistical and unit root tests;
- Identify the presence of trend in the streamflow gage stations of the NSO
- Evaluate which frequencies best represent the series using the explained variance;
- Identify the wet and dry periods of the streamflow time series from the Sobradinho's reservoir and compare the results from different methodologies such as Standard Runoff Index (SRI), changepoint detection and an HMM
- Develop forecast models using an Autoregressive model and an HMM.

3 STATIONARITY ANALYSIS AND PRINCIPAL MODES OF INTERANNUAL AND DECADAL VARIABILITY OF THE HISTORICAL HYDROLOGICAL TIME SERIES OF THE NATIONAL SYSTEM OPERATOR (ONS)[1]

3.1 Abstract

Climatic changes and anthropogenic actions questioned the stationary model which is widely diffused in hydrological modeling. Thus, the present work intends to make a diagnosis of the time series of the National Interconnected System (SIN) in Brazil. First, the stationarity was analyzed by statistical and unit root tests. In order to evaluate the presence of trend in the series was used the Mann-Kendall and Sen's slope tests. Then, the time series were decomposed by Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) and Wavelet Transform. Results between the statistical and unit root tests showed similar results for stationarity analysis, except for the south region, and for both types more than half of the series were classified as nonstationary. It was evaluated that the unit root tests are more indicated for modeling series with the presence of trend. When examining trend, it was noted that the south and southeast region both presented positive trends while the north and northeast region presented negative trends. Analysing the decomposed time series by the Morlet wavelet and the CEEMDAN method, most of the series have higher explained variance for the period of 2-8 years, ranging from 50-60%, while for the medium frequency (8-16 years) the explained variance is around 10-20% and for the low frequency (16-32 years) it represents 5-10% of the time series. Those decomposition methods have proved to be important tools for the identification of climatic variability of an important sector for the economy of the country, forming a diagnostic of the behaviour of the streamflow time series. In conclusion, most of the station gage analyzed are not stationary, although they are still model as such. The time series presented the trend component in 90% of the evaluated time series and the high frequency (2-8 years) variability is responsible for most of the explained variance of the whole series. So, a complete diagnosis of the time series is paramount for the modelling of the water resource systems, especially one economically important as the hydroelectric sector. The methodologies presented in this work are useful for hydrological studies in regions affected by climatic variability and assist in modeling the behavior of hydrological series.

[1] This paper was submitted to the Journal of Hydrology in January, 2019.

3.2 Introduction

In most hydroclimatic time series can be detected the presence of periodicities and trends in variables such as precipitation, streamflow, temperature and others. Furthermore, the better understanding of the different frequencies present in those time series can improve hydrological system modeling (e.g., improve long-range forecast). Assessing river discharge records can be a resourceful alternative to precipitation record due to the large space-time variability of that hydrological variable, thus, needing less data from weather stations caused by the need to proper characterization of rainfall (Coulibaly and Burn, 2004; Nalley et al., 2016). Coulibaly and Burn (2004) explain that river flows can serve as an appropriate index of interannual hydroclimatic variability at a local or regional scale.

When evaluating variables such as streamflow, which presents key information regarding hydrological changes, an important part of the process is to decompose the original data and assess its variability modes. Many different methods have been used to accomplish this goal, such as the complete ensemble mode decomposition with adaptative noise (CEEMDAN) or wavelet transform. This climatic variability directly affects the hydroclimatological time series and, consequently, the water availability and the activities that depend on the water. There are also impacts that can affect sea level, agriculture and hydropower generation, which can produce great pressure on hydrosystems (Marengo and Valverde, 2007).

One of the determinants of the modes of variation in hydrological time series are the atmospheric systems and their interactions in different time scales, being thus modulators of the climate and consequently of hydrological variables. According to Marengo and Valverde (2007), the presence of a variability in the interannual and interdecadal scales of precipitations and flows was observed in the Amazon basin and Northeast region of Brazil. Anjos (2015) evaluated the low frequency relation of historical inflows of the SIN with climatic indices using wavelet transforms. Alves et al. (2013) analyzed the modes of variation of the SIN stations using the wavelet transforms.

Another key component in the time series analysis is the analysis of stationarity. In models that apply hydrologic time series is paramount the familiarity with its statistical properties, mainly due the fact that such series tend to vary with time and can be characterized as nonstationary if these variations are considered significant. The climate on the planet is changing due to anthropogenic actions and natural climatic processes. These changes can be

observed through changes in the hydrological cycle with a direct impact on the parameters of hydrological changes, such as precipitation, evapotranspiration and flows (IPCC, 2013). It is speculated that the responsibility of the high nonstationarity of the hydrological series is related to the climatic changes and anthropogenic actions (Milly et al., 2008; Detzel et al., 2016; Serinaldi and Kilsby, 2015; Chen et al., 2017). The stationarity hypothesis is still used in many design and planning situations and the violation of the hypothesis in the flow series may lead to the underestimation or overestimation of the statistics which can generate poor planning and management of the water resources system.

Many studies have been developed for the investigation of the stationarity of the Brazilian hydrological series using, for the most of them, statistical tests. Müller et al. (1998) analyzed the series of precipitation and flow of the Itaipu basin through statistical tests and concluded that the fluviometric and pluviometric series are nonstationary. Batista et al. (2009) evaluated the hydrological series of the South-Southeastern Brazil, where the stationarity of the series of the Brazilian Southeast and the non-stationarity of the series of the South region were verified from 1970. Detzel et al. (2011) conducted a study to verify the stationarity of the series of affluence of Brazilian hydroelectric plants, concluding that the series of the South subsystem do not present any statistically stationary series, while for the series of the North and Northeast, the series were considered statistically stationary. Pedrosa and Souza (2009) conducted a study of the mean and minimum flows of the Paraíba river, where they evaluated that there is no statistical evidence to reject the hypothesis of stationarity in the region.

Most studies, in order to evaluate stationarity using hydrological series apply statistical tests, such as the Mann-Kendall or Student's t-tests, however, unit root tests also have been used to evaluate stationarity in hydrological time series. In the study by Karamouz et al. (2015), the stationarity of a series of precipitation was verified using the Augmented Dickey-Fuller (ADF) test. Zhao and Chen (2015) applied the ADF test to analyze stationarity of annual runoff series for four hydrological stations. Modarres and Ouarda (2014) used the Dickey-Fuller unit root test and the Phillips-Perron test to verify the seasonality of climatic indexes and drought indexes.

The trend component, in a time series, is the result of the gradual changes of nature in the hydrological environment or anthropogenic modifications such as land-use changes, urbanization and land-cover. The trend analysis and its potential impacts are important components in the planning and management of water resources, especially under temporal changes in hydrologic regimes. Therefore, a careful analysis of the flow regime and the

identification of temporal changes in the hydrologic cycle is an important topic in water resources studies (Alves et al., 2013; Sethi et al., 2015; Bayazit, 2015).

For the identification of trend many parametric and nonparametric methods have been applied, where the nonparametric tests, such as Mann-Kendall, are the most used due to the need of fewer assumptions in its applications (Bayazit, 2015). Many authors have evaluated trends in hydrological time series in Brazil, such as Krusche et al. (1997), who performed a statistical analysis using the Mann-Kendall test to evaluate the presence of trend in the Piracicaba river basin that show increase in the evapotranspiration and precipitation, while there was a decrease in the streamflow. Silveira et al. (2016), carried out a trend and variability assessment in the São Francisco basin applying moving averages for a ten-year horizon, linear regression and Mann-Kendall-Sen method for precipitation and temperature variables. Silva et al. (2013) verified the presence of trend in the precipitation and flow series in the Upper São Francisco region using the Mann-Kendall test. The analysis or detection of trends is an important aspect in research and studies with hydrological variables when considering the impacts under climate change.

Thus, to evaluate the behavior of such an extensive system as the National Interconnected System, the National System Operator (NSO), which is responsible for coordinating and controlling the operation of electricity generation and transmission facilities in the SIN, it adopts the forecasting of flows for a subset of hydroelectric stations considering 88 representative stations from different basins in Brazil known as Base Posts (BP). Hence, this paper aims to make a diagnostic of the 88 base stations that compose the NIS applying monthly time series ranging from 1931-2016, totaling 86 years of records. This study proposes to use classical statistical test, such as Student's t-test, Wilcoxon, Mann-Kendal and Cox-Stuart and tests widely used in macroeconomics like Augmented Dickey-Fuller, KPSS, Phillips-Perron and Dickey Fuller-GLS to evaluate the stationarity/nonstationary in the studied time series. Also, it is applied trend tests to verify the presence of a positive or negative tendency in the time series. Complementarily, it is proposed a decomposition methodology using wavelet transform and CEEMDAN to identify patterns of variation of flow series that are not explained by simple stationarity models, allowing to make a diagnostic of streamflow time series in an important Brazilian system.

3.3 Studied Area: National Interconnected System (NIS)

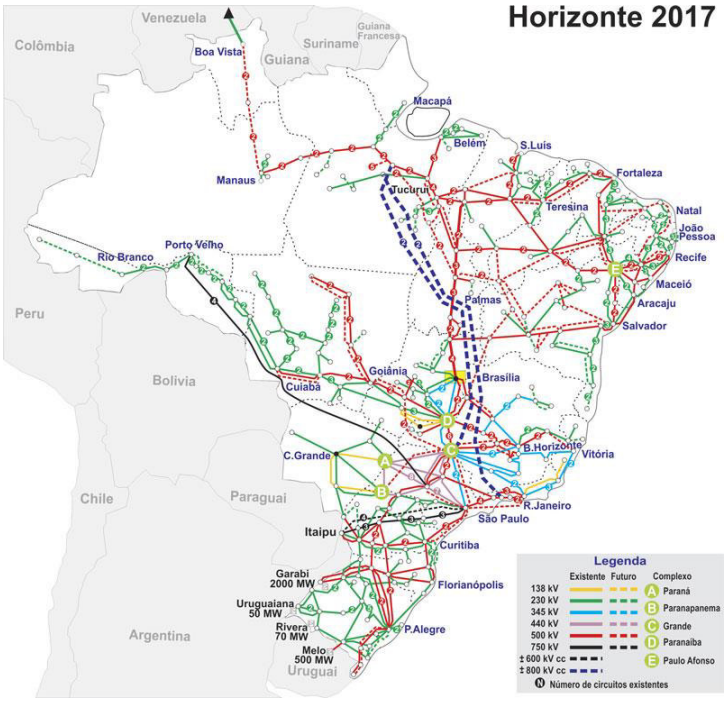
The world's energy generation is mostly composed of combustible fuel (67,3%), hydropower generation responds to 16,6%, nuclear corresponds to 10,4% and solar/wind/geothermal/others represents 5,7% of the electricity production (International Energy Agency, 2016). However, in Brazil, most of its electricity is produced from hydropower plants (67,8%).

The largest consumer of electricity in Brazil is the industry sector with 36% followed by residential (29%) and commercial (19%) uses. Energy usages, such as public, agricultural and transport sectors correspond to only 16% of the total according to the National Energy Balance of 2017 prepared by the Ministry of Mines and Energy (MME) with the Energy Research Company (EPE) (MME / EPE, 2018).

The National Interconnected System (Figure 2) is responsible for the production and transmission of energy being characterized as a hydro-thermo-eolic system, with predominance of generation by hydroelectric plants. Most of the hydropower plants are located in the Southeast-South region of the country, mainly due to the large consumption centers. Only 1.7% of the energy demanded by the country is outside the SIN, in small isolated systems located in the Amazon region (ONS, 2018).

Hydroelectric power generation holds a fundamental importance in Brazil's electricity production. The four production subsystems (North, Northeast, South and Southeast) are interconnected by thousands of kilometers of distribution networks, allowing the optimization of distribution and avoiding possible regional restrictions in order to supply the country's electricity demand.

Figure 2 - National Interconnected System

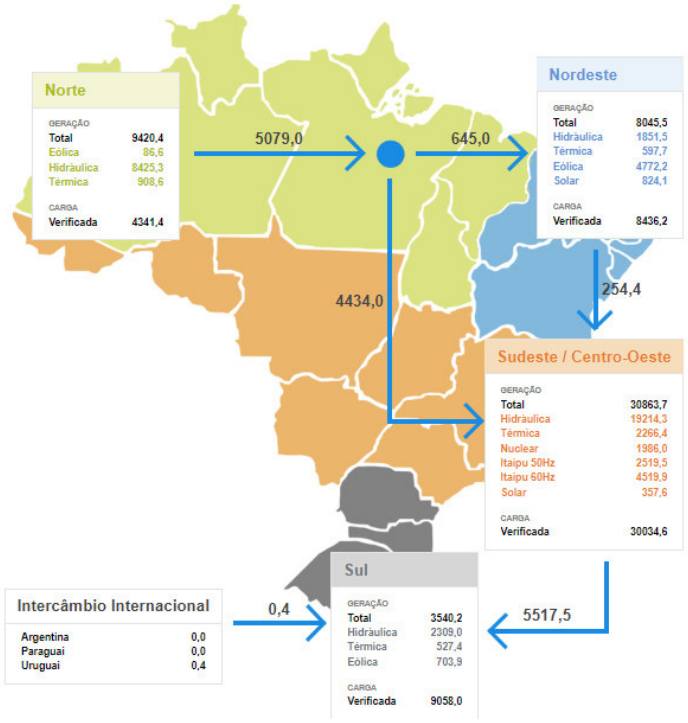


Source: ONS (2018)

In order to exemplify production dynamics, load requirements and transmission between the different regions, values presented in Figure 3, which it corresponds to observed energy production in mega-watts (MW) for the year 2018 are described hereafter. The North region produces 18,2% of the national energy in MW and demand 8,4%, the Northeast region produces 15,5% and demand 16,4%, the South region produces 6,8% and demand 17,7%, the Southeast/Central West region produces 59,5% and demand 58,2%.

Regarding interregional transfers, the North region transmits 645,0 MW of its production to the Northeast region and 4434, MW to the Southeast / Midwest region; the Southeast / Midwest region transmits 5517,5 MW of its production to the South region, in addition to receiving 254,4 MW of the production in the Northeast region.

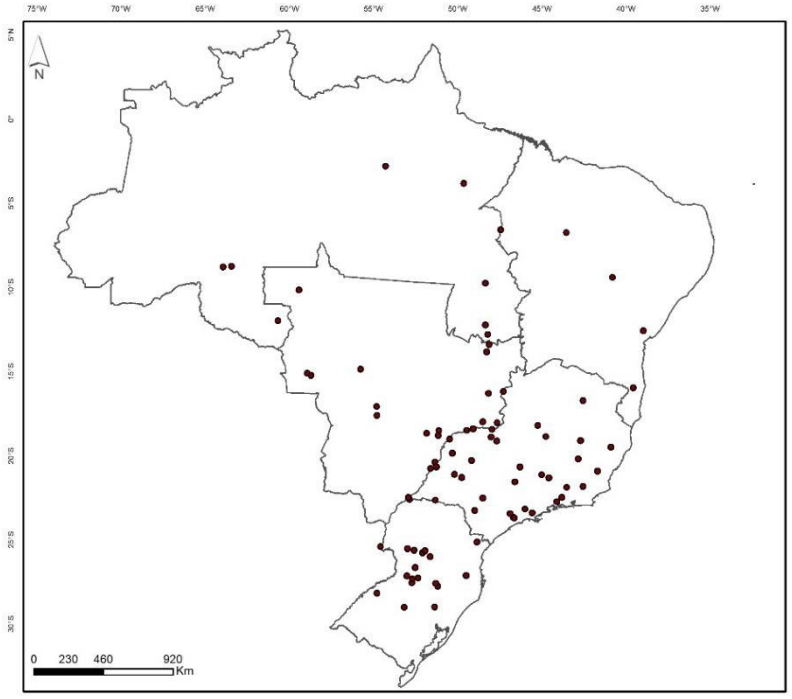
Figure 3 - Interregional transfers of the National Interconnected System



Source: ONS (2018)

The forecast of flows and generation of scenarios of inflows defined by the National System Operator (NSO) establishes the processes for the forecasting of monthly, weekly and daily flows, and for the generation of scenarios of monthly average natural inflows used in the preparation of the Monthly Program of the Energy Operation (MPEO). In order to do so, it is generally adopted the performance of flow forecasts for a subset of gauge stations also know as Base Posts. The rest of the gauge stations are predicted through monthly linear regressions based on the data provided in the BP to complement the forecasts of flows for the whole SIN. NSO currently works with a total number of 88 BP representative of the various regional hydrographic regimes found in Brazil (Figure 4). The dataset was obtained through the NSO and it refers to the monthly averages of the naturalized historical streamflow series, in cubic meters per second, of the 88 posts of the National Interconnected System (NIS). The time series varies from January of 1931 to December of 2016.

Figure 4 - Spatial distribution of the Base Posts

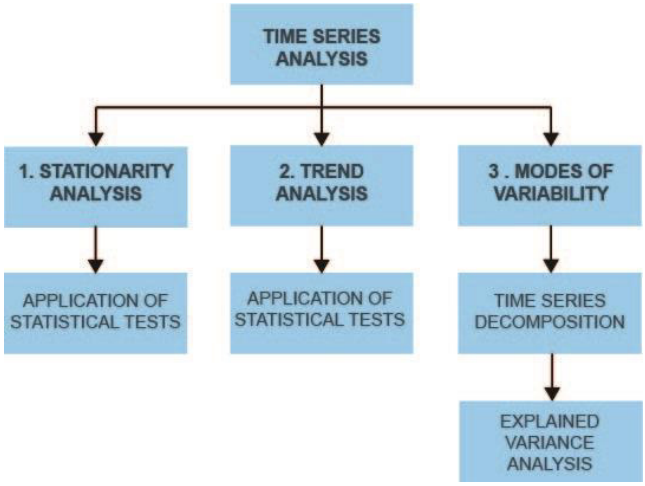


Source: Prepared by the author

3.4 Methods

Methods used in this study to analyze the hydrological data as regard to stationarity, trend and patterns of variability are described below. The methodology applied in this study is illustrated in Figure 5 and divided into three phases: (1) test application to evaluate stationarity; (2) test application for trend analysis and (3) time series decomposition and analysis of its variability patterns.

Figure 5 - Methodology used in this article



Source: Prepared by the author

3.4.1 Structural Analysis of Time Series

Most hydrological time series are composed of a stochastic component superimposed on a deterministic component. The time series used in this study were decomposed by the different methods given below.

3.4.1.1 Complete EEMD With Adaptative Noise

The Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) method is an adaptation to the empirical mode decomposition method (EMD) due to the presence of limitation for the decomposition of signals in non-linear and nonstationary processes. In the EMD method the first step is to identify the local extrema points and fitting the spline functions that connects them, next, computes the mean series and find the residue series in an iteratively process, until the resulting series became a zero-mean series. A complete description of the EMD method can be found in Huang et al. (1998). Huang et al. (1998) states that an Intrinsic Mode Function (IMF) shall complied two conditions: (i) have the number of extrema and the number of zero crossings equal or differ at most by one, and (ii) the mean value of the upper and lower envelopes is zero at any point.

To overcome some limitation exhibited in the EMD process, Wu and Huang (2009) proposed an ensemble approach called EEMD, which forms the signs by adding randomly generated white noise series with the original series. It was noted that limitations, such as mode

mixing issues in the reconstructed series may still prevail, thus, a new method named CEEMDAN was proposed by Torres et al. (2011). In the CEEMDAN method, the first EMD mode is computed over an ensemble of residue plus different realizations of a given noise, and IMF_2 is obtained by averaging. The procedure continues until a stopping criterion is attained (Antico et al., 2014). Let us define the operator $Ej\ (\cdot)$ which, given a signal, produces the $j-$th mode obtained by EMD. Let w^i be white noise with $\mathcal{N}\ (0, 1)$.

The CEEMDAN algorithm can be described as follows: obtain the first EMD mode of I realizations and calculate:

$$\widetilde{IMF_1} = \frac{1}{I}\sum_{i=1}^{I} IMF_1^i = \overline{IMF_1} \quad \text{(Eq. 04)}$$

Then, calculate the first residue and obtain the first EMD mode for each realization, then, define the second mode as:

$$\widetilde{IMF_2} = \frac{1}{I}\sum_{i=1}^{I} E_1(r_1 + \beta_1 E_1(w^i)) \quad \text{(Eq. 05)}$$

for k = 2,...,K calculate the *k-th* residue:

$$r_k = r_{(k-1)} - IMF_k \quad \text{(Eq. 06)}$$

Decompose the realizations until their first EMD mode and define the (k+1)-*th* mode as:

$$\widetilde{IMF_{(k+1)}} = \frac{1}{I}\sum_{i=1}^{I} E_1\left(r_k + \beta_k E_k(w^i)\right) \quad \text{(Eq. 07)}$$

Calculate the residue for the next k and decompose the next realization until the obtained residue can no longer be decomposed. The final residue satisfies the following:

$$R = x - \sum_{k=1}^{k} \widetilde{IMF_k} \quad \text{(Eq. 08)}$$

In CEEMDAN, the level of added noise is determined by the parameter β_k. Although different values of β_k can be set for different modes (see Equation 7), we use a fixed value of this parameter to obtain all modes. We set a noise level of 6.4e-07 and a maximum of 100 shifting iterations in the CEEMDAN applications to time series of streamflow. The variability modes decomposition was performed using the package *hht* (Bowman and Lees, 2013) in the software R.

3.4.1.2 Wavelet Analysis

Wavelet analysis is a wildely used technique in periodic phenomenon in nonstationary time series, specially in time variant frequencies. The wavelet function $\psi(t)$ is defined mathematically by $\int_{-\infty}^{+\infty} \psi(t)dt = 0$ (Huo et al., 2016). This function oscillates during a certain time period and then decreases to zero. In the wavelet method the model input is divided into a subset of continuous or discrete wavelet methods, each subsignal plays a different role and has a unique behaviour. In wavelet theory, the mother wavelet should represent the deterministic characteristics and its components (periodicity, trends, cycles, etc.) in an accurate manner for hydrological time series (Percival and Walden, 2000).

The wavelet transform decomposes a time series into different resolutions by analyzing its frequency of a uni or bi-variate time series. There are two different ways of modelling a time series with wavelet transform: discrete (DWT) and continuous (CWT) method. The DWT method decomposes a series into subsignals given proper wavelet and temporal scale, and the result can guide wavelet threshold denoising and wavelet decomposition. The CWT may be applied to find the scale contents of a signal and their variance in time (Pathak et al., 2016).

The choice of the mother wavelet is a very important one, where the kind of wavelet transform chosen depends of the type of output information needed. There are many mother wavelet functions from which to choose, such as Haar wavelet, Daubechies wavelet, Mexican Hat wavelet, Morlet wavelet, and others. The Morlet wavelet is commonly used in hydrological time series because describes the time series well and has a better time–frequency localization when compared to the other commonly used wavelets, such as the Mexican Hat and the Daubechies wavelets (Huo et al., 2016; Nalley et al., 2016). Due to the advantages presented in this type of model this study used the Morlet wavelet as well.

The mother Morlet wavelet can be implemented by the Equation 9:

$$\psi(t) = \pi^{-1/4} e^{i\omega t} e^{-t^2/2} \qquad \text{(Eq. 09)}$$

Where ω is the angular frequency and is set to 6 because it makes the Morlet wavelet approximately analytic and t is the time.

The Morlet wavelet transform (x_t) was defined as the convolution of the series with a set of (x_t) generated by the mother wavelet by translation in time by τ and scaling by s:

$$Wave\ (\tau, s) = \sum_t x_t \frac{1}{\sqrt{s}} \psi^* \left(\frac{t-\tau}{s}\right) \qquad \text{(Eq. 10)}$$

With * denoting the complex conjugate. The position of the particular daughter wavelet in the time domain is determined by the localizing time parameter τ being shifted by a

time increment of dt. The choice of the set of scales s determines the wavelet coverage of the series in the frequency domain.

The local amplitude of a periodic component in a time series can be obtained by the modulus of its wavelet transform. It is adopted the rectified version (Liu et al., 2007), since the modulus can produce biased wavelet amplitude.

$$Ampl(\tau, s) = \frac{1}{s^{\frac{1}{2}}} |Wave(\tau, s)|^2 \quad \text{(Eq. 11)}$$

The wavelet power spectrum is the interpretation of the wavelet energy density in the time-frequency domain and can be written as:

$$Power(\tau, s) = \frac{1}{s} . |Wave(\tau, s)|^2 \quad \text{(Eq. 12)}$$

In this work, it was used the package *WaveletComp* in R (Rosch and Schmidbauer, 2014).

3.4.2 Stationarity Assessment

The stationary concept revolves around the idea that a natural system oscillates randomly within an unchanging envelope of variability. The variables of this system have time-invariant probability density functions (PDF) and those variables can be estimated through historical data available. Due to climate change, many authors are using models to that account for a nonstationary framework. Thus, this study proposed to verify the streamflow in Brazil and evaluate it using four unit root tests and four statistical tests to determine whether those time series variables are nonstationary (Milly et al., 2008; Serinaldi and Kilsby, 2015; Bayazit, 2015).

3.4.2.1 Augmented Dickey-Fuller (ADF)

The Augmented Dickey Fuller test aims to determine whether $\beta = 1$ in a model $y_t = a + \beta y_{t-1} + \varepsilon_t$. Dickey and Fuller (1979) consider three differential form of autoregressive equations to test for the occurrence of a unit root, where it is used penalty criteria like AIC and BIC/SBC (Barros, 2017).

$$\Delta y_t = \gamma y_{t-1} + \sum_{t=1}^{p} \beta \Delta y_{t-1} + \varepsilon_t \quad \text{(Eq. 13)}$$

$$\Delta y_t = a + \gamma y_{t-1} + \sum_{t=1}^{p} \beta \Delta y_{t-1} + \varepsilon_t \quad \text{(Eq. 14)}$$

$$\Delta y_t = a + \gamma y_{t-1} + \beta t + \sum_{t=1}^{p} \beta \Delta y_{t-1} + \varepsilon_t \quad \text{(Eq. 15)}$$

The three different equations are related to the presence of the deterministic components. Equation 13 is a pure random walk model, Equation 14 adds an intercept or drift term, and Equation 15 includes both a drift and a linear trend (Enders, 2008). After performing the test, the results are compared with the statistics presented in Dickey-Fuller table allowing to determine if the null hypothesis is rejected or accepted.

