



SENAI CIMATEC

PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM
COMPUTACIONAL E TECNOLOGIA INDUSTRIAL
Mestrado em Modelagem Computacional e Tecnologia Industrial

Dissertação de mestrado

**Análise Espectral Singular: Modelagem de séries
temporais através de estudos comparativos usando
diferentes estratégias de previsão.**

Apresentada por: Renata de Miranda Esquivel
Orientador: Dr. Valter de Senna
Co-orientador: Dra. Gecynalda Soares da Silva Gomes

Maio de 2012

Renata de Miranda Esquivel

Análise Espectral Singular: Modelagem de séries temporais através de estudos comparativos usando diferentes estratégias de previsão.

Dissertação de mestrado apresentada ao Programa de Pós-graduação em Modelagem Computacional e Tecnologia Industrial, Curso de Mestrado em Modelagem Computacional e Tecnologia Industrial do SENAI CIMATEC, como requisito parcial para a obtenção do título de **Mestre em Modelagem Computacional e Tecnologia Industrial**.

Área de conhecimento: Interdisciplinar

Orientador: Dr. Valter de Senna
SENAI CIMATEC

Salvador
SENAI CIMATEC
2012

Nota sobre o estilo do PPGMCTI

Esta dissertação de mestrado foi elaborada considerando as normas de estilo (i.e. estéticas e estruturais) propostas aprovadas pelo colegiado do Programa de Pós-graduação em Modelagem Computacional e Tecnologia Industrial e estão disponíveis em formato eletrônico (*download* na Página Web http://ead.fieb.org.br/portal_faculdades/dissertacoes-e-teses-mcti.html ou solicitação via e-mail à secretaria do programa) e em formato impresso somente para consulta.

Ressalta-se que o formato proposto considera diversos itens das normas da Associação Brasileira de Normas Técnicas (ABNT), entretanto opta-se, em alguns aspectos, seguir um estilo próprio elaborado e amadurecido pelos professores do programa de pós-graduação supracitado.

SENAI CIMATEC

Programa de Pós-graduação em Modelagem Computacional e Tecnologia Industrial

Mestrado em Modelagem Computacional e Tecnologia Industrial

A Banca Examinadora, constituída pelos professores abaixo listados, leram e recomendam a aprovação da Dissertação de mestrado, intitulada “Análise Espectral Singular: Modelagem de séries temporais através de estudos comparativos usando diferentes estratégias de previsão. ”, apresentada no dia (30) de (maio) de (2012), como requisito parcial para a obtenção do título de **Mestre em Modelagem Computacional e Tecnologia Industrial**.

Orientador:

Prof. Dr. Valter de Senna

SENAI CIMATEC

Co-orientador:

Profa. Dra. Gecynalda Soares da Silva Gomes

UFBA- DEPARTAMENTO DE ESTATÍSTICA

Membro externo da Banca:

Prof. Dr. Marccone Lopes da Silva

FTC- Faculdade de Tecnologia e Ciências

Membro interno da Banca:

Prof. Dr. Gilney Figueira Zebende

SENAI CIMATEC

Resumo

Os métodos estatísticos para análise de séries temporais encontram-se amplamente desenvolvidos na literatura e vários modelos clássicos preditivos estão implementados em programas estatísticos genéricos ou específicos. Contudo, cada modelo clássico exige suposições relacionadas às características dos dados e o uso adequado dos modelos exigirá verificações dessas suposições, que podem demandar muitos esforços na etapa de identificação do padrão de comportamento da série temporal. Como alternativa, pode-se utilizar uma técnica poderosa e relativamente moderna conhecida como Análise Espectral Singular (SSA, do inglês *Singular Spectrum Analysis*). A SSA visa realizar uma decomposição da série temporal em poucos componentes independentes e interpretáveis. Este método não exige o conhecimento sobre o modelo paramétrico da série temporal e pode ser aplicado em qualquer série com alguma potencial estrutura. Apesar da existência de uma quantidade razoável de publicações sobre a SSA, esse método ainda é pouco usado pela comunidade científica brasileira. A presente dissertação objetiva apresentar o desenvolvimento teórico e aplicações da SSA visando compará-la com os modelos clássicos para séries temporais, além de divulgar uma abordagem não usual na solução de problemas práticos que envolvem séries temporais. Com esta finalidade, analisou-se duas séries temporais com características distintas, sendo uma da área da meteorologia e outra do campo financeiro. E adicionalmente, utilizou-se séries simuladas provenientes de processos estocásticos bem discutidos na literatura especializada. De uma forma geral, observou-se que o algoritmo recorrente de previsão SSA consegue representar melhor as variações existentes nos dados, como flutuações sazonais e picos, características encontradas com frequência em séries históricas. A previsão SSA apresentou um comportamento global mais condizente com a realidade das séries analisadas. Uma grande vantagem em aplicar a SSA em detrimento aos modelos clássicos refere-se à sua simplificação no entendimento das séries temporais e conseqüente diminuição da intervenção do analista. Sua utilização produziu, em geral, resultados tão bons ou superiores aos gerados pelos métodos clássicos considerados nesta investigação científica.

PALAVRAS-CHAVE: Análise Espectral Singular; Séries temporais; Previsão.

Abstract

Statistical methods for analysis of time series are largely developed in the literature and several classical predictive models are implemented in general statistical packages or specific. However, each classical model requires assumptions related to the characteristics of the data and appropriate use of models will require verification of these assumptions, which may require many efforts in the identification stage of the behavior pattern of the series. Alternatively, one can use a relatively modern and powerful technique known as Singular Spectral Analysis (SSA). The SSA aims at a decomposition of the series in a few independent components and interpretable. This method does not require knowledge of the parametric model of the series and can be applied to any series with a potential structure. Despite the existence of a reasonable amount of publications on the SSA, this method is rarely used by the Brazil's scientific community. This dissertation aims to present the theoretical development and applications of SSA in order to compare it with the classical models for time series, moreover, disclose an unusual approach in solving practical problems involving time series. For this purpose, we analyzed two time series with different features, that one area of meteorology and another from the financial field. Additionally, was used simulated series from stochastic processes, much discussed in the literature. In general, it was observed that the SSA recurrent forecasting algorithm can better represent the variation found in the data, for example, seasonal fluctuations and peaks, characteristics often found in historical series. The SSA forecasting algorithm showed an overall behavior more suited to the reality of the series analyzed. One advantage of applying the SSA instead of the classic models refers to simplifying the understanding of time series and the consequent reduction of the intervention of the analyst. In general, its use has produced results as good or greater than those generated by classical methods considered in this research.

