



DÉBORA ALVES DE OLIVEIRA VITOI

**COEFICIENTE DE HURST NA COMPARAÇÃO DE
SÉRIES DE NÍVEL DO MAR**

LAVRAS - MG

2016

DÉBORA ALVES DE OLIVEIRA VITOI

**COEFICIENTE DE HURST NA COMPARAÇÃO DE SÉRIES DE NÍVEL
DO MAR**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Mestre.

Orientadora
Dra. Thelma Sáfiadi

**LAVRAS - MG
2016**

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Vitói, Débora Alves de Oliveira.

Coeficiente de Hurst na comparação de séries de nível do mar /
Débora Alves de Oliveira Vitói. - 2016.

69 p. : il.

Orientador(a): Thelma Sáfadi.

.
Dissertação (mestrado acadêmico) - Universidade Federal de
Lavras, 2016.

Bibliografia.

1. Séries Temporais. 2. Memória Longa. 3. Análise de
agrupamento. I. Sáfadi, Thelma . . II. Título.

DÉBORA ALVES DE OLIVEIRA VITOI

**COEFICIENTE DE HURST NA COMPARAÇÃO DE SÉRIES DE NÍVEL
DO MAR**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Mestre.

APROVADA em 16 de setembro de 2016.

Dra. Carla Guimarães Brighenti

UFSJ

Dr. Renato Ribeiro de Lima

UFLA

Orientadora
Dra. Thelma Sáfydi

**LAVRAS - MG
2016**

Sonho que se sonha junto, não é sonho é realidade. E assim é a vida, cheia de sonhos efêmeros e cheia de sonhos reais. Quantos sonhos são deixados pelo caminho pela falta de cumplicidade, mas dedico a vocês, meus pais, minha irmã, meu marido, minha afilhada e minha filha Maria Clara, razão principal de minha luta e à toda minha família; hoje posso desfrutar a felicidade de mais essa conquista. Dedico a todos meus amigos que me permitiram fazer parte dessa nova caminhada. À minha eterna companheira vovó, (in memoriam) que me ensinou lições que jamais esquecerei.

AGRADECIMENTOS

A diferença da qualidade de germinação da semente plantada está diretamente ligada à qualidade da terra em que ela foi lançada. Se hoje colho as frutas de mais essa conquista, devo às pessoas diretamente envolvidas no processo. Por isso, ficam aqui meus agradecimentos.

Aos meus pais pelo apoio e confiança. À minha irmã, Ana Paula, que me proporcionou todo carinho e amor mesmo estando longe, minha afilhada Beatriz, pelo abraço e sorriso quando mais precisei. E todos os familiares que, de alguma forma, me ajudaram a chegar até aqui.

Ao meu marido, Rodrigo, por ter acreditado e me dado esperança nos momentos mais difíceis. E a toda sua família, por ter me acolhido quando foi necessário.

Aos padrinhos da minha filha, que me ajudaram de uma forma singela e carinhosa.

Colégio Papagaio Amarelo, não sabem como sou grata por terem participado dessa conquista. Vocês proporcionaram à minha filha a alegria e o carinho quando não estava presente. Obrigada por terem feito os dias da Maria mais felizes e cumprirem compromisso de ensiná-la.

Maria Clara, minha princesa, momentos de choros, de luta e de conquistas. Não foi fácil levar às aulas, deixá-la tão nova na escola. Mas, acredito que este sonho se realizou. Todos os dias sabia que estaria me esperando com o melhor abraço de filha.

A todos os professores do DEX, pelo carinho e paciência. Esta conquista não foi só minha, e sim nossa.

À secretaria do DEX, obrigada pela dedicação. Em especial à Nádia, pela atenção.

Dra. Maria do Carmo, me lembro como se fosse ontem, suas cobranças e puxões de orelhas, bem merecidos. Obrigada, pelo carinho e por ter acreditado em mim. Pessoas especiais, fazem parte de momentos especiais. Sou grata pelo seu ensinamento.

À CNPQ, por ter me proporcionado essa conquista. Sem a ajuda financeira, não teria realizado meu sonho.

Dra. Thelma Sáfadi, obrigada pela orientação. Por ter dado a mão quando pensei que iria cair e pela força quando queria desistir. Tenho orgulho de poder lhe chamar de mãe.

Aos novos e os antigos amigos da UFLA. Amigos que fizeram a diferença nessa conquista.

Janaína Marques, Kelly Lima, Lilian Oliveira, Sidcleide Barbosa e Ricardo Almeida, quero agradecê-los, por essa conquista. Acreditem, vocês souberam sorrir, abraçar, chorar, ajudar, amar a "filha filha", como irmãs, mães e amigas. Eu não conseguiria sozinha, vocês sempre estavam ao meu lado em todos os momentos. Janaína, com seu jeito sábio de resolver os problemas. Kelly, com as melhores merendas durante o dia a dia. Lilian, com seu sorriso encantador que me fortalecia. Sidcleide, acredite, você me ajudou a ser mãe. Ricardo, pelo carinho e apoio de sempre. Só tenho a agradecer a vocês, essa conquista é NOSSA.

"A menos que modifiquemos a nossa maneira de pensar, não seremos capazes de resolver os problemas causados pela forma como nos acostumamos a ver o mundo."

Albert Einstein

RESUMO

O coeficiente de Hurst é uma medida utilizada para determinar a memória longa e similaridades em séries temporais, podendo ser estimado pela Estatística R/S, através do Método da Variância Agregada e pelo Método da Diferenciação da Variância Agregada. Neste trabalho, o coeficiente de Hurst foi aplicado em séries mensais do nível do mar de 10 países obtido em um período de 26 anos, para verificar a existência do efeito memória longa. Além disso, a partir da estimação dos coeficientes obtidos pelos diferentes métodos gerar agrupamentos para detectar a similaridade existente entre as séries temporais. Observou-se similaridade entre níveis do mar da Noruega e Nova Zelândia; Tailândia, Malásia e Hong-Kong e também entre Brasil, Alasca, Japão, Austrália e Singapura.

Palavras-chave: séries temporais, memória longa, análise de agrupamento.

ABSTRACT

The Hurst coefficient is a measure used to determine long memory and similarities in time series, which can be estimated by the R / S Statistics, the Aggregate Variance Method and the Aggregate Variance Differentiation Method. In this work, the Hurst coefficient was applied in monthly series of the level of the but of 10 countries obtained in a period of 26 years, to verify the existence of the long memory effect. In addition, from the estimation of the coefficient to generate clustering to detect the similarity existing between the time series. In view of this, you may note that the following groups were formed: by Norway and New Zealand; Thailand, Malaysia and Hong Kong and the third group Brazil, Alaska, Japan, Australia and Singapore.

Keywords: time series, long memory, cluster analysis.

LISTA DE FIGURAS

Figura 1	Nível do mar de dado local durante dois dias.	18
Figura 2	Gráfico da série do Japão (a), gráfico da série estacionária (b). . .	44
Figura 3	Gráfico da série Austrália (a), gráfico da série estacionária (b). . .	45
Figura 4	Gráfico da série Brazil (a), gráfico da série estacionária (b). . . .	46
Figura 5	Gráfico da série Alasca (a), gráfico da série estacionária (b). . . .	47
Figura 6	Gráfico da série Malásia (a), gráfico da série estacionária (b). . .	48
Figura 7	Gráfico da série Hong Kong (a), gráfico da série estacionária (b). .	49
Figura 8	Gráfico da série Singapura (a), gráfico da série estacionária (b). .	50
Figura 9	Gráfico da série Tailândia (a), gráfico da série estacionária (b). .	51
Figura 10	Gráfico da série Noruega (a), gráfico da série estacionária (b). . .	52
Figura 11	Gráfico da série Nova Zelândia (a), gráfico da série estacionária (b).	53
Figura 12	Função de autocorrelação e função de autocorrelação parcial. . .	54
Figura 13	Função de autocorrelação e função de autocorrelação parcial. . .	55
Figura 14	Função de autocorrelação e função de autocorrelação parcial. . .	56
Figura 15	Função de autocorrelação e função de autocorrelação parcial. . .	57
Figura 16	Função de autocorrelação e função de autocorrelação parcial. . .	58
Figura 17	Função de autocorrelação e função de autocorrelação parcial. . .	59
Figura 18	Função de autocorrelação e função de autocorrelação parcial. . .	60
Figura 19	Função de autocorrelação e função de autocorrelação parcial. . .	61
Figura 20	Função de autocorrelação e função de autocorrelação parcial. . .	62
Figura 21	Função de autocorrelação e função de autocorrelação parcial. . .	63
Figura 22	Análise de agrupamento: no gráfico (a) agrupamentos com 3 gru- pos e no gráfico (b) agrupamentos com 4 grupos.	66

LISTA DE TABELAS

Tabela 1	Local em que a série do nível do mar foi analisado.	40
Tabela 2	Valores estimados de d para cada país.	44
Tabela 3	Valores estimados do coeficiente de Hurst para cada país.	64

SUMÁRIO

1	INTRODUÇÃO	13
2	REFERENCIAL TEÓRICO	15
2.1	Revisão Bibliográfica	15
2.2	Séries Temporais	16
2.3	Modelos de séries temporais	18
2.3.1	Tendência	19
2.3.2	Sazonalidade	21
2.3.3	Estacionariedade	22
2.4	Análise de agrupamento	24
2.5	Coefficiente de Hurst	27
2.6	Memória longa	29
2.7	Modelo ARFIMA	30
2.8	Estatística R/S	34
2.9	Estimadores para o coeficiente de Hurst	35
2.9.1	Método da Estatística R/S	35
2.9.2	Método da variância agregada	36
2.9.3	Método da diferenciação da variância agregada	38
2.10	Nível de mar	39
3	METODOLOGIA	40
3.1	Dados do nível mar	40
3.2	Análise dos dados	41
4	RESULTADOS E DISCUSSÃO	43
5	CONCLUSÃO	67
	REFERÊNCIAS	68

1 INTRODUÇÃO

Existem fatores climáticos, como índices pluviométricos e nível do mar, nos quais estudos e avaliações são de interesse econômico por possuírem sazonalidade (ou estão em nível agregado) em um determinado intervalo de tempo. Um desses fatores é o aumento do nível do mar causado pelo clima da Terra, problema que vem preocupando a população, pois o nível do mar vem aumentando a cada ano. Esse aumento advém do derretimento das geleiras, da expansão térmica, em consequência da mudança climática e, principalmente da queima de combustíveis fósseis.

Ciente dos danos que podem ser causados pelo aumento do nível do mar, Harold Edwin Hurst trabalhou no projeto para a construção de uma represa, com o objetivo de deixá-la com um nível de água estável, de forma que o rio não transbordasse e nem esvaziasse totalmente. Diante disso, Hurst, criou uma medida, conhecida como coeficiente de Hurst (H) para calcular a quantidade de água apropriada.

Nesse projeto, Hurst percebeu que os fenômenos naturais, tais como o fluxo de água dos rios, o regime das chuvas e a temperatura, seguiam um passeio aleatório viesado, ou seja, uma tendência com ruído que poderiam ser mensurados pelo valor do coeficiente H.

O interesse pelo coeficiente de Hurst tem motivado muitos pesquisadores, pois tem como objetivo modelar e analisar o processo de memória longa, que é determinado pela presença de dependência entre as observações distantes por um longo período de tempo, ou seja, eventos presentes influenciam eventos futuros distantes.

O objetivo principal deste trabalho é obter o coeficiente de Hurst em séries do nível do mar e utilizar o valor do coeficiente estimado na análise de agrupa-

mento para detectar a similaridade existente entre as séries do nível do mar. Além disso, comparar o resultado obtido com o encontrado em Sáfadi (2014) utilizando componentes independentes.

2 REFERENCIAL TEÓRICO

A seguir, serão apresentados os principais conceitos de Séries Temporais, Análise de Agrupamento e Memória Longa. Além disso, os principais métodos para o cálculo do coeficiente de Hurst são descritos.

2.1 Revisão Bibliográfica

Os estudos sobre o coeficiente de Hurst, vêm sendo uma hipótese de estudos para as áreas financeiras, biológicas, físicas, geográficas, entre outras. A seguir são apresentados alguns artigos que utilizaram o coeficiente de Hurst, para uso em séries temporais.

Rosso e Taquq (1997), apresentaram um estudo sobre os dados mensais e diários do Lago Maggiore, na Itália. Utilizaram o modelo autorregressivo diferenciado fracionado médias móveis (ARFIMA) em contraste com os modelos tradicionais ARIMA. Essa abordagem permite a modelagem de curto e longo prazo em uma série temporal, sem a necessidade dos dados serem gaussianos. Apresentaram ainda alguns métodos para a estimação dos valores do coeficiente de Hurst, sendo comparados entre eles.

Carbone, Castelli e Satanley (2004), calcularam o coeficiente de Hurst de várias séries temporais utilizando a diferenciação de médias móveis. A fim de avaliar a precisão da técnica, foram calculados os valores do expoente de Hurst para a série artificial, simulando caminhos *brownianos*. Foi descoberta uma maior variabilidade do coeficiente com a série local em comparação com as artificiais. O algoritmo diferenciado médias móveis, permite o cálculo do valor do expoente sem qualquer hipótese sobre o processo estocástico e sobre a função de probabilidade das variáveis aleatórias. A presente técnica examina o coeficiente de H em torno de um dado instante.

Tupayagale et al. (2012) realizaram um estudo baseado nos métodos de *wavelets*, que utiliza estruturas fractais nos dados de frequência, a fim de testar se a componente de longo prazo pode ser explorada para melhores previsões de volatilidade e fornecer previsões mais precisas para medir, monitorar e gerenciar o risco do mercado financeiro.

Mitra et al. (2012) utilizaram dados diários de série de índice de ações dos últimos 10 anos, que foram subdivididos em séries menores de 60 períodos, sendo comparados e examinados pelo o coeficiente de Hurst para previsão das séries financeiras. Foram observados períodos que obtiveram alto valores do coeficiente de Hurst e que produziram melhores lucros comerciais utilizando análises de fluídos diferenciados.

Safadi (2014) propôs uma análise de agrupamentos para séries temporais de níveis do mar em diferentes países durante o período de 26 anos utilizando componentes independentes. Foram considerados diferentes números de componentes independentes. A análise permitiu não só o agrupamento das séries, como também forneceu informações sobre as características comuns aos grupos, a partir da análise das componentes.

2.2 Séries Temporais

Na Estatística existem situações de estudo onde o conjunto de dados pode ser obtido por meio de observações periódicas do evento de interesse, que pode ser, por exemplo, a cada minuto ou hora, a cada dia, a cada mês ou ano, ou num período contínuo de tempo. Esses conjuntos de dados ordenados no tempo são denominados séries temporais.

Segundo Morettin e Toloí (2006), uma série temporal é um conjunto de observações coletadas de forma sequencial ao longo do tempo. Os principais objetivos de uma análise de séries temporais são construir modelos, descrever o com-

portamento da série, investigar o mecanismo gerador da série temporal, fazer previsão de valores futuros e procurar periodicidades relevantes nos dados.

Ao observar uma série temporal Z_t , contempla-se, na verdade, uma trajetória que descreve o comportamento de um determinado fato. Essa trajetória pode ser composta pela frequência com que o fato se repete em um intervalo de tempo, assim como variáveis métricas ao auferir e registrar variáveis em intervalos regulares (KASMIM et al., 2014). Para melhor compreensão, considere uma situação em que o interesse é coletar dados do nível do mar em uma determinada localidade, em dois dias distintos. Nesse caso, o marégrafo pode fornecer dois níveis diferentes, devido a diversos fatores que podem estar causando variações nessa medida. Assim, pode-se obter duas curvas que não apresentam o mesmo comportamento, como mostra a Figura 1.