3.4.2.2 KPSS

Due to the reduced statistical power of the ADF test, which might have created a bias in verifying the existence of the unit root. Another recommended test is the KPSS test (Kwiatkowski et al. 1992). It is used to measure the null hypothesis that a time series is stationary around a deterministic trend (Um et al 2018; Barros, 2017).

$$y_t = d_t + r_t + \varepsilon_t \quad \text{(Eq. 16)}$$

Where d_t is the deterministic trend, r_t is the random walk and ε_t is the stationary error.

3.4.2.3 Phillips-Perron

Phillips and Perron (1988) unit root test refers to the same null hypothesis as the ADF test without the moving average term, however the main different is that it addresses the serial correlations and heteroscedasticity in the error (Um et al 2018; Barros, 2017). The test regression can be described as:

$$\Delta y_t = \beta D_t + \pi y_{t-1} + u_t \quad \text{(Eq. 14)}$$

Where π is a least-squared estimate.

3.4.2.4 Dickey Fuller-GLS (ERS)

The ADF stationarity test is pointed out as having a low statistical power and sensibility to deterministic components. The DF-GLS (Elliot et al. 1996) test is computed by least square regression and locally de-trends the series to properly estimate the deterministic

parameters of the series, then uses the altered data to perform a typical ADF unit root test. The test is considered optimal among tests that use OLS to estimate parameters and increases the test's statistical power (Choi, 2001; Barros, 2017).

3.4.2.5 Mann-Kendall

The Mann-Kendall (Mann, 1945; Kendall, 1975) is a non-parametric test which is widely used in environmental and hydrological time series to detect the presence of monotonic trends. The null hypothesis employed in the test assumes that the data came from a population with independent and identically distributed realizations (IID), which is the premise to a stationary process. The Mann-Kendall test is calculated according to:

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} sgn(X_j - X_k) \qquad \text{(Eq. 15)}$$

With

$$sgn(x) = \begin{cases} 1 & if\ x > 0 \\ 0 & if\ x = 0 \\ -1 & if\ x < 0 \end{cases} \qquad \text{(Eq. 16)}$$

n is the sample size. In the test, positive statistics indicates an increasing trend, while a negative statistic indicates a decreasing trend.

3.4.2.6 Wilcoxon test

The Wilcoxon test (Gehan, 1965) is a classic non-parametric test used to test the equality between population averages, which can follow any statistical distribution. Thus, the null hypothesis of the test is H_0: $\mu_1 = \mu_2$.

The series is divided into sets of sub-samples, so the series are sorted and it is assigned indices of same length. For long samples the distribution is approximated from a normal distribution by the equation below, and then the value is compared with the test statistic according to the level of significance determined.

$$z = \frac{W - n_1(n_1 + n_2 + 1)/2}{\sqrt{n_1.n_2(n_1 + n_2 + 1)/12}} \qquad \text{(Eq. 17)}$$

3.4.2.7 Student's t-test

The Student's t-test stands out for its popularity and simplicity. This is a parametric test on means that computes the statistic for a t-distribution, and then the confidence interval is defined for the acceptance or rejection of the null hypothesis of the test, comparing the value obtained with the test statistic to the said level of significance. The test verifies the equality of two population means, assuming the normality of the samples, so it is necessary the normalization using the log-normal model, which is the most indicated distribution because it does not allow negative flows (Detzel et al., 2011).

The test is performed by two sub-samples in which the null hypothesis requires that the sub-samples have statistically similar averages. The test statistic can be calculated by:

$$t = \frac{(\overline{x_1} - \overline{x_2})}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{(Eq. 18)}$$

Where $\overline{x_1}\ and\ \overline{x_2}$ represent the means of each sub-sample; n_1 and $n_2$represent the number of elements of the sub-samples and $s_p$is the combined variance estimator.

3.4.2.8 Cox-Stuart test

The Cox-Stuart test, or signal test, is a non-parametric test used to verify the contrast between the position of the mean and identify the presence of monotonous trends, whether they are increasing or decreasing. The null hypothesis of the test is that the number of positive and negative signals are equal, that is, there is no trend. The test calculates the difference between pairs formed from sub-samples, extracted from the original sample. Positive or negative signs may be associated with pairs, with ties eliminated.

For samples with a number of elements equal or greater than 35 (n≥35), the statistic is approximated to a normal distribution, calculating the standard variable according to:

$$z = \frac{2x \mp 1 - n}{\sqrt{n}} \quad \text{(Eq. 19)}$$

The calculated variable is compared with the tabulated values and, thus, the null hypothesis is rejected or accepted.

The tests were performed using the packages *randtests* (Caeiro and Mateus, 2014), *trend* (Pohlert, 2018) and *urca* (Pfaff, 2008) in the software R.

3.4.3 Trend Analysis

The Mann-Kendall test was also used to verify the presence of trend in the time series and the Sen's slope test was applied to analyse the magnitude of such trends.

3.4.3.1 Sen's Slope

The Sen's Slope test calculates the slope and its intercept according to Sen's method (Sen, 1968). The linear slopes are determined by:

$$d_k = \frac{X_j - X_i}{j - i} \quad \text{(Eq. 20)}$$

for ($1 \leq i < j \leq n$), where d is the slope, X denotes the variable, n is the number of data, while i and j are the indices.

Then, the median from all slope is calculated and the intercepts are computed for each time step t according to the following equation:

$$a_t = X_t - bt \quad \text{(Eq. 21)}$$

3.5 Results and Discussion

In this section, results were obtained when analysing the series of streamflows by applying (1) statistical and unit root test for evaluate the presence of stationarity in the studied data, (2) trend analysis and (3) modes of variation of the described data using decomposition methods described above.

3.5.1 Statistical tests for stationarity analysis

The time series of the base posts were evaluated using four statistical tests, which were: Student's t-test, Wilcoxon, Mann-Kendal and Cox-Stuart, the p-value of such tests are described in the Table 1. The p-values for the bilateral tests are interpreted as the probability of observing, in another withdrawn sample from the same population, a higher (or lower) statistic than that observed with the sample tested. Thus, smaller the p-values means evidence of nonstationarity. Results are displayed in terms of p-values since all tests were performed with the same level of significance ($\alpha = 5\%$) and the values that rejected the null hypothesis (p-value

> 0.05) are highlighted in Table 1. The time series classification criteria was based on the rejection of the null hypothesis in most of the statistical tests (3 or higher), and when it was a tie, it was given a greater importance to the non-parametric tests, like Mann-Kendal, deciding if the series was classified as stationary/nonstationary. The spatial representation of the stationary and nonstationary stations is illustrated in Figure 6.

Table 1 – P-value results of the statistical tests applied to the Base Posts. The results that rejected the null hypothesis (p-value>0,05) are highlighted, thus defining if the series was classified as stationary/nonstationary.

Code	Station	Region	Student's[2] t	Wilcoxon[2]	Cox-Stuart[2]	Mann-Kendall[2]
145	RONDON II	NORTE	**5.1E-01**	**1.2E-01**	**4.5E-01**	**9.9E-01**
275	TUCURUI	NORTE	**3.2E-01**	**7.7E-01**	**2.5E-01**	3.2E-02
277	CURUA-UNA	NORTE	**4.1E-01**	**9.5E-01**	**7.2E-01**	**1.8E-01**
279	SAMUEL	NORTE	**7.4E-01**	**1.2E-01**	1.1E-02	2.0E-02
287	STO ANTONIO	NORTE	**2.1E-01**	**2.8E-01**	**3.6E-01**	**2.9E-01**
291	DARDANELOS	NORTE	**2.1E-01**	**2.6E-01**	2.0E-03	**7.5E-01**
168	SOBRADINHO	NORDESTE	4.7E-04	9.1E-07	8.4E-17	3.6E-14
188	ITAPEBI	NORDESTE	**1.0E-01**	1.4E-03	4.8E-05	7.0E-08
190	B. ESPERANCA	NORDESTE	**9.9E-01**	**6.9E-01**	**2.0E-01**	7.4E-03
254	P.CAVALO	NORDESTE	**5.2E-02**	1.6E-06	9.5E-05	3.8E-09
271	ESTREITO	NORDESTE	**4.1E-01**	**4.4E-01**	1.1E-02	2.3E-03
1	CAMARGOS	SUDESTE	**1.8E-01**	**7.5E-02**	3.5E-03	6.1E-05
6	FURNAS	SUDESTE	**8.0E-01**	**9.6E-01**	**7.6E-01**	3.7E-02
14	CACONDE	SUDESTE	**1.8E-01**	2.5E-02	8.7E-03	**7.2E-01**
17	MARIMBONDO	SUDESTE	1.3E-02	1.2E-03	1.2E-04	**3.0E-01**
18	A. VERMELHA	SUDESTE	1.5E-02	4.9E-04	1.9E-05	**2.8E-01**
24	EMBORCACAO	SUDESTE	**3.3E-01**	**1.8E-01**	**5.2E-02**	8.0E-03
25	NOVA PONTE	SUDESTE	2.1E-02	4.6E-02	2.1E-04	1.2E-03
31	ITUMBIARA	SUDESTE	**1.2E-01**	**1.8E-01**	9.3E-03	7.1E-03
32	CACH.DOURADA	SUDESTE	**1.9E-01**	**3.4E-01**	**7.1E-02**	2.2E-02
33	SAO SIMAO	SUDESTE	2.2E-03	1.4E-04	1.4E-07	1.5E-02
34	I.SOLTEIRA	SUDESTE	3.2E-04	2.2E-05	7.5E-10	3.0E-02
47	A.A.LAYDNER (JURUMIRIM)	SUDESTE	1.4E-14	2.0E-17	9.8E-15	6.6E-19
61	CAPIVARA	SUDESTE	4.5E-18	1.8E-24	0.0E+00	2.0E-25
63	ROSANA	SUDESTE	4.5E-17	1.7E-25	0.0E+00	2.2E-24
117	GUARAPIRANGA	SUDESTE	4.1E-07	1.7E-05	2.6E-03	4.8E-06
119	BILLINGS_PED	SUDESTE	**1.0E-01**	**3.7E-01**	**7.5E-01**	**6.8E-02**
120	JAGUARI	SUDESTE	4.7E-10	2.0E-11	4.2E-24	2.6E-16
121	PARAIBUNA	SUDESTE	**9.6E-01**	**8.0E-01**	**3.8E-01**	**1.6E-01**
125	STA CECILIA	SUDESTE	**5.7E-01**	**8.3E-01**	**5.1E-01**	2.1E-02

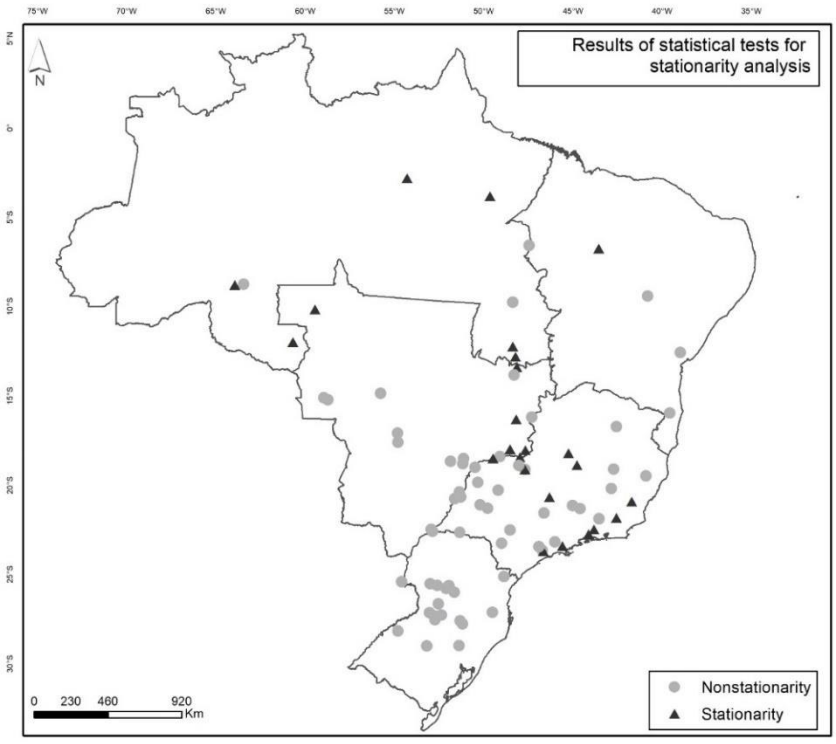
[2] The results that rejected the null hypothesis (p-value>0,05) are highlighted in Table 1, thus defining if the series was classified as stationary/nonstationary.

Code	Station	Region	Student's[2] t	Wilcoxon[2]	Cox-Stuart[2]	Mann-Kendall[2]
130	I. POMBOS	SUDESTE	**1.2E-01**	**4.2E-01**	**2.7E-01**	2.9E-02
134	SALTO GRANDE	SUDESTE	2.1E-02	3.6E-05	3.7E-09	2.5E-13
144	MASCARENHAS	SUDESTE	8.9E-03	7.3E-05	3.5E-06	8.7E-15
149	CANDONGA	SUDESTE	**8.5E-02**	3.0E-02	3.1E-03	2.5E-07
155	RETIRO BAIXO	SUDESTE	**1.2E-01**	**5.7E-01**	**8.6E-01**	1.1E-02
156	TRES MARIAS	SUDESTE	**8.8E-01**	**1.9E-01**	**1.0E-01**	1.4E-03
158	QUEIMADO	SUDESTE	1.2E-02	3.3E-03	1.9E-03	2.7E-08
160	ALTO TIETÊ	SUDESTE	**3.7E-01**	**2.9E-01**	**1.1E-01**	3.3E-02
191	CANA BRAVA	SUDESTE	**5.0E-01**	**1.0E+00**	**8.9E-01**	8.6E-03
196	ROSAL	SUDESTE	**1.1E-01**	**2.5E-01**	1.3E-03	**5.5E-01**
197	PICADA	SUDESTE	3.9E-03	4.5E-03	9.4E-07	5.4E-06
201	TOCOS	SUDESTE	**5.8E-01**	**2.5E-01**	**3.4E-01**	3.6E-02
205	CORUMBA IV	SUDESTE	**9.7E-02**	**2.7E-01**	2.1E-02	**1.5E-01**
206	MIRANDA	SUDESTE	**6.7E-02**	**2.5E-01**	1.5E-03	1.5E-02
209	CORUMBA I	SUDESTE	**1.9E-01**	**3.7E-01**	3.8E-02	**2.6E-01**
220	MONJOLINHO	SUDESTE	4.2E-06	1.3E-05	1.4E-05	5.3E-09
237	BARRA BONITA	SUDESTE	5.6E-07	5.9E-10	6.7E-09	7.3E-06
240	PROMISSAO	SUDESTE	1.7E-10	3.2E-17	0.0E+00	2.6E-10
242	NAVANHANDAVA	SUDESTE	5.3E-11	6.0E-18	0.0E+00	1.8E-10
243	T.IRMAOS	SUDESTE	6.2E-11	3.0E-18	0.0E+00	2.0E-10
245	JUPIA	SUDESTE	3.1E-07	2.9E-10	0.0E+00	2.8E-05
246	P.PRIMAVERA	SUDESTE	1.1E-10	7.9E-14	0.0E+00	3.7E-07
247	CACU	SUDESTE	4.5E-06	1.0E-05	8.9E-16	**7.3E-02**
251	SERRA FACAO	SUDESTE	**5.6E-01**	**2.9E-01**	9.2E-03	2.6E-02
253	SAO SALVADOR	SUDESTE	**5.3E-01**	**9.1E-01**	**1.0E+00**	1.1E-02
255	IRAPE	SUDESTE	**1.0E-01**	3.9E-04	1.3E-08	5.2E-07
257	PEIXE ANGIC	SUDESTE	**2.7E-01**	**6.1E-01**	**1.2E-01**	9.3E-04
259	ITIQUIRA I	SUDESTE	1.2E-14	1.0E-20	0.0E+00	4.3E-09
270	SERRA MESA	SUDESTE	**2.3E-01**	**4.2E-01**	1.5E-02	1.1E-03
273	LAJEADO	SUDESTE	**2.8E-01**	**3.9E-01**	3.4E-02	6.4E-04
278	MANSO	SUDESTE	4.3E-03	**1.1E-01**	8.9E-03	**5.4E-01**
281	PONTE PEDRA	SUDESTE	1.5E-51	1.4E-46	0.0E+00	1.3E-20
283	STA CLARA MG	SUDESTE	**2.2E-01**	8.4E-04	1.9E-02	1.3E-08
294	SALTO	SUDESTE	1.9E-07	8.8E-08	0.0E+00	1.4E-02
295	JAURU	SUDESTE	1.3E-12	4.3E-13	3.2E-22	1.2E-29
296	GUAPORE	SUDESTE	1.3E-04	4.5E-07	5.0E-10	2.4E-02
71	STA CLARA PR	SUL	6.1E-06	1.2E-06	2.9E-04	2.7E-10
72	FUNDAO	SUL	4.0E-06	6.4E-07	1.3E-04	8.4E-11
73	JORDAO	SUL	4.1E-08	2.6E-09	1.6E-06	6.4E-14
74	G.B.MUNHOZ	SUL	1.7E-06	2.5E-06	2.7E-05	5.8E-09
76	SEGREDO	SUL	2.2E-09	4.5E-10	9.3E-07	6.1E-14
77	SLT.SANTIAGO	SUL	8.0E-10	2.0E-10	9.3E-07	3.2E-14
78	SALTO OSORIO	SUL	8.4E-10	2.8E-10	9.3E-07	3.1E-14
92	ITA	SUL	7.2E-06	9.0E-08	8.8E-08	5.0E-10

Code	Station	Region	Student's[2] t	Wilcoxon[2]	Cox-Stuart[2]	Mann-Kendall[2]
93	PASSO FUNDO	SUL	2.9E-04	1.2E-04	1.6E-04	1.2E-07
94	FOZ CHAPECO	SUL	1.4E-06	3.0E-08	2.3E-09	3.8E-11
98	CASTRO ALVES	SUL	2.2E-10	8.5E-12	1.2E-11	4.9E-13
99	ESPORA	SUL	1.3E-54	4.5E-54	0.0E+00	9.1E-34
101	SALTO PILAO	SUL	3.1E-15	9.7E-19	7.5E-15	1.5E-21
102	SAO JOSE	SUL	2.5E-10	2.6E-12	9.5E-10	5.7E-14
111	PASSO REAL	SUL	1.7E-03	3.4E-06	1.2E-04	2.2E-07
115	G.P.SOUZA	SUL	1.3E-07	2.4E-08	1.7E-07	5.4E-15
164	E.SOUZA	SUL	1.0E-25	3.9E-28	0.0E+00	2.7E-25
211	FUNIL-GRANDE	SUL	**6.5E-01**	**5.7E-01**	3.8E-02	1.3E-02
215	BARRA GRANDE	SUL	2.5E-05	1.6E-07	1.7E-09	3.9E-09
216	CAMPOS NOVOS	SUL	8.1E-07	3.1E-07	3.7E-09	5.0E-11
266	ITAIPU	SUL	1.1E-29	1.4E-32	0.0E+00	2.1E-26
286	QUEBRA QUEIX	SUL	1.4E-08	4.1E-12	2.7E-09	4.8E-15

Source: Prepared by the author

Figure 6 - Results of statistical tests for stationarity analysis



Source: Prepared by the author

3.5.2 Unit-root tests for stationarity analysis

It was applied four different unit root test widely used in macroeconomics to verify if the streamflow analyzed are stationary or not. Interpreting the value of the statistics given for

the test (Table 2) and comparing to the critical values of each test, the time series were evaluated and classified as stationary or nonstationary. The values highlighted in Table 2 represent the station classified as stationary in each test. The criterion of stationarity analysis of the series was based on the rejection of the null hypothesis of the DF-GLS and the acceptance of the null hypothesis of the KPSS tests, simultaneously, thus, if both tests accepted the conditions, the series would be stationary. This criterion was chosen because the ADF test has a low statistical power and the PP test needs other tests to corroborate your result, depending on the size of available data (Barros, 2017). The spatial representation of the results is presented in the Figure 7.

Table 2 – Results of the unit-root tests applied to the Base Posts time series

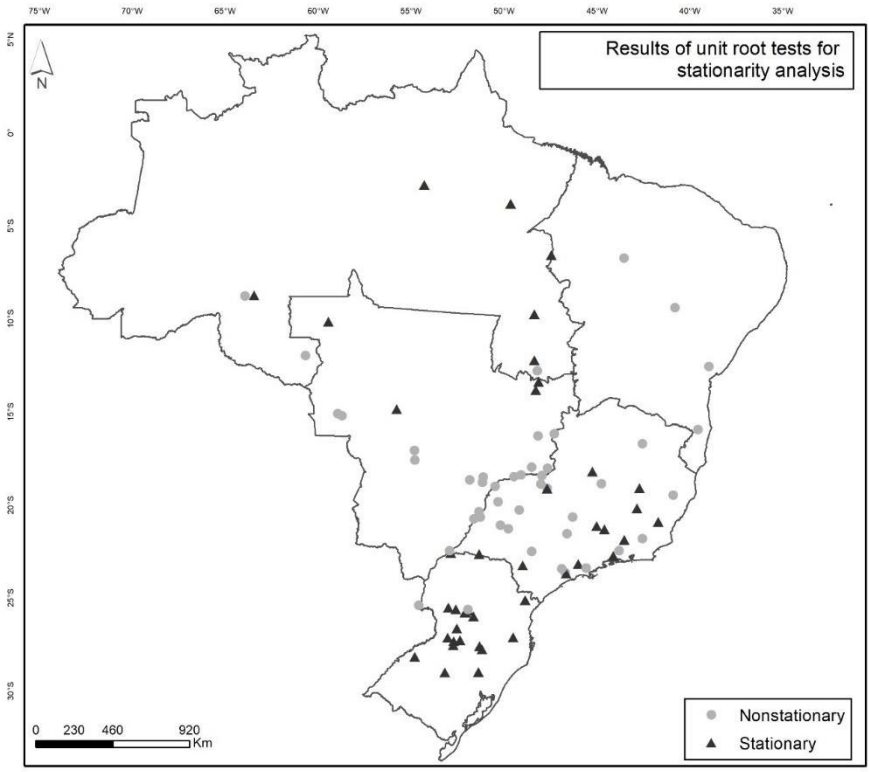
Cod	Station	Region	ADF	KPSS	PP	DF-GLS
145	RONDON II	NORTE	**-3.8**	0.2	**-13.0**	-2.3
275	TUCURUI	NORTE	**-5.4**	**0.1**	**-10.4**	**-3.7**
277	CURUA-UNA	NORTE	**-5.3**	**0.1**	**-10.9**	**-4.3**
279	SAMUEL	NORTE	**-6.7**	**0.0**	**-10.5**	**-6.1**
287	STO ANTONIO	NORTE	**-4.9**	0.2	**-9.9**	-2.5
291	DARDANELOS	NORTE	**-6.4**	**0.0**	**-10.4**	**-6.4**
168	SOBRADINHO F	NORDESTE	**-4.6**	0.2	**-12.1**	**-3.2**
188	ITAPEBI	NORDESTE	**-5.0**	0.4	**-16.3**	**-4.7**
190	B. ESPERANCA	NORDESTE	**-4.2**	0.3	**-11.9**	**-3.9**
254	P.CAVALO	NORDESTE	**-7.7**	0.2	**-20.9**	**-5.7**
271	ESTREITO TOC	NORDESTE	**-5.5**	**0.1**	**-11.2**	**-4.7**
1	CAMARGOS	SUDESTE	**-5.5**	**0.1**	**-12.5**	**-4.4**
6	FURNAS	SUDESTE	**-4.4**	0.2	**-12.2**	**-3.1**
14	CACONDE	SUDESTE	**-5.0**	**0.1**	**-12.5**	-1.7
17	MARIMBONDO	SUDESTE	**-4.7**	0.2	**-11.4**	-2.4
18	A. VERMELHA	SUDESTE	**-4.8**	0.2	**-11.4**	-2.2
24	EMBORCACAO	SUDESTE	**-5.0**	**0.1**	**-12.1**	-2.8
25	NOVA PONTE	SUDESTE	**-4.8**	**0.1**	**-11.5**	-2.4
31	ITUMBIARA	SUDESTE	**-5.0**	**0.1**	**-11.5**	-2.4
32	CACH.DOURADA	SUDESTE	**-5.1**	**0.1**	**-11.5**	-2.4
33	SAO SIMAO	SUDESTE	**-4.8**	**0.1**	**-11.3**	-2.4
34	I.SOLTEIRA	SUDESTE	**-4.7**	0.2	**-11.1**	-2.2
47	A.A.LAYDNER (JURUMIRIM)	SUDESTE	**-6.3**	**0.1**	**-15.5**	**-4.0**
61	CAPIVARA	SUDESTE	**-7.1**	**0.1**	**-17.1**	**-4.8**
63	ROSANA	SUDESTE	**-7.0**	**0.1**	**-17.4**	**-4.7**
117	GUARAPIRANGA	SUDESTE	**-5.5**	0.2	**-16.4**	-2.2
119	BILLINGS_PED	SUDESTE	**-6.6**	**0.1**	**-17.8**	**-4.9**
120	JAGUARI	SUDESTE	**-5.1**	**0.0**	**-12.3**	**-3.3**