KEYWORDS: Singular Spectrum Analysis; Time Series; Forecasting.

Sumário

1	Introdução	1
1.1	Objetivo	2
1.2	Importância da pesquisa	3
1.3	Limites e limitações	3
1.4	Organização da dissertação de mestrado	4
2	Análise de Séries Temporais	5
2.1	Noções	5
2.2	Breve história sobre a análise de séries temporais	8
2.3	Definições básicas	10
2.3.1	Processos estocásticos	10
2.3.2	Autocovariância e autocorrelação	14
2.3.3	Testes estatísticos para avaliar características em séries temporais	15
2.3.3.1	Testes para a detecção de estacionariedade	15
2.3.3.2	Testes para a detecção de tendência	17
2.3.3.3	Testes para sazonalidade	18
2.3.3.4	Comentários gerais	18
2.4	Estratégias de modelagem e previsão clássica	18
2.4.1	Modelos de suavização exponencial	19
2.4.1.1	Suavização exponencial simples (SES)	19
2.4.1.2	Suavização exponencial de Holt (SEH)	21
2.4.1.3	Suavização exponencial de Holt-Winters (H-W)	22
2.4.1.4	Comentários gerais	25
2.4.2	Modelos de Box e Jenkins	26
2.4.2.1	Modelos autorregressivos (AR)	28
2.4.2.2	Modelo de média móvel (MA)	29
2.4.2.3	Modelo autorregressivo de média móvel (ARMA)	31
2.4.2.4	Modelo autorregressivo integrado de média móvel (ARIMA)	32
2.4.2.5	Modelo autorregressivo integrado de média móvel sazonal (SARIMA)	34
2.4.2.6	Previsão	35
2.4.2.7	Comentários gerais	37
2.4.3	Medidas de erro	37
3	<i>Singular Spectrum Analysis</i>	39
3.1	Introdução	39
3.1.1	Estado da arte	40
3.2	Descrição da SSA básica	41
3.2.1	Decomposição	42
3.2.1.1	<i>Embedding</i>	42
3.2.1.2	<i>SVD</i>	43
3.2.2	Reconstrução	45
3.2.2.1	<i>Grouping</i>	45
3.2.2.2	<i>Diagonal averaging</i>	46
3.3	Informações adicionais	49

3.3.1	Separabilidade	49
3.3.2	Comprimento da janela (L)	50
3.3.3	Escolha dos autotriples	52
3.3.3.1	Valores singulares	53
3.3.3.2	Correlação ponderada	55
3.4	Previsão via SSA	56
3.4.1	Algoritmo recorrente de previsão	58
3.4.2	Exemplos de aplicações bem sucedidas	60
4	Desenvolvimento da pesquisa	64
4.1	Materiais e métodos	64
4.1.1	Cenários- Descrição dos dados	64
4.1.1.1	Estudo empírico: Meteorologia	64
4.1.1.2	Estudo empírico: Setor Financeiro	66
4.1.1.3	Estudo simulado: Séries artificiais	67
4.1.2	Métodos	68
5	Resultados	72
5.1	Estudos empíricos: Análise exploratória	72
5.1.1	VENTO	72
5.1.2	IBOVESPA	77
5.2	Estudos empíricos: Modelagem e previsão	82
5.2.1	Clássica	82
5.2.1.1	Alisamento Exponencial de Holt-Winters	82
5.2.1.2	Método de Box-Jenkins	85
5.2.2	Enfoque SSA	91
5.2.3	Comparação dos métodos preditivos	102
5.3	Estudos de simulação	107
5.3.1	Simulação 1	107
5.3.2	Simulação 2	116
6	Considerações finais	124
6.1	Conclusões	124
6.2	Contribuições	126
6.3	Atividades futuras de pesquisa	126
A	Testes estatísticos em séries temporais	128
A.1	Testes de Box-Pierce e Ljung-Box	128
A.2	Testes de Dickey-Fuller e KPSS	128
A.3	Testes de Cox-Stuart e Spearman	130
A.4	Testes de <i>Kruskal-Wallis</i> e <i>Friedman</i>	131
B	Exemplos simulados- comportamento da fac e facp	133
C	Álgebra de matrizes e Estatística Multivariada	135
C.1	Autovalores e autovetores	135
C.1.1	Autovalores e autovetores de uma matriz	135
C.2	Estatística Multivariada	136
C.2.1	Vetor e matriz aleatória	136
C.2.1.1	Médias, covariâncias e correlações	137
C.2.2	Decomposição espectral	139

C.2.3	Análise de componentes principais	141
C.2.3.1	Componentes principais exatas	142
C.2.3.2	Determinação do número de componentes principais	144
D	Modelagem de séries temporais via <i>Singular Spectrum Analysis</i>- Uma otimização do <i>window length</i>	147
D.1	Introdução	147
D.2	Materiais e Métodos	148
D.2.1	Breve descrição da SSA básica	148
D.2.1.1	Decomposição	148
D.2.1.2	Reconstrução	150
D.2.2	Otimização do <i>window length</i>	150
D.2.3	Aplicações: Características das séries temporais	151
D.3	Resultados	151
D.4	Considerações finais	152