Cada uma destas curvas, $Z^1(t)$ e $Z^2(t)$, apresentadas na Figura 1, é denominada de trajetória de um processo físico observado durante os tempos t . Pode-se considerar, que o lugar é fixo e quem varia é o tempo. O conjunto de todas as trajetórias possíveis que podem ser observada é chamado de processo estocástico, o qual pode ser definido como um conjunto de variáveis aleatórias que dependem de outras variáveis. Vale ressaltar que, em séries temporais, existe a dependência do valor em um instante t qualquer com as observações vizinhas. Diante disso, a técnica de séries temporais tem como intuito analisar e modelar essa dependência.

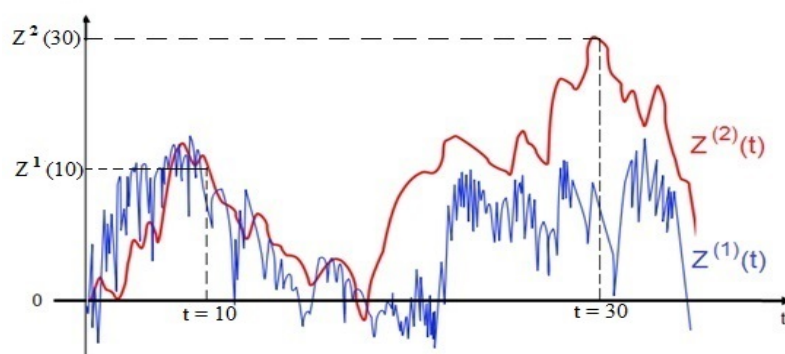


Figura 1 Nível do mar de dado local durante dois dias.

As séries temporais podem ser classificadas em dois casos: discretas, quando as observações são feitas em tempos específicos $T = \{t_1, t_2, \dots, t_n\}$. Nesse caso a série será denotada por $\{Z_t : t \in T\}$. Outro caso é o de séries contínuas, quando as observações são feitas em um intervalo de tempo contínuo $\{T \in t_i < t < t_j\}$ e a série será denotada por $\{Z(t) : t \in T\}$.

2.3 Modelos de séries temporais

As séries temporais podem ser caracterizadas ou descritas de acordo com modelos. Os modelos podem ser construídos em função de vários fatores, tais como: comportamento do fenômeno ou conhecimento empírico da natureza e do objeto de análise. A descrição desses modelos podem contar com o auxílio de métodos de estimação e da disponibilidade de *software* adequado (RIHBANE, 2014).

Para a composição de um modelo em séries temporais somam-se os componentes periódicos e a variação aleatória. De acordo com Oliveira e Favero (2002), são conhecidas como componentes de uma série temporal a tendência, a sazonalidade e o ciclo.

A questão crucial do modelo clássico é decidir como será a equação que relaciona as componentes com a variável, sendo através do modelo aditivo ou do modelo multiplicativo. No modelo aditivo o valor da série Z_t será o resultado da soma dos valores das componentes:

$$Z_t = T_t + S_t + a_t,$$

em que T_t é a componente de tendência, S_t é a componente de sazonalidade e a_t é a componente aleatória com média zero e variância finita. Por outro lado, o modelo multiplicativo considera que a série temporal Z_t é resultante do produto das componentes, dado por:

$$Z_t = T_t S_t a_t.$$

É importante ressaltar, que nem sempre, uma série temporal apresentará toda as componentes. A decomposição da série permitirá identificar quais componentes estão atuando naquele conjunto em particular, além de possibilitar a obtenção de índices ou de equações para realizar previsões para períodos futuros da série. Lembrando que, independente do objetivo da análise para a série, o modelo precisa ser o mais simples possível e parcimonioso.

É imprescindível que se escolha um modelo adequado na modelagem da série temporal, pois este pode revelar algumas características particulares que irão auxiliar na previsão dos valores futuros ou apenas para descrever o comportamento da mesma.

2.3.1 Tendência

Morettin e Toloi (2006) definem a tendência como aumento ou diminuição gradual das observações ao longo do período. Assim, a tendência dos dados per-

mite avaliar se os valores da série crescem, decrescem ou permanecem estáveis. Normalmente, a tendência é o elemento mais importante de uma série temporal, podendo ser linear, polinomial ou exponencial.

A estimação da tendência pode ser realizada por diversos métodos. Segundo Morettin e Tolo (2006), a estimação pode ser feita por meio do ajuste de uma função do tempo, como um polinômio, uma exponencial ou outra função suave de t , por meio da suavização (ou filtragem) dos valores da série ao redor de um ponto, para estimar a tendência naquele ponto ou suavização dos valores da série através de sucessivos ajustes de retas de mínimos quadrados ponderados ("lowes").

Para determinar a tendência, supõe-se que a componente sazonal não esteja presente e o modelo é aditivo, ou seja,

$$Z_t = T_t + a_t$$

em que, a_t é o ruído branco.

Ao estimar a tendência T_t , a série ajustada Y_t é livre de tendência e pode ser descrita por

$$Y_t = Z_t - \hat{T}_t,$$

em que, Z_t é a série temporal, \hat{T}_t é o estimador da tendência e Y_t é a série ajustada.

Outro modo de eliminar a tendência consiste em tomar diferenças sucessivas na série original até conseguir uma série estacionária, ou seja, que se desenvolva de forma aleatória no tempo em torno de uma média constante. A primeira diferença é dada por

$$\Delta Z_t = Z_t - Z_{t-1}.$$

A segunda diferença é,

$$\begin{aligned}
 \Delta^2 Z_t &= \Delta(\Delta Z_t) \\
 &= \Delta(Z_t - Z_{t-1}) \\
 &= Z_t - Z_{t-1} - (Z_{t-1} - Z_{t-2}) \\
 &= Z_t - 2Z_{t-1} + Z_{t-2}.
 \end{aligned}$$

Generalizando temos,

$$\Delta^n Z_t = \Delta\{\Delta^{n-1} Z_t\}$$

em que, $t = 1, 2, \dots, n$.

A verificação da tendência deve ser feita antes de realizar a sua estimação; para isso, existem alguns testes, como teste de Wald - Wolfowitz, teste de Cox-Stuart e o teste de Mann-Kendall, que são apresentados com detalhes em Morettin e Tolo (2006).

2.3.2 Sazonalidade

Uma série temporal é definida como sazonal quando os fenômenos que ocorrem durante o tempo se repetem a cada período idêntico de tempo. Assim, a sazonalidade corresponde às oscilações de subida e queda que sempre ocorrem em um determinado período do ano, do mês, da semana ou do dia.

Geralmente, as séries sazonais apresentam alta correlação em *lags* sazonais. Morettin e Tolo (2006) relatam que os métodos mais comuns para estimar a sazonalidade são os procedimentos de regressão, usado para sazonalidade determinística, e de médias móveis, apropriado para séries cuja componente sazonal varia com o tempo.

Obtida a estimativa de S_t , se o modelo for aditivo, a série pode ser apresentada por

$$Z_t^{SA} = Z_T - \hat{S}_t.$$

No caso do modelo for multiplicativo,

$$Z_t^{SA} = \frac{Z_T}{\hat{S}_t},$$

em que, Z_t^{SA} é a série sazonalmente ajustada e \hat{S}_t é o estimador de S_t .

Um dos procedimentos mais usuais para eliminar a sazonalidade, é tomar médias sazonais. Porém, é um procedimento em que muitos dados são perdidos. Para verificar a sazonalidade existem alguns testes de hipóteses estatísticos, como teste de Fisher, teste de Kruskal-Wallis, Teste de Friedman e o teste de análise de variância, vistos com mais detalhes em Morettin e Toloí (2006).

2.3.3 Estacionariedade

Uma das principais conjecturas para uma série temporal é a que ela seja estacionária, ou seja, que ela se desenvolva no tempo aleatoriamente ao redor de uma média constante, reproduzindo uma forma de equilíbrio estável (MORETTIN; TOLOI, 2006). Existem dois tipos de estacionariedade: a estacionariedade fraca (ou de segunda ordem) e a estacionariedade estrita (forte).

A ideia de estacionariedade é determinada por um processo estocástico de forma matemática. Morettin e Toloí (2006) definem um processo como fortemente estacionário e fracamente estacionário. Sendo considerado fraco se:

- (i) $E[Z(t)] = \mu$;
- (ii) $Cov[Z(t), Z(t + \tau)] = \gamma(\tau)$;

em que, τ é um valor qualquer e γ é a função de autocovariância.

Um processo com estacionariedade forte consiste numa sequência de variáveis aleatórias (a_t) , mutuamente independentes e identicamente distribuídas, tendo média e variância constante. Porém, raramente se consegue séries com estacionariedade forte, por isso se define as duas e assume a fraca como fundamental.

Na prática, a maioria das séries não apresenta um comportamento de estacionariedade, dado que uma série pode apresentar tendência ou ainda um comportamento explosivo. Por isso, é necessária a identificação e eliminação de tais componentes para a obtenção de um processo estacionário.

A função de autocorrelação (fac) mede o grau de correlação de uma variável em um dado instante, ou seja, quanto o valor de uma variável aleatória é capaz de influenciar seus vizinhos. É uma importante ferramenta para identificar as propriedades de uma série, pois analisa a correlação entre as observações de uma mesma variável em diferentes horizontes de tempo.

Definição 1. A função de autocorrelação de um processo Z_t com variância σ^2 e média μ , é dada por:

$$\rho(\tau) = \frac{\gamma_\tau}{\gamma_0}$$

em que, γ_τ , é a função de autocovariância, $\gamma_0 = \sigma^2$. Se $\tau = 0$, então, a função de autocorrelação $\rho(0) = 1$.

São aceitas as seguintes propriedades para funções de autocorrelação:

(i) A correlação entre $Z(t)$ e $Z(t + \tau)$ é a mesma para $Z(t)$ e $Z(t - \tau)$, ou seja,
 $\rho(\tau) = \rho(-\tau)$;

(ii) $-1 < \rho(\tau) < 1$.

A demonstração das propriedades encontram-se em Morettin e Toloi (2006).

As propriedades de autocorrelação (fac) são análogas ao coeficiente de autocorrelação, pois, r_k é o estimador da função de autocorrelação e c_k é o estimador da função de autocovariância. E, são ferramentas importantes para acessar as propriedades da série temporal.

Considere as observações, (X_1, X_2, \dots, X_n) e o estimador da função de autocorrelação r_k é dado por:

$$r_k = \frac{c_k}{c_0}, k = 0, 1, \dots, N - 1,$$

em que, $c_k = \sum_{t=1}^{n-k} \frac{(x_t - \bar{x})(x_{t+k} - \bar{x})}{n}$, o estimador da função de autocovariância

e $\bar{x} = \frac{1}{N} \sum_{t=1}^N x_t$ a média amostral.

O valor da autocorrelação está entre $-1 \leq r_k \leq 1$, em que -1 indica a anti-correlação perfeita, 1 a correlação perfeita e 0 significa à ausência de correlação.

É possível encontrar vários processos estocástico com a mesma função de autocorrelação, o que dificulta a interpretação do correlograma. O correlograma é um gráfico que representa os primeiros coeficientes de autocorrelação e pode ser uma ferramenta poderosa para identificar as características de uma série temporal.

2.4 Análise de agrupamento

A análise de agrupamento é um conjunto de técnicas computacionais cujo propósito é classificar objetos baseando-se em suas características. Além de ser uma ferramenta útil para a análise de dados em muitas situações diferentes, pode ser utilizada para reduzir a dimensão de um conjunto de dados.

A ideia básica da análise de agrupamento consiste em colocar em um mesmo grupo, os objetos que sejam similares de acordo com algum critério pré-

determinado (LIDEN, 2009). Segundo Ferreira (2011), a semelhança entre os objetos é quantificada por meio de uma medida de proximidade, que engloba tanto as medidas de similaridades quanto as de dissimilaridades.

A busca por um agrupamento perfeito, ou seja, aquele em que os objetos dentro dos grupos são semelhantes, mas os grupos são dissimilares entre si, não é uma tarefa simples e requer a utilização de alguns métodos específicos (FERREIRA, 2011).

Tendo em vista a dificuldade de se obter a homogeneidade entre os grupos e heterogeneidade entre eles, foram desenvolvidas diversas técnicas capazes de auxiliar na formação dos agrupamentos.

Os métodos de agrupamentos são divididos em métodos hierárquicos e não hierárquicos. Os métodos hierárquicos são os mais comumente utilizados (FERREIRA, 2011) e consistem em iniciar representando cada objeto como um grupo e, a cada passo, um grupo ou elemento é ligado a outro de acordo com sua similaridade, até o último passo, onde é formado um grupo único com todos os elementos (OLIVEIRA, 2004).

Dentre os métodos hierárquicos, temos o agrupamento hierárquico de Ward, também conhecido como método do incremento das somas de quadrados. As somas de quadrados são distâncias quadráticas (FERREIRA, 2011). Nesse método, a soma de quadrado dentro dos grupos é utilizada como critério de agrupamento, determinando a homogeneidade entre os objetos. Inicialmente, tem-se n grupos, cada um contendo um dos objetos, ou seja, temos tantos grupos quanto objetos. Posteriormente, cada objeto ou grupo é agrupado com outro grupo.

A soma dos erros quadráticos para cada agrupamento é definida como:

$$\begin{aligned} SSE_l &= \sum_{i=1}^{n_l} (y_i^{(l)} - \bar{y}_{\cdot}^{(l)})^T (y_i^{(l)} - \bar{y}_{\cdot}^{(l)}) \\ &= \sum_{i=1}^{n_l} \sum_{j=1}^p (y_{ij}^{(l)} - \bar{y}_{\cdot j}^{(l)})^2 \end{aligned}$$

em que n_l é o número de objetos no l -ésimo grupo, $n = \sum_{l=1}^k n_l$ é o número total de objetos, $\bar{y}_{\cdot}^{(l)}$ é o valor de médias do l -ésimo grupo dada por

$$\bar{y}_{\cdot}^{(l)} = \frac{1}{n_l} \sum_{i=1}^{n_l} y_i^{(l)}$$

e $\bar{y}_{\cdot j}^{(l)}$ j -ésimo componente de $\bar{y}_{\cdot}^{(l)}$, ou seja, a média do l -ésimo grupo relativa a j -ésima variável.

Considerando dois grupos quaisquer R e S, aglomera-se o par que minimiza a troca SSE, provocada pela junção dos grupos R e S em um único grupo, o qual será denominado de ΔSSE . As somas de quadrados dos demais grupos, exceto R e S, permanecem inalteradas. Logo, a troca SSE provocada pela junção dos grupos R e S é dada por:

$$\Delta SSE = SSE_{rs} - SSE_r - SSE_s, \quad (2.1)$$

sendo,

$$SSE_{rs} = \sum_{i=1}^{n_{rs}} (y_i^{(rs)} - \bar{y}_{\cdot}^{(rs)})^T (y_i^{(rs)} - \bar{y}_{\cdot}^{(rs)}),$$

em que $n_{rs} = n_r + n_s$ e $\bar{y}_{\cdot}^{(rs)} = \frac{(n_r \bar{y}_{\cdot}^{(r)}) + (n_s \bar{y}_{\cdot}^{(s)})}{(n_r + n_s)}$ são o tamanho e o centróide do novo grupo RS respectivamente.