Cod	Station	Region	ADF	KPSS	PP	DF-GLS
121	PARAIBUNA	SUDESTE	**-5.2**	0.2	**-12.6**	**-3.2**
125	STA CECILIA	SUDESTE	**-5.2**	**0.1**	**-12.3**	-2.4
130	I. POMBOS	SUDESTE	**-4.9**	**0.0**	**-11.9**	-2.7
134	SALTO GRANDE	SUDESTE	**-5.5**	**0.1**	**-13.9**	**-3.9**
144	MASCARENHAS	SUDESTE	**-5.3**	0.1	**-13.4**	**-4.0**
149	CANDONGA	SUDESTE	**-5.3**	**0.1**	**-13.1**	**-3.8**
155	RETIRO BAIXO	SUDESTE	**-4.6**	0.3	**-13.3**	**-4.6**
156	TRES MARIAS	SUDESTE	**-4.9**	**0.1**	**-12.7**	**-4.4**
158	QUEIMADO	SUDESTE	**-4.3**	0.2	**-13.3**	**-3.9**
160	ALTO TIETÊ	SUDESTE	**-7.0**	**0.0**	**-16.1**	**-3.6**
191	CANA BRAVA	SUDESTE	**-5.0**	**0.1**	**-12.5**	**-4.6**
196	ROSAL	SUDESTE	**-5.6**	**0.1**	**-14.0**	**-4.8**
197	PICADA	SUDESTE	**-4.5**	**0.0**	**-12.6**	**-4.4**
201	TOCOS	SUDESTE	**-6.0**	**0.0**	**-12.7**	**-3.5**
205	CORUMBA IV	SUDESTE	**-4.9**	0.3	**-11.5**	**-3.0**
206	MIRANDA	SUDESTE	**-4.7**	**0.1**	**-11.5**	-2.4
209	CORUMBA I	SUDESTE	**-4.8**	0.2	**-11.5**	**-2.9**
220	MONJOLINHO	SUDESTE	**-7.1**	**0.1**	**-17.3**	**-6.8**
237	BARRA BONITA	SUDESTE	**-6.0**	**0.1**	**-13.8**	-2.3
240	PROMISSAO	SUDESTE	**-5.8**	**0.1**	**-13.5**	-2.0
242	NAVANHANDAVA	SUDESTE	**-5.7**	0.2	**-13.4**	-2.0
243	T.IRMAOS	SUDESTE	**-5.7**	0.2	**-13.4**	-2.0
245	JUPIA	SUDESTE	**-4.8**	0.2	**-11.1**	-2.0
246	P.PRIMAVERA	SUDESTE	**-4.5**	0.2	**-10.9**	-2.0
247	CACU	SUDESTE	**-4.7**	0.2	**-11.0**	-2.6
251	SERRA FACAO	SUDESTE	**-4.9**	**0.1**	**-12.2**	-2.8
253	SAO SALVADOR	SUDESTE	**-5.0**	0.1	**-12.3**	**-4.5**
255	IRAPE	SUDESTE	**-5.2**	0.3	**-16.1**	**-4.7**
257	PEIXE ANGIC	SUDESTE	**-5.4**	**0.1**	**-12.5**	**-4.5**
259	ITIQUIRA I	SUDESTE	-3.1	0.3	**-11.5**	**-3.0**
270	SERRA MESA	SUDESTE	**-5.0**	**0.1**	**-12.5**	**-4.8**
273	LAJEADO	SUDESTE	**-5.6**	**0.1**	**-11.8**	**-4.4**
278	MANSO	SUDESTE	**-5.2**	**0.1**	**-12.7**	**-4.8**
281	PONTE PEDRA	SUDESTE	-2.5	0.8	**-7.4**	-2.1
283	STA CLARA MG	SUDESTE	**-5.2**	0.2	**-17.5**	**-4.6**
294	SALTO	SUDESTE	**-4.5**	0.3	**-11.1**	-2.8
295	JAURU	SUDESTE	**-3.8**	0.4	**-11.8**	**-3.2**
296	GUAPORE	SUDESTE	**-4.2**	0.2	**-11.5**	-2.1
71	STA CLARA PR	SUL	**-7.4**	**0.0**	**-20.5**	**-7.0**
72	FUNDAO	SUL	**-7.4**	**0.0**	**-20.5**	**-7.0**
73	JORDAO	SUL	**-7.4**	**0.0**	**-20.1**	**-7.0**
74	G.B.MUNHOZ	SUL	**-7.8**	**0.0**	**-18.6**	**-7.1**
76	SEGREDO	SUL	**-7.8**	**0.0**	**-18.6**	**-7.0**
77	SLT.SANTIAGO	SUL	**-7.7**	**0.0**	**-18.7**	**-7.0**

Cod	Station	Region	ADF	KPSS	PP	DF-GLS
78	SALTO OSORIO	SUL	**-7.7**	**0.0**	**-18.7**	**-7.0**
92	ITA	SUL	**-7.3**	**0.0**	**-18.6**	**-6.8**
93	PASSO FUNDO	SUL	**-7.0**	**0.1**	**-17.0**	**-6.6**
94	FOZ CHAPECO	SUL	**-7.3**	**0.0**	**-18.3**	**-6.8**
98	CASTRO ALVES	SUL	**-6.7**	**0.1**	**-19.9**	**-5.5**
99	ESPORA	SUL	**-3.5**	0.6	**-10.6**	-2.3
101	SALTO PILAO	SUL	**-17.4**	**0.1**	**-20.0**	**-7.0**
102	SAO JOSE	SUL	**-7.4**	**0.1**	**-17.6**	**-7.1**
111	PASSO REAL	SUL	**-6.5**	**0.0**	**-16.2**	**-5.8**
115	G.P.SOUZA	SUL	**-6.4**	**0.0**	**-17.6**	**-4.3**
164	E.SOUZA	SUL	**-5.2**	0.3	**-13.3**	-2.2
211	FUNIL-GRANDE	SUL	**-4.5**	**0.1**	**-12.5**	**-3.1**
215	BARRA GRANDE	SUL	**-7.6**	**0.0**	**-20.0**	**-7.1**
216	CAMPOS NOVOS	SUL	**-17.2**	**0.0**	**-18.3**	**-7.3**
266	ITAIPU	SUL	**-5.4**	0.2	**-11.1**	-2.2
286	QUEBRA QUEIX	SUL	**-7.0**	**0.1**	**-20.0**	**-6.7**

Source: Prepared by the author

Figure 7 - Results of unit root tests for stationary analysis

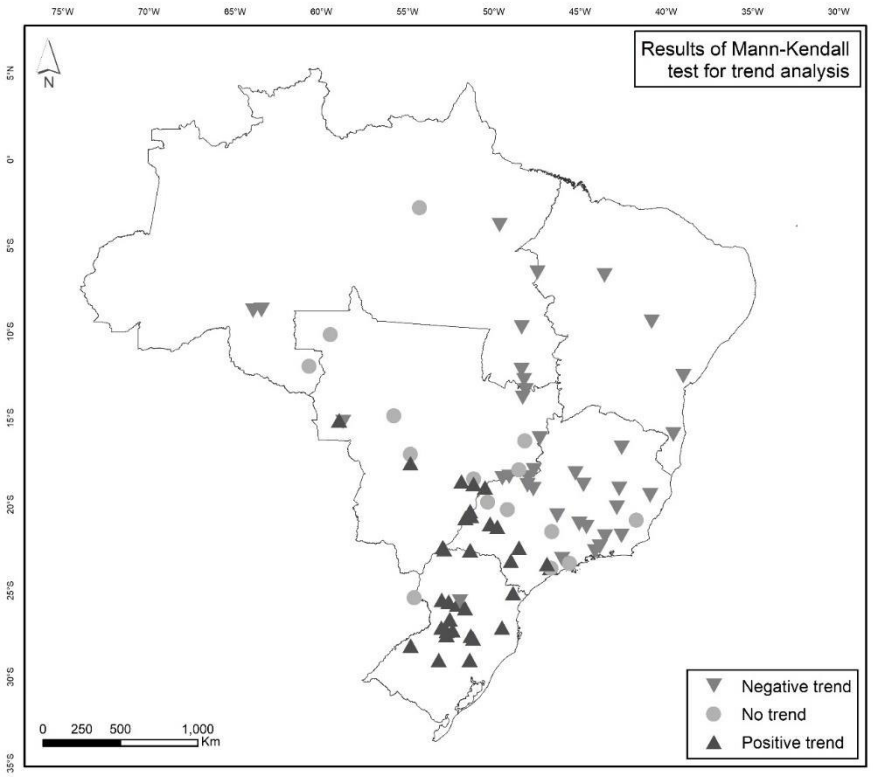


Source: Prepared by the author

3.5.3 Mann-Kendall and Sen's Slope

The Mann-Kendall test was also used to identify the presence of trend and the Sen's slope was applied to verify their magnitude in the studied region. Its spatial results are shown in Figure 8.

Figure 8 - Results of trend analysis



Source: Prepared by the author

All the northeast station presented a significant trend with negative pattern. Marengo and Valverde (2007) observed that Sobradinho, a reservoir located in the northeast region, presented a positive trend from 1931 until approximately 1979, when the trend became negative. That shift may be associated with the rainfall pattern or the multiple uses of water from that reservoir.

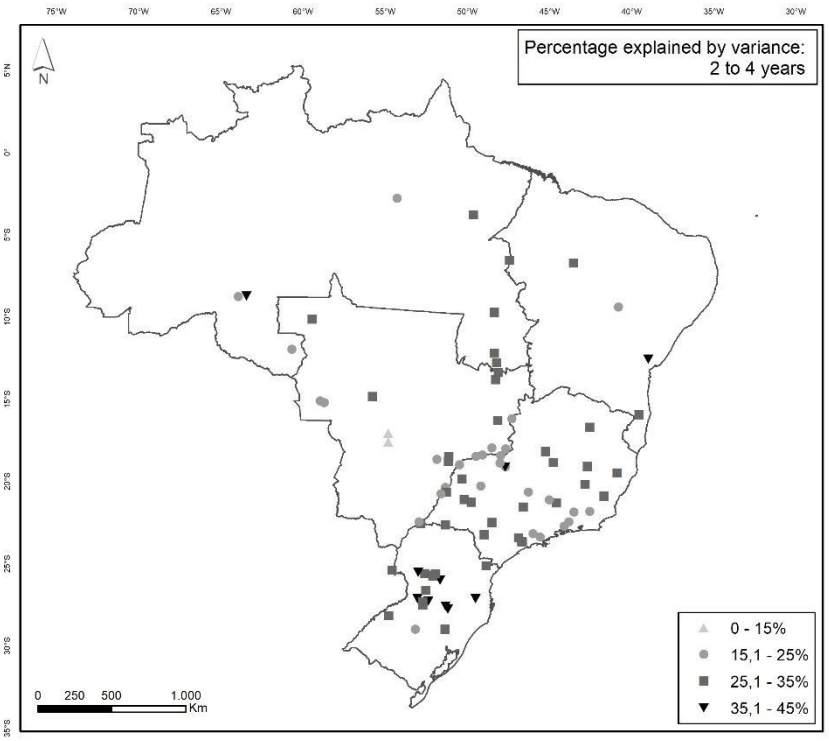
When evaluating the trend in South region streamflow almost all station presented the p-value lower then the significance level adopted, which it was 95% (α=0,05), indicating the presence of a trend in the time series, also, most of the time series presented a positive trend.

These results agree with many previous studies (Pasquini and Depetris, 2007; Antico et al., 2014; Castino et al., 2017) in which the presence of an upward trend was verify.

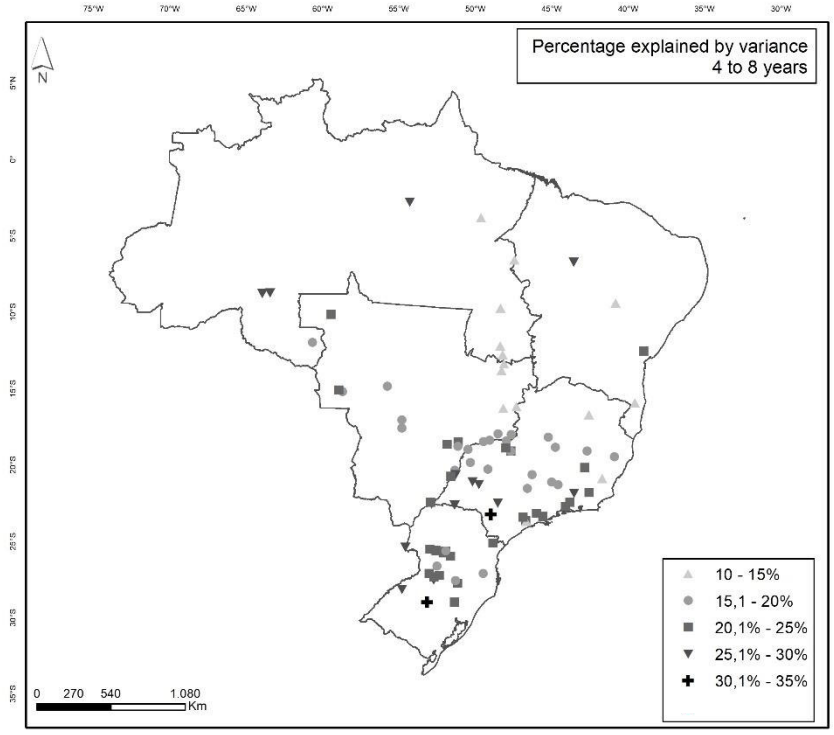
3.5.4 Wavelet Analysis and Complete EEMD With Adaptative Noise

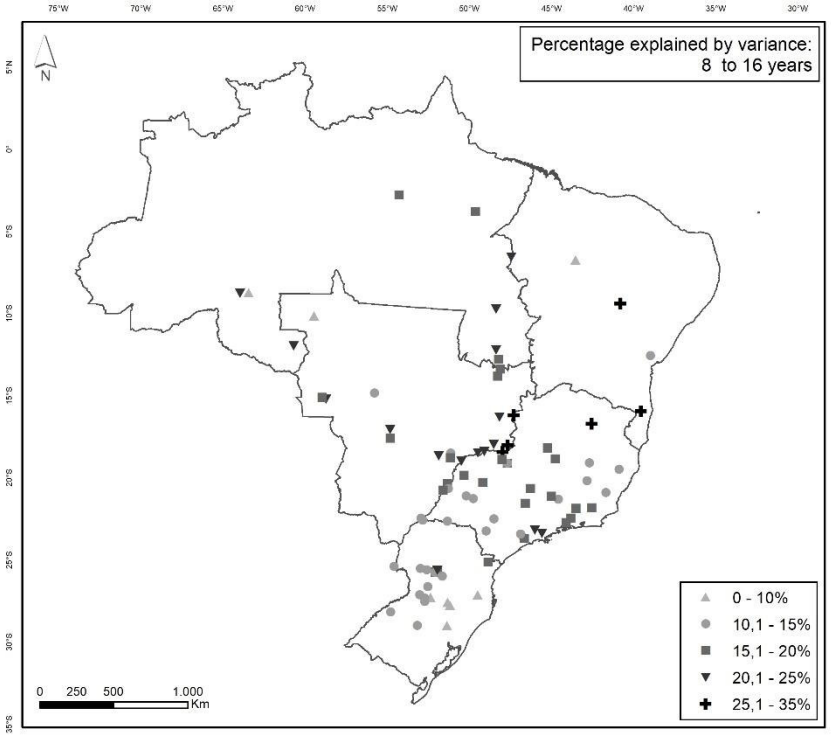
The time series were decompose applying the wavelet transform and the Figure 9 shows the spatial distribution of the explained variance in the high (2-4 years and 4-8 years), medium (8-16 years), low (16-32 years) frequency and the residue, which is represented with the frequency higher than 32 years.

Figure 9 - Percentage of explained variance for the analyzed stations using wavelet transform. (a) Explained variance for the 2-4 years period; (b) Explained variance for the 4-8 years period; (c) Explained variance for the 8-16 years period; (d) Explained variance for the 16-32 years period and (e) Explained variance for the residue (higher than 32 years)

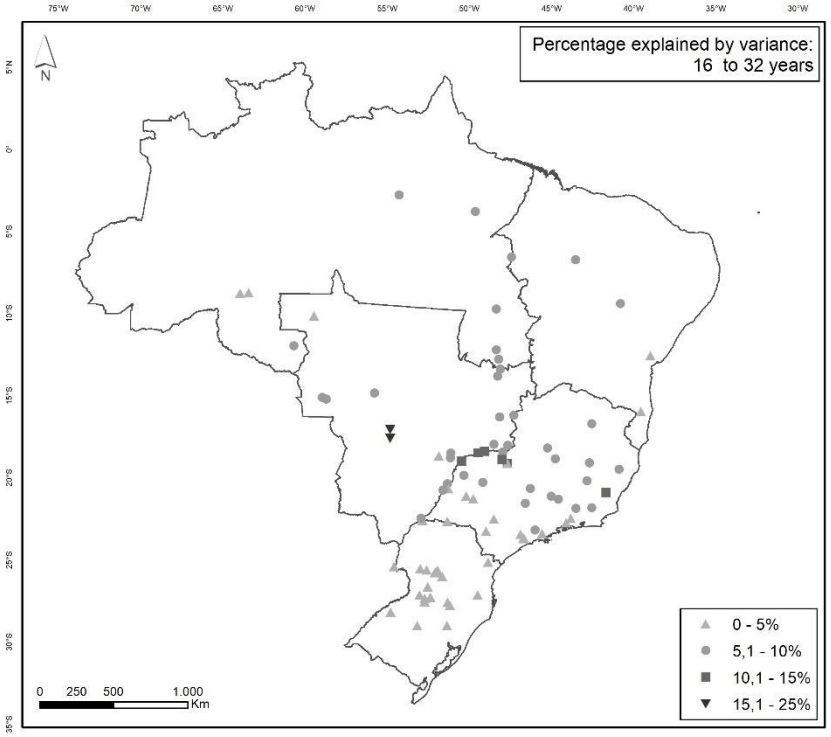


(a)

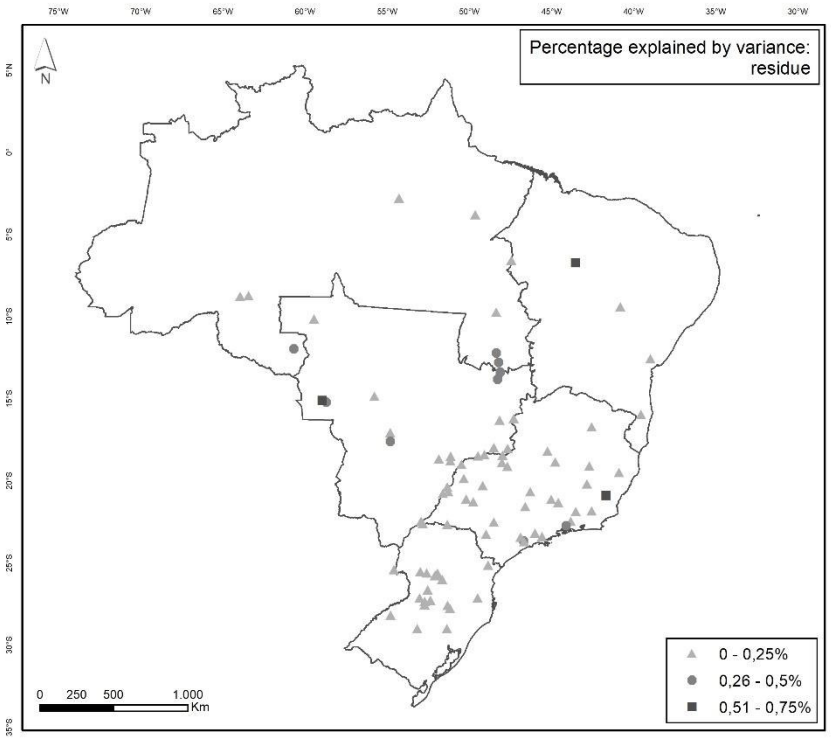

Percentage explained by variance
4 to 8 years
10 - 15%
15,1 - 20%
20,1% - 25%
25,1% - 30%
30,1% - 35%
0 270 540 1.080
Km


Percentage explained by variance:
8 to 16 years
0 - 10%
10,1 - 15%
15,1 - 20%
20,1 - 25%
25,1 - 35%
0 250 500 1.000
Km

(c)



(d)



(e)

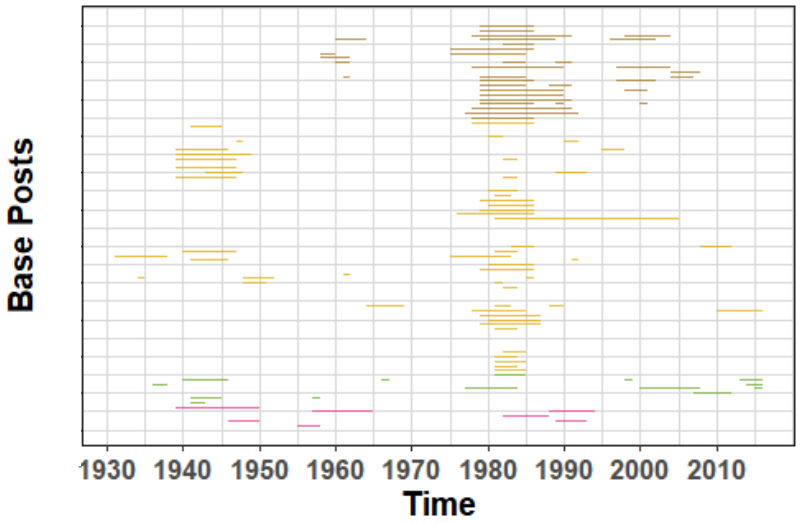
Source: Prepared by the author

It was evaluated each frequency separately. For more than half of the stations analyzed (47 stations), the 2-4 years period represents approximately 25-35% of the entirely time series. In the 4-8 years band, the majority of the stations (31 stations) have an explained variance of 20-25%. For the 8-16 years band, most of the stations (30 stations) were between

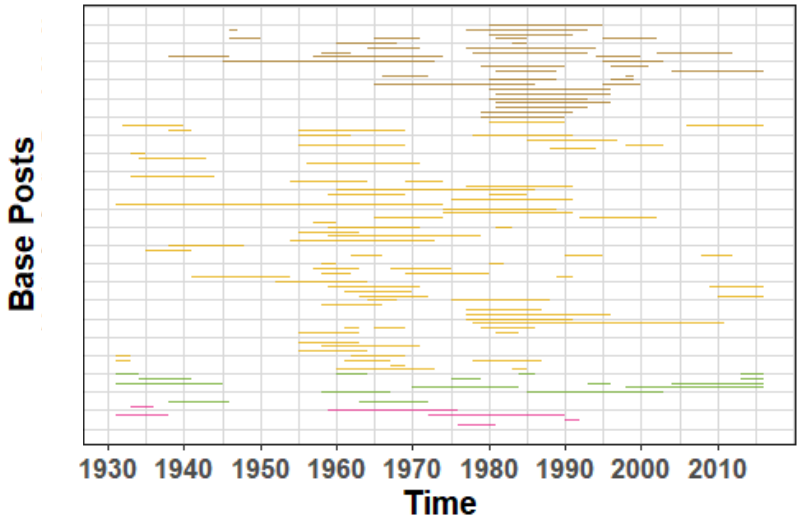
10-15%. For the low frequency band, the explained variance for most of the stations range from 5-10%, and the percentage explained by the residue is under 2.5% in 76 stations.

Using the wavelet transform was evaluated the wavelet power spectrum and verified the years that presented significance (p-value < 0.05) in the analyzed frequencies. The results are summarized in Figure 10 and are organized by regions from South (Top), then Southeast, Northeast and North (Bottom). For the 2-4 years frequency, there is a great significance in the 1980's for almost all the posts. In the 4-8 years frequency, the majority of the posts presented significance between the years of 1960 and 1990, and, for the 8-16 years frequency, there were large periods of years presenting significance. In the low frequency (16-32 years), many reservoirs presented periods with significance, which indicates that this frequency was important, especially for reservoirs in the southeast until the 1980s.