Lista de Tabelas

3.1	Previsões na clássica série “mortes acidentais”	61
3.2	Precisão das previsões para algumas séries da Alemanha.	62
3.3	Escores para os países.	63
5.1	Previsões provenientes da forma aditiva do Holt-Winters (VENTO).	83
5.2	Previsões obtidas por Holt-Winters aditivo e multiplicativo (IBOVESPA).	84
5.3	Previsões de Box e Jenkins (VENTO).	86
5.4	Previsões obtidas através do SARIMA (IBOVESPA).	90
5.5	Previsões via SSA (VENTO).	96
5.6	Previsões obtidas através da SSA (IBOVESPA).	101
5.7	Comparação entre as técnicas consideradas (VENTO).	102
5.8	Medidas descritivas das distribuições das previsões (VENTO).	103
5.9	Comparação entre as técnicas clássicas (IBOVESPA).	105
5.10	Comparação entre os diferentes agrupamentos da SSA (IBOVESPA).	106
5.11	Comparação entre as previsões (Processo AR - Simulação 1).	111
5.12	Comparação entre as previsões (Processo MA - Simulação 1).	112
5.13	Comparação entre as previsões (Processo ARMA - Simulação 1).	113
5.14	Comparação entre as previsões (Processo ARIMA - Simulação 1).	113
5.15	Escores de desempenho preditivo SSA (Simulação 1).	114
5.16	Comparação entre as previsões (Processo AR - Simulação 2).	120
5.17	Comparação entre as previsões (Processo MA - Simulação 2).	120
5.18	Comparação entre as previsões (Processo ARMA - Simulação 2).	121
5.19	Comparação entre as previsões (Processo ARIMA - Simulação 2).	122
5.20	Escores de desempenho preditivo SSA (Simulação 2).	122

Lista de Figuras

3.1	Número mensal de mortes acidentais nos EUA entre 1973 e 1978	54
3.2	Autovalores resultantes da decomposição da série em 24 componentes	55
3.3	Matriz w-correlação (correlações ponderadas)	56
4.1	Potencial Eólico (Fonte: Atlas do Potencial Eólico do Estado da Bahia)	65
5.1	Velocidade média do vento ao longo do tempo.	72
5.2	Distribuição da velocidade média do vento (histograma à esquerda e <i>box-plot</i> à direita).	73
5.3	Correlograma (lag= 20 e lag= 3).	74
5.4	Resíduos oriundos da decomposição clássica (modelo aditivo e multiplicativo).	74
5.5	Decomposição da série VENTO ao longo do tempo.	75
5.6	Sazonalidade da série VENTO para os primeiros 336 e 1344 instantes de tempo.	75
5.7	Função autocorrelação (fac) e função autocorrelação parcial (facp) da série VENTO.	76
5.8	Diferenciação na série VENTO.	77
5.9	Número de pontos do Ibovespa no final do primeiro dia útil de cada mês ao longo do tempo (01 de fevereiro de 1996 a 03 de março de 2011).	78
5.10	Resíduos para os modelos aditivo e multiplicativo.	79
5.11	Separação dos resíduos provenientes do modelo aditivo.	80
5.12	Decomposição da série IBOVESPA ao longo do tempo.	80
5.13	Correlograma (fac estimada) e a função de autocorrelação parcial amostral (facp estimada) da série IBOVESPA.	81
5.14	Primeira e segunda diferença da série IBOVESPA.	82
5.15	Velocidade média observada e predita pela forma aditiva do algoritmo de H-W (VENTO).	83
5.16	Previsões via Holt-Winters comparadas com os valores reais (IBOVESPA).	85
5.17	Previsões do SARIMA 1 e SARIMA 2 comparadas com os valores reais (VENTO).	87
5.18	Diagnóstico dos modelos SARIMA 1 e SARIMA 2 (VENTO).	87
5.19	Diagnóstico do SARIMA(1,1,0)×(1,1,2) (a) e SARIMA(3,1,2)×(1,1,1) (b) (IBOVESPA).	89
5.20	Diagnóstico do modelo SARIMA(2,1,3)×(1,1,2) (IBOVESPA).	89
5.21	Previsões advindas dos modelos SARIMA (IBOVESPA).	91
5.22	Logaritmo dos autovalores (VENTO).	92
5.23	Primeiros pares de autovetores (VENTO).	92
5.24	Matriz w-correlação (VENTO).	93
5.25	Decomposição da série via SSA (VENTO).	94
5.26	Reconstrução da série via SSA (VENTO).	94
5.27	Previsão da série via SSA (VENTO).	95
5.28	Decomposição via SSA (IBOVESPA).	96
5.29	Logaritmo dos autovalores da SVD (IBOVESPA).	97
5.30	Primeiros pares de autovetores (IBOVESPA).	98
5.31	Primeiros autovetores (IBOVESPA).	98
5.32	Reconstrução da série IBOVESPA.	99
5.33	Matriz de correlações ponderadas (IBOVESPA).	100
5.34	Sinal e ruído: Correlação ponderada (IBOVESPA).	100
5.35	Comparação geral (VENTO)	104
5.36	Comparação para alguns instantes	104
5.37	Comparação geral das previsões [Agrupamento SSA: (a) 1-7; (b) 1-10 e (c) 1-28 (IBOVESPA)].	106

5.38	Processos AR (a), MA (b), ARMA (c) e ARIMA (d) (Simulação 1).	108
5.39	Logaritmo dos autovalores: AR (a), MA (b), ARMA (c) e ARIMA (d) (Simulação 1).	109
5.40	Correlações entre o sinal e o ruído: AR, MA, ARMA e ARIMA (Simulação 1).	110
5.41	Comparação das previsões (AR (a) e MA (b) - Simulação 1).	115
5.42	Comparação das previsões (ARMA (a) e ARIMA (b) - Simulação 1).	115
5.43	Processos AR (a), MA (b), ARMA (c) e ARIMA (d) (Simulação 2).	116
5.44	Logaritmo dos autovalores: AR (a), MA (b), ARMA (c) e ARIMA (d) (Simulação 2).	117
5.45	Correlação ponderada entre o Sinal e o Ruído: AR, MA, ARMA e ARIMA (Simulação 2).	119
5.46	Comparação gráfica (AR (a), MA (b), ARMA (c) e ARIMA (d) - Simulação 2).	123
B.1	Gráficos das séries simuladas, modelo AR(1) e $\phi_1=0,7$ (a); fac do AR (b) e facp do processo AR(1) (c); modelo MA(1) e $\theta_1 = 0,1$ (d); fac do MA (e) e facp do MA (f).	133
C.1	<i>Scree plot</i> ($\hat{\lambda}_i$ versus a <i>i</i> -ésima componente principal).	146