Ward (1963), mostrou que, a troca da soma de quadrados dos erros, provocada pela junção dos grupos R e S, é diretamente proporcional à distância quadrática entre os centroides dos grupos correspondentes, que pode ser simplificada em:

$$\Delta SSE = \frac{n_r n_s}{n_r + n_s} (\bar{y}^{(r)} - \bar{y}^{(s)})^T (\bar{y}^{(r)} - \bar{y}^{(s)}).$$

Contudo, Ward demonstrou que $\Delta SSE = 0,5d_{rs}^2$, ou seja, a metade da distância quadrática entre os centróides dos dois grupos. Minimizar ΔSSE é equivalente a minimizar uma quantidade proporcional às distâncias entre grupos.

2.5 Coeficiente de Hurst

Segundo Favaretto (2004), as inundações anuais do rio Nilo, no Egito, tiveram grande importância no sustento de uma grande parte da civilização antiga que dependia da agricultura desenvolvida às suas margens.

Além de favorecer a agricultura, as águas do rio eram utilizadas para beber, pescar, como meio de transporte de pessoas e mercadorias, além de aproveitadas para a construção de diques, reservatórios e canais de irrigação. O Lago de Méris, reservatório construído no Baixo Egito, é um exemplo desse trabalho que, durante os períodos de estiagem, abasteciam as terras localizadas ao norte do território.

Visto que as inundações ocorridas eram importantes para a população, seria de fundamental importância que o rio mantivesse um nível de água elevado. Diante disso, em 1906, o engenheiro Harold Edwin Hurst, foi designado para trabalhar no projeto de uma represa no rio Nilo.

De acordo com Favaretto (2004), Hurst gastou quase que toda a sua vida estudando os problemas relacionados com o armazenamento de água do rio e, no decorrer dos estudos, deparou-se com o problema de controle da vazão do reservatório, pois, um reservatório ideal nunca deve transbordar e nem esvaziar. Sendo

assim, houve a necessidade de estabelecer um plano de controle para a construção de um reservatório ideal, levando em consideração o comportamento aleatório (randômico) da influência das águas das chuvas que neste caso, é considerado como incontrolável.

Determinado em avaliar este fato, Hurst observou que, se um ano de muita enchente for conseqüente de outro ano de muita enchente e um ano de pouca enchente for seguido por outro ano de pouca enchente, esse fenômeno apresentava um comportamento persistente, ou seja, um comportamento não aleatório. Caso contrário se um ano de muita enchente fosse acompanhado por outro ano de pouca enchente, esse fenômeno era considerado como anti-persistente.

Diante dessas situações, Hurst propôs o coeficiente de Hurst (H), sendo bastante utilizado em estudos de séries temporais que apresentam comportamentos naturais como, enchentes e temperaturas. O objetivo do coeficiente de Hurst é quantificar a tendência em uma série, ou seja, encontrar e estabelecer a intensidade da tendência.

Para analisar o coeficiente H , Hurst estabeleceu valores do coeficiente de modo que o valor de H variasse de 0 a 1. Para os seguintes valores, seguem alguns resultados:

- (i) Para $0 < H < 0,5$: tem-se uma série anti-persistente, em que um aumento tenderá a ser seguido por um decréscimo, ou vice-versa.
- (ii) Para $H = 0,5$: a série tem um comportamento aleatório e os valores da série atual não se torna dependente dos valores da série do passado, ou seja, os desvios dos níveis das inundações, em relação ao seu valor médio são independentes (aleatórios);
- (iii) Para $0,5 < H < 1$: descreve um comportamento de persistência, em que um

decrecimento tenderá a seguir de um decrecimento. Assim, os desvios dos níveis das inundações são seguidos de inundações e níveis baixos, seguidos de níveis baixo.

No entanto, através dos estudos realizados sobre o rio Nilo, Hurst encontrou um valor do expoente de $H > 0,91$, o que implica uma maior variação do fluxo de água do que poderia ser possível. Mais detalhes podem ser vistos em Mitra (2012).

Alguns estimadores para o coeficiente de Hurst serão apresentados na seção 2.9.

2.6 Memória longa

O efeito de memória longa foi identificado pela primeira vez por Hurst em (1951) ao medir as variações de cheias e secas do rio Nilo. Assim, pôde-se constatar que a variação do volume seguia números similares de tempos em tempos, ou seja, um comportamento passado teria impacto e se repetiria posteriormente (MALAVOGLIA, 2009). À vista disso, a presença de memória longa em séries temporais retrata a dependência temporal persistente presente nos dados, ou seja, apresenta uma dependência significativa entre observações separadas por um longo tempo e apresentam alguma memória dos eventos passados.

Segundo Malavoglia (2009), o conceito de memória longa é definido como sendo aquele em que existe uma correlação entre os valores em determinada data e valores futuros a estes ou, ainda, se um evento causado em uma data possa ter sua repercussão percebida em demais eventos posteriores.

Uma definição mais formal, para memória longa é dada por:

Definição 2. *Suponha que Z_t tenha autocorrelação ρ_j . Dizemos que Z_t possui*

memória longa se:

$$\lim_{n \rightarrow \infty} \sum_{j=-n}^n |\rho_j|$$

é não finita.

Quando a série temporal tem um comportamento de memória longa, a sua função de autocorrelação decai durante o tempo, este decaimento segue uma lei de potência (BATISTA, 2006). Assim, um processo de memória longa é um processo estacionário em que a função de autocorrelação decresce hiperbolicamente para zero, isto é,

$$\rho \sim Cj^{2d-1}, j \rightarrow \infty, \quad (2.2)$$

em que, $C > 0$ e $0 < d < 0,5$.

Outra característica para uma série com memória longa, é a função densidade espectral, que diverge para o infinito na frequência zero. Isso corresponde a dizer que a função não é limitada na frequência zero, definida como:

$$f(\lambda) \rightarrow C_f \lambda^{d-1},$$

em que $\lambda \rightarrow 0$ e C_f é uma constante maior que zero.

A função $f(\lambda)$, é a transformada de Fourier da função de autocorrelação de uma série temporal (VEDOVATTO, 2014). Assim, podemos caracterizar que, a função de autocorrelação é análoga à função densidade espectral, pois o que as difere é que uma está definida no domínio de tempo e a outra no domínio da frequência, respectivamente.

2.7 Modelo ARFIMA

Para uma série temporal Z_t , um ajuste de um modelo é necessário para descrever e organizar informações de uma série. Os modelos clássicos (ARMA,

ARIMA), descritos por Box e Jenkins e Brockwell e Davis, tem tido uma grande aplicação em hidrologia. Nas décadas de 1970 e 1980, foram propostas algumas modificações desses modelos, a fim de ser capaz de modelar séries temporais sazonais e alguns dados não gaussianos (ROSSO; TAQQU, 1997). Uma extensão importante, foi elaborada por Granger e Joyeux e Hosking, que propôs uma diferenciação fracionada no modelo ARIMA, descrito como ARFIMA. Esse modelo, permite que a diferenciação do modelo seja fracionada, reproduzindo o fenômeno de Hurst, pois, o modelo exhibe memória longa.

Para qualquer número real $d > -1$, define-se o operador de diferença fracionária, como sendo:

$$\begin{aligned} (1 - B)^d &= \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k \\ &= 1 - dB + \frac{1}{2!}d(d-1)B^2 - \frac{1}{3!}d(d-1)(d-2)B^3 + \dots \end{aligned}$$

Definição 3. Dizemos que $\{Z_t\}$ é um processo auto-regressivo fracionário integrado de média móvel, ou ARFIMA(p, d, q) com $d \in (-\frac{1}{2}, \frac{1}{2})$, se $\{Z_t\}$ for estacionária e satisfazer a equação:

$$\phi(B)(1 - B)^d Z_t = \theta(B)a_t, \quad (2.3)$$

em que, $\{a_t\}$ é um ruído branco com média zero, $\phi(B)$ é o polinômio auto-regressivo de ordem p estimado, $\theta(B)$ é o polinômio de média móvel de ordem q estimados, d é a ordem de diferenciação fracionária.

A função é definida dessa forma, pois, os polinômios $\phi(B)$ e $\theta(B)$, explicam a estrutura de correlação de ordem baixa da série e decaem exponencialmente. O parâmetro d explica a estrutura de correlação das ordens mais altas da série e decaem hiperbolicamente conforme a distância aumenta (MORETTIN; TOLOI,

2006).

A partir da equação 2.3, Hosking (1981) demonstra que o processo ARFIMA (p,d,q) é:

(i) estacionário se $d < \frac{1}{2}$ e todas as raízes de $\phi(B) = 0$ tiverem fora do círculo unitário;

(ii) invertível se $d > -\frac{1}{2}$ e todas as raízes de $\theta(B) = 0$ estiverem fora do círculo unitário;

(iii) modelo ARIMA se $d = 0$;

O valor de d está relacionado com o expoente de Hurst, pois,

$$d = H - \frac{1}{2}$$

em que, H é o expoente de Hurst e d a diferença fracionária. Quanto maior o valor de d , maior é a intensidade de memória longa em uma série temporal (ROSSO; TAQQU, 1997).

O modelo ARFIMA, é capaz de reproduzir a estrutura de autocorrelação das séries temporais que indicam a persistência de curto ou longo prazo. Essa é uma vantagem da estrutura desse modelo, pois outros modelos de longa memória, que são capazes de reproduzir o fenômeno de Hurst, não têm qualquer flexibilidade na escolha da estrutura de autocorrelação de memória curta.

Observa-se ainda que, se Z_t , for estacionário e invertível e $f(\lambda)$ for a função de densidade espectral de Z_t então:

* $\lim_{\lambda \rightarrow 0} \lambda^{2d} f(\lambda)$ existe e é finito;

* $\lim_{k \rightarrow \infty} k^{1-2d} \rho_k$ existe e é finito.

Portanto as funções, que representam o modelo são dadas, respectivamente por,

(i) função de autocorrelação:

$$\rho_h = \frac{(-d)!(h+d-1)!}{(d-1)!(h-d)!} \forall h = \{1, 2, \dots\},$$

em que, d é a diferenciação fracionada;

(ii) função de autocorrelação parcial:

$$\phi_{hh} = \frac{d}{h-d} \forall h = \{1, 2, \dots\},$$

em que, d é a diferenciação fracionada;

(iii) densidade espectral:

$$f(\lambda) = \begin{cases} \frac{\sigma_a^2}{2\pi} (2 \sin(\frac{\lambda}{2}))^{-2d} & 0 < \lambda \leq \pi \\ \lambda^{-2d} & \lambda \rightarrow 0 \end{cases}$$

em que, σ_a^2 a variância, d é a diferenciação fracionada.

Através do comportamento das funções densidade espectral, quando $\lambda \rightarrow 0$, indica que para $d > 0$, Z_t é um processo de memória longa, que também pode ser caracterizado pelo decaimento hiperbólico da função de autocorrelação (MORETTIN; TOLOI, 2006).

2.8 Estatística R/S

A análise de reescalonamento, conhecida como Estatística R/S, foi definida por Hurst em 1951, com o propósito de testar a existência de memória longa, fornecer uma estimativa para o coeficiente de H em uma série, classificar as séries temporais entre aleatórias e não aleatórias e fornecer informações sobre o grau de auto-correlação da série analisada.

Para explicar esse método, Hurst tomou como exemplo o Lago Albert, onde registrou as medidas das descargas anuais como função do tempo. O problema é determinar um modelo de um reservatório ideal baseado nos registros das descargas provenientes do lago. Um reservatório ideal nunca deve transbordar ou esvaziar (FAVARETTO, 2004).

O método da Análise R/S é bem entendido no contexto para o qual foi criado, que é o modelamento de um reservatório (CARNIER NETO, 2006). Dessa forma, considerando que em um dado ano t , a descarga de água no reservatório é representada por $X(t)$, então a quantidade média de descargas de água que deve ser eliminada do reservatório num período de τ anos é dada por:

$$\bar{X}_\tau = \frac{1}{\tau} \sum_{t=1}^{\tau} X(t),$$

e o valor médio será igual ao volume usado por ano.

Por consequência, a soma dos desvios da média das descargas, ou seja, a quantidade de descargas armazenadas $X(i)$, abatendo a quantidade de descargas média, é representada por:

$$Z(t, \tau) = \sum_{i=1}^t X(i) - \bar{X}_\tau.$$

A diferença entre o valor máximo e o valor mínimo das descargas acumu-

ladas é a medida do reservatório que nunca estará vazio e nunca será transbordado. Portanto, a quantidade de água contida no reservatório, representado por $R(\tau)$, pode ser definida como

$$R(\tau) = \max_{1 < t < \tau} Z(t, \tau) - \min_{1 < t < \tau} Z(t, \tau),$$

com $1 \leq t \leq T$.

Assim, a Estatística R/S é uma razão dimensional entre a diferença de máximo e mínimo do reservatório (R) e o desvio padrão (S) das observações, dada por:

$$\frac{R}{S} = \frac{\max_{1 < t < \tau} Z(t, \tau) - \min_{1 < t < \tau} Z(t, \tau)}{\sqrt{\frac{1}{\tau} \sum_{t=1}^{\tau} (X(t) - \bar{X}_{\tau})^2}}.$$

A Estatística R/S permite comparar o reescalonamento de séries dos mais variados tipos de fenômenos naturais (FAVARETTO, 2004).

2.9 Estimadores para o coeficiente de Hurst

Existem vários métodos utilizados para estimar o coeficiente de Hurst. Nesta seção serão introduzidas algumas importantes funções, nas quais suas performances podem diferenciar-se substancialmente, dependendo da forma como são definidas e no contexto em que são aplicadas.

2.9.1 Método da Estatística R/S

O expoente de Hurst é estimado pelo cálculo da média do reescalonamento sobre as múltiplas regiões dos dados. Para a Estatística, o valor esperado representa o valor médio de um conjunto de dados. Usando essa notação, o valor esperado de R/S , converge sobre a função potência de expoente de Hurst (BATISTA, 2006).

Considere um conjunto com τ observações de uma série temporal. O valor

de H é obtido de:

$$E \left[\frac{R(\tau)}{S(\tau)} \right] = c\tau^H$$

onde, a esperança é calculada para uma variável t e um τ fixo.

Para obter H , é necessário fazer uma regressão linear sobre o conjunto de pontos composto por $\log \tau$ e o log do reescalonamento médio de um conjunto de regiões τ , em que a inclinação da curva é a estimativa do expoente de Hurst, ou seja:

$$\begin{aligned} E \left[\frac{R(\tau)}{S(\tau)} \right] &= c\tau^H \\ \log\left(\frac{R(\tau)}{S(\tau)}\right) &= \log(c\tau^{\hat{H}}) \\ \log\left(\frac{R(\tau)}{S(\tau)}\right) &= \log c + \log \tau^{\hat{H}} \\ \log\left(\frac{R(\tau)}{S(\tau)}\right) &= \log c + \hat{H} \log \tau. \end{aligned}$$

Esse método foi proposto por Mandelbri em 1968. Além de estimar o valor de H , traz informações importantes sobre as características geométricas da série.

2.9.2 Método da variância agregada

Uma propriedade importante para séries de memória longa está relacionada com a variância da média amostral, pois converge para zero lentamente e é definida como sendo:

$$Var(\bar{Z}) \sim cn^{2H-2}, \quad (2.4)$$

em que n é o tamanho da amostra suficientemente grande, $c > 0$ e \bar{Z} é a média amostral.