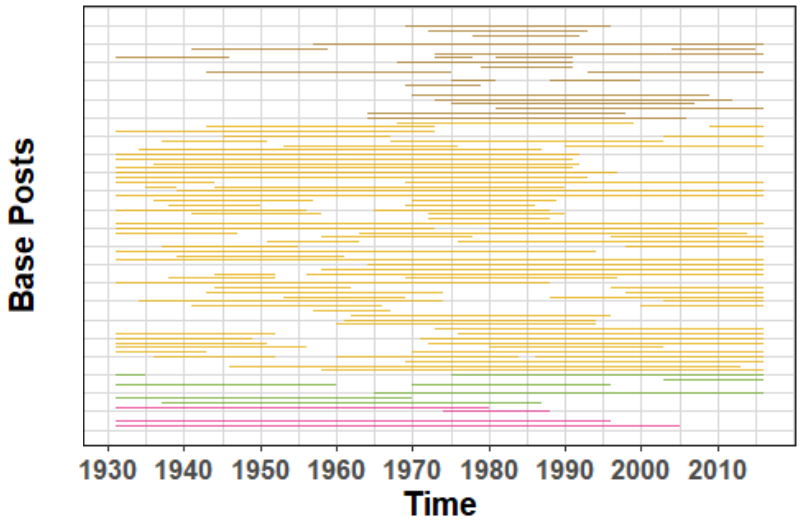
Figure 10 - (a) Years with significance in the 2-4 years frequency, (b) Years with significance in the 4-8 years frequency, (c) Years with significance in the 8-16 years frequency, (d) Years with significance in the 16-32 years frequency. The figure is organized by region from South (Top), then Southeast, Northeast and North (Bottom)



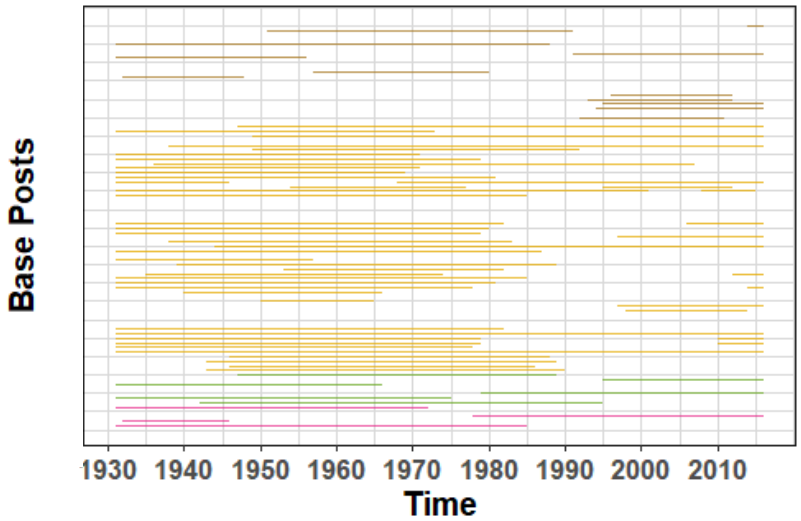
(a)



(b)



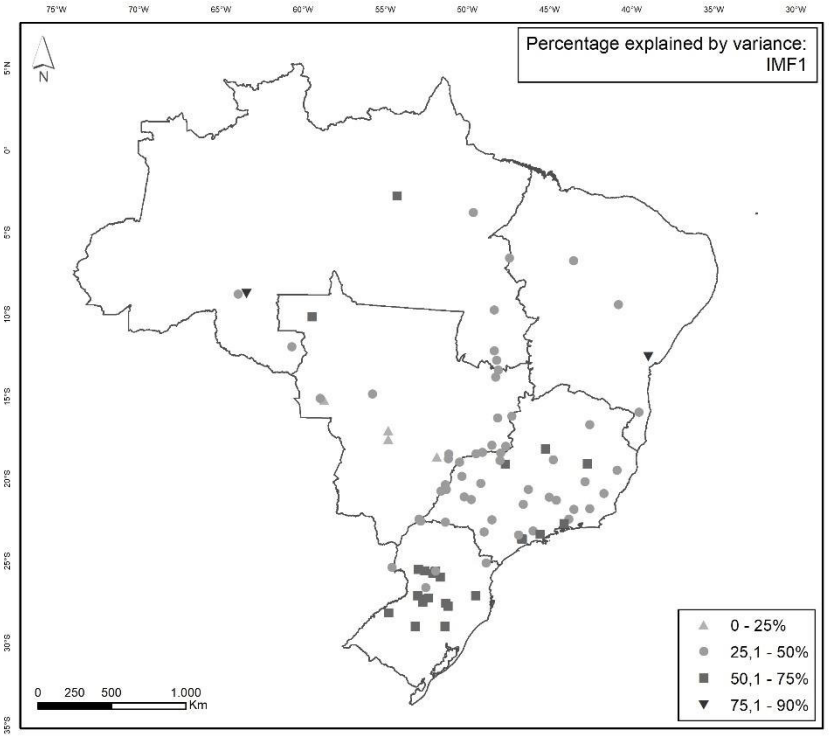
(c)



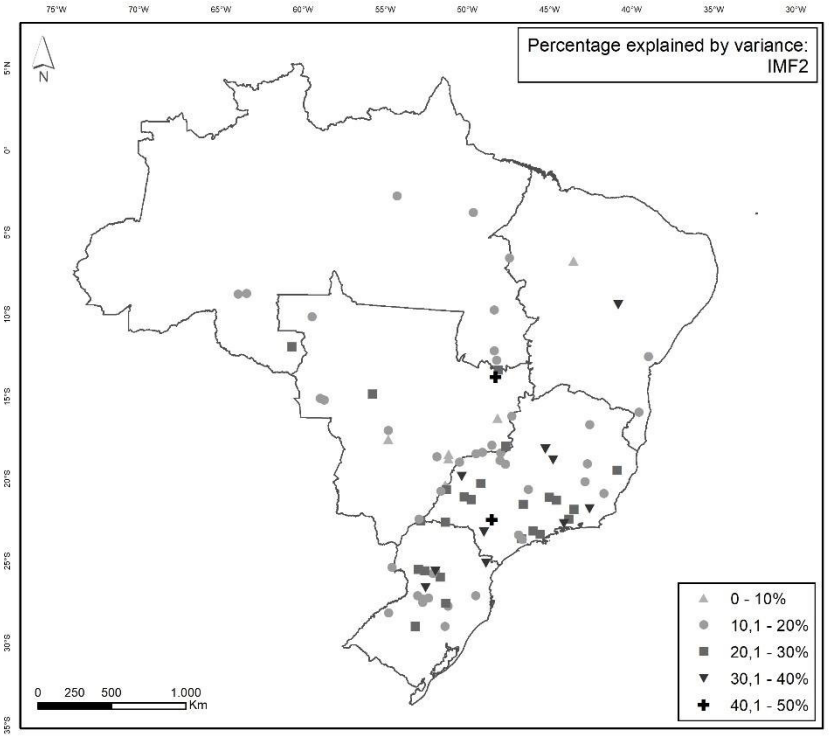
(d)

Source: Prepared by the author

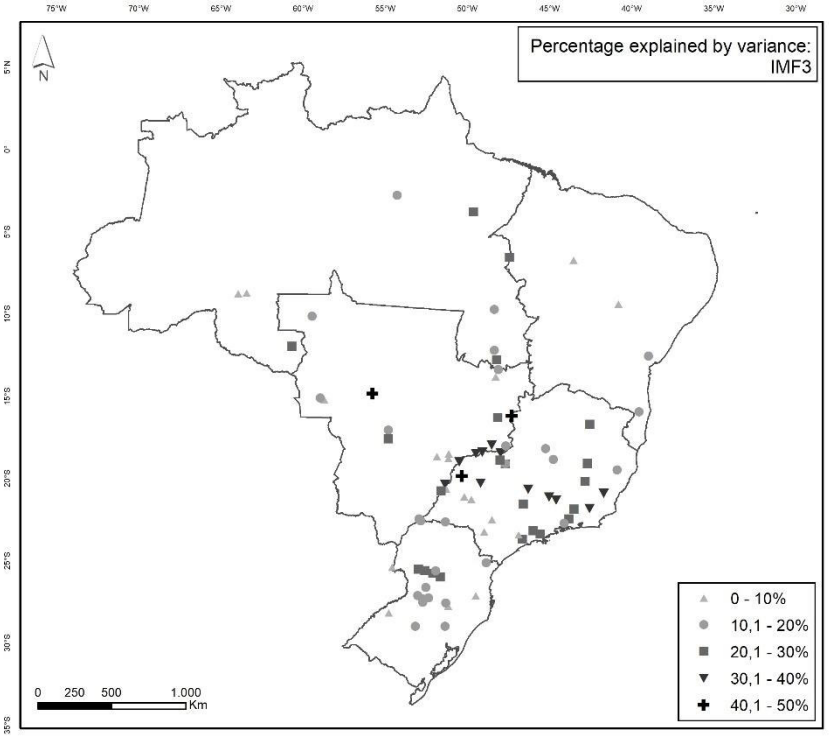
Figure 11 - Percentage of explained variance using CEEMDAN decomposition. (a) Explained variance for IMF1 (2-4 years); (b) Explained variance for the IMF2 (5-13 years); (c) Explained variance for IMF3 (13-20 years); (d) Explained variance for IMF4 (16-25 years) and (e) Explained variance for IMF5 (higher than 25 years)



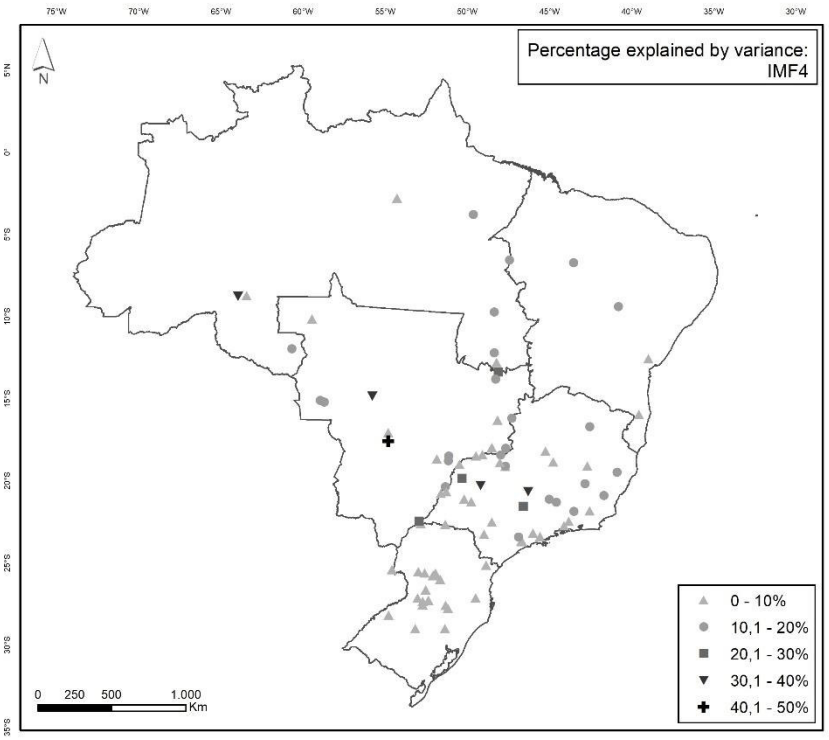
(a)



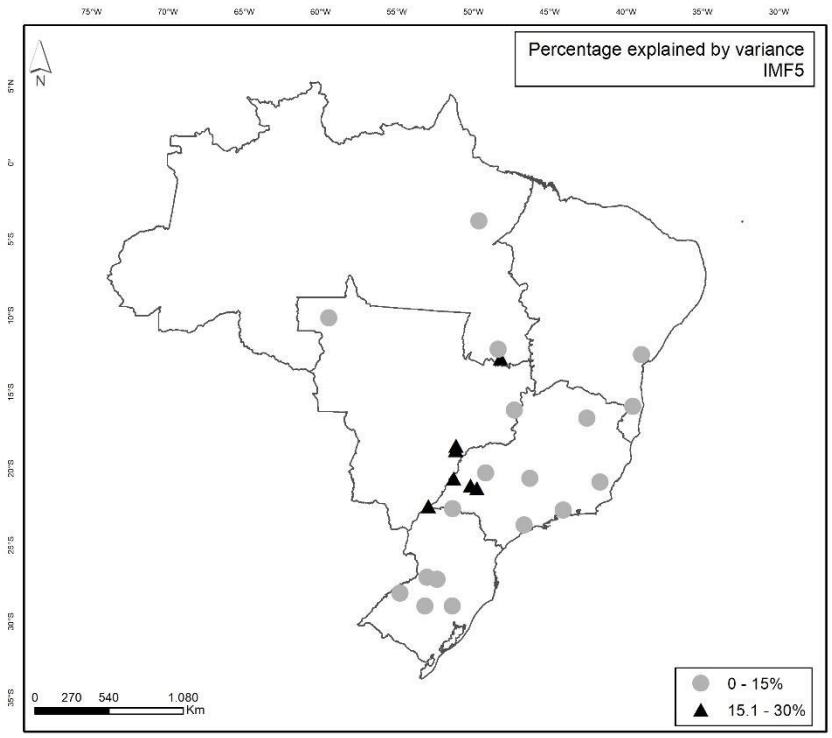
(b)


75°W
70°W
65°W
60°W
55°W
50°W
45°W
40°W
35°W
30°W
5°N
0°
5°S
10°S
15°S
20°S
25°S
30°S
35°S
N
Percentage explained by variance:
IMF3
0 - 10%
10,1 - 20%
20,1 - 30%
30,1 - 40%
40,1 - 50%
0 250 500 1.000
Km

(c)


75°W
70°W
65°W
60°W
55°W
50°W
45°W
40°W
35°W
30°W
5°N
0°
5°S
10°S
15°S
20°S
25°S
30°S
35°S
N
Percentage explained by variance:
IMF4
0 - 10%
10,1 - 20%
20,1 - 30%
30,1 - 40%
40,1 - 50%
0 250 500 1.000
Km

(d)



(e)

Source: Prepared by the author

In the CEEMDAN decomposition, most stations resulted in 4 IMFs and the percentage of each explained variance is spatially shown in Figure 11. The IMF are not necessary linked to a specific period, however it was used the wavelet transform through the power spectrum, verifying the most significant period, thus the periodicity of the series is the result of the wavelet analysis and each IMF range from similar periodicity in this analysis. IMF 1 corresponds to a period of 2-4 years, IMF 2 ranges from 5-13 years, IMF 3 ranges from 13-20 years, IMF 4 ranges from 16-25 years and IMF 5 has 25 years as the most dominant period.

In the IMF1 analysis, the majority of the stations (57 stations) have an explained variance ranging from 25-50%. For 46 of the stations, the IMF2 explains 10-20% of the time series. In the IMF3, 31 of the stations have an explained variance between 10-20% and in the IMF4, for 46 stations, this period explains about less than 10% of the series. Only 25 stations generated a fifth IMF and for most of them the variance explained less than 10% of the series.

3.6 Analysis of the Results

Each station was analyzed individually and then a more general analysis was made in order to make a diagnosis by geographic regions. The south region of Brazil did not present any stationarity series in the statistical tests. Similar results were found using statistical tests by

Detzel et al. (2011). The southeast region presented most of its series nonstationary. Northeast region presented mainly stationary time series. Using the unit root tests, the results were similar to the statistical tests except for the south region, which present almost all station stationary. When evaluating trend, it was noted that the south and southeast region both presented positive trends, while the north and the northeast region presented negative trends.

The different results in the stationary analysis of the two types of tests in the south sector might be related to the form that unit root tests treat trends. In the unit root regression of the DF-GLS test, the time series was de-trended, so the lower p-value in the other tests might just be an indicative of the presence of strong positive trend in the south of the country.

Table 3 - Summary of the results from the stationarity and trend tests

Region	Statistical tests		Unit root tests		Trend Test		
	Non-Stationary	Stationary	Non-Stationary	Stationary	No trend	Positive Trend	Negative Trend
SOUTHEAST	37	18	36	19	7	18	30
NORTHEAST	4	1	4	1	0	0	5
NORTH	1	5	2	4	3	0	3
SOUTH	22	0	3	19	0	21	1

Source: Prepared by the author

Many authors (Marengo et al, 2007; Alves et al. 2013) observed a systematically change in hydrological time series in the late 70s and early 80's which may be more associated with the natural fluctuations like decennial variability than with the linear trend in time series, thus, the need to decompose the time series and analysis more closely its variations patterns.

Figure 9 and 11 shows that through the wavelet analysis and CEEMDAN decomposition the high frequency component, which ranges from 2-8 years, is responsible for more than 50% of the explained variance in all analyzed regions. Similar results were found by Alves et al (2013). In the CEEMDAN analysis, this percentage can reach even higher values such as 60 or 70% of the explained variance. Analysing the low frequency, the south region has the lowest values of explained variance and the southeast has the highest values in both types of decomposition.

3.7 Conclusion

Understanding the extent to which natural streamflow patterns have been modified is an important consideration for water resources studies of streams. Assessing hydrological time series requires that we quantify and analyse the attributes of the series for a deeply understating of its past behaviour and the causes that might be related to the changes in its patterns, such as climate fluctuations and anthropogenic modifications.

In this study were used monthly time series to evaluate the performance of statistical test (Mann-Kendal, Wilcoxon, Student's t, Cox-Stuart tests) and unit root test (DF, KPSS, PP and DF-GLS tests) to verify the presence of stationarity and trend analysis (Mann-Kendall and Sen's Slope test). Furthermore, the time series were decomposed into different band of frequency (2-4, 4-8, 8-16, 16-32 years and residue) using wavelet transform and using CEEMDAN method into 5 IMFs and evaluated the explained variance using both methods.

In general, the series located in the South and Southeast regions shown indications of nonstationary using the statistical tests. Using the unit root tests, the results were similar to the statistical tests except for the south region, which it presented almost all station stationary, which may be associated with the form that the unit root test treat the trend component of the time series, thus, indicating that unit root test are more appropriated to be used in the analysis of stationarity for hydrological time series. When evaluating trend by the Mann-Kendall test, it was noted that the south and southeast region both presented positive trends, while the north and northeast presented negative trends.

Analysing the decomposed time series by the Morlet wavelet and the CEEMDAN method, most of the series have higher explained variance for the period of 2-8 years, ranging from 50-60%, while for the medium frequency (8-16) the explained variance is around 10-20% and for the low frequency (16-32 years) it represents 5-10% of the time series. Those decomposition methods have proved to be important tools for the identification of climatic variability of an important sector for the country's economy, forming a diagnostic of the behaviour of the streamflow time series. The methodologies presented are useful for hydrological studies in regions affected by climatic variability and assist in modeling the comportment of hydrological series.

4 PLURIANNUAL STREAMFLOW FORECASTING USING HIDDEN MARKOV MODEL

4.1 Abstract

An improved water resources managing must rely on an accurate identification and analysis of the past dynamic of the hydrological time series. As records of time series lengths increased, the presence of low frequency structures of climate and hydrologic time series became an important feature in hydrological analysis. An ensemble of techniques was applied to identify extreme events in the low frequency (16-32 years), it was decomposed by the wavelet transform. It was used the Standard Runoff Index (SRI), changepoint detection and a Hidden Markov Model (HMM) in order to identify the different states of the Sobradinho's reservoir. Then, an Autoregressive model and a HMM were applied to develop a forecast model. The analysis with the changepoint showed a better methodology in the identification of extreme events in relation to SRI, because it can identify significant changes in the series statistics. It was evaluated the cumulative distribution function using the original time series and it was observed that they have different distributions, hence the proposes of identifying the states and forecast the next state was justified. A model for forecasting was developed adjusting an autoregressive model to the low frequency time series. It demonstrated good results for the next state forecast, however it worked poorly for longer periods, with a relative error higher than 50%. In the HMM model the future state were calculated based on the posterior probability. The results shown a reduction of the wet period between the years 2010-2016 and a greater probability of a normal period was verified. It is possible to identify the next most probable state for the Sobradinho's reservoir and the next state was a good indication for the streamflow distribution compared to the historical time series. Thus, HMM forecast proved to be an important tool to assist in the management and operation of the reservoir, especially for the one in this study due to its importance of generating energy for the region.

4.2 Introduction

In the face of a changing environment, many water resources studies applied stochastic methods where temporal uncertainties require to be quantified (Kwon et al., 2007). The early time series models assumed (Thomas and Fiering, 1962; Yevjevich, 1987; Salas,

1980), as a simplification, that the time series were stationary, therefore the stochastic hydrology revolved around the families of autoregressive and moving average models. However, many recent papers questions that assumption and comment that the past will no longer resemble the future, thus the use of models that account for nonstationary analysis are necessary (Milly et al., 2008; Salas and Obeysekera, 2013; Cheng and Aghakouchak; 2014).

One important characteristic in historical hydroclimatic time series analysis is to identify the past dynamics aiming to understand the behaviour of the series and make future predictions. An approach that can be used is to verify the presences of periodicity and cycles that can be reflected throughout the historical series (Hoek and Bos 2007), so if an event has a high probability of occurring one can better understand its dynamical background and find a deterministic behavior that can describe its future value. As records lengths increased, the presence of low frequency structures of climate and hydrologic time series became an important feature in hydrological analysis (Kwon et al., 2007). Especially in streamflow analysis, due to the extremely nonuniform temporal distribution of global runoff.

The investigation of hydrological process such as low frequency oscillations and their direct impacts around the globe have motivated several studies showing that these oscillations are driven primarily by the climate (Kalra et al., 2013). Climate variations in this time frame have increase the occurrence of extreme events, such as droughts and floods, also, these variations control water availability, affect ecosystems and modulate higher frequency variability, thus it have a major social and economical impact (Jain and Lall, 2000; Milly et al., 2002; Kwon et al., 2008; Grimm and Sabóia, 2015).

Drought is an extreme event of considerable importance, especially in the Brazilian Northeast, where precipitations are scarce and it can lead to diminishing in water supply, water quality deterioration and reduced power generation (Mishra and Singh, 2010). Thus, the evaluation of droughts is an important part in planning and operating a water resources system. Drought indices are a key component for monitoring extreme events. There is a large number of drought indicators that have been developed in the past decades (Mishra and Singh, 2010; Liu et al.,2012; Hao et al, 2016), such as Standardized Precipitation Index (SPI) (McKee et al., 1993), which is commonly used for meteorological drought monitoring, and its derivation Standardized Precipitation Evapotranspiration Index (SPEI), Standardized Soil Moisture Index (SSI), Standardized Runoff Index (SRI) (Hao et al., 2014, Mo, 2011, Shukla and Wood, 2008, Vicente-Serrano et al., 2010), based on the distribution of various hydro-climatic variables. The SRI index is a variation of the SPI index, using as an input variable the monthly flows, aiming

to define and classify events of hydrological droughts, and can be used in a complementary way to SPI, which contemplates only meteorological aspects (Shukla and Wood, 2008).

Besides drought indexes, another methodology may be applied to identify extreme events, that consist in estimating the points where the statistical properties of a series of observations change, also known as changepoint detection (Kickick and Eckley, 2014). Hydrological time series are in constant change and there is a necessity in efficiently identifying and accurately estimate the location of multiple change point along historical time series. Fritier et al. (2012) applied the changepoint detection using the PELT segmentation method to evaluate long-term changes in the North Atlantic Oscillation (NAO) index and in a precipitation series. Li et al. (2017) analyzed a streamflow series using changepoint detection method to evaluate significant changes in the series.

Applications of methods that reproduce the spectral signature of time series and can assimilate those low frequencies, such as wavelet analysis, which are paramount for hydrological time series studies. Many papers applied continuous wavelet analysis for understanding multi-temporal scale characteristics of hydrological time series (Torrence and Compo, 1998; Coulibaly and Burn, 2004; Massei et al., 2010) and coupled those models with climate indices to better their analysis and forecast of future time series (Anctil and Coulibaly, 2004; Massei et al., 2010; Erkyihun et al., 2016). Studies showed that the inclusion of such indices can improve the ability of explain the different variability regimes (Souza Filho and Lall, 2003; Silveira et al., 2017). It is well known that climate indices, such as El Niño-Southern Oscillation (ENSO), the Pacific Decadal Oscillation (PDO) and Atlantic Multidecadal Oscillation (AMO) are linked to rainfall pattern in Northeast Brazil (Kayano and Andreoli, 2004; Knight et al., 2006; Kayano and Andreoli, 2009).

In this paper, it was proposed to investigate the low frequency of a streamflow time series, which was decompose using a continuous wavelet analysis. The approach of this paper is separated into two phases: (1) exploratory analysis of the low frequency variability and classification of dry/wet periods using SRI, the changepoint detection and a Hidden Markov Model and (2) the development of a forecast model using an Autoregressive model and a Hidden Markov Model.

4.2 Data and methods

The Sobradinho reservoir is located at the São Francisco hydrographic basin (BHSF). The basin has an area of approximately 639.219 km², which is equivalent to about 8% of the country. The São Francisco River has a length of 2,863 km, stretching across six Brazilian states (Minas Gerais, Goiás, Bahia, Pernambuco, Alagoas e Sergipe) and the Federal District. The São Francisco River originates in Minas Gerais and runs 2,863 km to the mouth of the Atlantic Ocean, where its flow is greater than 854 m³/s 95% of the time (ANA, 2016). For planning purposes, the basin was divided into four physiographic regions: high (16% of the basin area), medium (63%), sub-medium (17%), where Sobradinho is located, and low (4%).

The water and climate characteristics of the basin are highly variable. The Sobradinho reservoir presents critical periods of prolonged droughts, as a result of low rainfall and high evapotranspiration. The rainfall season starts approximately in November and ends in April. Due to the high intra- and interannual rainfall variations in the basin, reservoirs with large storage capacity were built to accumulate water and generate energy. The three largest reservoirs are Três Marias, Sobradinho e Itaparica (Mendes et al., 2015; ANA, 2018).

Regarding uses, there is a predominance of withdrawal for irrigation (213.7 m³/s), which represents 77% of total demands in the Region. Irrigation is followed by urban demand, with 31.3 m³/s (11%), concentrated mainly in the Metropolitan Region of Belo Horizonte, and industrial with 19.8 m³/s (7%). Animal demand in the region is 10.2 m³/s (4%) and rural demand is 3.7 m³/s (1%) (ANA, 2016).

The São Francisco Region plays an important role in generating electricity, with a potential installed in 2013 of 10,708 MW (12% of the country's total). The hydroelectric exploitation of the São Francisco River represents the energy supply base of the Northeast region (ANA, 2016).

Figure 12 - Location of the Sobradinho Reservoir

Source: Prepared by the author

Monthly mean discharges, ranging from January 1931 to December 2016, measured at a gauging station located along the São Francisco River was acquired from the National System Operator (NSO). Time series of climate oscillation indices used, such as Pacific Decadal Oscillation (PDO) and Atlantic Multidecadal Oscillation (AMO), were acquired in the National Oceanic and Atmospheric Administration website (https://www.esrl. noaa.gov/psd/gcos_wgsp/ Timeseries/). The definitions of climate indices are briefly presented below, and detailed definition can be found in the references.

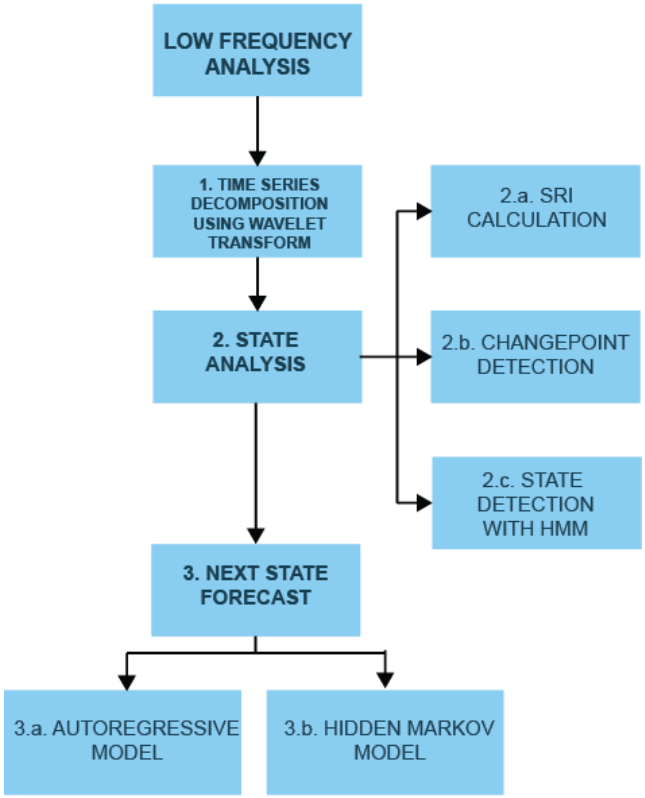
The phenomenon detected at the Pacific Ocean at the interdecadal scale is known as the Pacific Decadal Oscillation (PDO), when SST of the North Pacific Ocean is used to detect it (Mantua et.al, 1997). This index is the principal component of the North Pacific SST. Alves et al. (2013) analysed historical series of NSO inflows, El Niño events, La Niña and the PDO index, concluded that there is evidence of a correlation between the index and the flows of the National Interconnected System (NIS), significantly affecting the flows of the main hydroelectric system of the Brazil.

The Atlantic Multi-Decadal Oscillation (AMO) is based on the average Atlantic North SST anomaly. The index is usually calculated by removing the SST trend in order to

remove the signals resulting from the changes. The AMO is the main low frequency climate change mechanism in the Atlantic Ocean (Enfield et al., 2001).