Lista de Siglas

ACP	Análise de Componentes Principais
CP	Componente Principal
AIC	Akaike Information Criterion (sigla em inglês para Critério de Informação de Akaike)
AR	Modelo Autorregressivo
ARIMA	Modelo Autorregressivos Integrado de Média Móvel
ARMA	Modelo Autorregressivo de Média Móvel
ARMAV	Modelo Autorregressivo de Média Móvel Multivariado
BIC	Bayesian Information Criterion (sigla em inglês para Critério de Informação Bayesiano)
CIMATEC ..	Centro Integrado de Manufatura e Tecnologia
DF	Dickey-Fuller
EOF	<i>Empirical Orthogonal Functions</i>
H-W	Suavização Exponencial de Holt-Winters
IBOVESPA .	Índice da Bolsa de Valores de São Paulo
KPSS	Kwiatkowski-Phillips-Schmidt-Shin
LRF	<i>Linear Recurrent Formulae</i>
MA	Modelo de Média Móvel
MSE	<i>Mean Square Error</i> (sigla em inglês para Erro Quadrático Médio)
PPGMCTI ..	Programa de Pós-graduação em Modelagem Computacional e Tecnologia Industrial
SARIMA ...	Modelo Autorregressivo Integrado de Média Móvel Sazonal
SEH	Suavização Exponencial de Holt
SENAI	Serviço Nacional de Aprendizagem Industrial
SES	Suavização Exponencial Simples
SSA	<i>Singular Spectrum Analysis</i> (sigla em inglês para Análise Espectral Singular)
STL	<i>Seasonal Decomposition of Time Series by Loess</i> (sigla em inglês para Decomposição Sazonal por Loess)
SVD	<i>Singular Value Decomposition</i> (sigla em inglês para Decomposição do Valor Singular)
TPE	<i>Total Percentage Error</i> (sigla em inglês para Erro Percentual Total)

Introdução

O desenvolvimento de métodos estatísticos para análise de dados obtidos em situações em que as observações são dependentes tem apresentado crescimento vertiginoso nas últimas décadas e, em particular, na análise de dados provenientes de séries temporais.

Ao trabalhar com séries temporais o objetivo mais usual é a predição de valores futuros. A necessidade em obter previsões precisas de eventos futuros ou suas consequências, sejam climáticas, econômicas, epidemiológicas ou de qualquer natureza, tem levado ao desenvolvimento de muitas técnicas de previsão em séries temporais.

Os clássicos métodos estatísticos para análise de séries temporais encontram-se bem desenvolvidos na literatura. Contudo, muitos desses métodos requerem um conhecimento especializado para sua boa aplicação, como, por exemplo, caso a série apresente tendência pode-se aplicar a suavização exponencial de Holt, e se além dessa componente, a série possuir sazonalidade esta poderá ser modelada através da suavização de Holt-Winters. Sendo assim, o uso adequado dos modelos clássicos exigirá verificações das suas suposições, que podem demandar muitos esforços na análise exploratória dos dados. Como alternativa, pode-se utilizar a Análise Espectral Singular (SSA, do inglês *Singular Spectrum Analysis*).

A SSA é um método não-paramétrico usado em análise de séries temporais que não exige o conhecimento prévio do comportamento da série. Essa técnica investiga o comportamento das séries históricas através de uma decomposição e reconstrução dos seus componentes constitutivos, caracterizando os estágios da SSA.

A ferramenta SSA tem-se mostrado poderosa e útil nas análises de séries das áreas de meteorologia, geofísica, física, climatologia, economia, saúde, entre outros campos do conhecimento. Essa ferramenta pode ser aplicada em séries curtas ou longas, séries não estacionárias ou estacionárias, ruidosas ou não, ou seja, em qualquer série temporal com alguma estrutura (GOLYANDINA, NEKRUTKIN e ZHIGLJAVSKY, 2001; HASSANI, 2007).

Vale salientar que, a SSA não se detém a realizar apenas previsões de valores futuros, mas a técnica também tem a finalidade de identificar e extrair padrões geradores da série temporal.

No decorrer da presente dissertação serão abordadas a fundamentação teórica do processo de modelagem e previsão, clássica e via SSA; além da apresentação e discussão dos

resultados. Os métodos que comumente apresentam bom poder preditivo são analisados, a saber: os algoritmos de alisamento exponencial de Holt e de Holt-Winters, os modelos autorregressivos integrados de médias móveis (ARIMA) e a classe de modelos autorregressivos integrados de médias móveis sazonais (SARIMA). Tais métodos foram contrastados com a modelagem resultante da metodologia SSA.

Dois ambientes computacionais foram considerados para a realização das análises, o R e o *Caterpillar*. O processo de modelagem das séries temporais e a previsão clássica ocorreram no ambiente R, enquanto que a previsão através da SSA foi sucedida no *software* nomeado como *Caterpillar*.

O R é conhecido como um ambiente para computação estatística e gráfica e possui uma linguagem de programação de mesmo nome, que é similar à linguagem e ambiente S. A grande vantagem do *software* R é a possibilidade de criar funções e adicionar funcionalidades pré-definidas do programa, assim como a sua flexibilidade na expansão através da criação de pacotes; já que o R é um *software* livre (código aberto). O programa R pode ser executado no sistema operacional Linux, Macintosh e Windows. Nas análises do presente trabalho utilizou-se a versão 2.13.0 do R. Detalhes sobre o R podem ser consultados na página <http://www.R-project.org/>.

O *Caterpillar* foi desenvolvido por um grupo de pesquisadores do Departamento de Matemática da Universidade de São Petersburgo na Rússia denominado como *GistaT Group*. Informações sobre o desenvolvimento da técnica SSA e do correspondente *software* são disponibilizadas na página oficial do grupo. No presente trabalho fez-se uso de uma licença temporária da versão mais atual (3.4) disponível para download na *homepage* do *GistaT Group* <http://www.gistatgroup.com/>.