Desse modo é possível estimar o valor de H , pelo método proposto por Beran.

Considere $\tau \in \mathbb{Z}$. Construa uma série temporal com n observações sucessivas, com diferentes valores de τ . Suponha que uma série de tamanho τ contenha

m_τ subséries. Assim, as médias amostrais são $\overline{Z_1(\tau)}, \overline{Z_2(\tau)}, \dots, \overline{Z_{m_\tau}}$ e a média total da série é então, representada como:

$$\overline{Z} = \frac{1}{n} \sum_{j=1}^n Z_j \quad (2.5)$$

em que, $j = 1, \dots, m_\tau$. Para cada τ , calcule a variância amostral das médias amostrais $\overline{Z_j(\tau)}$, descrita como:

$$\widehat{Var}[Z(\tau)] = \frac{1}{m_\tau - 1} \sum_{j=1}^{m_\tau} [\overline{Z_j(\tau)} - \overline{Z}]^2.$$

Assim, pela a equação 2.4, pode-se reescrever a equação $\widehat{Var}[Z(\tau)]$ da forma:

$$\begin{aligned} \widehat{Var}(\overline{Z(\tau)}) &= c\tau^{2\hat{H}-2} \\ \log \widehat{Var}(\overline{Z_\tau}) &= \log(c\tau^{2\hat{H}-2}) \\ \log \widehat{Var}(\overline{Z(\tau)}) &= \log c + \log \tau^{2\hat{H}-2} \\ \log \widehat{Var}(\overline{Z(\tau)}) &= \log c + (2\hat{H} - 2) \log \tau. \end{aligned}$$

Fazendo um ajuste pelo método de mínimos quadrados, pode-se obter o gráfico $\log \widehat{Var}(\overline{Z(\tau)}) \times \log \tau$, o qual representa pontos espalhados ao longo de uma linha reta com inclinação $\hat{\beta} = (2\hat{H} - 2)$, e permite uma avaliação gráfica de H . Assim, um valor estimado para H , é dado pela forma:

$$\begin{aligned} \hat{\beta} &= 2\hat{H} - 2 \\ 2\hat{H} &= \hat{\beta} + 2 \\ \hat{H} &= \frac{\hat{\beta} + 2}{2}. \end{aligned}$$

2.9.3 Método da diferenciação da variância agregada

Teverosky e Taquq propuseram um método para detectar a presença de memória longa em séries temporais, utilizando o método da variância agregada. Tomando o logaritmo da diferença de primeira ordem, da equação 2.5, o estimador da variância, é dado como:

$$\log \Delta \widehat{Var}(\overline{Z(\tau)}) \simeq \log \frac{d}{dm} \widehat{Var}(\overline{Z(\tau)}) + \log \Delta \tau,$$

em que

$$\begin{aligned} \widehat{Var}(\overline{Z(\tau)}) &= c\tau^{2\widehat{H}-2} \\ \frac{d}{dm} \widehat{Var}(\overline{Z(\tau)}) &= (2\widehat{H} - 2)c\tau^{2\widehat{H}-3}. \end{aligned}$$

Por outro lado, uma vez que os valores de τ são espaçados logaritmicamente, obtém-se um valor do $\log(\tau) = \text{constante}$ ou seja, $\log \Delta \tau = \log \tau + c_1$.

Assim,

$$\begin{aligned} \log \Delta \widehat{Var}(\overline{Z(\tau)}) &\simeq \log \frac{d}{dm} \widehat{Var}(\overline{Z(\tau)}) + \log \Delta \tau \\ \log \Delta \widehat{Var}(\overline{Z(\tau)}) &\simeq \log(2\widehat{H} - 2)c\tau^{2\widehat{H}-3} + \log \tau + c_1 \\ \log \Delta \widehat{Var}(\overline{Z(\tau)}) &\simeq \log(2\widehat{H} - 2)c + \log \tau^{2\widehat{H}-3} + \log \tau + c_1 \\ \log \Delta \widehat{Var}(\overline{Z(\tau)}) &\simeq \log(2\widehat{H} - 2)c + (2\widehat{H} - 3) \log \tau + \log \tau + c_1 \\ \log \Delta \widehat{Var}(\overline{Z(\tau)}) &\simeq (2\widehat{H} - 3 + 1) \log \tau + c_2 \\ \log \Delta \widehat{Var}(\overline{Z(\tau)}) &\simeq (2\widehat{H} - 2) \log \tau + c_2. \end{aligned}$$

Assim, quando foi feita a diferenciação, houve uma dispersão dos pontos ocasionando uma diminuição da variância. Fazendo um ajuste pelo método de mínimos quadrados, que minimiza a soma dos resíduos quadrados observa-se um gráfico de $\log \Delta \widehat{Var}(\overline{Z(\tau)}) \times \log \tau$, que representa pontos espalhados ao longo de uma linha reta com inclinação $\widehat{\beta} = (2\widehat{H} - 2)$, que permite uma avaliação gráfica

de H .

2.10 Nível de mar

Há uma grande complexidade para o nível do mar embora pareça uma questão simples de ser solucionada, não é todo constante. Ele sofre variações por diversos fatores cíclicos e ondulações das marés, pressão atmosférica, temperatura da água, entre outros, podendo atuar em períodos de segundos a anos.

A medição do nível médio do mar foi tradicionalmente feita com base nas leituras dos marégrafos, instrumentos que permitem medir a variação do nível das águas num determinado local. Eliminando dos dados recolhidos as flutuações por ondas, a fatores meteorológicos, a marés e a outros fatores astronômicos, obtém-se uma leitura do nível médio do mar durante determinado período.

As medições assim obtidas, incorporam os efeitos eustáticos (resultante de qualquer fator que faça variar o volume de água) e isostáticos (que representa o equilíbrio no deslocamento do volume de água).

Um dos principais riscos associados ao aquecimento global da Terra, devido às alterações climáticas, é a elevação do nível dos oceanos. Alguns países desenvolvidos vêm sendo afetados pelo aumento do nível do mar.

Em 1988 foi criado o IPCC (Painel Intergovernamental sobre mudanças climáticas), para estudar o problema das mudanças climáticas. Órgão criado pelo Programa das Nações Unidas do Meio Ambiente (PNUMA) e pela Organização Meteorológica Mundial (OMM).

O IPCC, relata em uma nota, que durante os anos de 1961 a 2003, houve um aumento médio do nível do mar, de $1,8 \pm 0,5$ mm / ano, calculado através dos medidores de dados.

3 METODOLOGIA

As seguintes seções apresentam os dados e as técnicas de pesquisa para o desenvolvimento do presente trabalho.

3.1 Dados do nível mar

Neste trabalho, foram considerados dados de nível do mar de 10 cidades em diferentes países. É importante destacar que, esses dados são resultados de décadas de observações relacionadas às características físicas das águas oceânicas e inúmeros outros programas de investigação oceanográfica.

Foram considerados dados mensais a partir de janeiro de 1986 a dezembro de 2011. Os dados foram padronizados, ou seja, consiste em subtrair o valor de uma variável a sua média e dividir o resultado pelo desvio padrão e os dados ausentes foram estimados por média móvel.

Tabela 1 Local em que a série do nível do mar foi analisado.

Cidade	Latitude	Longitude	Local
Abashiri	4401N	14117E	Japão
Brisbane	2722N	15310E	Austrália
Ilha Fiscal	2254S	04310W	Brasil
Kodiak Island	5744N	15231W	Alasca
Langkawi	0626N	09945E	Malásia
Quarry Bay	2218N	11413E	Hong-Kong
Tanjog Pagar	0116N	10351E	Singapura
Ko Taphao Noi	0750N	09826E	Tailândia
Vardo	7020N	03106E	Noruega
Wellington	4117S	14447E	Nova Zelândia

As cidades e suas respectivas localizações são mostradas na Tabela 1. As cidades Abashiri (Japão), Kodiak Island (Alasca) e Quarry Bay (Hong-kong) são banhadas pelo Oceano Pacífico Norte; Brisbane (Austrália) e Wellington (Nova Zelândia) são banhadas pelo Oceano Pacífico Sul. O Oceano Atlântico Sul banha

a cidade de Ilha Fiscal (Brasil). Langkawi (Malásia) e Ko Taphao Noi (Tailândia) são banhada pelo Oceano Índico. A cidade de Tanjog Pagar (Singapura) é banhada por dois Oceanos: Índico e Pacífico Norte. E por fim, a cidade de Vardo (Noruega) é banhada pelo Oceano Ártico.

Os dados foram obtidos do GLOSS / CLIVAR (anteriormente conhecido como o WOCE), um programa internacional que visa o estabelecimento de redes globais e regionais do nível do mar de alta qualidade para aplicação no clima, oceanografia e pesquisa do nível do mar costeiro. Foram utilizados dados do nível do mar considerados como “rápido”, que são distribuídos por horas, dias e meses e estão disponibilizados, no endereço eletrônico do órgão GLOSS (Sistema Global de Observação do Nível do Mar).

3.2 Análise dos dados

Na análise dos dados será verificada a existência de memória longa para cada país. Para observar a memória longa e a componente tendência e sazonalidade, será utilizado o gráfico da série. Para verificar se as séries não são estacionárias, será aplicada a diferença fracionada, d , valor obtido com o pacote “fracdiff”. Os correlogramas das séries diferenciadas serão elaborados para observar o comportamento de cada série. Com o valor estimado da diferença fracionada, será analisada a presença de memória longa e sua persistência.

Para cada série serão estimados os valores de H , utilizando o método R/S, método da variância agregada e o método da diferenciação da variância. Os valores de H estimados para cada série serão utilizados na análise de agrupamentos. Esses agrupamentos serão comparados com os resultados obtidos em Safadi (2014), que utilizou análise de agrupamento usando componentes independentes.

As análises foram realizadas utilizando o software RStudio (R Core Team, 2015). Para obter esses resultados, foram utilizados os pacotes “fArma”, “fracd-

iff”, “stats”, “tseries”, “arfima”, “pracma” e “randtests”.

As cidades serão identificadas pelos respectivos países.

4 RESULTADOS E DISCUSSÃO

Nas Figuras 2 a 11 são apresentados os gráficos das 10 séries de nível do mar durante o período de janeiro de 1986 a dezembro de 2011. Observa-se em (a) que as séries não são estacionárias.

Nas Figuras 8(a) e 10(a), Singapura e Noruega apresentam um comportamento semelhante, com um caimento do crescimento do nível do mar entre os anos 1990 e 2000 e logo após, no ano 2000, um crescimento do nível do mar. Na Figura 7(a), Hong-Kong apresenta inicialmente um comportamento distinto dos anteriores, pois denota um equilíbrio até o ano de 2000, e, a partir daí, um crescimento do nível do mar. Malásia e Tailândia, representadas nas Figura 6(a) e 9(a) respectivamente, apresentam um comportamento decrescente até o ano de 1992, após 1992 um crescimento até o ano de 1998 e, a partir desse ano, houve um equilíbrio. Japão, Austrália, Brasil, Alasca e Nova Zelândia apresentam comportamento semelhante.

Para todas as séries, foram obtidos os valores do parâmetro d , Tabela 2. Observa-se que, com exceção do Brasil, Austrália e Alasca, que apresentam valores de $d < 0,5$, os demais países apresentam valores superiores a $0,5$, indicando séries com memória longa.

Utilizando os valores de d , as séries foram diferenciadas usando o método da diferenciação visto na seção 2.3.1, com o comando “diffseries”. Assim, a estacionariedade de todas as séries é observada nas Figuras 2(b) a 11(b).

Tabela 2 Valores estimados de d para cada país.

Países	Diferença fracionada
Japão	0,6699
Austrália	0,4287
Brasil	0,3413
Alasca	0,2407
Malásia	0,7022
Hong-Kong	0,7463
Singapura	0,7659
Tailândia	0,6280
Noruega	1,0917
Nova Zelândia	0,5981

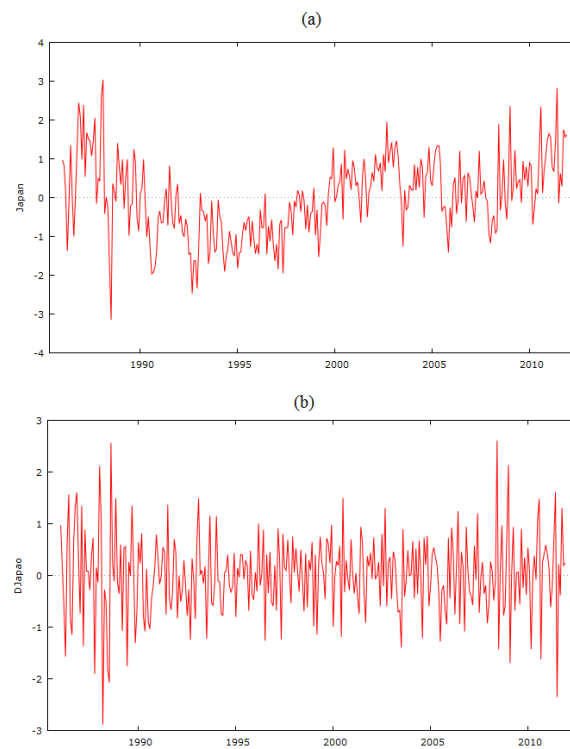


Figura 2 Gráfico da série do Japão (a), gráfico da série estacionária (b).

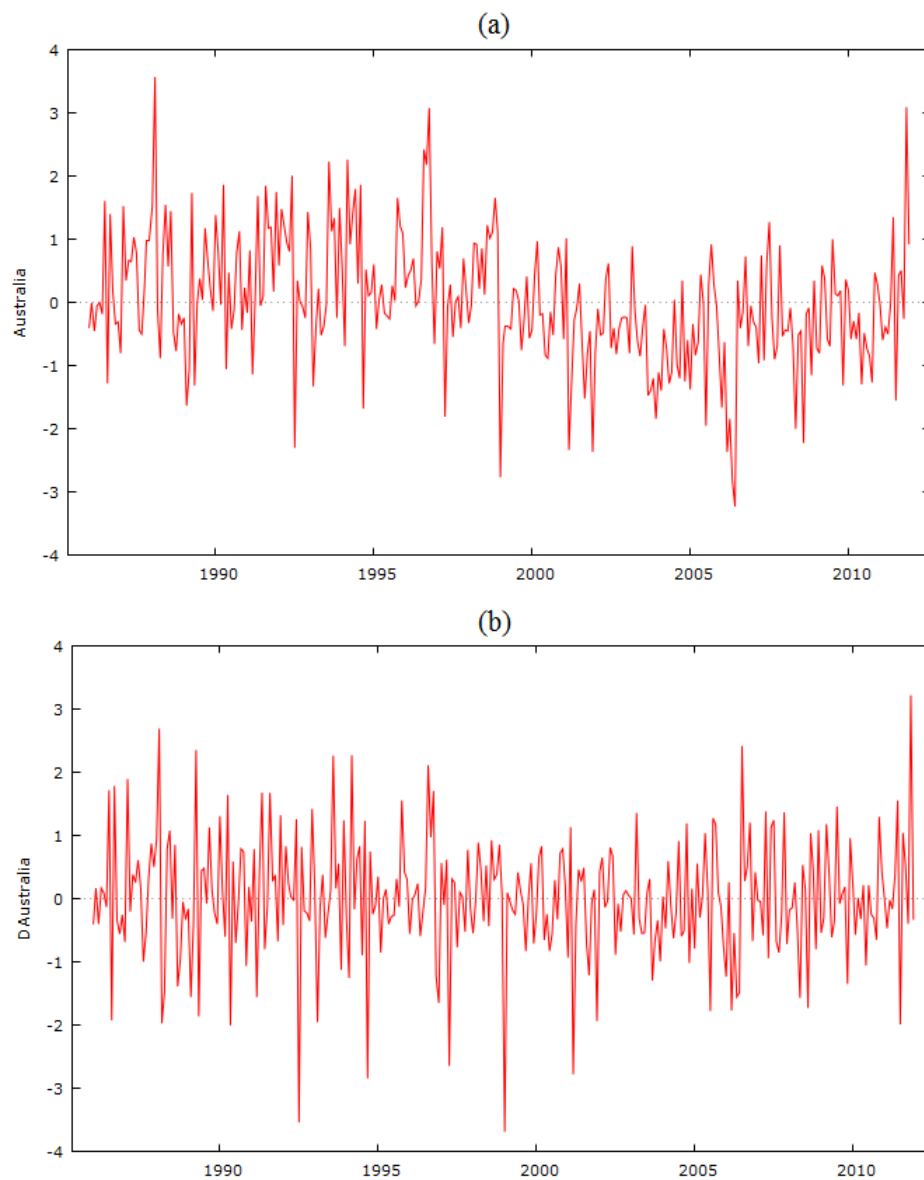


Figura 3 Gráfico da série Austrália (a), gráfico da série estacionária (b).