The methodology of the paper is divided into: (1) identification of the change in state in the low frequency of the time series by using the SRI, changepoint detection and HMM and (2) the prediction of the next state by using an autoregressive model and a hidden Markov model.

Figure 13 - Methodology applied in the paper.



Source: Prepared by the author

4.2.1 Wavelet transforms

A widely used method in extracting the low frequency portion of a time series and a better understanding of its fluctuations is the wavelet transform (Torrence and Compo, 1998; Labat 2005; Nourani, 2013), which decompose the series in the time-frequency domain and identify the dominant modes of variability.

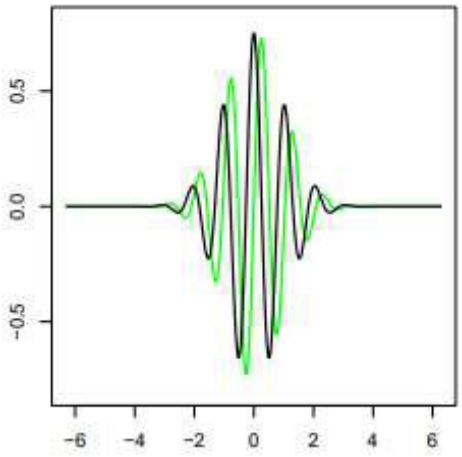
A wavelet transform decomposes a time series into a set of functions $\psi\,(t,s)$, also known as the "daughter wavelet", derived from the translation in time (t) and scaling by (s) of the "mother wavelet" $\psi_0(t)$:

$$\psi\,(t,s) = \frac{1}{s^{\frac{1}{2}}}\psi_0\left(\frac{t'-t}{s}\right) \qquad \text{(Eq. 25)}$$

where the dilation parameter s (> 0) corresponds to scale or temporal period and, hence, connects the wavelet size to the resolutions of particular frequencies; and the translation parameter t controls the locations of wavelet in the time domain. The term $s^{\frac{1}{2}}$ is the energy normalization factor to keep the energy of daughter wavelets similar to the energy of the mother wavelet (Sivakumar, 2017).

In this study, it is applied the Morlet wavelet, which is commonly used in hydrological time series because describes the time series well and has a better time–frequency localization (Huo et al., 2016; Nalley et al., 2016).

Figure 14 - Morlet wavelet



Source: Rosch and Schmidbauer (2014)

The mother Morlet wavelet can be implemented by the equation:

$$\psi(t) = \pi^{-1/4}e^{i\omega t}e^{-t^2/2} \qquad \text{(Eq. 26)}$$

Where ω is the angular frequency is set to 6 because it makes the Morlet wavelet approximately analytic and t is the time.

The Morlet wavelet transform (x_t) was defined as the convolution of the series with a set of (x_t) generated by the mother wavelet by translation in time by τ and scaling by s:

$$Wave\,(\tau,s) = \sum_t x_t \frac{1}{\sqrt{s}}\psi^*\left(\frac{t-\tau}{s}\right) \qquad \text{(Eq. 27)}$$

With * denoting the complex conjugate. The position of the particular daughter wavelet in the time domain is determined by the localizing time parameter τ being shifted by a time increment of dt. The choice of the set of scales s determines the wavelet coverage of the series in the frequency domain.

In this work, it was used the package *WaveletComp* in R (Rosch and Schmidbauer, 2014).

4.2.2 Standard Runoff index

When applying a drought index, such as SPI a long-term record is fitted to a probability distribution, which is then transformed to a normal distribution so that the mean SPI for the location and desired period is zero. A variety of probability distributions (e.g., gamma, generalized extreme value (GEV), generalized logistic distribution, and beta distributions) have been used to fit monthly observations of different hydroclimatic variables for the propose of calculating drought indices (McKee et al., 1993; Stagge et al., 2015; Vicente-Serrano et al., 2010).

The SRI is based on the concept of standardized precipitation index (SPI) (McKee et al., 1993), discussed by Shukla and Wood (2008). Although the indexes show similarities, SRI incorporates hydrological processes that determine the seasonal loss in streamflow due to the influence of the climate, which can be used to describe hydrological aspects of droughts. The series was set to a Gamma distribution. McKee et al. (1993) used this distribution for precipitation data, however it can be applied to other variables relevant to drought, for example, reservoir flow or volume.

Due to the existence of any null values, a function of accumulated probability is applied to includes these events:

$$H(x) = q + (1 - q)G(x) \qquad \text{(Eq. 28)}$$

Where q is the probability of a zero value.

The series was fitted to the Gamma distribution, then the cumulative probability calculated was transformed into a standardized variable (SRI).

After standardized variables, their values are classified according to the categories in Table 4. It was used the package *SCI* in R (Gudmundsson and Stagge, 2016) for the calculation of the SRI.

Table 4 - SPI values

SPI Values	(McKee et al, 1993)
≥2,00	Extremely wet
1,5 a 1,99	Very wet
1,00 a 1,49	Moderately wet
-0,99 a +0,99	Near normal
-1,00 a -1,49	Moderately dry
-1,50 a -1,99	Severely dry
≤-2,00	Extremely dry

Source: McKee et al. (1993)

4.2.3 Changepoint detection

A widely used approach to identify changes in the statistical properties of a time series is known as changepoint. More formally, a dataset is defined as being $y_{1:n} = (y_1, \ldots, y_n)$. In a model with m changepoints and their positions $\tau_{1:m} = (\tau_1, \ldots, \tau_m)$, each changepoint is an integer between 1 and $n-1$. It is defined that $\tau_0 = 0$ and $\tau_{m+1} = n$. Consequently, the m changepoints will split the data into $m+1$ segments. It is considered a common approach in the methodology to detect multiple changepoints the minimization of a cost function.

$$\sum_{i=1}^{m+1} \left[C\left(y_{\tau_{i-1}+1):\tau_i}\right) \right] + \beta f(m) \qquad \text{(Eq. 29)}$$

where C is a cost function for a segment e.g., negative log-likelihood and $\beta f(m)$ is a penalty to guard against over fitting (a multiple changepoint version of the threshold c).

In the penalty parameter, the most common approach is one which is linear in the number of changepoints, that is, $\beta f\ (m)\ =\ \beta m$, such as Akaike's information criterion (AIC; Akaike 1974) (β = 2p) and Schwarz information criterion (SIC, also known as BIC; Schwarz 1978).

Many changepoint algorithm have been proposed to accomplish this objective, such as the binary segmentation algorithm (Scott and Knott, 1974; Sen and Srivastava; 1975), that performs a single changepoint search for the entire dataset, and if a point is detected, the data set is divided into two segments, if false, then no change point is detected, and the method stops. This procedure is repeated until there are no more changepoints in the data. The binary segment is computationally efficient, but it does not guarantee to find the global minimum.

The segmentation neighborhood (SN) proposed by Auger and Lawrence (1989) searches the entire segmentation space using dynamic programming. The method set a maximum number of changepoints that is required and then computes the cost function for all possible segments. One of the drawbacks in using the segmentation neighborhood is that the method has significant computational cost.

The optimal partitioning method (OP) (Jackson et al., 2005) also based on dynamic programming, consists of a process which compares the costs of segmentation to each iteration, which has managed to reduce the computational cost. The OP method improves the computational efficiency on the SN method, but it is not computational competitive with the BS method.

More recently, the PELT algorithm (Killick, Fearnhead, and Eckley 2012), which shows speed gains and increased accuracy over methods, such as binary segmentation. The method is an adaptation of the optimal partitioning and applies pruning for computational efficiency by removing points that can never be minima from the minimization performed at each iteration by the cost function (Killick and Eckley, 2014).

The function used in this study was Pruned Exact Linear Time (PELT), which reduces the computational cost of the method, but does not affect the cost applied to the portion of the segmented series. The changepoint analysis was performed using the package *changepoint* in R (Killick and Eckley, 2014).

4.2.4 Autoregressive model

Autoregressive (AR) models have been extensively used in hydrology and waters resources applications (Box and Jenkins, 1970; Thomas and Fiering, 1962; Salas, 1980). AR models describe how an observed variable depends directly on one or more previous measurements plus white noise. When an observation Z_t, measured at time t depends only on the previous observation plus a white noise, this process is known as first-order or AR (1) autoregressive or even Markov process, being mathematically represented as:

$$Z_t - \mu = \phi_1(Z_{t-1} - \mu) + a_t \qquad \text{(Eq. 30)}$$

Where μ is the average of the process, $\phi_1$is the autoregressive parameter and a_t is the white noise at time t is identically independently distributes (iid) which has mean zero and a variance of σ_a^2. One important characteristic of AR model is its short-term memory; hence, future behaviour is related to previous information.

The general form of an autoregressive model is called AR(p). The Markovian process is a special case of the autoregressive model of order p. This can be expressed as:

$$Z_t - \mu = \phi_1(Z_{t-1} - \mu) + \phi_2(Z_{t-2} - \mu) + \cdots \phi_p\big(Z_{t-p} - \mu\big) + a_t \qquad \text{(Eq. 31)}$$

Manipulating the theoretical autocorrelation function and replacing the values of $k = 1,2,3\ldots, p$ it is obtained the Yule-Walker equations written as:

$$\begin{gathered}
\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \cdots + \phi_p\rho_{k-p}, \qquad k > 0 \\
\rho_1 = \phi_1 + \phi_2\rho_1 + \cdots + \phi_p\rho_{p-1} \\
\rho_2 = \phi_1\rho_1 + \phi_2 + \cdots + \phi_p\rho_{p-2} \\
. \quad . \quad . \quad .... \quad . \\
. \quad . \quad . \quad .... \quad . \\
. \quad . \quad . \quad .... \quad . \\
\rho_p = \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \cdots + \phi_p
\end{gathered} \qquad \text{(Eq. 32)}$$

Describing the equations in matrix form, we have the relation between the autoregressive parameters. This method is used for estimating the parameters of the AR(p) model by the method of moments as well as for determining the correlogram (ρ_k) for a given set of parameters:

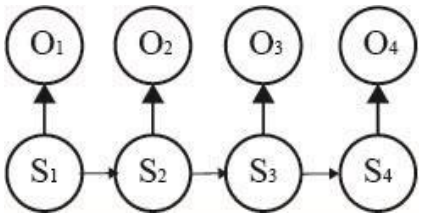
$$\phi = P_P^{-1}\rho_p \qquad \text{(Eq. 33)}$$

$$\phi = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_P \end{bmatrix}, \rho_p = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix}, P_p = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \rho_{p-2} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{p-1} & \rho_{p-1} & \rho_{p-3} & 1 \end{bmatrix} \qquad \text{(Eq. 34)}$$

4.2.5 Hidden Markov Models

Hidden Markov models (HMMs) (Rabiner, 1989) is a statistical model in which the distribution that generates an observation are assumed to be conditioned on the state of an underlying and unobserved Markov process (Mallya et al., 2012; Zucchini et al., 2016). The HMM was developed for speech recognition and since then has been successfully used in many knowledge areas, including hydrology (Robertson et al., 2003; Mallya et al., 2012; Bracken et al., 2014; Liu et al., 2018)

Figure 15 - Example of a basic Hidden Markov Model



Source: Adapted from Zucchini et al. (2016)

A hidden Markov Model is comprised of state variables $\boldsymbol{S}_{1:T} = (S_1, \dots, S_T)$, observation variables $\boldsymbol{O}_{1:T} = (O_1^1, \dots, O_1^m, O_2^1, \dots, O_2^m, \dots, O_T^1, \dots, O_T^m)$, which are dependent on the state variables. Thus, the distribution of $\boldsymbol{O}_t$ can be written as $f_i(\boldsymbol{O}_t) = f(\boldsymbol{O}_t|S_t = i)$ and the marginal distribution, for a discrete number of states, can be described as mixture distribution with n components (Visser, 2011; Zucchini et al., 2016) and it is written as:

$$f(\boldsymbol{O}_t) = \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} p_i f_i(\boldsymbol{O}_t) \qquad \text{(Eq. 35)}$$

Where $\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} p_i = 1$, $p_i \geq 0$ and $f_i()$ is the conditional distribution of the data.

Another important characteristic in HMM are their dependence of the data over time. Thus, there is a dependence between consecutive sates S_t and the current state depends only on the previous states.

$$P(S_t|S_1, \dots, S_{t-1}) = P(S_t|S_{t-1}) \qquad \text{(Eq. 36)}$$

The transition between states are governed by probabilities described by transition probabilities and are denoted by the matrix $\mathbf{A}(t)$, where the first row (a_{1j}) contains the probabilities from moving from state $S_t = 1$ to S_{t+1}**.** When dealing with the transition's parameters in **A** one must define the initial state or the prior probabilities $\boldsymbol{\pi}$ which define where the process begins.

$$P(S_t = 1, S_t = 2, \dots, S_t = \boldsymbol{\pi}\mathbf{A}^{t-1} \qquad \text{(Eq. 37)}$$

The HMMs have their evolution of states over time governed by a Markov process. So, to generate their response distribution one must follow this sequence: (1) choose the initial state of the hidden Markov process by drawing from the initial state probability vector $\boldsymbol{\pi}$; (2) draw an observation from the state-dependent distribution $f_{S_t}()$; (3) generate a transition from the appropriate row of the transition matrix $\mathbf{A}(t)$, which provides the next value of the state variable S_{t+1}**;** and (4) repeat steps 2 and 3 until $t = T - 1$ (Visser, 2011).

The greatest distinction between Markov Models and hidden Markov Models is that its distributions function from the states are not deterministic, but rather a probabilistic density

function, due to the fact that the states are hidden and cannot be observed directly. So, the observed sequence of states is probabilistically related to the hidden process (Visser, 2011).

To compute how likely is a given sequence in an HMM, the joint likelihood of observations $\boldsymbol{O}_{1:T}$ and the latent states $\boldsymbol{S}_{1:T} = (S_1, \dots, S_T)$ can be written as:

$$L(\boldsymbol{O}_{1:T}, \boldsymbol{S}_{1:T}|\boldsymbol{\lambda}) = \pi s_1 f_{S_1}(\boldsymbol{O}_1) \prod_{t=1}^{T-1} a_{\boldsymbol{S}_t \boldsymbol{S}_{t+1}} f_{\boldsymbol{S}_{t+1}}(\boldsymbol{O}_{t+1}) \quad \text{(Eq. 38)}$$

Where $\boldsymbol{\lambda}$ is called the parameter vector of the HMM and it consist of three submodels, which are parameters for the prior model (initial state π), for the transition model ($\boldsymbol{A}$) and for the response model ($f_i(\boldsymbol{O}_t)$). For the joint likelihood of the data independent of a specific state one must sum possible states sequences:

$$L(\boldsymbol{O}_{1:T}|\boldsymbol{\lambda}) = \sum_{\text{all } \boldsymbol{S}_{1:T}} \pi s_1 f_{S_1}(\boldsymbol{O}_1) \prod_{t=1}^{T-1} a_{\boldsymbol{S}_t \boldsymbol{S}_{t+1}} f_{\boldsymbol{S}_{t+1}}(\boldsymbol{O}_{t+1}) \quad \text{(Eq. 39)}$$

Due to the impracticality of calculating the likelihood by the above equation, it is used the forward algorithm (Lystig and Hughes, 2002)

$$\phi_1(j) \coloneqq P(\boldsymbol{O}_1|S_1 = j) = \pi_j f_j(\boldsymbol{O}_j) \quad \text{(Eq. 40)}$$

$$\phi_t(j) \coloneqq P\big(\boldsymbol{O}_t, S_t = j|\boldsymbol{O}_{1:(t-1)}\big) = \sum_{i=1}^{n} [\phi_{t-1}(i)\, a_{ij} f_j(\boldsymbol{O}_t)] \times (\Phi_{t-1})^{-1} \quad \text{(Eq. 41)}$$

Where $\Phi_t = \sum_{i=1}^{n} \phi_t(i)$. We get the following expression for the log-likelihood:

$$l_T = \sum_{t=1}^{T} \log \Phi_t \quad \text{(Eq. 42)}$$

Another important part in HMM is knowing which sequence of states might have generated a given time series and at which point there is a switching between of states. This process is also called the decoding process. In the decoding process it is computed by the Viterbi algorithm (Viterbi, 1967) the hidden state sequence $S_{1:T}$, so that the sequence has maximum probability $P(S_{1:T}|\boldsymbol{O}_{1:T}, \boldsymbol{\lambda})$. Then, it is obtained the posterior state sequence. The probabilities calculated by the Viterbi algorithm differ from those calculated in the forward algorithm since they represent the probability of the most probable path to a state at a time t, and not a total. One of the drawbacks when using this algorithm is that it does not provide information of similar sequence, thus offering only one sequence.

To estimate the parameters of HMM it is used the expectation-maximization (EM) algorithm. In this algorithm the parameters are obtained with the maximization of the expected joint log-likelihood given the observations and states thought an iteratively process. So first, it

is computed the expected hidden state based on the current parameters values and then re-calculate the parameters conditional on the current estimate of the hidden state sequence repeating this interactively process until convergence. Further detail may be found in Baum and Petrie (1966) and Zucchini et al. (2016).

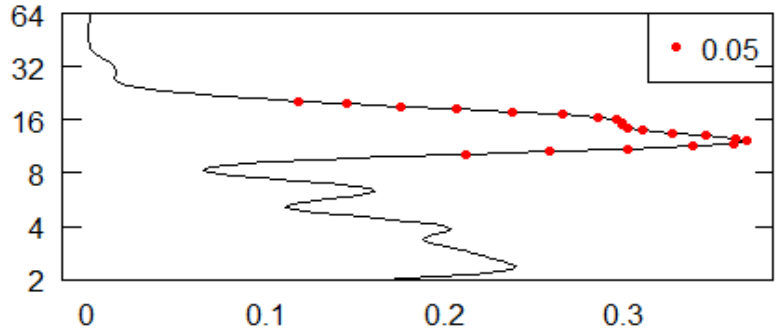
4.3 Results

In this section are presented the results organized according to the topics of the methodology used.

4.3.1 Time series decomposition

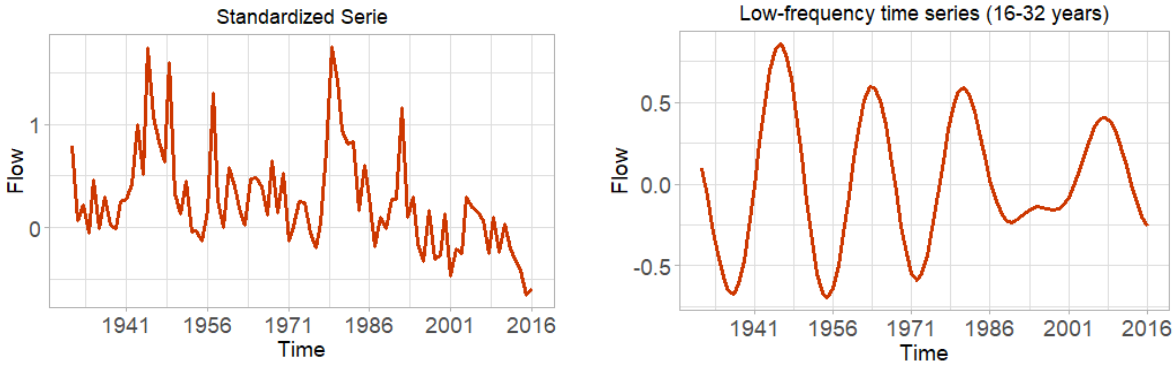
The time series from Sobradinho's reservoir inflow was decomposed and the low frequency was extracted from the original series using a wavelet transform. This frequency was chosen since it presented significance in its power spectrum and average wavelet power (Figure 16).

Figure 16 - Average wavelet power



Source: Prepared by the author

Figure 17 - (a) Standardized Time Series; (b) Low frequency time series (16-32 years)

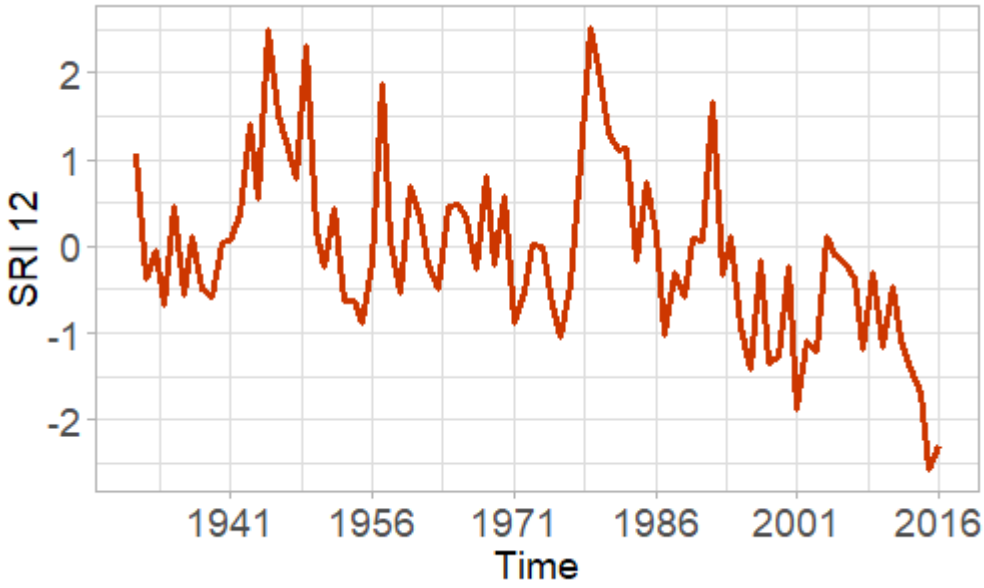


Source: Prepared by the author

4.3.2 State detection

Proceeding to an analyse of the low frequency time series using the changepoint detection, SRI classification and the HMM aiming to identify and characterize the "states" present in the time series. In the SRI analysis, it was evaluated the streamflow time series and its low frequency to identify wet/drought periods.

Figure 18 - SRI 12 of the streamflow time series

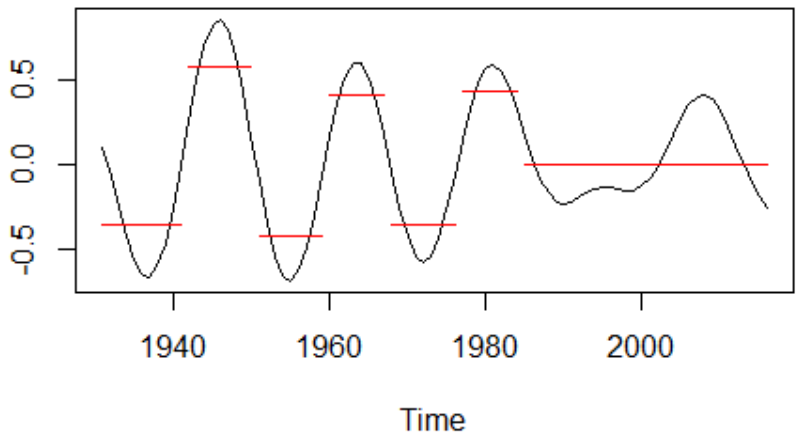


Source: Prepared by the author

For SRI 12 of the streamflow time series, the years of 2015 and 2016 were classified as extremely dry years; 2001 and 2014 were classified as severely dry and all the moderately dry years occur after the 1984 year. As the low frequency series decompose by the wavelet transform is already standardized, it was compared the values of the series to thresholds of the SRI and all the years were classified as near normal.

As the low frequency plot showed a clear separation of high and low states, it was proceeded with the changepoint analysis to look for any significant statistical changes in the time series. The changepoint was applied to the low frequency time series using the penalty values of 1 (Figure 19) and 0.1 (Figure 20).

Figure 19 - Changepoint detection using the penalty value of 1



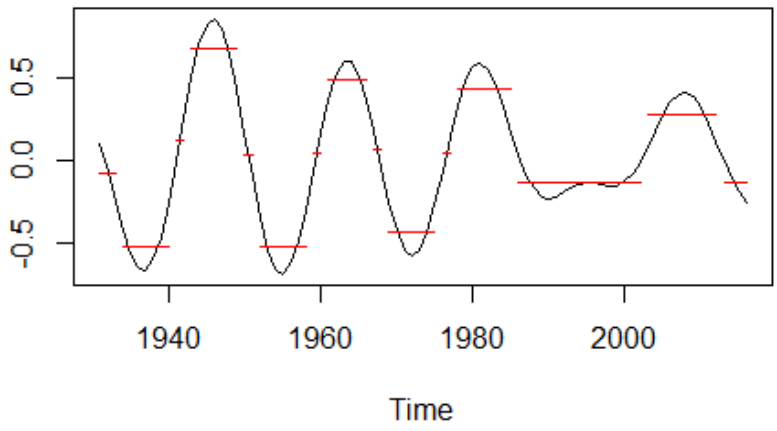
Source: Prepared by the author

Table 5 - Changepoint segmentation with penalty value of 1

Period	Mean	Duration (years)
1931 - 1941	-0.355	11
1942 - 1950	0.575	9
1951 - 1959	-0.429	9
1960 - 1967	0.410	8
1968 - 1976	-0.354	9
1977 - 1984	0.430	8
1985 - 2016	-0.002	32

Source: Prepared by the author

Figure 20 - Changepoint detection using the penalty value of 0.1



Source: Prepared by the author

Table 6 - Changepoint segmentation with penalty value of 0.1

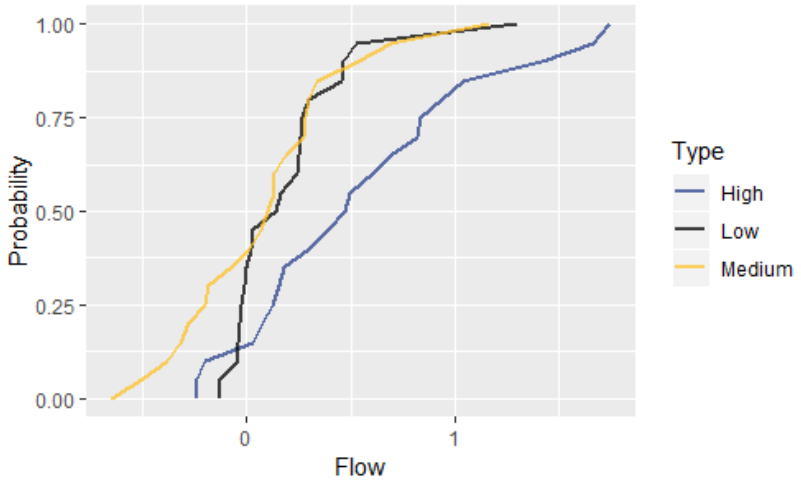
Period	Mean	Duration (years)	Period	Mean	Duration (years)
1931 - 1933	-0.077	3	1967 - 1968	0.063	2
1934 - 1940	-0.522	7	1969 - 1975	-0.439	7
1941 - 1942	0.120	2	1976 - 1977	0.037	2
1943 - 1949	0.680	7	1978 - 1985	0.433	9
1950 - 1951	0.034	2	1986 - 2002	-0.141	15
1952 - 1958	-0.527	7	2003 - 2012	0.272	10
1959 - 1960	0.043	2	2012 - 2016	-0.142	5
1961 - 1966	0.491	6			

Source: Prepared by the author

With the identified periods by the changepoint algorithm and a visual inspection, the periods were separated into three states: high, normal and low period. For the first analysis, it was noted a clear variation patter between the years of 1931 and 1984: the series starts with a value close to the low state (-0.355) and it is succeeded with variation of high (0.48) and low (-0.38) states in an approximately 10-year period. The first analyse did not capture the observed cycle in Figure 19 after 1984, then the penalty value was reduced, aiming to a better characterization of that period. It was observed that the cycles before 1984 were reduce to an approximately 7 years period and after that year there was a 15-year period in the normal state (-0.14) and another high point in the series for another 10-year period.