1.1 Objetivo

O presente trabalho de dissertação objetiva apresentar o desenvolvimento teórico, as vantagens e as aplicações da Análise Espectral Singular visando compará-la com os principais modelos preditivos clássicos para séries temporais, além de divulgar uma abordagem não usual na solução de problemas práticos que envolvem séries temporais. Com esta finalidade, analisou-se duas séries temporais com características distintas, sendo uma da área da meteorologia e outra do campo financeiro. E adicionalmente, motivou-se utilizar séries simuladas provenientes de processos estocásticos bem discutidos na literatura especializada.

1.2 Importância da pesquisa

O presente trabalho de dissertação traz uma contribuição diferenciada com o uso da técnica SSA, no sentido de difundir e disponibilizar aos pesquisadores alternativas de métodos que consideram os aspectos não triviais das séries temporais. Já que, existem pouquíssimos grupos de pesquisa no Brasil que utilizam a metodologia SSA, além de desconhecermos, até o momento, pesquisas brasileiras que investigam e/ou aplicam a previsão SSA. E entre os poucos grupos brasileiros existentes, nenhum utiliza um enfoque Estatístico.

Outro fator motivador para o uso dessa técnica refere-se à sua potencialidade na identificação e extração de componentes de uma série temporal, além da sua capacidade preditiva. Desta forma, as questões de investigação dos pesquisadores poderão ser discutidas em um contexto ainda não explorado em diversas áreas do conhecimento, possibilitando melhores previsões e, por conseguinte, intervenções humanas mais eficientes.

Portanto, pelo amplo espectro de aplicações da SSA, a divulgação e acesso à utilização deste método, certamente trará enormes contribuições a todos que precisam se antecipar a eventos futuros.

1.3 Limites e limitações

Apesar da existência de técnicas para a análise de séries temporais multivariadas, como, por exemplo, o enfoque multivariado dos modelos Box-Jenkins - como a classe de modelos autorregressivos de médias móveis multivariados (ARMAV) e a SSA para séries multivariadas - MSSA (do inglês, *Multivariate Singular Spectrum Analysis*), optou-se em restringir a pesquisa ao campo metodológico de séries univariadas. Fato este justificável em razão do uso frequente de séries univariadas pela comunidade científica brasileira, sendo assim conveniente adotar a forma mais simples dos métodos para análise de séries temporais.

A metodologia SSA pode ser automatizada, de modo a não exigir do usuário muita familiaridade com o processo, para sua aplicação. Essa pesquisa de dissertação apresenta limitações no que tange a implementação da metodologia de previsão no *software* R, já que não é do nosso conhecimento, até o momento, a disponibilidade da previsão SSA em ambiente de código aberto. Todavia, em virtude da complexidade na fundamentação teórica da técnica SSA, devido a grande quantidade de ferramentas envolvidas nos estágios da técnica SSA e no seu processo de previsão, não houve tempo hábil para desenvolver um sistema automatizado de análise via SSA.

1.4 Organização da dissertação de mestrado

Este documento apresenta 5 capítulos e está estruturado da seguinte forma:

- **Capítulo 1 - Introdução:** Contextualiza o âmbito, no qual a pesquisa proposta está inserida. Apresenta, portanto, a definição do problema, objetivos e justificativas da pesquisa e como esta dissertação de mestrado está estruturada;
- **Capítulo 2 - Análise de Séries Temporais:** Apresenta um apanhado geral da Análise de Séries Temporais e a metodologia das principais técnicas clássicas para séries temporais;
- **Capítulo 3 - *Singular Spectrum Analysis*:** Aborda a história, a descrição e a fundamentação teórica do processo de decomposição, reconstrução e previsão através da Análise Espectral Singular;
- **Capítulo 4 - Desenvolvimento da pesquisa:** Exibe a metodologia do trabalho de dissertação, a qual contempla a descrição das séries utilizadas, os materiais, e do processo de modelagem das séries, os métodos;
- **Capítulo 5 - Resultados:** Descreve as principais características das séries temporais analisadas e exibe os resultados da modelagem clássica e via a Análise Espectral Singular, além da apresentação das discussões;
- **Capítulo 6 - Considerações Finais:** Apresenta as conclusões, contribuições e algumas sugestões de atividades de pesquisa a serem desenvolvidas no futuro.

Análise de Séries Temporais

2.1 Noções

A Análise de Séries Temporais é um ramo da Ciência Estatística dedicado ao tratamento analítico de uma série de instâncias ou observações dependentes. Informalmente, pode-se definir uma série temporal como um conjunto de observações $\{Y_k(t), t \in T \text{ e } k \in \mathbb{N}\}$ obtidas sequencialmente ao longo de um intervalo de tempo ou outro parâmetro físico, como profundidade, por exemplo. Em que Y_k denota as k variáveis de investigação, T o conjunto índice ou conjunto de instantes de tempo e \mathbb{N} o conjunto dos números naturais (MORETTIN e TOLOI, 2006). A definição formal de série temporal será apresentada na Seção (2.3). Nesta seção são abordadas as fundamentações probabilísticas sobre Processos Estocásticos.

As séries temporais podem ser classificadas em termos do conjunto índice e em relação ao conjunto dos possíveis valores. Em ambos os contextos as possíveis classificações são: discretas ou contínuas. No que tange o conjunto índice, as séries são ditas contínuas se $T = \{t: t_1 < t < t_2\}$, ou seja, as observações são coletadas de forma ininterrupta. Por outro lado, o conjunto índice das séries temporais discretas é denotado por $T = \{t_1, t_2, \dots, t_N\}$, representando que a coleta das instâncias ocorre em instantes pontuais, podendo ser igualmente espaçados ou não. Contudo, uma série temporal discreta pode ser obtida através da amostragem de uma série contínua limitada por instantes específicos (MORETTIN e TOLOI, 2006).

Assim como ocorre nos procedimentos desenvolvidos para observações independentes, as séries temporais podem ser univariadas ou multivariadas. As séries temporais consideradas no presente trabalho representam apenas uma característica de interesse Y_t , séries univariadas; em tempo discreto $T = \{1, 2, \dots, N\}$; com observações no conjunto contínuo e em intervalos equidistantes no tempo.