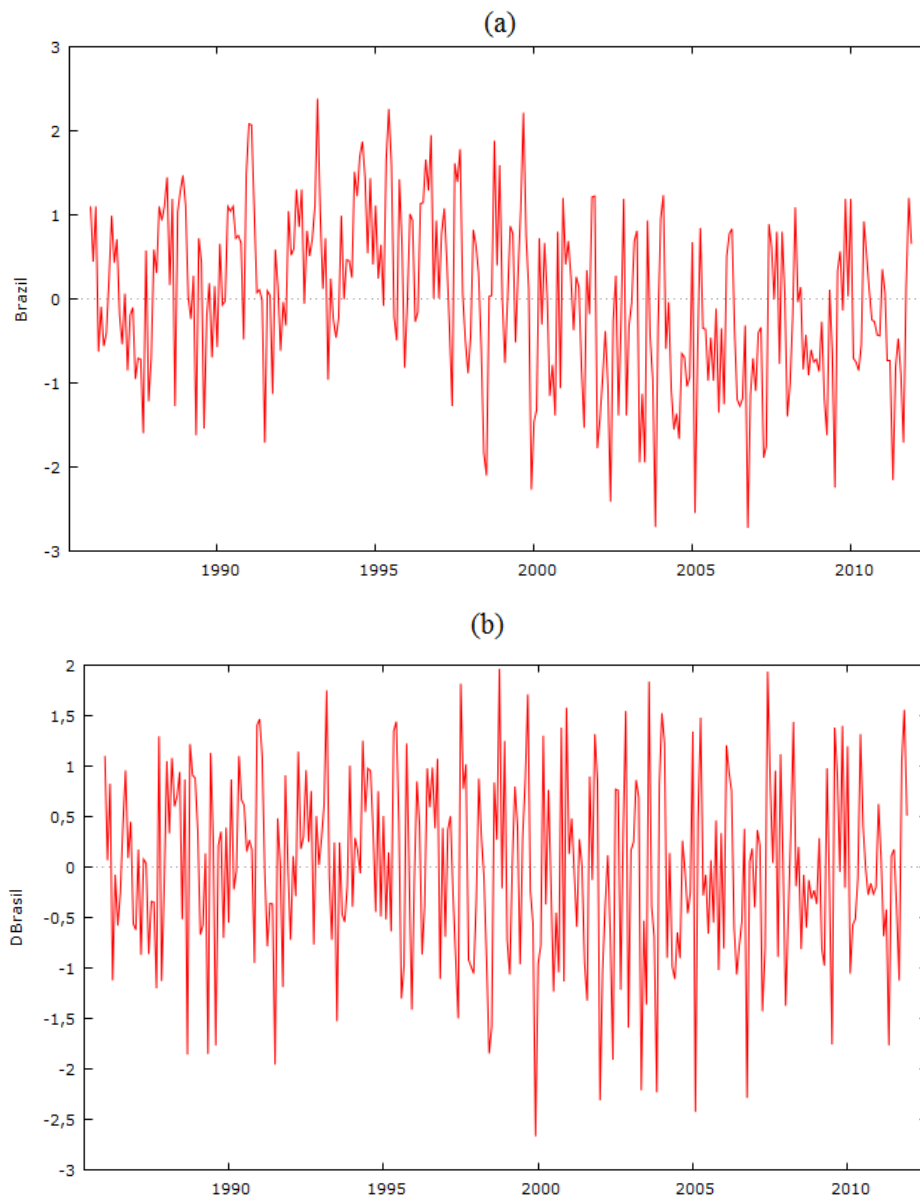


Figura 4 Gráfico da série Brazil (a), gráfico da série estacionária (b).

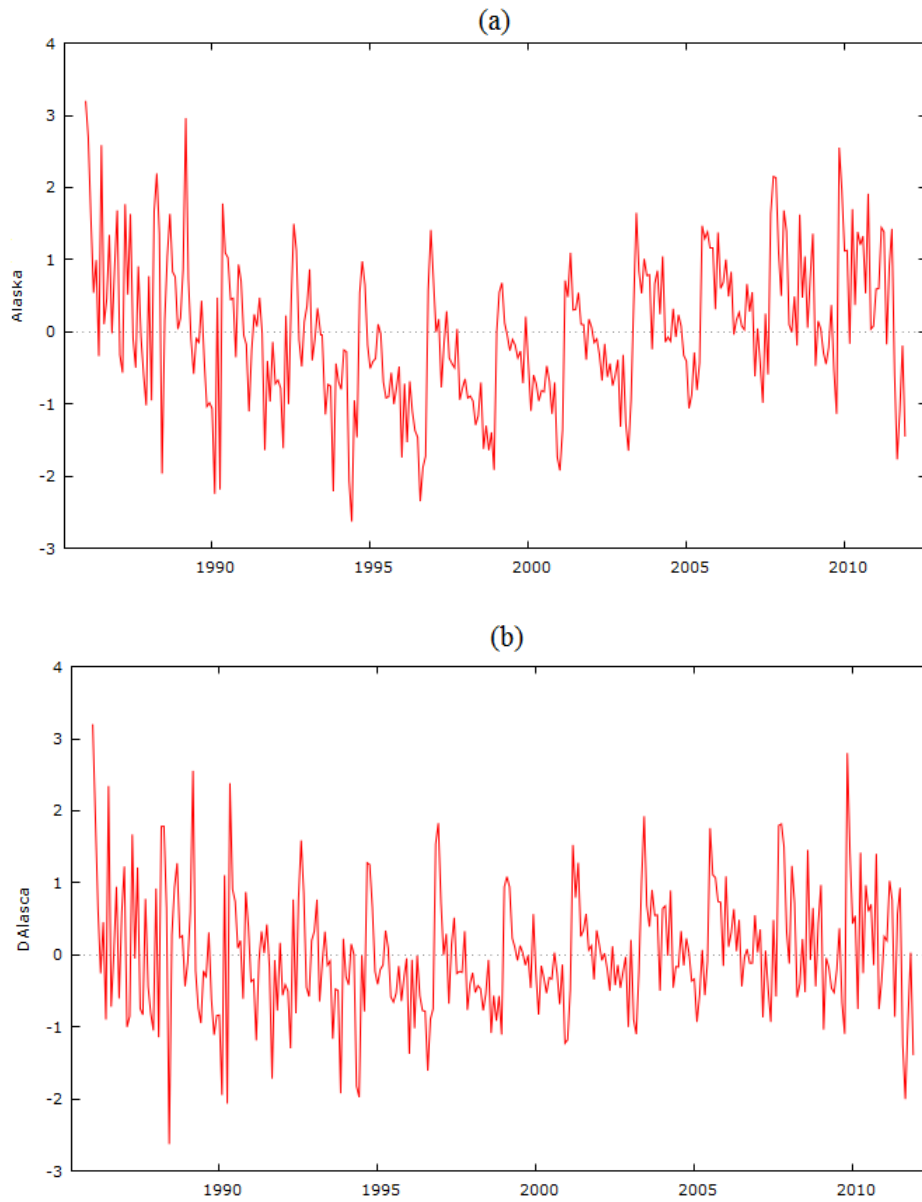


Figura 5 Gráfico da série Alasca (a), gráfico da série estacionária (b).

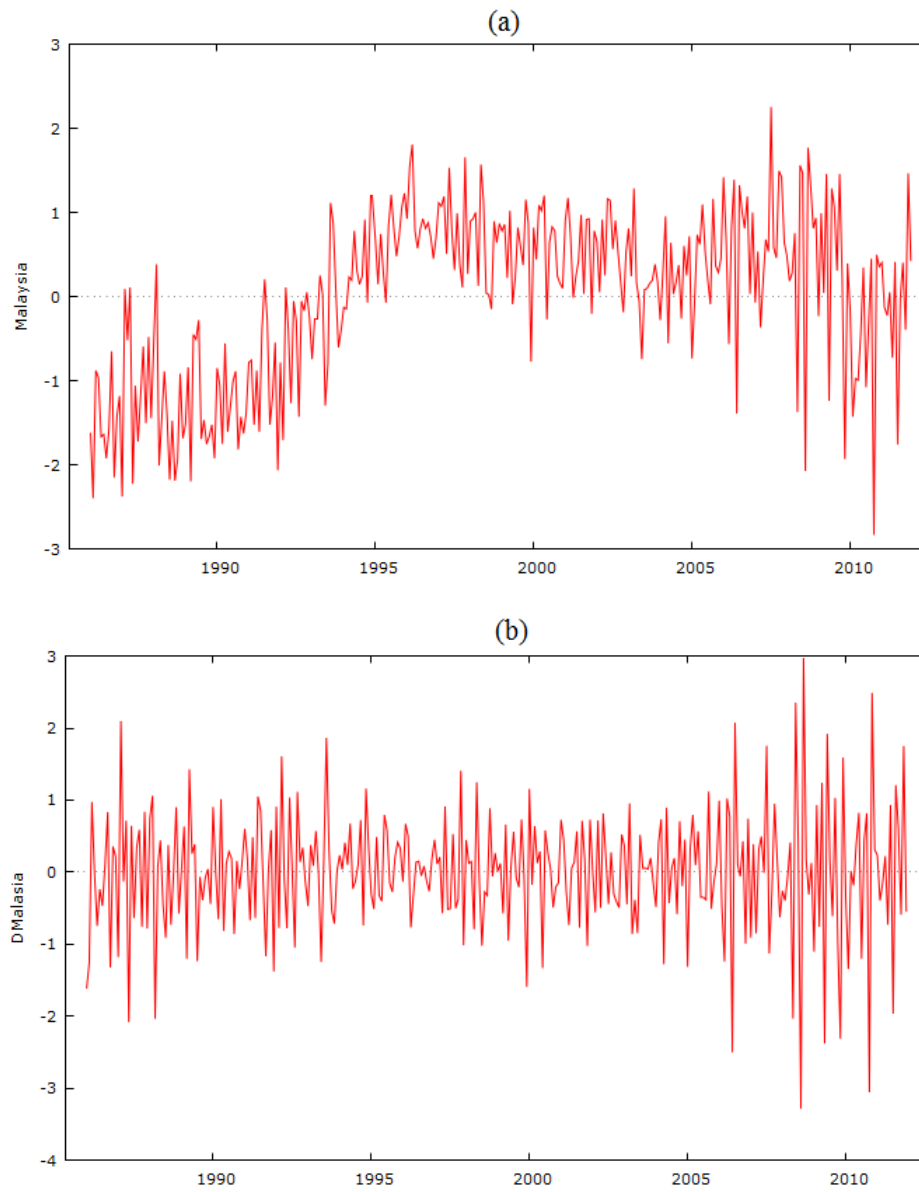


Figura 6 Gráfico da série Malásia (a), gráfico da série estacionária (b).

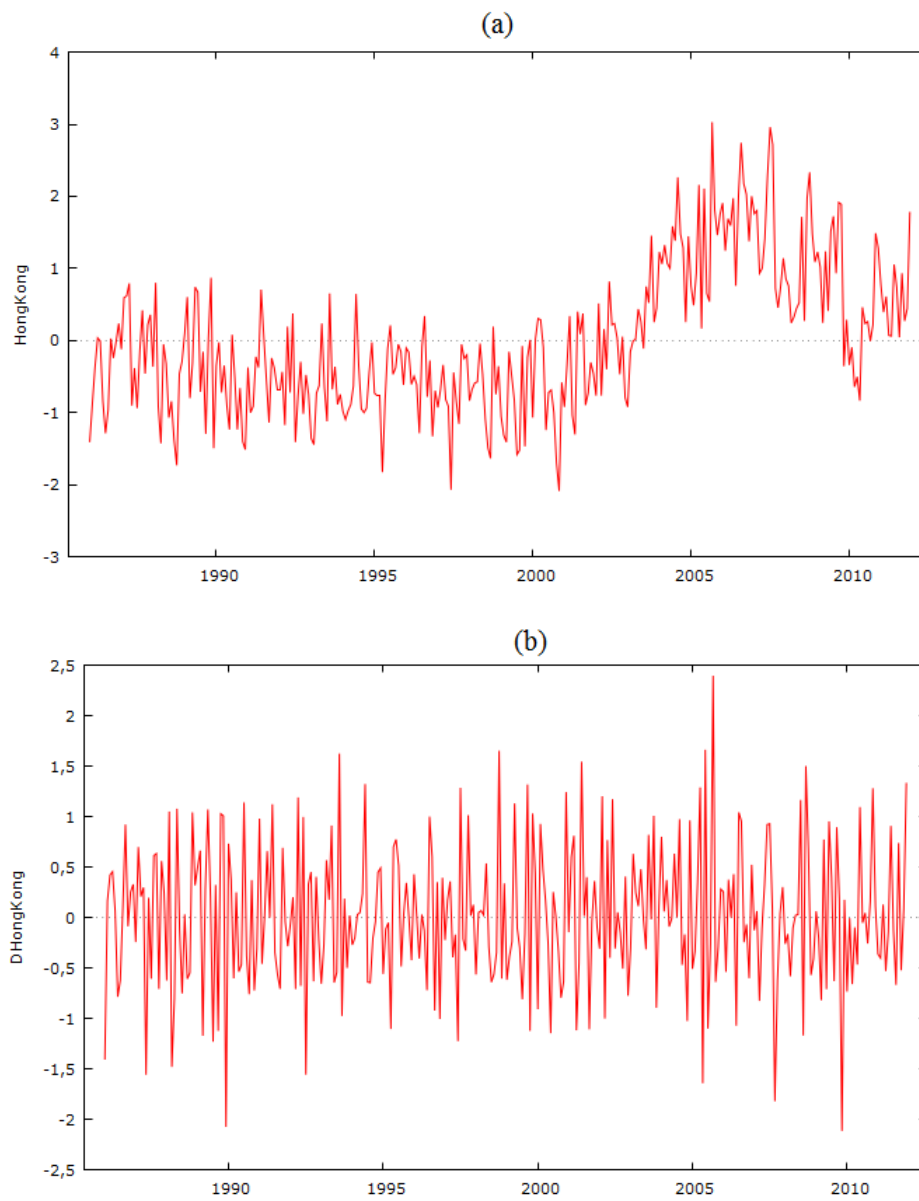


Figura 7 Gráfico da série Hong Kong (a), gráfico da série estacionária (b).