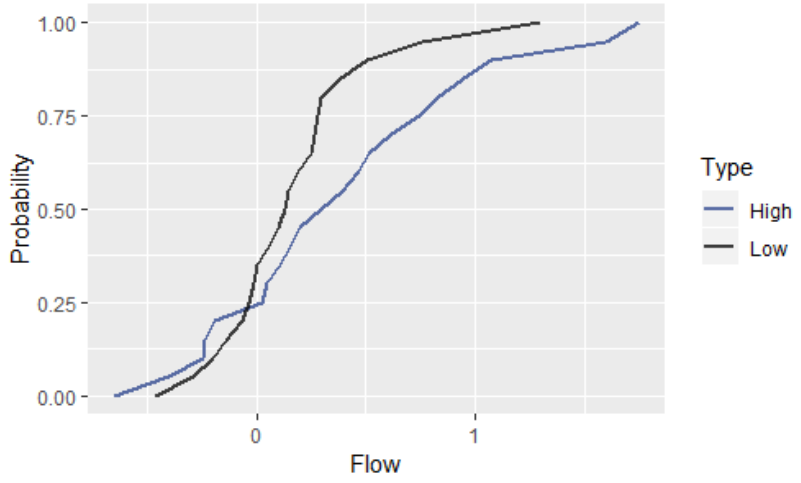
Although there were several changepoints detected in the series, the SRI classification was not able to capture the changes in the statistical properties of the low frequency series. Thus, being the changepoint a better classification for abrupt changes and classification between wet/dry periods, due to its powers to identify statistical changes in the series. Then, for the years that fall into the same category, the cumulative probability distribution was plotted using the data of the original series to identify if those periods follow a similar statistical distribution.

Figure 21 - Cumulative distribution function of the streamflow time series using the three-state classification based on the changepoint detection



Source: Prepared by the author

Figure 22 - Cumulative distribution function of the streamflow time series using the two-state classification based on the changepoint detection



Source: Prepared by the author

The curves for the medium and low state were very similar, so it was decided to plot only the low and high state of the original series. The Figure 22 demonstrates that the dry and wet years do not follow the same distribution and that in the high state the probability for wet periods is larger than in the low state.

It was observed that the low frequency shows a clear effect in the pattern of the whole time series, which justifies its study as one of the influencing components of the time series behavior. It was also applied a Hidden Markov Model to evaluate the states of the low frequency and try to correlate its states with dry and wet periods.

The HMM was fitted for 2, 3, 4 and 5 states and the statistics of the model are represented in Table 7. The selection of number of hidden states is an important aspect of the HMM. Information criterion, such as BIC (or AIC), can be used to select the number of states in the model and the model with the lowest parameter is chosen. It was used BIC as the model selection criterion (Bellone et al., 2000).

Table 7 - Performance of HMM with different number of states K

K (states)	Log Likelihood	AIC	BIC
2	-11.002	36.003	53.184
3	12.875	2.250	36.611
4	27.137	-8.274	48.176
5	35.140	-2.279	81.169

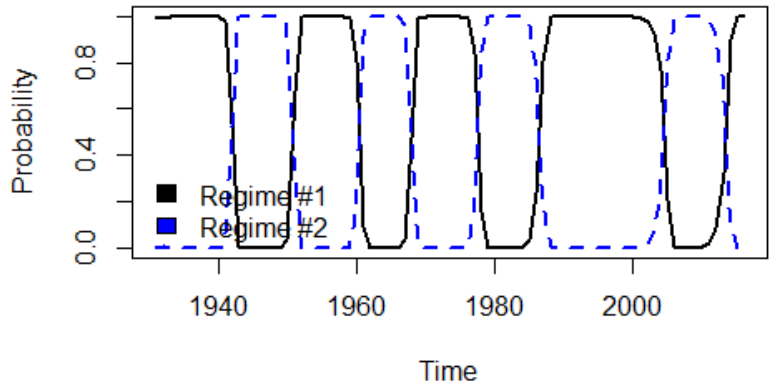
Source: Prepared by the author

The two-state model of the low frequency data has the following parameters. The initial state probability vector has values 0 and 1, indicating the process starts in state 2. The state dependent mean response times are μ_1 = 0.418 (σ_1= 0.209) and μ_2 = -0.263 (σ_2= 0.230). State 1 represents the wet periods and State 2 represents the dry periods with mean values similar to those found by the changepoint detection. The dynamic part of the model which captures the switching between states is represented in matrix A and was estimated by the model as:

$$A = \begin{pmatrix} 0.883 & 0.117 \\ 0.080 & 0.920 \end{pmatrix}$$

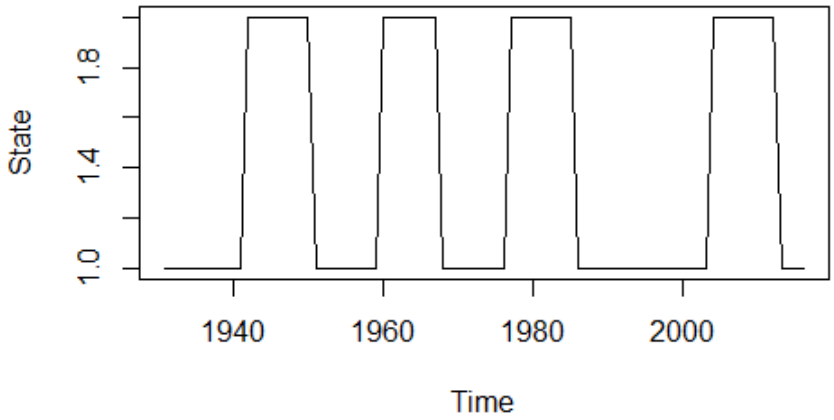
The probabilities of the diagonal matrix indicated that the states are stable, that is, the probability in continuing in the same state is higher, thus, the low frequency presents a stable behaviour. And the probability of transition of the state 1 to 2 is higher than transitioning from state 2 to 1, hence the probability of having a dry period after a wet period is higher. The Figure 23 illustrates the probability regime of the wet state (Regime 1) and dry state (Regime 2). The Viterbi algorithm was used to present the most probable path to generate the given time series of observations (Figure 24).

Figure 23 - Regime posterior probabilities of a two-state model



Source: Prepared by the author

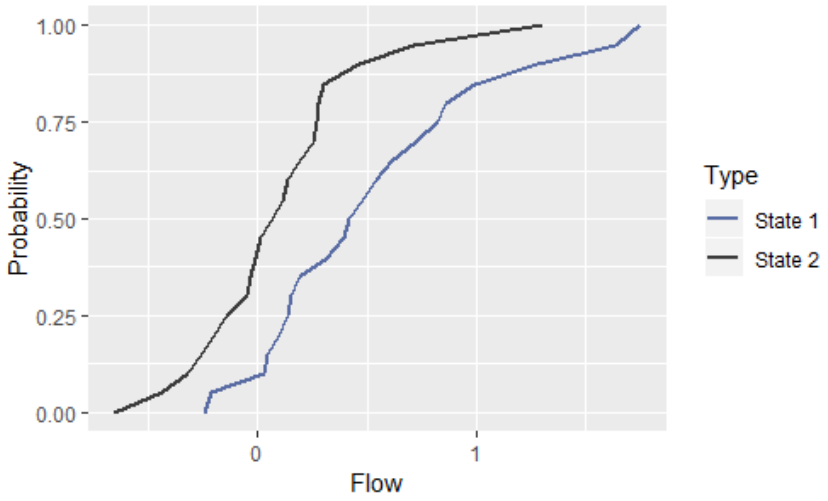
Figure 24 - Most probable path given by the Viterbi algorithm for the two-state model



Source: Prepared by the author

From the results of the path given by Viterbi algorithm, it was adjusted the cumulative distribution function Figure 25 and it showed similar results with the CDF of the changepoint detection, using a two-state classification.

Figure 25 - Cumulative distribution function of the streamflow time series using the two-state HMM



Source: Prepared by the author

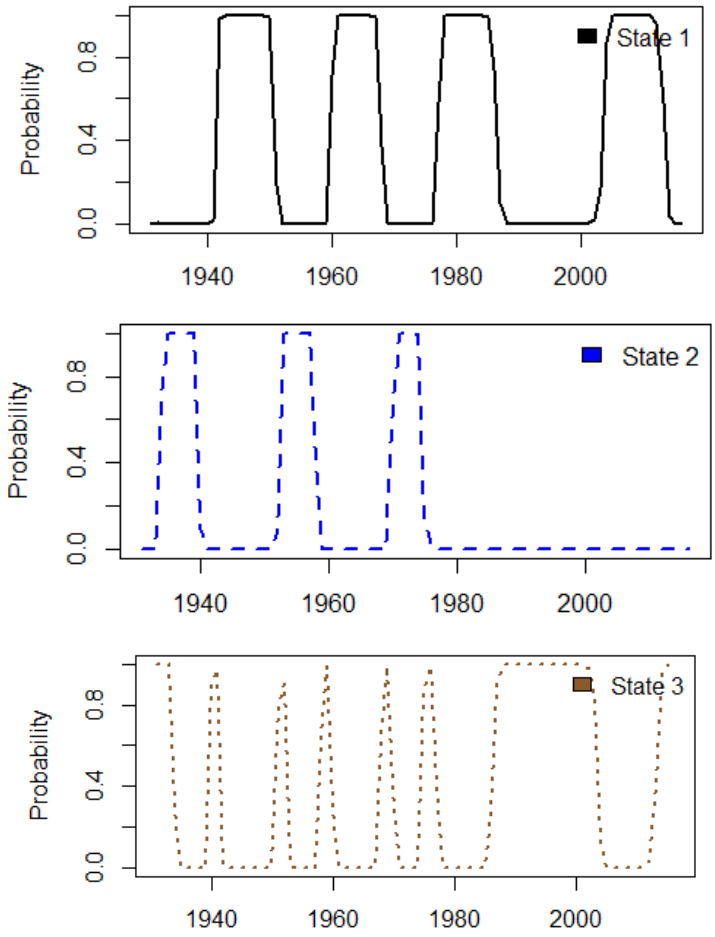
For the 3-state model, the initial state probabilities were 0, 0 and 1 for the states 1,2 and 3, respectively. The state dependent mean response times are μ_1 = 0.407 (σ_1= 0.210), μ_2 = -0.553 (σ_2= 0.097) and μ_3 = -0.140 (σ_3= 0.104). State 1 represents the wet periods, while State 2 represents the dry periods and State 3 represents a near normal state period with a mean close to 0. The transition matrix A was estimated by the model as:

$$A = \begin{pmatrix} 0.889 & 0.000 & 0.111 \\ 0.000 & 0.819 & 0.181 \\ 0.124 & 0.093 & 0.783 \end{pmatrix}$$

The probabilities were higher in the diagonal of the matrix indicated that the probability in remaining in the same state is elevated, thus the low frequency presents a stable behaviour for a 3-state model as well. The probability of transition of the state 1 and 2 to state 3 is about 11% and 18%, respectively. There is also a 12,4% from transitioning from state 3 to state 1. We can draw from the transition matrix that the model presents a higher probability of switching from dry and wet periods to a near normal state, and switching from a near normal state to a wet state.

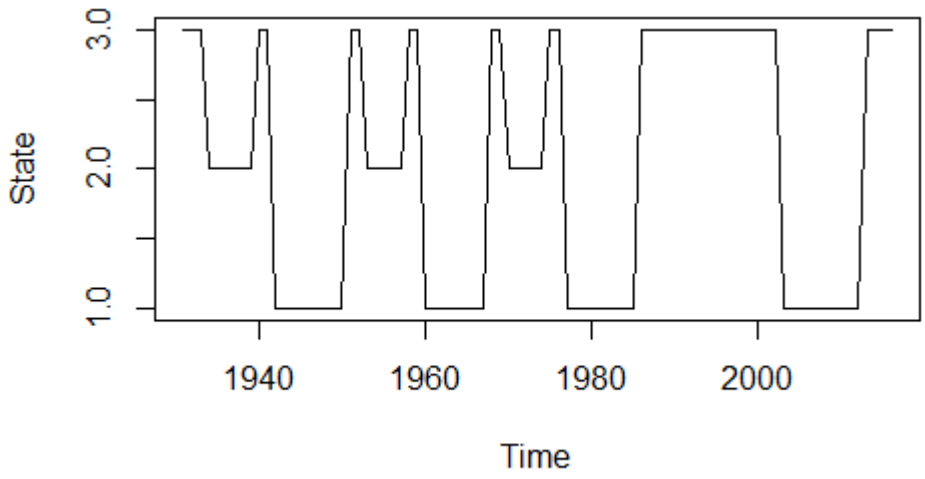
The Figure 26 illustrates the probability regime of the wet state (State 1), dry state (State 2) and near normal (State 3). The Viterbi algorithm was used to present (Figure 27) the most probable path to generate the given time series of observations, while the cumulative density function was calculate with the original data (Figure 28).

Figure 26 - Regime Posterior Probabilities of a three-state HMM



Source: Prepared by the author

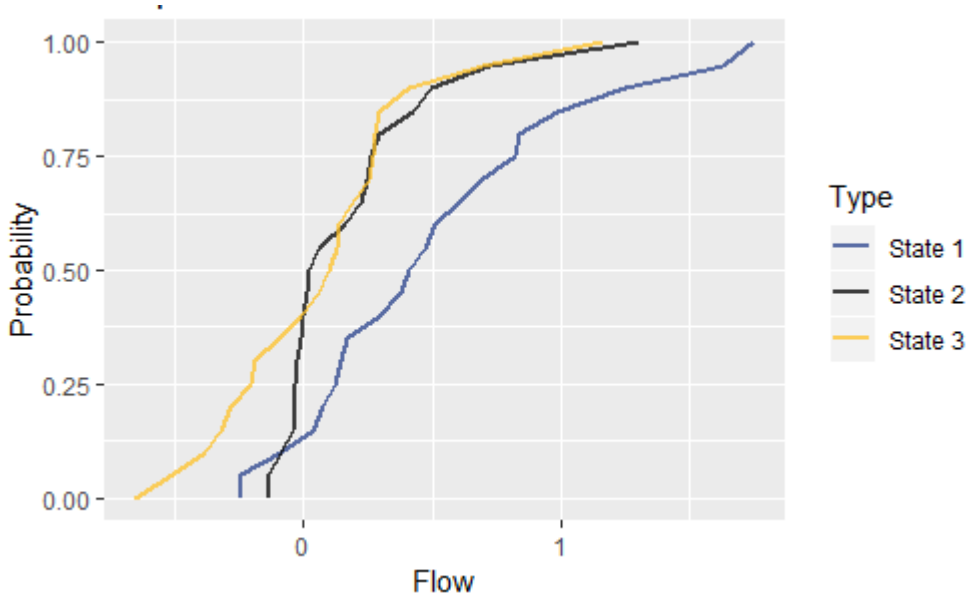
Figure 27 - Most probable path given by the Viterbi algorithm for the three-state model



Source: Prepared by the author

The distribution shows a very similar behaviour between dry and the near normal state, although the near normal state (State 3) presents drier events than the dry state (State 2) these results are also similar to the changepoint detection when divided into three classes.

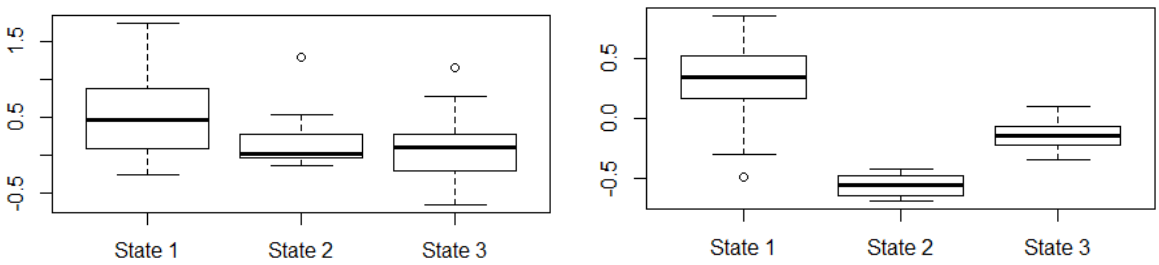
Figure 28 - Cumulative distribution function of the streamflow time series using the three-state HMM



Source: Prepared by the author

Due to the results of the CDF plot, it was evaluated the statistics of the whole series and the low frequency (Table 8) with a box-plot. The box-plot of the data is shown in Figure 29. For the complete time series (left) the State 3, which was classified as nears normal using the mean value, has lower values and high variability than the State 2 (dry state). In the box-plot for the low frequency, the State 2 is the one with lower values and a low variability.

Figure 29 - State variation of the streamflow series (left) and the low frequency series (right)



Source: Prepared by the author

Table 8 - Statistical properties of the low frequency series

	State 1	State 2	State 3
Minimum	-0.4986	-0.6944	-0.34946
First quartile	0.1700	-0.6407	-0.22233
Median	0.3415	-0.5588	-0.14964
Mean	0.3374	-0.5613	-0.14514
Third quartile	0.5111	-0.4908	-0.07514
Maximum	0.8585	-0.4273	0.09557

Source: Prepared by the author

Due to the different behaviour in the two time series (original series and low frequency), it was investigated a period higher than 32 years and introduced a multivariate model to evaluate, if this coupling would improve the model to characterize the states, also it was used a multivariate HMM to introduce the PDO and AMO as well.

The multivariate HMM using the frequencies of 16-32 and higher than 32 years was fitted for 2, 3, 4 and 5 states and the statistics of the model are represented in Table 9. The model that presented the lowest BIC was the 3-state model.

Table 9 - Performance of multivariate HMM with different number of states K

K (states)	Log Likelihood	AIC	BIC
2	-92.913	207.826	234.824
3	-49.048	138.097	187.184
4	-29.023	120.046	196.131
5	-4.923	97.847	205.838

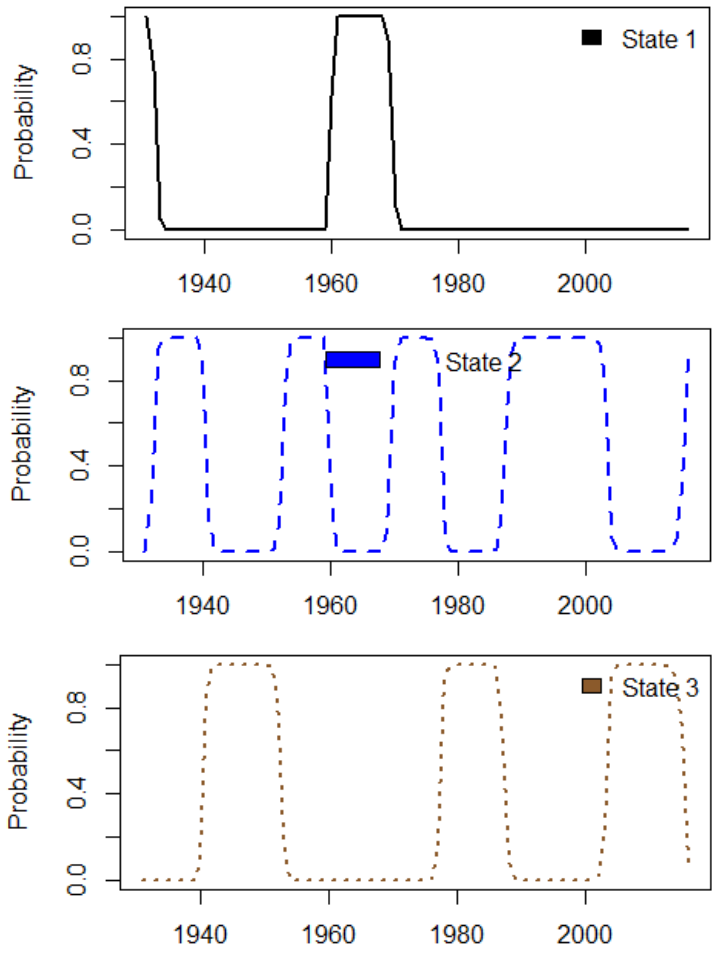
Source: Prepared by the author

For the 3-state multivariate model, the initial state probabilities were 1, 0 and 0 for the states 1, 2 and 3, respectively. The state dependent mean response times for the first variable (16-32 years) was μ_1 = 0.433 (σ_1= 0.368), μ_2 = -0.451 (σ_2= 0.269) and μ_3 = 0.453 (σ_3= 0.359), for the second variable (higher than 32 years) were μ_1 = -0.895 (σ_1= 0.121), μ_2 = -0.168 (σ_2= 0.303) and μ_3 = 0.544 (σ_3= 0.257). For the State 1, there is a high range between the mean of the two used variables. In the State 2 both variable present negative mean, which can indicate the dry period, while in State 3 both means are positive indicating a wet period. The transition matrix A was estimated by the model as:

$$A = \begin{pmatrix} 0.820 & 0.180 & 0.000 \\ 0.025 & 0.900 & 0.075 \\ 0.000 & 0.088 & 0.912 \end{pmatrix}$$

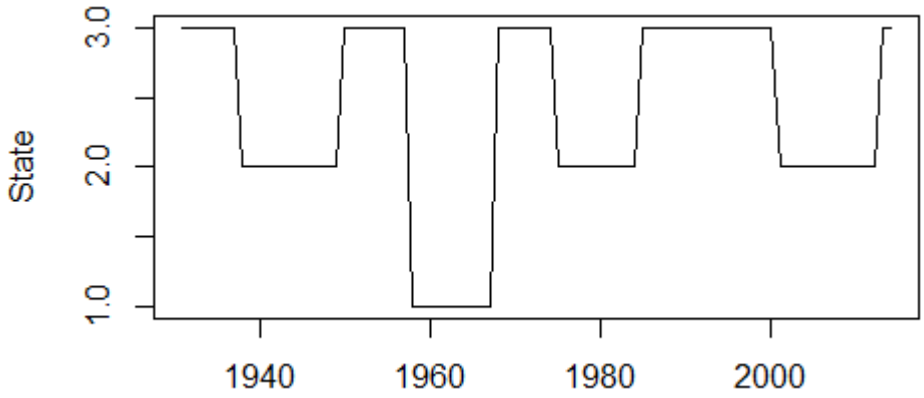
The probabilities were higher in the diagonal of the matrix indicated that the probability in remaining in the same state is higher. The probability of transition of the state 1 to state 2 is about 18%. The Figure 30 illustrates the probability regime of the three regimes influenced by the period higher than 32 years. The Viterbi algorithm was used to illustrate the most probable path to generate the given time series of observations.

Figure 30 - Regime posterior probabilities of a multivariate three-state HMM



Source: Prepared by the author

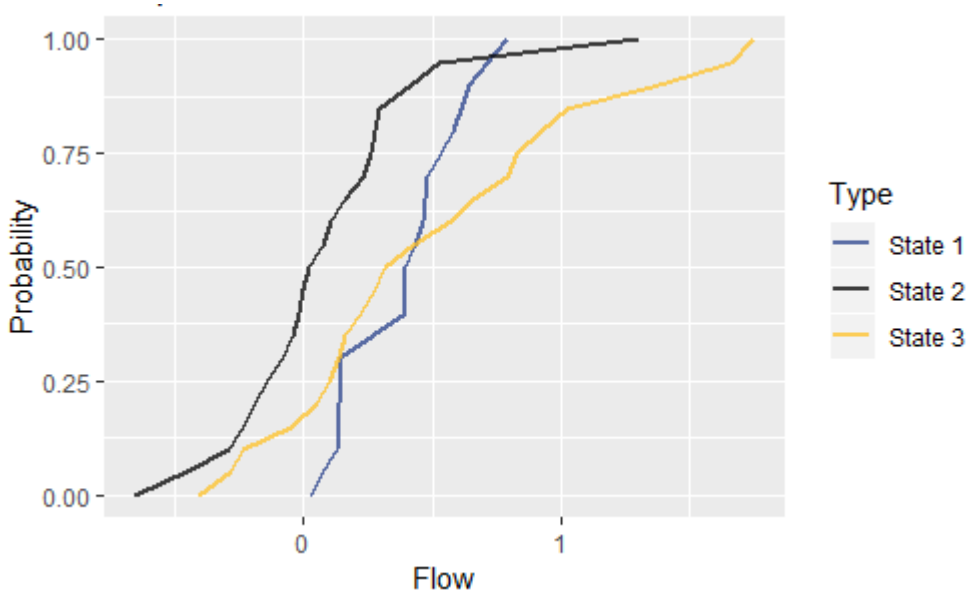
Figure 31 - Most probable path given by the Viterbi algorithm for a multivariate three-state model



Source: Prepared by the author

The cumulative distribution of the three-state model using both frequencies of the streamflow (Figure 32) identify that we may have a dry state, a wet state and a state that can be either dry or wet.