Geralmente as séries temporais possuem variações ou componentes específicos que caracterizam os seguintes comportamentos:

- **Tendência:** Representa o efeito de longo prazo ao redor da média, ou seja, movimento dominante em uma série temporal. Além de indicar a velocidade em que a série desenvolve-se ao longo do tempo. Tal efeito pode ser de crescimento/decrescimento linear ou não-linear, por exemplo, a média pode ser uma função polinomial do tempo.

- Sazonalidade: Um componente em que Morettin e Tolo (2006) caracteriza-o como difícil de definir do ponto de vista conceitual e estatístico, mas que pode ser definido empiricamente como os fenômenos que ocorrem com periodicidade anual. Em outras palavras, a variação sazonal refere-se a movimentos similares ou repetição de um padrão que ocorre na série em anos sucessivos. Como exemplo pode-se citar que, em média as temperaturas no verão são mais altas do que no inverno em qualquer que seja o ano. Assim como, é de praxe que as vendas de produtos de praia sejam menores no inverno do que no verão para todos os anos.
- Ciclo: Variação com comportamento similar à componente sazonal, também periódica ou oscilatória, mas não relaciona-se diretamente a uma medida temporal e está susceptível a influências de fatores externos. Por esse motivo os ciclos não possuem duração uniforme. Tipicamente, o movimento cíclico possui um comprimento maior do que o observado na variação sazonal (períodos superiores a um ano). Pode-se citar como exemplos os ciclos relacionados à economia, durações entre dois a cinquenta anos; ciclos de epidemias e o ciclo solar, que indica a atividade do sol em intervalos de aproximadamente 11 anos.
- Variações aleatórias: Movimentos de grande instabilidade que correspondem às flutuações de uma série temporal devido ao acaso ou fatores aleatórios. Tal movimento é também denominado como “erro aleatório” e representa as variações não capturadas pelos demais componentes.

Na análise de séries temporais, é comum realizar a identificação das componentes descritas acima. Essa identificação pode ser realizada através da decomposição dos componentes da série investigada. Na literatura há sugestões sobre a forma de decomposição da série em componentes constituintes, permitindo melhor compreensão sobre a série ao isolar as componentes não-observáveis da série. A maneira clássica de decomposição fundamenta-se em um modelo aditivo ou multiplicativo envolvendo os componentes da série. A decomposição aditiva pode ser representada através da expressão

$$Y_t = m_t + S_t + \varepsilon_t, \quad (2.1)$$

e caso o modelo seja multiplicativo, pode-se representar a decomposição como

$$Y_t = m_t S_t \varepsilon_t, \quad (2.2)$$

em que, m_t e S_t representam os componentes de tendência e sazonalidade (incluindo a variação cíclica), respectivamente; e ε_t retrata o erro aleatório.

Geralmente, utiliza-se a decomposição multiplicativa quando a amplitude da variação da série temporal cresce no decorrer do tempo. Ou seja, há indicativo de que as flutuações sazonais são proporcionais ao nível médio da série.

Conforme apresentado por Morettin e Tolo (2006) e Kunst (2009), os principais objetivos ao estudar séries temporais podem ser sintetizados a partir dos seguintes itens:

- **Descrição:** descrever o padrão de comportamento da série através de gráficos, medidas sumárias, verificação de tendência, ciclos e variações sazonais, entre outras ferramentas descritivas. Geralmente este objetivo é parte integrante dos estudos.
- **Modelagem:** Construção de modelos que expliquem o mecanismo gerador da série temporal; uso da variação de uma série para explicar a variação de outra série; investigar a relação de uma série com outras de interesse. A construção desses modelos envolve a estimação de parâmetros e a avaliação da qualidade estatística do ajuste.
- **Previsão:** estimação de valores futuros da série temporal considerando os valores em instantes anteriores. As previsões podem ser a curto, médio ou a longo prazo. Em muitos estudos de séries temporais, especialmente econômicos, a previsão é crucial, já que possibilita conhecer a provável evolução da série e, conseqüentemente, tomar decisões apropriadas.
- **Controle de processos:** Uso específico para monitorar os valores da série que representam a qualidade de um processo, o que preconiza a detecção de mudanças estruturais ou periodicidades na série. Estas alterações estruturais podem indicar, no âmbito do controle estatístico de qualidade, que o modelo adotado para retratar o funcionamento do processo já não é válido.

Vale salientar que, em uma investigação pode-se explorar mais de um dos objetivos descritos acima.

Segundo Brillinger (2001), os conceitos de séries temporais são utilizados com diferentes enfoques, por exemplo: modelos probabilísticos, métodos não-paramétricos, estudos longitudinais, medidas repetidas, curvas de crescimento, métodos multivariados, processamento de sinais e análise espectral, modelos de redes neurais, modelos de espaço de estados, sistemas dinâmicos, entre outros.

As abordagens utilizadas em séries temporais são analisadas de acordo com dois domínios: temporal e de frequências. Quando a análise é realizada no domínio do tempo os modelos obtidos são classificados como paramétricos, enquanto que no domínio de fre-

quências os modelos construídos são ditos não-paramétricos (TSAY, 2000; MORETTIN e TOLOI, 2006).

A classe dos modelos paramétricos caracteriza-se pela sua quantidade finita de parâmetros. Os modelos dessa classe comumente usados são os modelos autorregressivos integrados de médias móveis (ARIMA) e as suas variações, modelos de regressão, modelos estruturais e modelos não-lineares.

Na classe dos modelos não-paramétricos, os mais conhecidos são a função de autocovariância \ autocorrelação e a sua transformada de Fourier, também conhecida como espectro de frequência da série temporal, e Wavelets (TSAY, 2000; MORETTIN e TOLOI, 2006; KUNST, 2009). Comumente, do ponto de vista clássico, quando os modelos não-paramétricos são aplicados, costuma-se denominar a investigação como análise espectral.