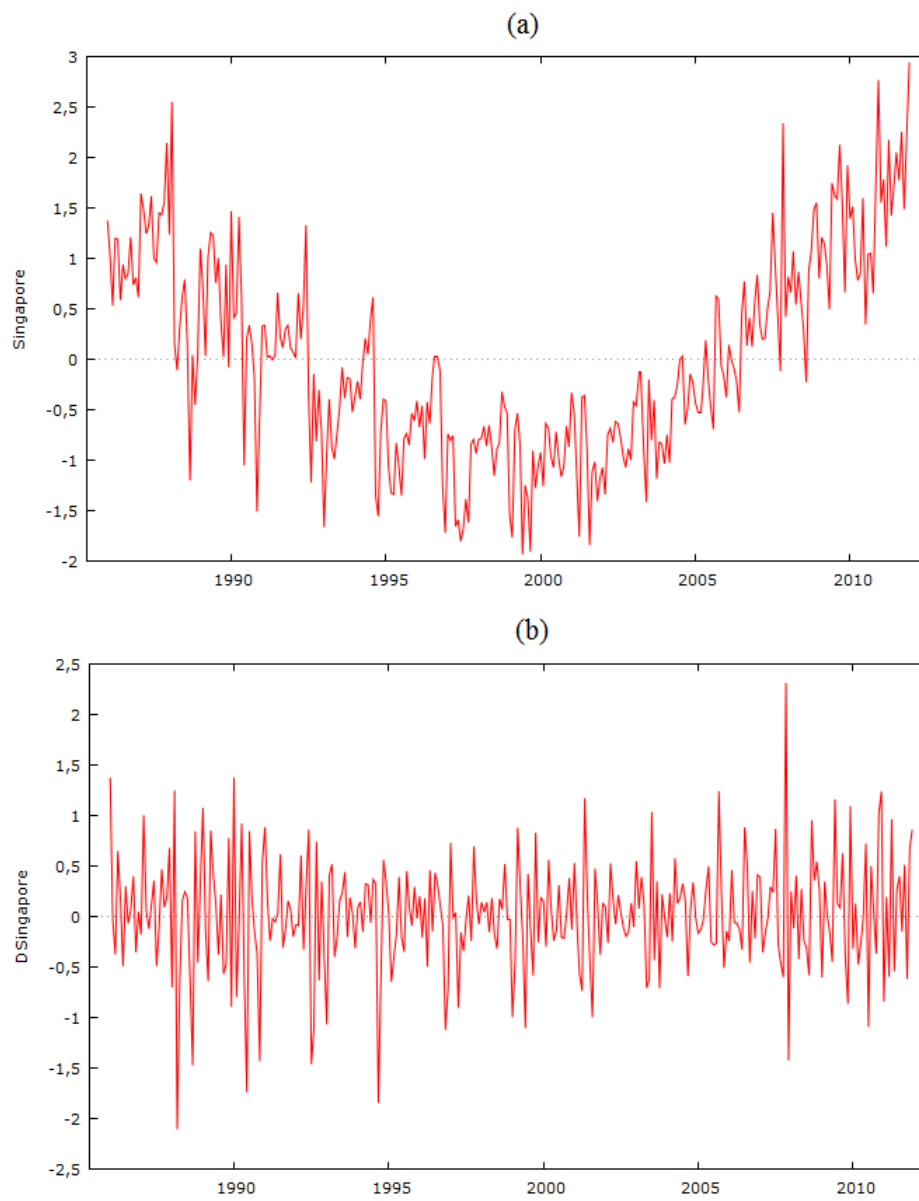


Figura 8 Gráfico da série Singapura (a), gráfico da série estacionária (b).

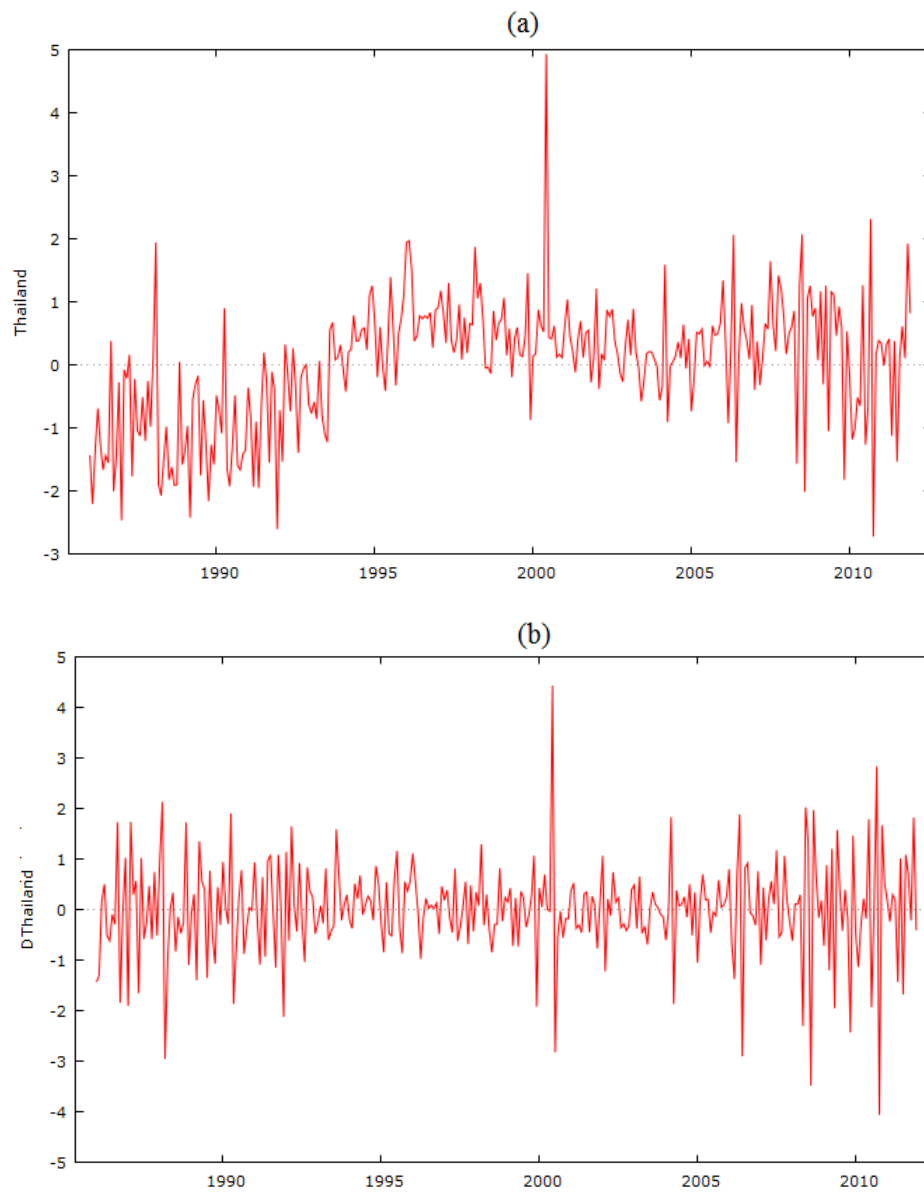


Figura 9 Gráfico da série Tailândia (a), gráfico da série estacionária (b).

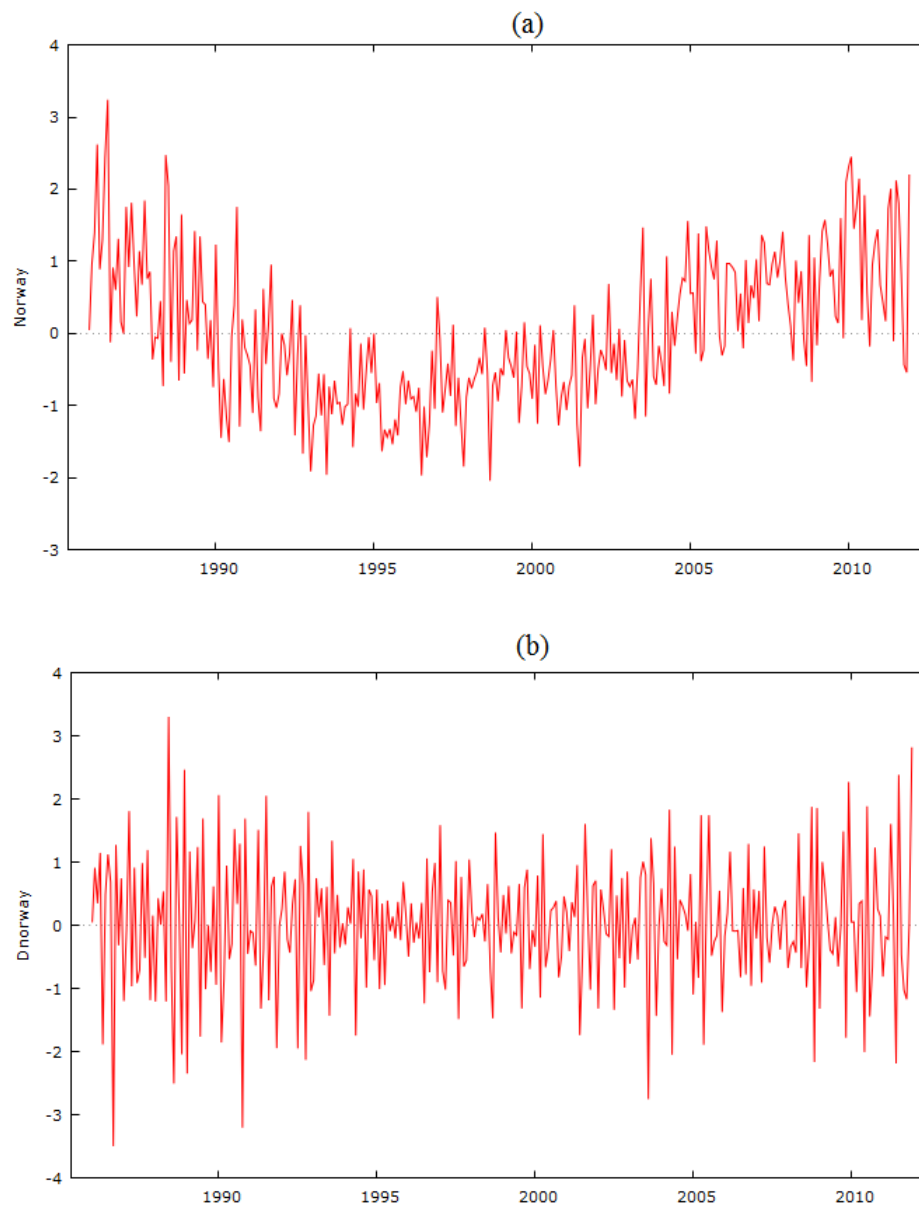


Figura 10 Gráfico da série Noruega (a), gráfico da série estacionária (b).

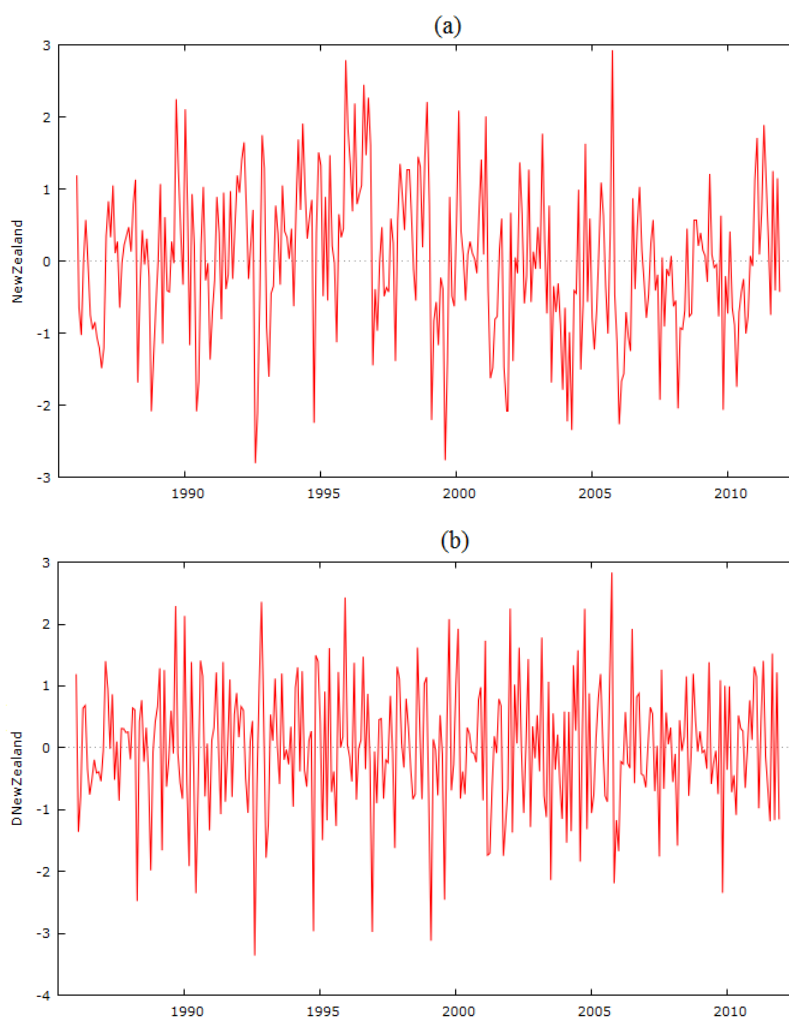


Figura 11 Gráfico da série Nova Zelândia (a), gráfico da série estacionária (b).

O gráfico das funções de autocorrelação e autocorrelação parcial para as séries diferenciadas são apresentados nas Figuras 11 a 18 (a) e (b), respectivamente. Na Figura (a) é possível observar que as séries são estacionárias e apresentam um componente sazonal de aproximadamente 2 anos, indicada em correlações significativas múltiplas de 24.

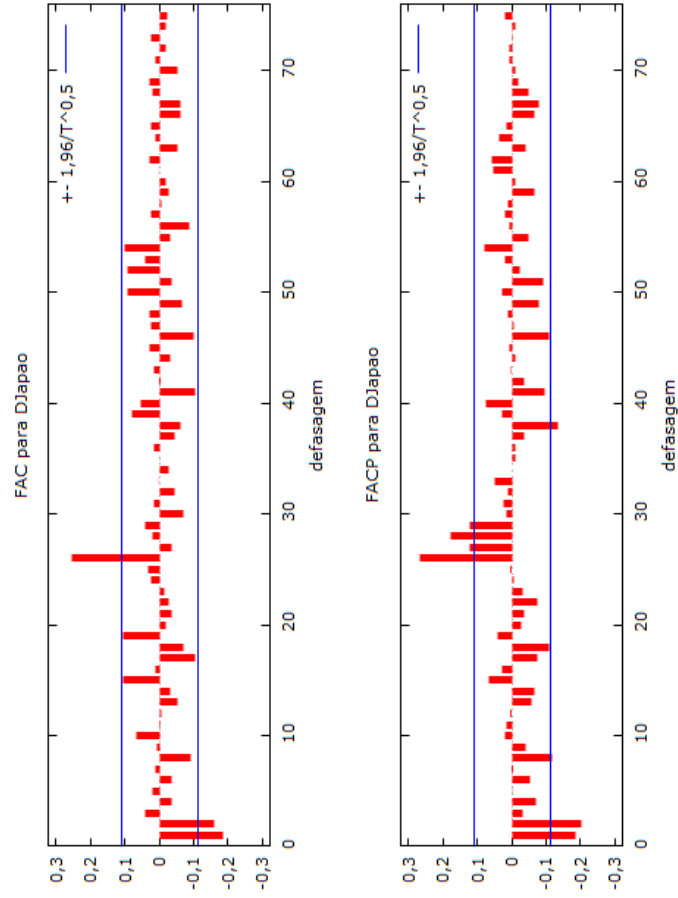


Figura 12 Função de autocorrelação e função de autocorrelação parcial.