Figure 32 - Cumulative distribution function of the streamflow time series using a multivariate three-state HMM



Source: Prepared by the author

The multivariate HMM using the low frequency of the streamflow and PDO time series was fitted for 2, 3, 4 and 5 states and the statistics of the model are represented in Table 10. The model that presented the lowest BIC was the 3-state model.

Table 10 - Performance of multivariate HMM (streamflow and PDO) with different number of states K

K (states)	Log Likelihood	AIC	BIC
2	-120.79	263.57	290.57
3	-94.59	229.18	278.26
4	-90.65	243.29	319.38
5	-72.30	232.60	340.59

Source: Prepared by the author

For the 3-state multivariate model, the initial state probabilities were 0, 0 and 1 for the states 1, 2 and 3, respectively. The state dependent mean response times for the first variable (streamflow) was μ_1 = -0.548 (σ_1= 0.101), μ_2 = 0.412 (σ_2= 0.208) and μ_3 = -0.136 (σ_3= 0.104), for the second variable (PDO) were μ_1 = -0.250 (σ_1= 0.956), μ_2 = -0.185 (σ_2= 0.746) and μ_3 = 0.360 (σ_3= 0.891). For the streamflow State 1 represents the dry periods, State 2 represents the wet periods and State 3 represents a near normal state period with a mean close to 0. For the PDO the mean of states 1 and 2 were very close, so both states are taken as dry periods and state 3 is the wet period.

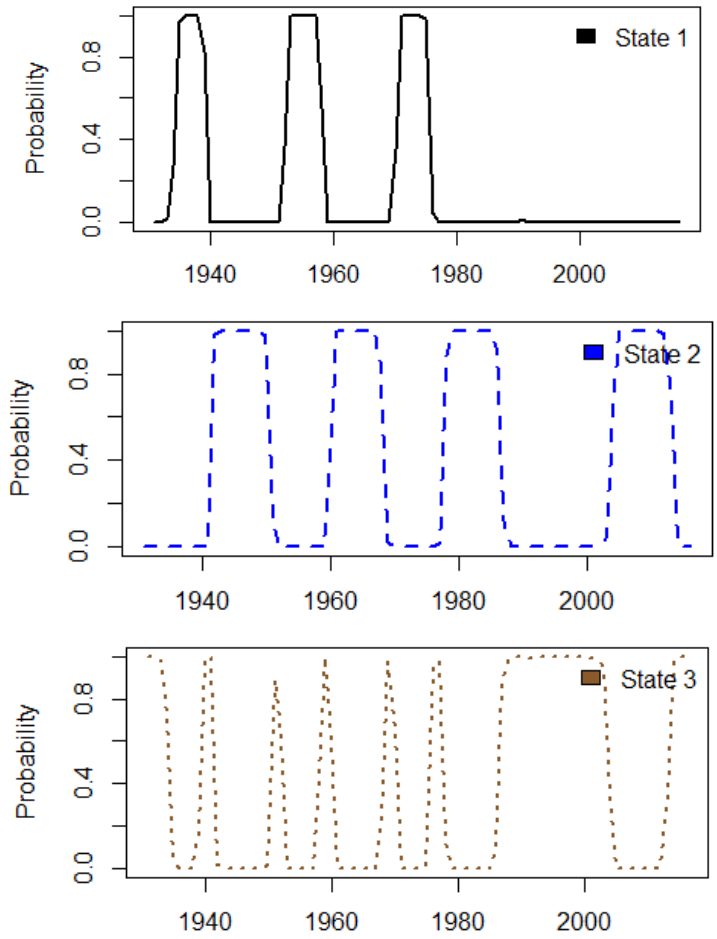
The transition matrix A was estimated by the model as:

$$A = \begin{pmatrix} 0.822 & 0.000 & 0.178 \\ 0.000 & 0.888 & 0.112 \\ 0.093 & 0.124 & 0.784 \end{pmatrix}$$

The probabilities were higher in the diagonal of the matrix indicated that the probability in remaining in the same state is higher, thus the low frequency influence by the PDO presents a stable behaviour for a multivariate 3-state model as well. The probability of transition of the state 1 to 3 and 2 to 3 is about 17,8% and 11,2%, respectively. There is also a 12,4% from transitioning from state 3 to state 2.

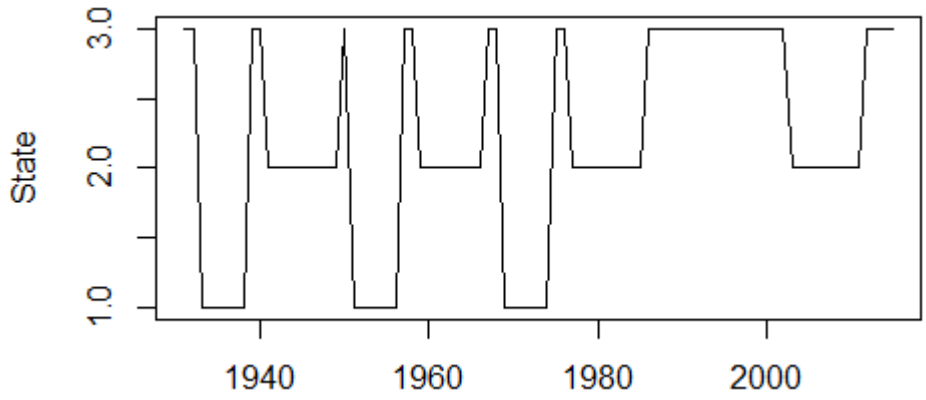
The Figure 33 illustrates the probability regime of the three regimes influenced by the PDO. The Viterbi algorithm was used to illustrate the most probable path to generate the given time series of observations.

Figure 33 - Regime posterior probabilities of a multivariate three-state HMM using streamflow and PDO



Source: Prepared by the author

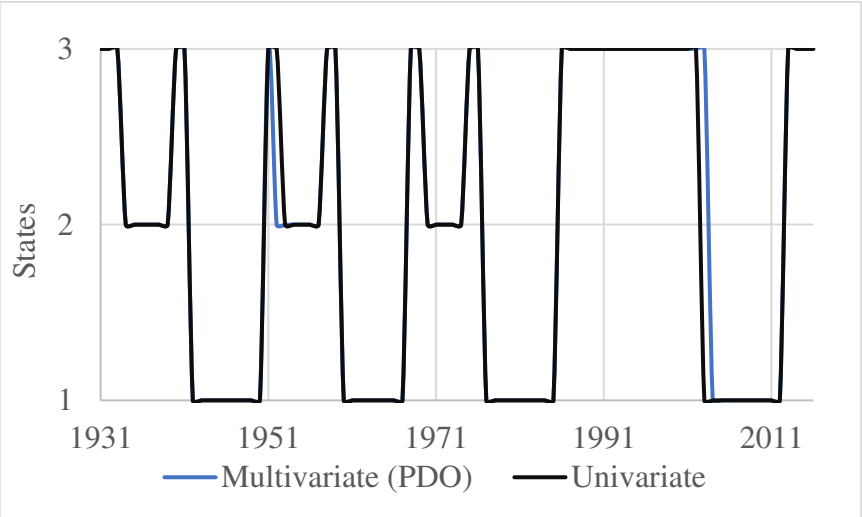
Figure 34 - Most probable path given by the Viterbi algorithm for a multivariate three-state model using streamflow and PDO



Source: Prepared by the author

Comparing the most likely path for the univariate and the multivariate model the results are almost the same except for the years of 1952 and 2003.

Figure 35 - Comparison between Viterbi results of univariate three-state model and the multivariate three-state model using PDO



Source: Prepared by the author

For the multivariate HMM with AMO was fitted for 2, 3, 4 and 5 states and the statistics of the model are represented in Table 11. The model that presented the lowest BIC was the 3-state model.

Table 11 - Performance of multivariate HMM (streamflow and AMO) with different number of states K

K (states)	Log Likelihood	AIC	BIC
2	12.635	-3.271	23.727
3	36.969	-33.938	15.149
4	53.213	-44.426	31.659
5	67.684	-47.368	60.624

Source: Prepared by the author

For the 3-state multivariate model, the initial state probabilities were 0, 0 and 1 for the states 1,2 and 3, respectively. The state dependent mean response times for the first variable (streamflow) was μ_1 = 0.408 (σ_1= 0.210), μ_2 = -0.553 (σ_2= 0.097) and μ_3 = -0.139 (σ_3= 0.104), for the second variable (AMO) were μ_1 = 0.028 (σ_1= 0.168), μ_2 = -0.003 (σ_2= 0.222) and μ_3 = 0.022 (σ_3= 0.182). For the streamflow State 1 represents the wet periods, State 2 represents the dry periods and State 3 represents a near normal state period with a mean close

to 0. For the AMO the mean of states 1,2 and 3 were very close so all states are taken as near normal periods.

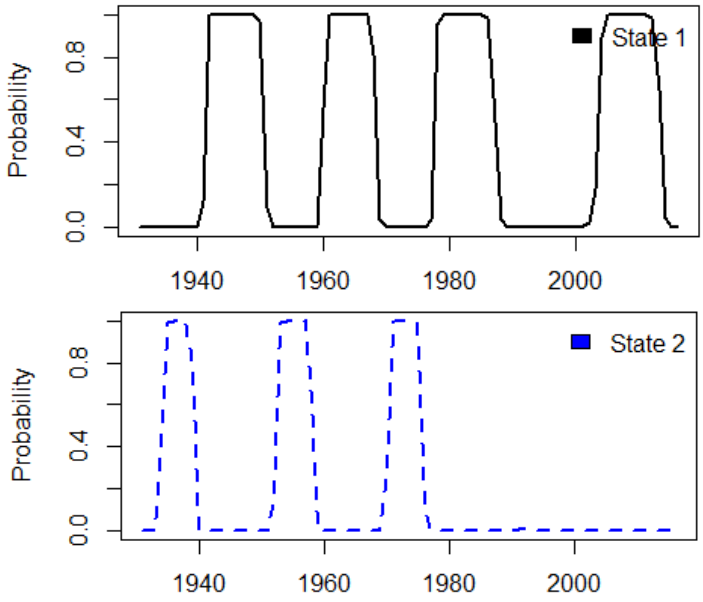
The transition matrix A was estimated by the model as:

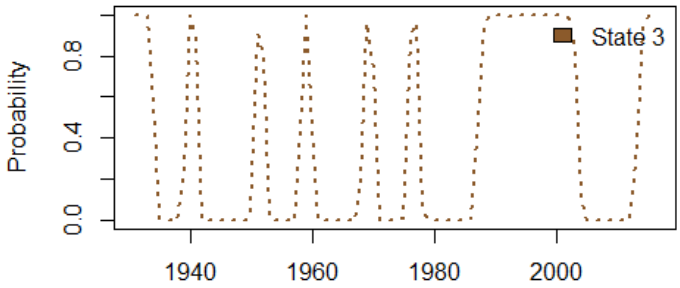
$$A = \begin{pmatrix} 0.889 & 0.000 & 0.111 \\ 0.000 & 0.819 & 0.181 \\ 0.124 & 0.093 & 0.784 \end{pmatrix}$$

The probabilities were higher in the diagonal of the matrix, indicating that the probability in remaining in the same state is higher, thus, the low frequency influence by the AMO also presents a stable behaviour for a multivariate 3-state model. The probability of transition of the state 1 to 3 and 2 to 3 is about 11,1% and 18,1%, respectively. There is also a 12,4% from transitioning from state 3 to state 1. Although the 3-state multivariate model has presented a higher likelihood than the univariate model, the results of the transition matrix of the univariate and multivariate with AMO 3-state model is equal, presenting that the influence of the AMO in this low frequency time series is very low.

The Figure 36 illustrates the probability regime of the three regimes influenced by the AMO. The Viterbi algorithm was used to illustrate the most probable path to generate the given time series of observations.

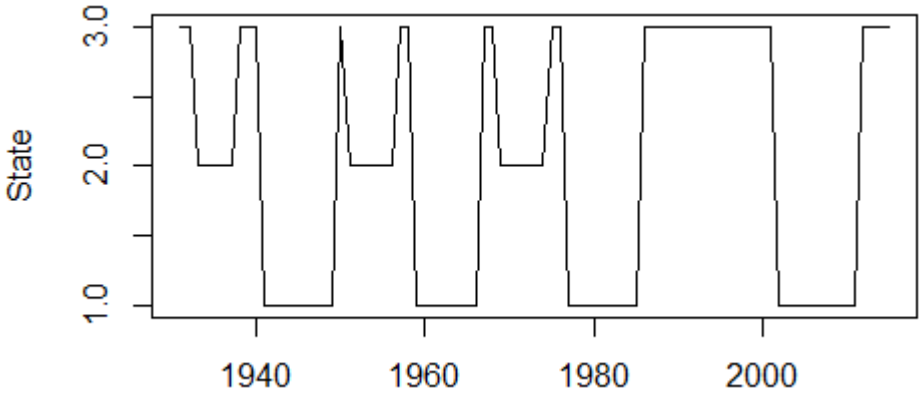
Figure 36 - Regime posterior probabilities of a multivariate three-state HMM using streamflow and AMO





Source: Prepared by the author

Figure 37 - Most probable path given by the Viterbi algorithm for a multivariate three-state model using streamflow and AMO

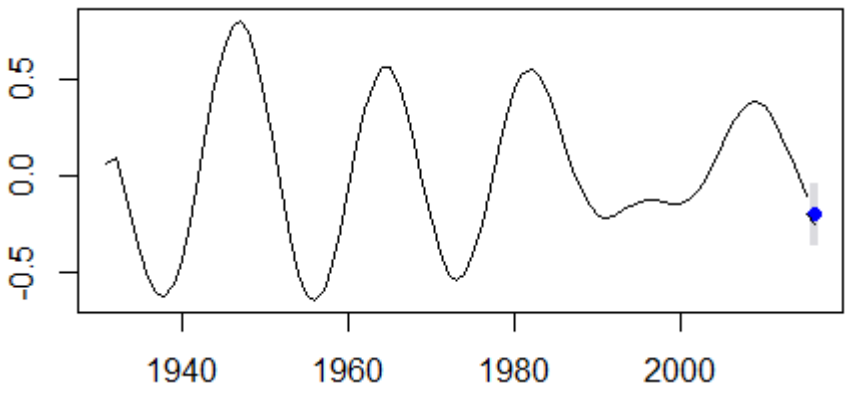


Source: Prepared by the author

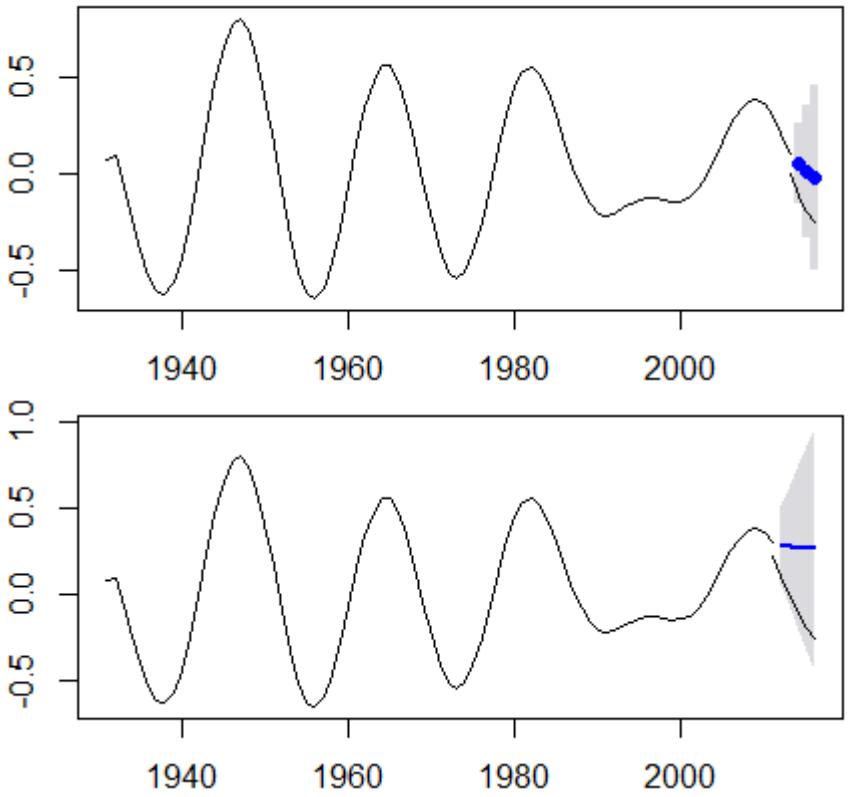
4.3.3 Low frequency forecast model

We applied and AR and an HMM model to predict the next observation of the time series. For the AR model it was adjusted a model with order one, aiming to identify the next 1, 3 and 5 states ahead.

Figure 38 - Forecast of streamflow time series using an AR model of order 1. (a) Next state forecast, (b) the next three states and (c) the next five states



(a)



Source: Prepared by the author

Table 12 - Results of the prediction using an AR model of order 1

Year	Observed Values	1-year ahead	Relative Error	3-year ahead	Relative Error	5-year ahead	Relative Error
2012	0.113	-	-	-	-	0.218	94%
2013	-0.003	-	-	-	-	0.213	-6303%
2014	-0.111	-	-	-0.045	-59%	0.210	-289%
2015	-0.197	-	-	-0.078	-60%	0.206	-205%
2016	-0.255	-0.245	-4%	-0.105	-59%	0.204	-180%

Source: Prepared by the author

Only the 1-year ahead forecast presented good results, the other models showed more than 50% of relative error, proving that they were not a good fit.

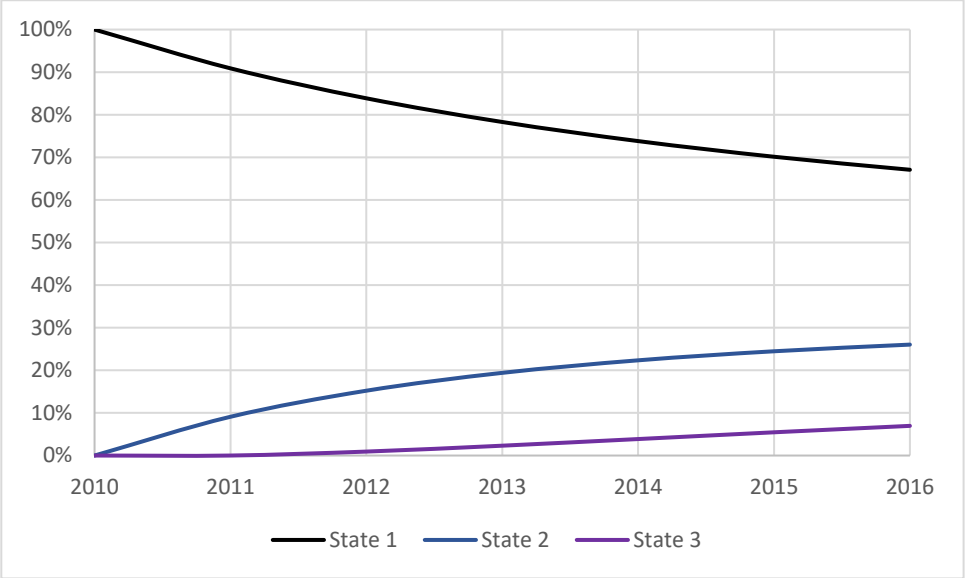
For the HMM model it was used the 3-state model to predict the next state. In the next state forecast, it was multiplied the last posterior probabilities and the transition matrix given by the model. The results are presented in Table 13.

Table 13 - Probability of the next state of the univariate HMM model

Year	State 1	State 2	State 3
2010	0.99E-01	1.60E-05	6.46E-25
2011	9.09E-01	9.10E-02	1.63E-06
2012	8.39E-01	1.52E-01	9.28E-03
2013	7.83E-01	1.94E-01	2.31E-02
2014	7.38E-01	2.24E-01	3.87E-02
2015	7.01E-01	2.45E-01	5.45E-02
2016	6.71E-01	2.60E-01	6.96E-02

Source: Prepared by the author

Figure 39 - Probability of the next state of the univariate HMM model



Source: Prepared by the author

The last state of trained model is state 1 and there is a decrease in the probability in being in State 1 (wet period). It was observed an increase in the probability in being in state 2 (near normal state). It was also verified that after a period the next state probability tends to become constant. Concluding that this approach of forecast would only be a good fit for a short-term prediction and for a long-term (approximately 30 years) the probability would be an average probability for each state. The State 1 would have an average probability of 50%, State 2 would be 32%, leaving State 3 with 18%.

4.4 Conclusion

In this paper we illustrated an ensemble of techniques to evaluate the low frequency of a streamflow time series. The series was decomposed using a wavelet transform and the low frequency was considered the period between 16 and 32 years. This time series was then analyzed using SRI, changepoint and a Hidden Markov model to try to identify the dry and wet periods.

The analysis with the changepoint showed a better result in the identification of extreme events in relation to SRI, because it can identify significant changes in the series statistics, while SRI provided thresholds that were not adequate to characterize only a partial variability of the whole time series.

From the years with significant changes found in the low frequency by the changepoint, it was used the complete series to calculate the cumulative distribution of dry years and wet years, and it was verified that they have different distributions, thus the initial idea to identify the states and forecast the next state was justified.

The HMM was able to identify the periods of extremes for two states well, presenting similar results to the changepoint. However, for a three-state analysis, the model presented a state that was classified as normal because it had a mean close to zero (-0.14964), but it had drier events than the state classified as dry by the mean. Then, it was proceeded to analyze the period superior to the low frequency, and it was verified that it presented a signal opposite to the one in the low frequency band and may have affected the distribution of these states.

It was introduced a multivariate model using the frequency higher than 32-years, PDO and AMO to analyze if these series would improve the model. The cumulative distribution of the three-state model using both frequencies of the streamflow presented a dry state, a wet state and a state that can be either dry or wet, thus this model may not exhibit an accurate base for the next state forecast. The PDO model showed a change in the most likely path in the years 1952 and 2003. The model with the AMO presented a higher likelihood than the model with the PDO, however it showed the same probability of the states over time of the univariate model using the Viterbi algorithm.

The low-frequency variability showed relevance in the example studied and its probability distribution conditions the flows of a given year. The recognition of the current state of the low frequency allows a risk assessment of extreme events, such as droughts and floods, as well as the generation of time series with greater adherence to the forecast data.

A model for forecasting was developed adjusting an autoregressive model to the low frequency time series, but it only presented good results for 1-year ahead forecast, however it worked poorly for a 3-year and 5-year ahead with a relative error higher than 50%. In the HMM model the future states were calculated based on the posterior probability. The results shown a reduction of the wet period between the years 2010-2016 and a greater probability of a normal period was verified. The observed series presented high mean values for the 2010 year, however from 2013 there was a transition from the normal state to the dry state. Due to the dominant diagonal of the matrix having high values of transition probability, all models presented a stable structure, presenting a greater probability of remaining in the same state. In this way, it was concluded that the forecast probability of next state to many years ahead would present an equilibrium, not providing a good model. However, for a short period it is possible to identify the next most probable state for the Sobradinho's reservoir and the next state was a good indication compared to the historical time series. The Autoregressive model of order 1, similar to the one applied by Kwon et al. (2007), could not represent the low frequency as well as the HMM did for the studied reservoir. Thus, HMM forecast model may be an important tool to assist in the management and operation of the reservoir, especially for the one in this study due to its importance of generating energy for the region. A recommendation of this paper is that future research try to identify how to combine the information contained in the various spectrum bands in an HMM.

5 CONCLUSIONS

Comprehending the modification that occurred in the streamflow patterns has been an important approach in water resources studies of streams. Assessing hydrologic time series requires that we understand its past dynamics and possible impacts that might be related to the changes in its patterns, such as climate fluctuations and anthropogenic modifications.

In this study were used monthly time series from January of 1931 to December of 2016 of the 88 base posts operated by the NSO to evaluate the performance of statistical test (Mann-Kendal, Wilcoxon, Student's t, Cox-Stuart tests) and unit root test (DF, KPSS, PP and DF-GLS tests) aiming to verify the presence of stationarity and trend analysis (Mann-Kendall and Sen's Slope test) in the studied time series.

It was made a spatial analysis of the results and the series located in the South and Southeast regions shown indications of nonstationary using the statistical tests. Using the unit root tests, the results were similar to the statistical tests except for the south region, which presented almost all station stationary, that may be associated with the form that the unit root test treat the trend component of the time series, indicating that unit root test are more appropriated to be used in the analysis of stationarity for hydrological time series. When evaluating trend by the Mann-Kendall test, it was noted that the south and southeast region both presented positive trends while the north and northeast region presented negative trends.

Furthermore, the series were decomposed into different frequency's (2-4, 4-8, 8-16, 16-32 years and residue) using wavelet transform and using CEEMDAN method into 5 IMFs and assessed the percentage of explained variance using both methods, intending to identify which frequency would represent better each station gauge. Analysing the decomposed time series by the Morlet wavelet and the CEEMDAN method, most of the series have higher explained variance for the period of 2-8 years, ranging from 50-60%, while for the medium frequency (8-16) the explained variance is around 10-20% and for the low frequency (16-32 years) it represents 5-10% of the time series. Those decomposition methods have proved to be important tools for the identification of climatic variability in an important sector, such as the hydropower sector.

Then, it was applied a wavelet transform to decompose the time series of Sobradinho's reservoir into different frequencies, focusing in identifying the low frequency, and it presented a significant result for the period of 16-32 years in its power spectrum. It was applied an ensemble of techniques to evaluate the low frequency of one streamflow time series

managed by the NSO, such as SRI, changepoint and a Hidden Markov model to try to identify the extreme events.

The analysis with the changepoint showed better results in the identification of wet and dry periods in relation to SRI because it can identify significant changes in the series statistics, while SRI provided thresholds that were not adequate to characterize only a partial variability of the whole time series.

From the years with significant changes found in the low frequency by the changepoint, it was used the complete series to calculate the cumulative distribution of dry years and wet years and verified that they have different distributions, thus, the initial proposal to recognise the states and predict the next state was reasonable.

The HMM was able to identify the periods of extremes for two states well, presenting similar results to the changepoint. However, for a three-state analysis, the model presented a state that was classified as normal because it had an average near zero (-0.14964) but had drier events than the state classified as dry by the mean. Then, we proceeded to analyze the period superior to the low frequency, and it was verified that it presented a signal opposite to the one in the low frequency band and may have affected the distribution of these states.