2.2 Breve história sobre a análise de séries temporais

Conforme apresentado em Tsay (2000), as aplicações contribuíram fundamentalmente para o desenvolvimento das metodologias de séries temporais. Há bastante tempo, realiza-se coletas e o acompanhamento de séries temporais. Uma das séries mais antigas, conhecida na literatura através de uma das publicações do Graunt (1662), intitulada por *Natural and Political Observations upon the Bills of Mortality*, retrata a distribuição das idades de morte no século XVII em Londres (KLEIN, 1997). Tufte (1983) apresenta curiosidades sobre o início do trabalho com séries temporais, este autor mostra informações sobre os movimentos dos planetas e do sol do século 10 ou 11 (BRILLINGER, 2001). Outra aplicação que motivou os estudos da análise de séries temporais, segundo os historiadores, refere-se a clássica série do número médio anual de manchas solares de Wolfer para o período de 1749 a 1924 (BRILLINGER, 2001; MORETTIN e TOLOI, 2006).

No que tange ao desenvolvimento metodológico, a análise espectral pode ser considerada como técnica pioneira. Tal técnica produziu a primeira aplicação datada em 1664 através do experimento de Newton, o qual envolvia a passagem da luz solar, decomposta em seus componentes, por um prisma (BRILLINGER, 2001).

Ao passar do tempo, tiveram outros desenvolvimentos técnicos na metodologia de séries temporais. Pode-se citar o *paper* do T.N. Thiele de 1880 que contém contribuições relacionadas à formulação e análise de um modelo para uma série temporal constituído de uma soma de componentes da regressão, movimento Browniano e um ruído branco; a partir do qual desenvolveu um procedimento recursivo para estimar a componente da regressão e prever o movimento Browniano, procedimento conhecido a partir de 1961 por filtro de Kalman (KALMAN e BUCY, 1961; LAURITZEN, 1981; TSAY, 2000).

Todavia, a explosão no uso da técnica ocorreu na década de 80 como evidencia os diversos artigos publicados nas revistas específicas de Estatística, cujo título possui o termo “filtro de Kalman”. E a partir daí, desenvolveu-se diversas formas de aplicabilidade da técnica, como por exemplo, tratamento de valores faltantes ou *missings*; no desenvolvimento de novos métodos para a extração de sinal; para alisamento e ajuste sazonal; entre outros (TSAY, 2000).

Vale ressaltar que, as abordagens utilizadas em séries temporais (domínio temporal e domínio de frequências) desenvolveram-se de forma independente a partir das contribuições dos seus defensores. Segundo Tsay (2000), no início da história ocorriam muitos debates e críticas entre as duas abordagens, contudo, a forte separação entre as escolas está diminuindo. E, atualmente, a escolha entre qual abordagem usar depende, principalmente, do objetivo da análise e da experiência do analista, podendo utilizar técnicas de ambas as abordagens.

De acordo com a literatura, o século XX proporcionou muitos avanços importantes na análise de séries temporais, principalmente na área da previsão. A partir de 1920 a previsão deixou de ser realizada através de extrapolações simples ao redor do valor médio e passou a levar em consideração os valores passados da série temporal. O desenvolvimento na área iniciou-se com Yule (1927), ao aplicar um modelo linear aos dados de manchas solares, chamando-o de autorregressivo (AR) (KLEIN, 1997; TSAY, 2000).

Com a evolução dos estudos na área da previsão de séries temporais e com a chegada do computador nos anos 50, houve uma popularidade nos modelos de previsão. É época em que foi criada uma técnica simples, eficiente e razoavelmente precisa para tratar as causas de flutuações e realizar previsões a partir do padrão básico identificado na série. Tal técnica ficou conhecida como Alisamento ou Suavização Exponencial. Sendo Brown o precursor da técnica, ele trabalhou em prol do desenvolvimento da metodologia de alisamento exponencial e publicou dois importantes livros da época, o *Statistical Forecasting for Inventory Control* de 1959 e o *Smoothing, Forecasting and Prediction of Discrete Time Series*, em 1963 (KLEIN, 1997; MORETTIN e TOLOI, 2006; GARDNER, 2006).

Em paralelo aos trabalhos de Brown, Charles Holt, com o apoio do *Office of Naval Research* (ONR), desenvolvia técnicas de suavização exponencial para tratar séries com tendências e sazonalidades. Inicialmente, o seu trabalho foi documentado no memorando do próprio ONR (HOLT, 1957) e bem posteriormente, em 2004, publicado no *International Journal of Forecasting* (HOLT, 2004a, 2004b). Apesar do registro do trabalho em 1957, a técnica desenvolvida por Holt apenas atingiu ampla divulgação a partir do trabalho do Winters (1960), que testou os métodos de Holt com dados empíricos. Desde então, a técnica ficou conhecida como suavização exponencial de Holt-Winters (KLEIN, 1997; GARDNER, 2006).

Na década de 70 houve um marco importante para a análise de séries temporais, uma nova classe de modelos vinha à tona. Esta classe de modelos ficou conhecida como metodologia de Box e Jenkins a partir da publicação, em 1970, do clássico livro *Time Series Analysis, Forecasting and Control*. Em linhas gerais, a metodologia de Box e Jenkins ou ARIMA trabalha com modelos probabilísticos lineares e considera um processo sistemático na modelagem das séries. Tal processo inicia-se na especificação e identificação do modelo mais adequado aos dados; passando pela estimação dos parâmetros; verificação da adequação do modelo; e a previsão dos valores futuros (CHAVES, 1991; TSAY, 2000; MORETTIN e TOLOI, 2006).

Com o passar das décadas e com a evolução dos estudos, os pesquisadores da época notaram a necessidade em considerar a não-linearidade na análise de séries temporais, haja vista que muitas das aplicações não apresentavam a linearidade suposta nos modelos criados até o momento (CHAVES, 1991; TSAY, 2000).

A partir da década de 80, com a melhora no acesso aos microprocessadores, os pesquisadores vêm utilizando técnicas alternativas para a detecção de padrões e tratamento de séries não-lineares, principalmente, de uma forma automatizada.

2.3 Definições básicas

Antes de apresentar os modelos clássicos para análise de séries temporais, faz-se necessário abordar algumas definições relacionadas aos processos estocásticos, dentro do contexto de séries temporais.

2.3.1 Processos estocásticos

Os modelos utilizados na representação dos fenômenos investigados na forma de séries temporais são ditos processos estocásticos ou aleatórios, ou seja, fenômenos controlados por leis probabilísticas.