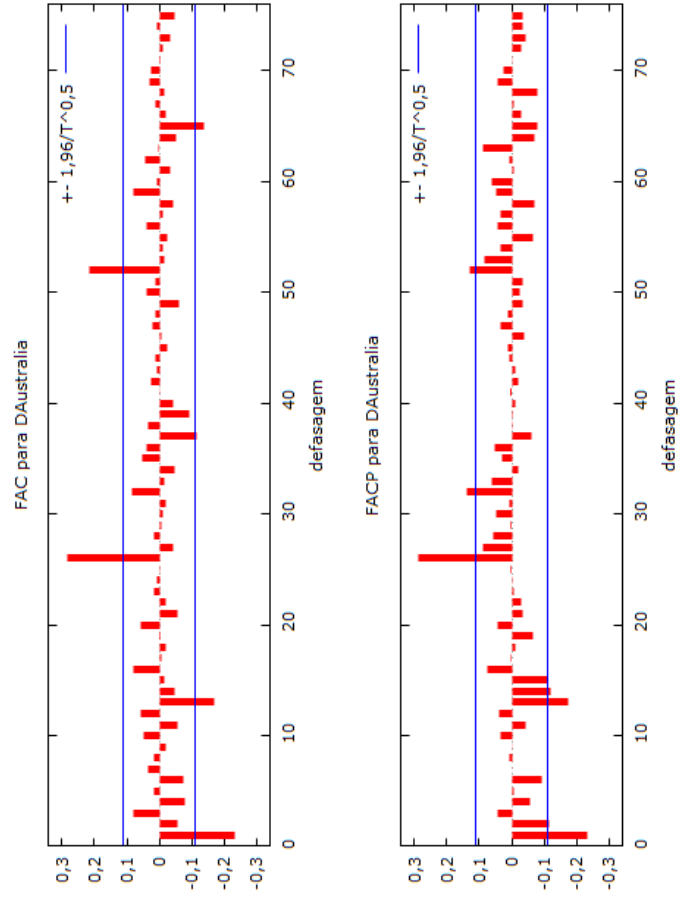


Figura 13 Função de autocorrelação e função de autocorrelação parcial.

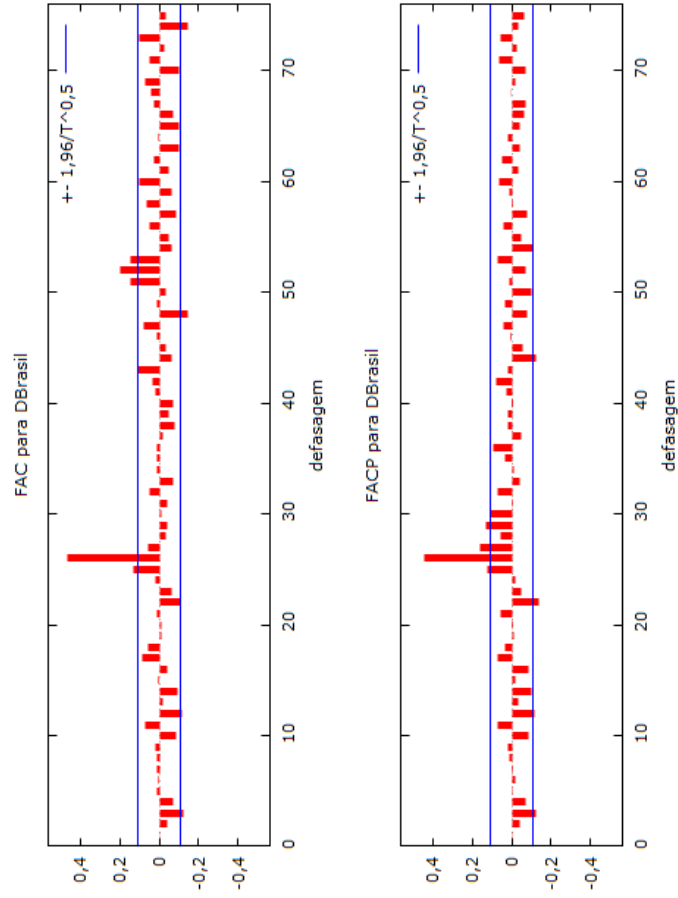


Figura 14 Função de autocorrelação e função de autocorrelação parcial.

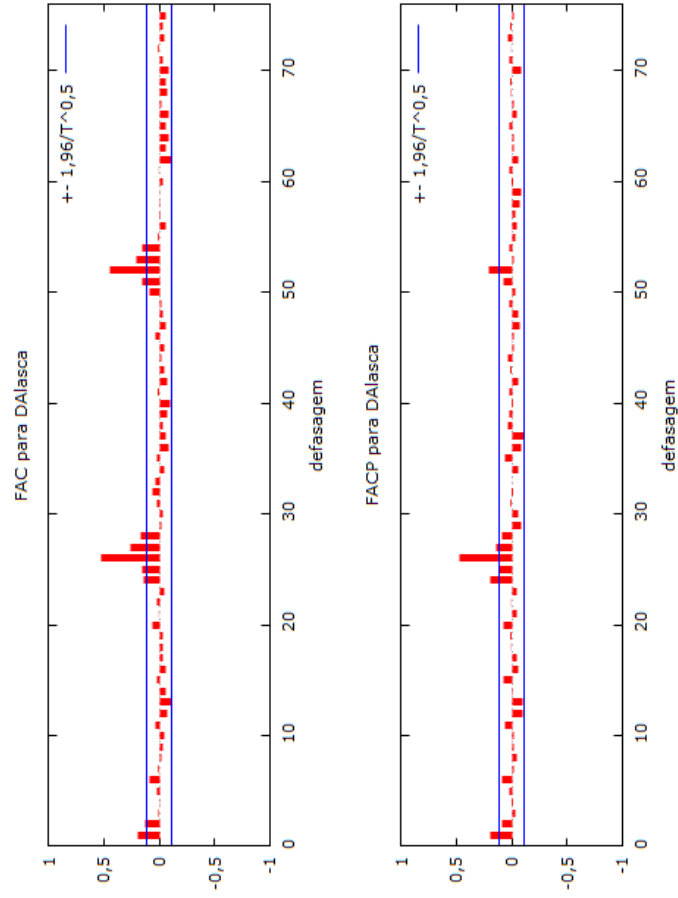


Figura 15 Função de autocorrelação e função de autocorrelação parcial.

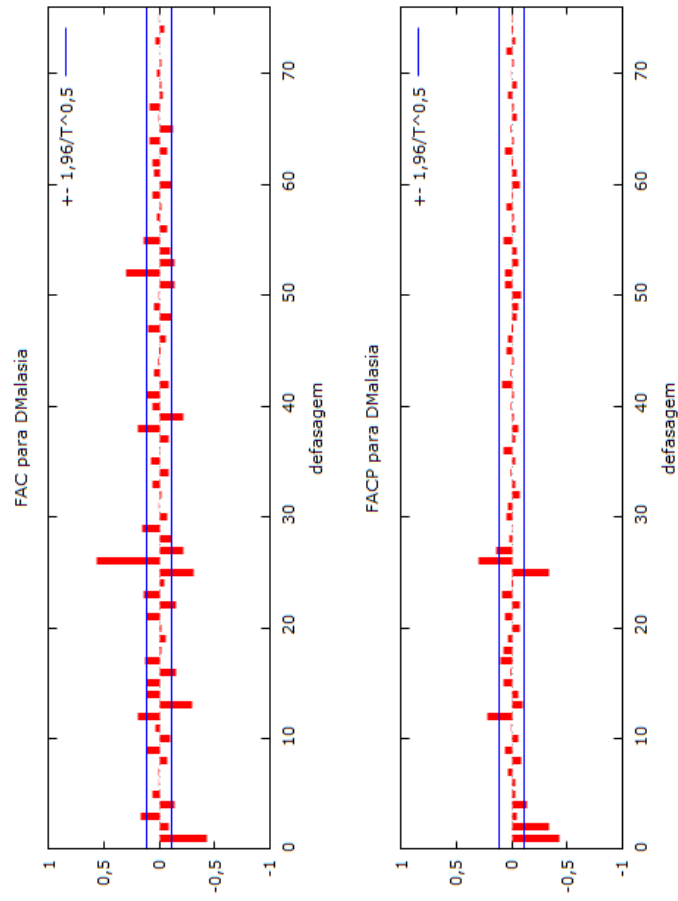


Figura 16 Função de autocorrelação e função de autocorrelação parcial.

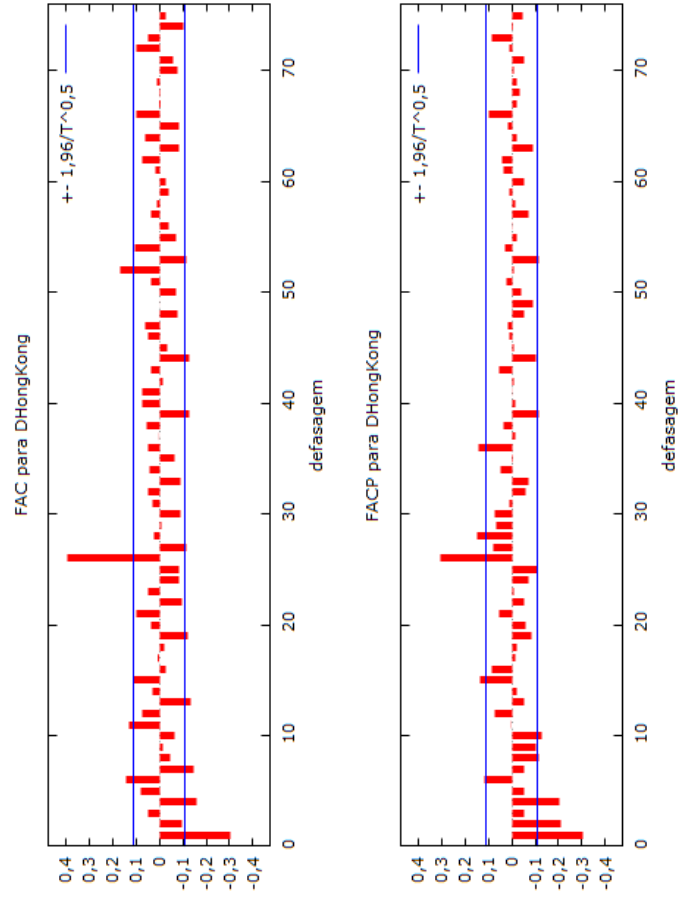


Figura 17 Função de autocorrelação e função de autocorrelação parcial.

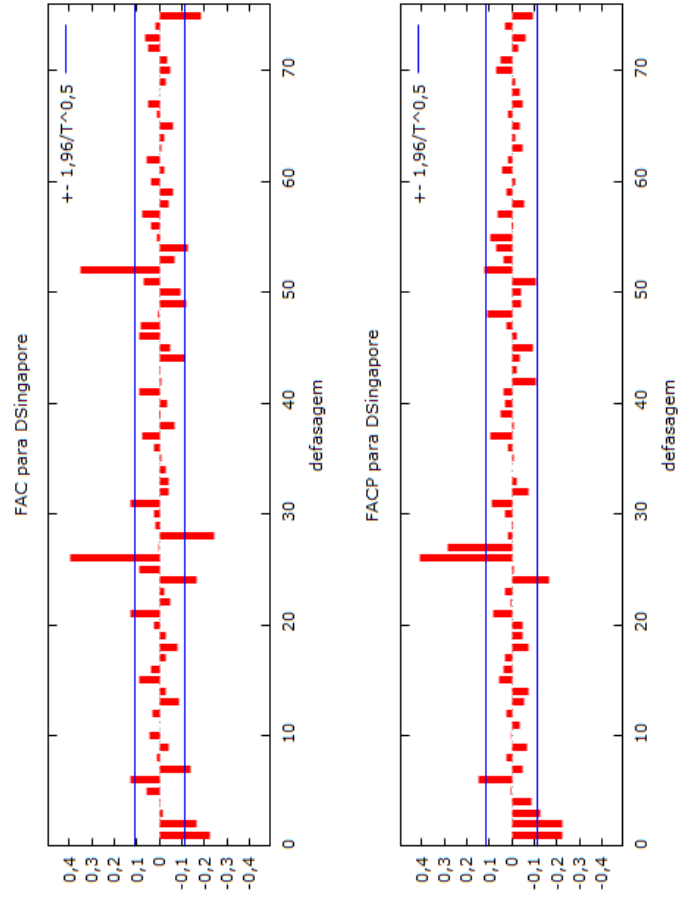


Figura 18 Função de autocorrelação e função de autocorrelação parcial.

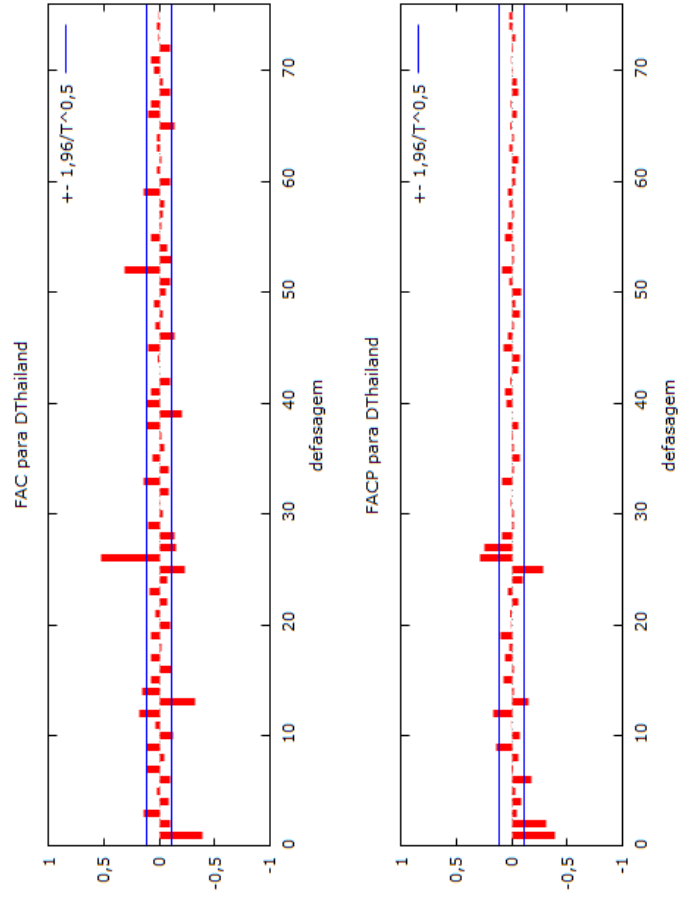


Figura 19 Função de autocorrelação e função de autocorrelação parcial.