It was introduced a multivariate model using the frequency higher than 32-years, PDO and AMO to analyze if these series would improve the model. The cumulative distribution of the three-state model using both frequencies of the streamflow presented a dry state, a wet state and a state that can be either dry or wet, thus this model may perform an accurate base for the next state forecast. The PDO model showed a change in the most likely path in the years 1952 and 2003. The model with the AMO presented a higher likelihood than the model with the PDO, however, it showed the same probability of the states over time of the univariate model using the Viterbi algorithm.

A model for forecasting was developed adjusting an autoregressive model and an HMM to the low frequency time series to address the prediction of the next probable state of the system. For the AR model, it only presented good results for the next state, however it worked poorly in identifying longer periods, with a relative error higher than 50%. In the HMM model the future state were calculated based on the posterior probability. The results shown a reduction of the wet period between the years 2010-2016 and a greater probability of a normal period was verified. The observed series presented a high mean value for the 2010 year, however from 2013 there was a transition from the normal state to the dry state. For a short period, it is possible to identify the next most probable state for the Sobradinho's reservoir.

Thus, HMM forecast model may be an important implement to assist in the management and operation of the reservoir, especially for the one in this study due to its importance of generating energy for the region.

It is recommended that future studies try to identify how to combine the information contained in the various spectrum bands in an HMM, allowing to implement a forecast model of all frequencies that compose the time series. Also, it is recommended the application of other contemporary forecasting methods, such as chaos theory, as well as classical methods, such as the Kalman Filter that can capture the nonstationary behaviour of the streamflow time series.

REFERENCES

AKAIKE, H. A new look at the statistical model identification. **IEEE transactions on automatic control**, v. 19, n. 6, p. 716-723, 1974.

ALVES, B. C. C.; SOUZA FILHO, F. A.; SILVEIRA, C. S. Análise de tendências e padrões de variação das séries históricas de vazões do operador nacional do sistema (ONS). **Revista Brasileira de Recursos Hídricos**, v.18, n.4, 2013. https://doi.org/10.21168/rbrh.v18n4.p19-34

ANA, Agência Nacional de Águas. **Plano de Recursos Hídricos da Bacia do Rio São Francisco**. 2016.

ANCTIL, F.; COULIBALY, P. Wavelet analysis of the interannual variability in southern Québec streamflow. **Journal of climate**, v. 17, n. 1, p. 163-173, 2004. https://doi.org/10.1175/1520-0442(2004)017

ANJOS, S. L. **Ondaletas aplicadas à análise de variabilidade de baixa frequência em séries de afluência aos reservatórios hidrelétricos brasileiros.** 2015. 142 f. Dissertação (Mestrado em Engenharia Civil e Ambiental) – Faculdade de Tecnologia, Universidade de Brasília, Brasília, 2015.

ANTICO, A.; SCHLOTTHAUER, G.; TORRES, M. E. Analysis of hydroclimatic variability and trends using a novel empirical mode decomposition: application to the Paraná River Basin. **Journal of Geophysical Research: Atmospheres**, v. 119, n. 3, p. 1218-1233, 2014. https://doi.org/10.1002/2013JD020420

AUGER, I. E.; LAWRENCE, C. E. Algorithms for the optimal identification of segment neighborhoods. **Bulletin of mathematical biology**, v. 51, n. 1, p. 39-54, 1989.

BARROS, A. C. **Análise de Séries Temporais**. Elsevier Brasil. 2017.

BATISTA, A. L.; FREITAS Jr, S, A de; DETZEL, D. H. M.; MINE, M. R. M; FILL, H. D. O. A.; FERNANDES, C.; KAVISKI, E. Verificação da estacionariedade de séries hidrológicas no Sul-Sudeste do Brasil. In.:_____ **Anais do XVIII Simpósio Brasileiro de Recursos Hídricos**. Campo Grande, p. 1-19, 2009.

BAUM, L. E.; PETRIE, T. Statistical inference for probabilistic functions of finite state Markov chains. **The annals of mathematical statistics**, v. 37, n. 6, p. 1554-1563, 1966.

BAYAZIT, M. Nonstationarity of hydrological records and recent trends in trend analysis: a state-of-the-art review. **Environmental Processes**, v. 2, n. 3, p. 527-542, 2015. https://doi.org/10.1007/s40710-015-0081-7

BELLONE, E. **Non-homogeneous Hidden Markov Models for downscaling synoptic atmospheric patterns to precipitation amount**. 2000. 134 f. PhD Thesis, University of Washington, Washington, 2000.

BOWMAN, D. C.; LEES, J. M. The Hilbert–Huang transform: A high resolution spectral method for nonlinear and nonstationary time series. **Seismological Research Letters**, v. 84, n. 6, p. 1074-1080, 2013. https://doi.org/10.1785/0220130025.

BOX, G. E. P; JENKINS, G. M. **Time series analysis: forecasting and control.** 1970.

BRACKEN, C.; RAJAGOPALAN, B.; ZAGONA, E. A hidden M arkov model combined with climate indices for multidecadal streamflow simulation. **Water Resources Research**, v. 50, n. 10, p. 7836-7846, 2014.

BROCKWELL, P. J.; DAVIS, R. A. **Introduction to Time Series and Forecasting**. Springer, Cham, 2016.

CAEIRO, F.; MATEUS, A. **randtests: Testing randomness in R**. R package version, v. 1, 2 014. package version 1.0. https://CRAN.R-project.org/package=randtests

CASTINO, F.; BOOKHAGEN, B.; STRECKER, M. R. Oscillations and trends of river discharge in the southern Central Andes and linkages with climate variability. **Journal of Hydrology**, v. 555, p. 108-124, 2017. https://doi.org/10.1016/j.jhydrol.2017.10.001

CHEN, P. C. et al. Comparison of methods for non-stationary hydrologic frequency analysis: case study using annual maximum daily precipitation in Taiwan. **Journal of hydrology**, v. 545, p. 197-211, 2017. https://doi.org/10.1016/j.jhydrol.2016.12.001

CHENG, L.; AGHAKOUCHAK, A. Nonstationary precipitation intensity-duration-frequency curves for infrastructure design in a changing climate. **Scientific reports**, v. 4, p. 7093, 2014.

CHOI, In. Unit root tests for panel data. **Journal of international money and Finance**, v. 20, n. 2, p. 249-272, 2001. https://doi.org/10.1016/S0261-5606(00)00048-6

COULIBALY, P.; BURN, D. H. Wavelet analysis of variability in annual Canadian streamflows. **Water Resources Research**, v. 40, n. 3, 2004. https://doi.org/10.1029/2003WR002667

DETZEL, D. H. M. et al. Estacionariedade das afluências às usinas hidrelétricas brasileiras. **Revista Brasileira de Recursos Hídricos**, v. 16, n. 3, p. 95-111, 2011.

DETZEL, D. H. M.; FERNANDES, C. V. S.; MINE, M. R. M. Nonstationarity in determining flow-duration curves aiming water resources permits. **Revista Brasileira de Recursos Hídricos**, v. 21, n. 1, p. 80-87, 2016. http://dx.doi.org/10.21168/rbrh.v21n1

DICKEY, D. A.; FULLER, W. A. Distribution of the estimators for autoregressive time series with a unit root. **Journal of the American statistical association**, v. 74, n. 366a, p. 427-431, 1979. https://doi.org/10.1080/01621459.1979.10482531

ELLIOTT, G.; ROTHENBERG, T. J.; STOCK, J. H. **Efficient tests for an autoregressive unit root.** 1992.

ENDERS, W. **Applied econometric time series**. John Wiley and Sons, 2008.

ENFIELD, D. B.; MESTAS-NUÑEZ, A. M.; TRIMBLE, P. J. The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental US. **Geophysical Research Letters**, v. 28, n. 10, p. 2077-2080, 2001.

ERKYIHUN, S. T. et al. Wavelet-based time series bootstrap model for multidecadal streamflow simulation using climate indicators. **Water Resources Research**, v. 52, n. 5, p. 4061-4077, 2016. https://doi.org/10.1002/2016WR018696

FRITIER, N.; MASSEI, N.; LAIGNEL, B.; DURAND, A.; DIEPPOIS, B.; DELOFFRE, J. Links between NAO fluctuations and inter-annual variability of winter-months precipitation

in the Seine River watershed (north-western France). **Comptes Rendus Geoscience**, 344(8), 396-405. 2012.

GÁRFIAS-SOLIZ, J.; LLANOS-ACEBO, H.; MARTEL, R. Time series and stochastic analyses to study the hydrodynamic characteristics of karstic aquifers. **Hydrological Processes: An International Journal**, v. 24, n. 3, p. 300-316, 2010. https://doi.org/10.1002/hyp.7487

GEHAN, E. A. A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. **Biometrika**, v. 52, n. 1-2, p. 203-224, 1965. https://doi.org/10.1093/biomet/52.1-2.203

GRIMM, A. M.; SABOIA, J. PJ. Interdecadal variability of the South American precipitation in the monsoon season. **Journal of Climate**, v. 28, n. 2, p. 755-775, 2015. https://doi.org/10.1175/JCLI-D-14-00046.1

GUDMUNDSSON, L.; STAGGE, J. H. (2016). **SCI: Standardized Climate Indices such as SPI, SRI or SPEIR**. package version 1.0-2.

HAO, Z. et al. A theoretical drought classification method for the multivariate drought index based on distribution properties of standardized drought indices. **Advances in water resources**, v. 92, p. 240-247, 2016.

HIPEL, K. W.; MCLEOD, A. I. **Time series modelling of water resources and environmental systems**. Elsevier, 1994.

HOEK, W. Z.; BOS, J. AA. Early Holocene climate oscillations---causes and consequences. **Quaternary Science Reviews**, v. 26, p. 1901-1906, 2007.

HUANG, N. E. et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. In: **Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences**. The Royal Society, 1998. p. 903-995. https://doi.org/10.1098/rspa.1998.0193

HUO, X. et al. Application of Wavelet Coherence Method to Investigate Karst Spring Discharge Response to Climate Teleconnection Patterns. **JAWRA Journal of the American Water Resources Association**, v. 52, n. 6, p. 1281-1296, 2016. https://doi.org/10.1111/1752-1688.12452

INTERNATIONAL ENERGY AGENCY. **Key world energy statistics**. 2016.

JACKSON, B. et al. An algorithm for optimal partitioning of data on an interval. **IEEE Signal Processing Letters**, v. 12, n. 2, p. 105-108, 2005. https://doi.org/10.1109/LSP.2001.838216

JAIN, S.; LALL, U. Magnitude and timing of annual maximum floods: Trends and large-scale climatic associations for the Blacksmith Fork River, Utah. **Water Resources Research**, v. 36, n. 12, p. 3641-3651, 2000. https://doi.org/10.1029/2000WR900183

KALRA, A. et al. Using large-scale climatic patterns for improving long lead time streamflow forecasts for Gunnison and San Juan River Basins. **Hydrological Processes**, v. 27, n. 11, p. 1543-1559, 2013. https://doi.org/10.1002/hyp.9236

KARAMOUZ, M.; AHMADVAND, F.; FERESHTEHPOUR, M. Flood scenarios determination using nonstationary flood frequency analysis in coastal areas. In: **9th world congress, Water Resources Management in a Changing World**. 2015.

KAYANO, M. T.; ANDREOLI, R. V. Clima da região Nordeste do Brasil. **CAVALCANTI, Iracema FA et al. Tempo e clima no Brasil. São Paulo: Oficina de textos**, 2009.

KAYANO, M. T.; ANDREOLI, R. V. Decadal variability of northern northeast Brazil rainfall and its relation to tropical sea surface temperature and global sea level pressure anomalies. **Journal of Geophysical Research: Oceans**, v. 109, n. C11, 2004. https://doi.org/10.1029/2004JC002429

KENDALL, M. G. **Rank correlation methods**. 1955.

KILLICK, R.; ECKLEY, I. changepoint: An R package for changepoint analysis. **Journal of statistical software**, v. 58, n. 3, p. 1-19, 2014.

KILLICK, R.; FEARNHEAD, P.; ECKLEY, I. A. Optimal detection of changepoints with a linear computational cost. **Journal of the American Statistical Association**, v. 107, n. 500, p. 1590-1598, 2012.

KNIGHT, J. R.; FOLLAND, C. K.; SCAIFE, A. A. Climate impacts of the Atlantic multidecadal oscillation. **Geophysical Research Letters**, v. 33, n. 17, 2006. https://doi.org/10.1029/2006GL026242

KRUSCHE, A. V. et al. Spatial and temporal water quality variability in the Piracicaba river basin, Brazil. **JAWRA Journal of the American Water Resources Association**, v. 33, n. 5, p. 1117-1123, 1997. https://doi.org/10.1111/j.1752-1688.1997.tb04129.x

KWIATKOWSKI, D. et al. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?. **Journal of econometrics**, v. 54, n. 1-3, p. 159-178, 1992. https://doi.org/10.1016/0304-4076(92)90104-Y

KWON, H.; BROWN, C.; LALL, U. Climate informed flood frequency analysis and prediction in Montana using hierarchical Bayesian modeling. **Geophysical Research Letters**, v. 35, n. 5, 2008. https://doi.org/10.1029/2007GL032220

KWON, H.; LALL, U.; KHALIL, A. F. Stochastic simulation model for nonstationary time series using an autoregressive wavelet decomposition: Applications to rainfall and temperature. **Water Resources Research**, v. 43, n. 5, 2007. https://doi.org/10.1029/2006WR005258

LABAT, D. Recent advances in wavelet analyses: Part 1. A review of concepts. **Journal of Hydrology**, v. 314, n. 1-4, p. 275-288, 2005. https://doi.org/10.1016/j.jhydrol.2005.04.003

LI, H.; ZHANG, Q.; SINGH, V. P.; SHI, P.; SUN, P. Hydrological effects of cropland and climatic changes in arid and semi-arid river basins: A case study from the Yellow River basin, China. **Journal of hydrology**, 549, 547-557. 2017.

LIU, L. et al. Hydro-climatological drought analyses and projections using meteorological and hydrological drought indices: a case study in Blue River Basin, Oklahoma. **Water resources management**, v. 26, n. 10, p. 2761-2779, 2012.

LIU, Y. et al. Monthly streamflow forecasting based on hidden Markov model and Gaussian Mixture Regression. **Journal of Hydrology**, v. 561, p. 146-159, 2018.

LIU, Y.; SAN LIANG, X.; WEISBERG, R. H. Rectification of the bias in the wavelet power spectrum. **Journal of Atmospheric and Oceanic Technology**, v. 24, n. 12, p. 2093-2102, 2007. https://doi.org/10.1175/2007JTECHO511.1

LYSTIG, T. C.; HUGHES, J. P. Exact computation of the observed information matrix for hidden Markov models. **Journal of Computational and Graphical Statistics**, v. 11, n. 3, p. 678-689, 2002.

MALLYA, G. et al. Probabilistic assessment of drought characteristics using hidden Markov model. **Journal of Hydrologic Engineering**, v. 18, n. 7, p. 834-845, 2012.

MANN, H. B. Nonparametric tests against trend. **Econometrica: Journal of the Econometric Society**, p. 245-259, 1945. doi:10.2307/1907187

MANTUA, N. J. et al. A Pacific interdecadal climate oscillation with impacts on salmon production. **Bulletin of the american Meteorological Society**, v. 78, n. 6, p. 1069-1080, 1997.

MARENGO, J. A.; VALVERDE, M. C. Caracterização do clima no Século XX e Cenário de Mudanças de clima para o Brasil no Século XXI usando os modelos do IPCC-AR4. **Revista Multiciência**, v. 8, p. 5-28, 2007.

MASSEI, N. et al. Long-term hydrological changes of the Seine River flow (France) and their relation to the North Atlantic Oscillation over the period 1950–2008. **International journal of Climatology**, v. 30, n. 14, p. 2146-2154, 2010. https://doi.org/10.1002/joc.2022

MCKEE, T. B. et al. The relationship of drought frequency and duration to time scales. In: **Proceedings of the 8th Conference on Applied Climatology**. Boston, MA: American Meteorological Society, 1993. p. 179-183.

MENDES, L. A. et al. Trade-Off Analysis among Multiple Water Uses in a Hydropower System: Case of São Francisco River Basin, Brazil. **Journal of Water Resources Planning and Management**, v. 141, n. 10, p. 4015014, 2015.

MILLY, P. C. D. et al. Stationarity is dead: Whither water management?. **Science**, v. 319, n. 5863, p. 573-574, 2008. https://doi.org/10.1126/science.1151915

MISHRA, A. K.; SINGH, V. P. A review of drought concepts. **Journal of hydrology**, v. 391, n. 1-2, p. 202-216, 2010. https://doi.org/10.1016/j.jhydrol.2010.07.012

MME/EPE - Balanço Energético Nacional 2018: Ano base 2017/Ministério de Minas e Energia. Empresa de Pesquisa Energética – EPE. Rio de Janeiro. 2018.

MO, K. C. et al. Drought indices based on the Climate Forecast System Reanalysis and ensemble NLDAS. **Journal of Hydrometeorology**, v. 12, n. 2, p. 181-205, 2011. https://doi.org/10.1175/2010JHM1310.1

MODARRES, R.; OUARDA, T. BMJ. Modeling the relationship between climate oscillations and drought by a multivariate GARCH model. **Water Resources Research**, v. 50, n. 1, p. 601-618, 2014. https://doi.org/10.1002/2013WR013810

MORETTIN, P. A.; TOLOI, C. Análise de séries temporais. In: **Análise de séries temporais**. 2006.

MÜLLER, I. I.; KRÜGER, C. M.; KAVISKI, E. Análise de estacionariedade de séries hidrológicas na bacia incremental de Itaipu. **Revista Brasileira de Recursos Hídricos**, v. 3, n. 4, p. 51-71, 1998.

NALLEY, D. et al. Inter-annual to inter-decadal streamflow variability in Quebec and Ontario in relation to dominant large-scale climate indices. **Journal of Hydrology**, v. 536, p. 426-446, 2016. https://doi.org/10.1016/j.jhydrol.2016.02.049

NOURANI, V. et al. Using self-organizing maps and wavelet transforms for space–time pre-processing of satellite precipitation and runoff data in neural network based rainfall–runoff modeling. **Journal of hydrology**, v. 476, p. 228-243, 2013. https://doi.org/10.1016/j.jhydrol.2012.10.054

ONS - OPERADOR NACIONAL DO SISTEMA ELÉTRICO. **O que é o SIN - Sistema Interligado Nacional**. Rio de Janeiro: ONS, 2016. Available in: <http://www.ons.org.br/paginas/sobre-o-sin/o-que-e-o-sin>. Access in: out. 2018.

PASQUINI, A. I.; DEPETRIS, P. J. Discharge trends and flow dynamics of South American rivers draining the southern Atlantic seaboard: An overview. **Journal of hydrology**, v. 333, n. 2-4, p. 385-399, 2007. https://doi.org/10.1016/j.jhydrol.2006.09.005

PATHAK, P. et al. Wavelet-aided analysis to estimate seasonal variability and dominant periodicities in temperature, precipitation, and streamflow in the Midwestern United States. **Water resources management**, v. 30, n. 13, p. 4649-4665, 2016. https://doi.org/10.1007/s11269-016-1445-0

PEDROSA, V. de A; SOUZA, R. C. de. Estacionariedade e estudo de vazões mínimas do rio Paraíba do Meio em Alagoas. In.: _. **Anais do XVI Simpósio Brasileiro de Recursos Hídricos**. 2009.

PERCIVAL, D. B.; WALDEN, A. T. **Wavelet methods for time series analysis**. Cambridge university press, 2000.

PFAFF, B. Analysis of integrated and cointegrated time series with R. **Springer Science and Business Media**, 2008.

PHILLIPS, P. CB; PERRON, P. Testing for a unit root in time series regression. **Biometrika**, v. 75, n. 2, p. 335-346, 1988. https://doi.org/10.1093/biomet/75.2.335

POHLERT, Thorsten. Non-parametric trend tests and change-point detection. CC BY-ND, v. 4, 2016.

RABINER, L. R. A tutorial on hidden Markov models and selected applications in speech recognition. **Proceedings of the IEEE**, v. 77, n. 2, p. 257-286, 1989.

ROBERTSON, A. W.; KIRSHNER, S.; SMYTH, P. Hidden Markov models for modeling daily rainfall occurrence over Brazil. **Information and Computer Science, University of California**, 2003.

RÖSCH, A.; SCHMIDBAUER, H. **WaveletComp 1.1: A guided tour through the R package**. 2014.

SALAS, J. D. **Applied modeling of hydrologic time series**. Water Resources Publication, 1980.

SALAS, J. D.; OBEYSEKERA, J. Revisiting the concepts of return period and risk for nonstationary hydrologic extreme events. **Journal of Hydrologic Engineering**, v. 19, n. 3, p. 554-568, 2013. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000820

SANG, Y. F. A review on the applications of wavelet transform in hydrology time series analysis. **Atmospheric research**, v. 122, p. 8-15, 2013. https://doi.org/10.1016/j.atmosres.2012.11.003

SCHWARZ, G. et al. Estimating the dimension of a model. **The annals of statistics**, v. 6, n. 2, p. 461-464, 1978.

SCOTT, A. J.; KNOTT, M. A cluster analysis method for grouping means in the analysis of variance. **Biometrics**, p. 507-512, 1974.

SEN, A.; SRIVASTAVA, M. S. On tests for detecting change in mean. **The Annals of statistics**, p. 98-108, 1975.

SEN, P. K. Estimates of the regression coefficient based on Kendall's tau. **Journal of the American statistical association**, v. 63, n. 324, p. 1379-1389, 1968.

SERINALDI, F.; KILSBY, C. G. Stationarity is undead: Uncertainty dominates the distribution of extremes. **Advances in Water Resources**, v. 77, p. 17-36, 2015. https://doi.org/10.1016/j.advwatres.2014.12.013

SETHI, R. et al. Performance evaluation and hydrological trend detection of a reservoir under climate change condition. **Modeling Earth Systems and Environment**, v. 1, n. 4, p. 33, 2015. https://doi.org/10.1007/s40808-015-0035-0

SHUKLA, S.; WOOD, A. W. Use of a standardized runoff index for characterizing hydrologic drought. **Geophysical research letters**, v. 35, n. 2, 2008. https://doi.org/10.1029/2007GL032487

SILVA, R. M. et al. Space-time variability of rainfall and hydrological trends in the Alto São Francisco River basin. **IAHS-AISH Publ**, v. 359, p. 48-54, 2013.

SILVEIRA, C. S.; SOUZA FILHO, F. A.; LÁZARRO, Y. M. Avaliação de desempenho dos modelos de mudança climático do IPCC-AR4 quanto à sazonalidade e os padrões de variabilidade interanual da precipitação sobre a Nordeste do Brasil, bacia da Prata e Amazônia. **Revista Brasileira de Recursos Hídricos**, 2011.

SILVEIRA, C. S. et al. Monthly streamflow forecast for National Interconnected System (NIS) using Periodic Auto-regressive Endogenous Models (PAR) and Exogenous (PARX) with climate information. **Revista Brasileira de Recursos Hídricos**, v. 22, 2017. http://dx.doi.org/10.1590/2318-0331.011715186.

SILVEIRA, C. S. et al. Mudanças climáticas na bacia do rio São Francisco: uma análise para precipitação e temperatura. **Revista Brasileira de Recursos Hídricos**, v. 21, n. 2, p. 416-428, 2016.

SIVAKUMAR, B. **Chaos in hydrology: bridging determinism and stochasticity**. Springer, 2016.

SOUZA FILHO, F. A.; LALL, U. Seasonal to interannual ensemble streamflow forecasts for Ceara, Brazil: Applications of a multivariate, semiparametric algorithm. **Water Resources Research**, v. 39, n. 11, 2003. https://doi.org/10.1029/2002WR001373

STAGGE, J. H. et al. Candidate distributions for climatological drought indices (SPI and SPEI). **International Journal of Climatology**, v. 35, n. 13, p. 4027-4040, 2015. https://doi.org/10.1002/joc.4267

STOCKER, T. F. et al. IPCC, 2013: **Climate Change 2013: The Physical Science Basis.** Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, 1535 pp. 2013.

THOMAS, H. A.; FIERING, M. B. **Design of Water Resources Systems**. 1962.

TORRENCE, C.; COMPO, G. P. A practical guide to wavelet analysis. **Bulletin of the American Meteorological society**, v. 79, n. 1, p. 61-78, 1998.

TORRES, M. E. et al. A complete ensemble empirical mode decomposition with adaptive noise. In: **Acoustics, speech and signal processing (ICASSP), 2011 IEEE international conference on**. IEEE, 2011. p. 4144-4147. https://doi.org/10.1109/ICASSP.2011.5947265

UM, M. J. et al. Performance Evaluation of four Statistical Tests for Trend and Non-stationarity and Assessment of Observed and Projected Annual Maximum Precipitation Series in Major United States Cities. **Water Resources Management**, v. 32, n. 3, p. 913-933, 2018. https://doi.org/10.1007/s11269-017-1846-8

VICENTE-SERRANO, S. M.; BEGUERÍA, S.; LÓPEZ-MORENO, J. I. A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. **Journal of climate**, v. 23, n. 7, p. 1696-1718, 2010. https://doi.org/10.1175/2009JCLI2909.1

VISSER, I. Seven things to remember about hidden Markov models: A tutorial on Markovian models for time series. **Journal of Mathematical Psychology**, v. 55, n. 6, p. 403-415, 2011.

VITERBI, A. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. **IEEE transactions on Information Theory**, v. 13, n. 2, p. 260-269, 1967.

WU, Z.; HUANG, N. E. Ensemble empirical mode decomposition: a noise-assisted data analysis method. **Advances in adaptive data analysis**, v. 1, n. 01, p. 1-41, 2009. https://doi.org/10.1142/S1793536909000047

YEVJEVICH, V. Stochastic models in hydrology. **Stochastic Hydrology and Hydraulics**, v. 1, n. 1, p. 17-36, 1987. https://doi.org/10.1007/BF01543907

ZHAO, X.; CHEN, X. Auto regressive and ensemble empirical mode decomposition hybrid model for annual runoff forecasting. **Water Resources Management**, v. 29, n. 8, p. 2913-2926, 2015. https://doi.org/10.1007/s11269-015-0977-z

ZUCCHINI, W; MACDONALD, I L.; LANGROCK, R. **Hidden Markov models for time series: an introduction using R**. Chapman and Hall/CRC, 2016.