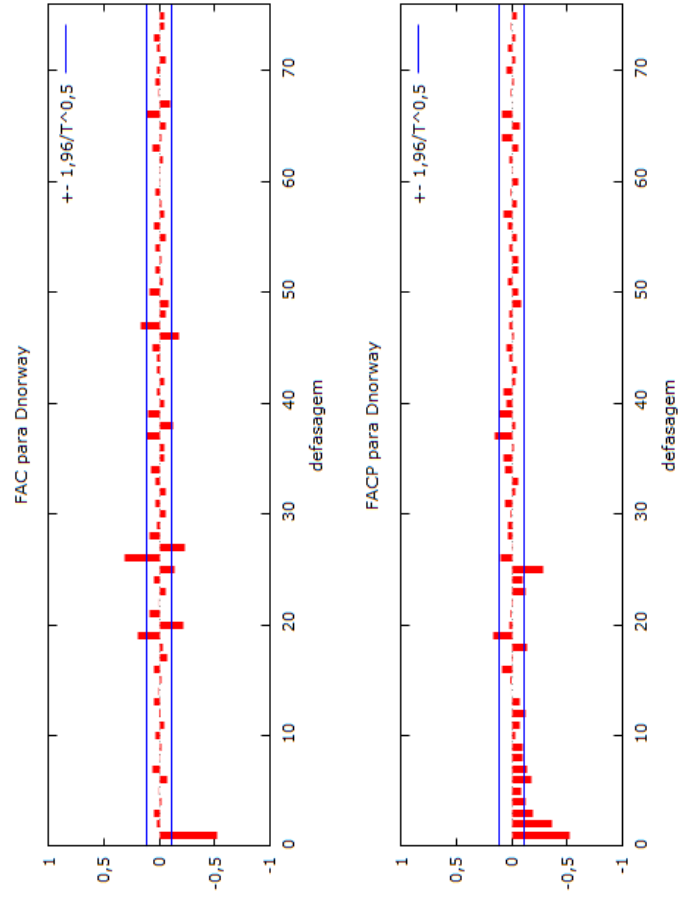


Figura 20 Função de autocorrelação e função de autocorrelação parcial.

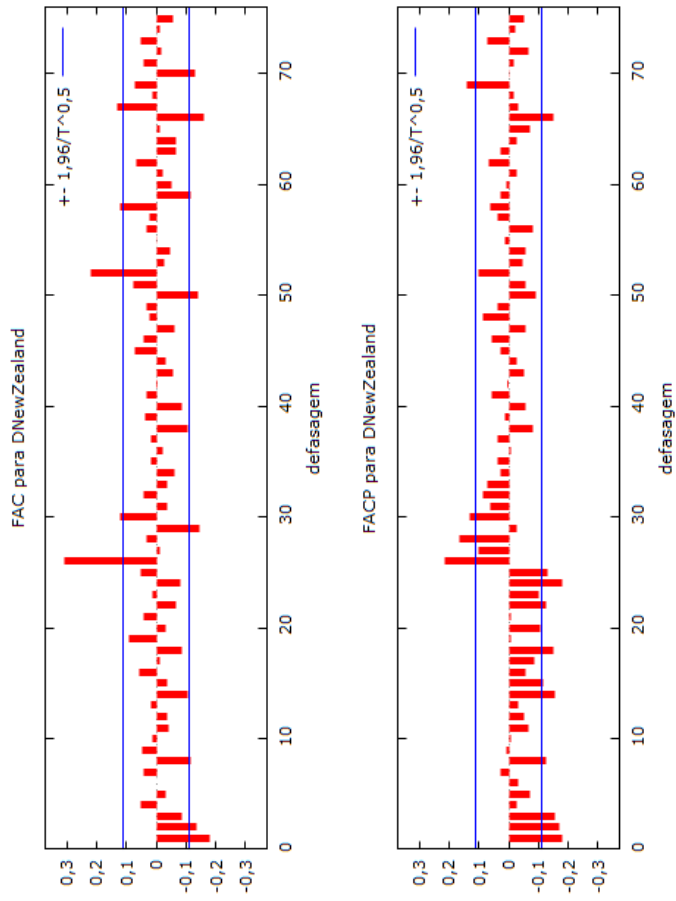


Figura 21 Função de autocorrelação e função de autocorrelação parcial.

O coeficiente de Hurst, utilizando os métodos R/S, variância agregada e a diferenciação da variância agregada para cada série foi obtido e os valores são apresentados na Tabela 3.

Tabela 3 Valores estimados do coeficiente de Hurst para cada país.

Países	Estatística <i>R/S</i>	Variância Agregada	Diferenciação da Variância Agregada
Japão	0,5369	0,3268	0,4697
Austrália	0,5668	0,5456	0,5242
Brasil	0,4687	0,6415	0,6123
Alasca	0,5705	0,6767	0,5630
Malásia	0,2934	0,5234	0,6226
Hong-Kong	0,2289	0,4981	0,5853
Singapura	0,5901	0,5340	0,5208
Tailândia	0,3403	0,5538	0,7687
Noruega	0,1377	-0,064	0,2411
Nova Zelândia	0,3650	0,0237	0,4195

Interessante observar que, os países que apresentaram um valor de $H > 0,5$ para os três métodos foram Austrália, Alasca e Singapura. Os países Brasil, Malásia e Tailândia, apresentaram apenas um valor de $H < 0,5$ mediante aos três métodos, e os países Japão e Hong-Kong, apresentaram apenas um valor para $H > 0,5$ entre os três métodos. E apenas Noruega e Nova Zelândia apresentaram valores de $H < 0,5$, nos três métodos.

Pode-se destacar ainda, que os métodos da Variância Agregada e o da Diferenciação da Variância Agregada apresentam valores semelhantes. Apenas os países Japão, Noruega e Nova Zelândia apresentam valores de $H < 0,5$ para esses métodos.

Considerando os três valores de H obtidos para cada país, foi feito o agrupamento utilizando o método hierárquico Ward, apresentado na Seção 2.4. O dendrograma é apresentado na Figura 22. Duas possibilidades de agrupamentos são

apresentadas:

- (i) Um com 3 grupos (Figura 22 (a)): grupo 1 com Noruega e Nova Zelândia, grupo 2 com Tailândia, Malásia e Hong-Kong e grupo 3 com Brasil, Alasca, Japão, Austrália e Singapura;
- (ii) Um com 4 grupos (Figura 22 (b)): grupo 1 com Noruega e Nova Zelândia, grupo 2 com Tailândia, Malásia e Hong-Kong, grupo 3 com Brasil, Alasca e grupo 4 com Japão, Austrália e Singapura.

Sáfadi (2014) utiliza as mesmas séries de nível do mar na análise de agrupamentos. Em sua análise, utilizou o método de componente independente com 2, 3, 4 e 5 componentes. Nos seus resultados é possível observar que grupos distintos são formados para cada número de componente independente. Apenas Brasil e Alasca formam um grupo nas 4 situações. Fato observado também ao utilizamos o coeficiente de Hurst.

Japão poderia ser considerado como um grupo, fato obtido em Sáfadi (2014), considerando 4 e 5 componentes independentes.

Tailândia e Malásia também têm similaridade em todas as situações analisadas em Sáfadi (2014), fato também obtido com a metodologia aqui empregada.

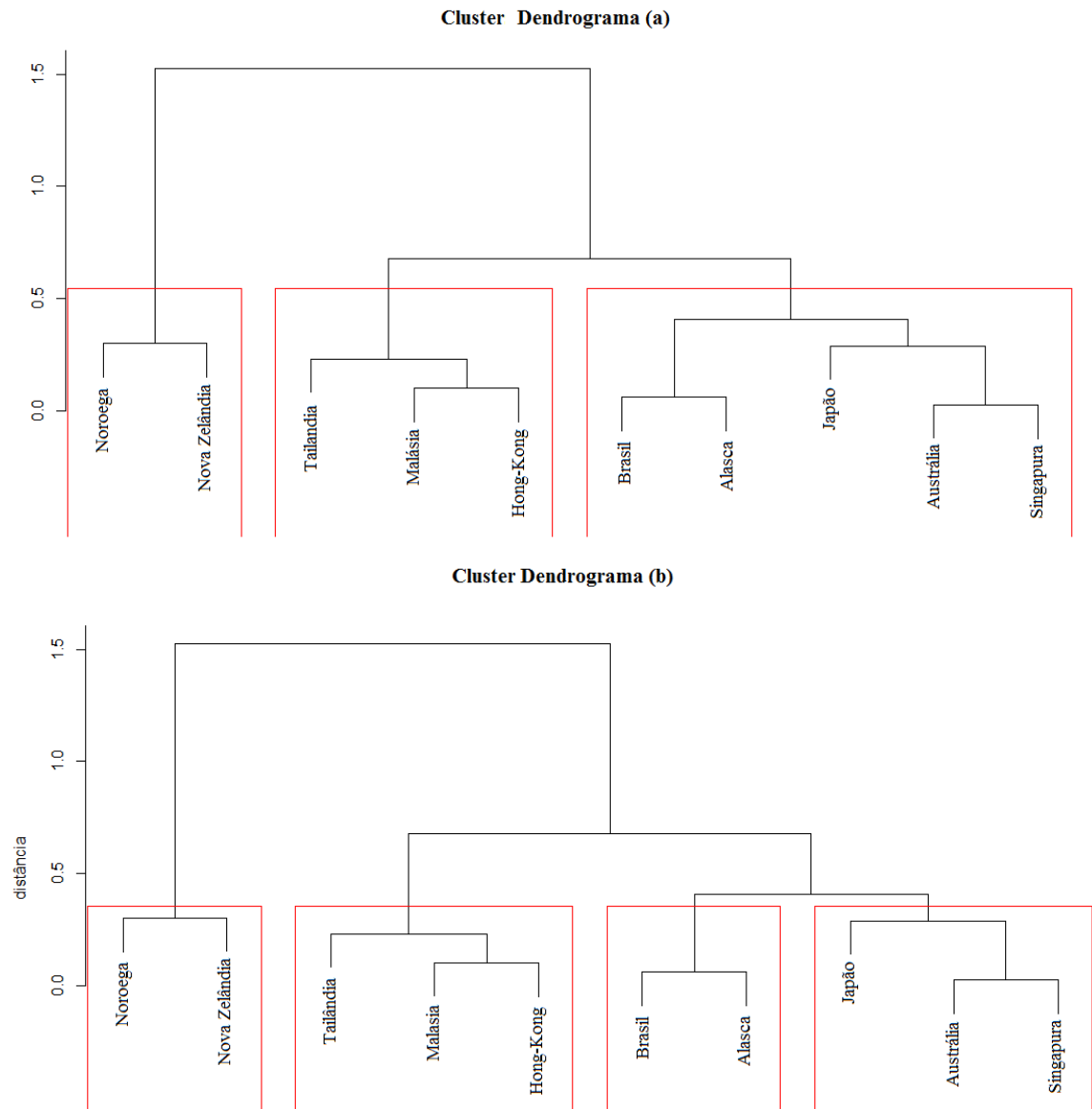


Figura 22 Análise de agrupamento: no gráfico (a) agrupamentos com 3 grupos e no gráfico (b) agrupamentos com 4 grupos.

5 CONCLUSÃO

Observou-se que as séries de nível do mar analisadas apresentaram memória longa. A metodologia proposta, de realizar agrupamentos considerando os valores do coeficiente de Hurst obtidos com diferentes métodos, mostrou-se adequada.

REFERÊNCIAS

BATISTA, C. A. **Métodos emergentes de física: estatística aplicados a séries temporais**. 2006. 69 f. Dissertação (Mestrado em Biometria) - Universidade Federal Rural de Pernambuco, Recife, 2006.

CARBONE, A; CASTELLI, G.; SATANLEY, H. E. Time-dependent Hurst exponent in financial time series. **Journal Physica A**, Warsaw, v. 344, p. 267-271, July 2004.

CARNIER NETO, D. **Análise das séries temporais de monitoramento de nível d'água em poços no aquífero Rio Claro**. 2006. 88 f. Dissertação (Mestrado em Geociências) - Universidade Estadual Paulista, Rio Claro, 2006.

FAVARETTO, A. S. **Estimativa do expoente de Hurst de séries temporais de chuvas no estado de São Paulo usando as transformadas de Fourier, Wavelets e Análise R/S**. 2004. 89 f. Dissertação (Mestrado em Física) - Universidade Estadual Paulista, Instituto de Geociências e Ciências Exatas, Rio Claro, 2004.

FERREIRA, D. **Estatística multivariada**. 2. ed. ver. e ampl. Lavras: UFLA, 2011.

KASMIM, M. A. et al. Métodos de análise em séries temporais e o planejamento de políticas públicas em saúde. In: CONGRESSO NACIONAL DE PESQUISA EM CIÊNCIAS SOCIAIS APLICADAS, 3., 2014, Francisco Beltrão. **Anais...** Francisco Beltrão: CONAPE, 2014. 11 f.

KUMAR, A. S. Testing for long memory in volatility in the indian forex market. **Economic Annals**, Belgrade, v. 59, n. 203, dezembro 2014.

LIDEN, R. Técnicas de agrupamento. **Revista de Sistemas de Informação da FSMA**, Visconde de Araújo, n. 4, p. 18-36, 2009.

MALAVOGLIA, R. C. **Verificação da presença de memória longa nos principais índices de bolsa de valores:** um estudo por meio da utilização da estatística R/S e o expoente de Hurst. 2009. 64 f. Dissertação (Mestrado em Administração de Organizações) - Universidade de São Paulo, Ribeirão Preto, 2009.

MITRA, S. K. Is hurst exponent value useful in forecasting financial time series? **Asian Social Science**, Toronto, v. 8, n. 8, p. 111, July 2012.

MORETTIN, P. A.; TOLOI, C. M. **Análise de séries temporais.** 2. ed. São Paulo: E. Blücher, 2006.

OLIVEIRA, M. A.; FAVERO, L. P. L. **Uma breve descrição de algumas técnicas para análise de séries temporais.** São Paulo: EDUSP, 2002.

OLIVEIRA, R. **Análise de Cluster:** métodos hierárquicos e de particionamento. 2004. 93 p. Graduação (Bacharel em Sistemas de Produção) - Universidade Presbiteriana Mackenzie, São Paulo, 2004.

R CORE TEAM. **R:** a language and environment for statistical computing. Vienna, 2015. Disponível em: <<http://www.r-project.org/>>. Acesso em: 10 ago. 2015.

RIHBANE, F. E. C. **Preenchimento de falhas aleatórias de séries temporais micrometeorológicas pela técnica de Monte Carlo.** 2014. 50 f. Dissertação (Mestrado em Física Ambiental) - Universidade Federal de Mato Grosso, Cuiabá, 2014.

ROSSO, A. M.; TAQQU, M. S. R. Fractionally differenced ARIMA models applied to hidrologic time series: Identification, estimation, and simulation. **Water Resources Research**, Washington, v. 33, n. 5, p. 1035-1044, maio 1997.

SAFADI, T. Using independent component for clustering of time series data. **Applied Mathematics and Computation**, New York, v. 243, p. 522-527, 2014.

TUPAYAGALE, P. **Long memory in the volatility of local currency bond markets:** evidence from Hong Kong, Mexico and South Africa. 2012. Disponível em: <<http://creativecommons.org/licenses/by/3.0/br>>. Acesso em: 20 ago. 2015.

VEDOVATTO, T. V. **Medidas de memória longa em séries temporais:** comparação de métodos de estimação do coeficiente de Hurst. 2014. 97 f. Dissertação (Mestrado em Estatística) - Instituto de Ciências Exatas da Universidade de Brasília, Brasília, 2014.