Um processo estocástico é geralmente interpretado como uma coleção de variáveis aleatórias dependentes de um ou mais parâmetros variáveis, como o tempo, por exemplo. Em outras palavras, um processo estocástico é o conjunto de todas as possíveis trajetórias de um fenômeno ou uma coleção de realizações do processo físico.

A ideia de um processo estocástico e uma de suas realizações é análoga a utilizada na descrição de população e amostra. Na investigação de muitos problemas práticos

coleta-se informações de uma amostra do fenômeno visando conhecê-lo na população. De forma semelhante ocorre nos estudos de séries temporais, em que observa-se uma realização do processo aleatório estudado, denominada série temporal, com intuito de conhecer o processo estocástico associado. Por exemplo, a temperatura do ar, em um dado local e intervalo de tempo, é considerada como uma realização ou trajetória do processo estocástico gerador das temperaturas (populacional) no local do experimento (GUJARATI, 2000; GNEDENKO, 2008; MORETTIN e TOLOI, 2006).

Considere um processo estocástico $\{Y(t), t \in T\}$, em que T indexa a evolução do parâmetro físico ou conjunto índice, em particular o tempo. Usualmente, descreve-se um processo estocástico através de suas funções média, autocovariância e variância. A função média de $Y(t)$ pode ser denotada como

$$E[Y(t)] = \mu_t ,$$

a autocovariância entre $Y(t)$ e $Y(s)$ por

$$\gamma_{t,s} = cov(Y(t), Y(s)) = E[(Y(t) - \mu_t)(Y(s) - \mu_s)],$$

sendo t e s dois instantes distintos. Ou ainda,

$$\gamma(h) = cov(Y(t), Y(t+h)) = E[(Y(t) - \mu_t)(Y(t+h) - \mu_{t+h})],$$

representando a autocovariância entre $Y(t)$ e $Y(t+h)$. Sendo h a defasagem ou *lag*, isto é, a distância entre dois instantes de tempo. E caso h seja zero tem-se a variância de $Y(t)$ dada por

$$\gamma(0) = \text{Var}[Y(t)] = \sigma_t^2.$$

Existem alguns processos estocásticos com comportamentos específicos, caracterizando os diferentes tipos de processos aleatórios. Por exemplo, pode-se citar os processos estocásticos conhecidos na literatura e que recebem atenção em muitos estudos de séries temporais, a saber: Processos estacionários; Processos Gaussianos ou normais; Ruído branco; Passeio aleatório; Processos Markovianos; Movimento browniano; entre outros.

Para o propósito do presente trabalho as ideias sobre os processos estacionários, ruído branco e passeio aleatório serão apresentadas.

Os processos estocásticos podem apresentar estacionariedade *fraca*, também denominada como “ampla” ou ainda de “segunda ordem”; ou estacionariedade “forte” (estrita). Contudo, nas situações práticas, restringe-se a análise ao caracterizar os processos aleatórios através dos seus dois primeiros momentos, o que possibilita investigar os “processos estacionários de segunda ordem” (MORETTIN e TOLOI, 2006).

Um processo estocástico é dito fracamente estacionário se e somente se

1. $E[Y(t)] = \mu_t = \mu$, constante, para todo $t \in T$;
2. $\text{Var}[Y(t)] = \sigma_t^2 = \sigma^2$, constante, para todo $t \in T$;
3. $\gamma(h) = \text{cov}(Y(t), Y(t+h))$, é uma função de h .

Em outras palavras, um processo aleatório é fracamente estacionário se as suas médias e variâncias forem constantes ao longo do tempo e a sua função de autocovariância depender apenas da defasagem h entre dois instantes distintos. Por convenção, um processo fracamente estacionário é conhecido simplesmente como *processo estacionário* (GUJARATI, 2000; MORETTIN e TOLOI, 2006; GNEDENKO, 2008).

Caso uma série não seja estacionária, esta pode transformar-se em estacionária ao ser diferenciada uma quantidade d finita de vezes. Ao realizar tal procedimento, a série também estará livre de tendência, ou seja, a componente de tendência será removida da série temporal.

Quando uma série temporal não é estacionária, mas torna-se estacionária após d diferenças, costuma chamá-la como série integrada de ordem d e denotá-la como $I(d)$. E por convenção, o processo $I(d=0)$ representa uma série estacionária (GUJARATI, 2000; GNEDENKO, 2008).

A diferenciação na série ocorre ao aplicar o conhecido operador diferença dado por

$$\Delta^d Y_t = \Delta[\Delta^{d-1} Y_t],$$

em que $\Delta Y_t = Y_t - Y_{t-1}$ é a primeira diferença da série temporal e denota a variação dos dados entre dois tempos consecutivos; e d representa o número de diferenças. Geralmente,

no máximo duas diferenças são suficientes para tornar a série temporal estacionária.

Todavia, para algumas séries temporais as diferenças não serão suficientes para atingir a estacionariedade. Costuma-se aplicar uma transformação não linear nos dados, antes de realizar as diferenças, de modo que a série temporal passe a ser estacionária. De uma forma geral, aplica-se uma transformação na série temporal com o intuito de alcançar a estabilidade da variância e/ou simetria nos dados, ou ainda, em séries sazonais, tornar o efeito sazonal aditivo. A escolha da transformação deve ser realizada de acordo com o comportamento da série temporal. Tipicamente, utiliza-se a conhecida transformação de Box-Cox; ver Morettin e Toloí (2006).

Segundo estudos teóricos, a transformação nos dados acarreta em previsões viesadas, já que, os métodos clássicos de previsão não consideram ajustes nos estimadores. Além disso, verificou-se também que a heterocedasticidade nos dados não influencia na qualidade da previsão, pois, no caso de séries temporais, os estimadores não são viesados (GRANGER e NEWBOLD, 1976 *apud.* MORETTIN e TOLOI, 2006). Outro estudo teórico interessante, refere-se a comparação de previsões através de modelos ARIMA não sazonais e um modelo sazonal. Neste estudo, verificou-se que é preferível obter previsões a partir de um modelo sazonal ao invés do ajustamento sazonal seguido pela aplicação de um modelo não sazonal na série temporal (PLOSSER, 1979 *apud.* MORETTIN e TOLOI, 2006).

Os modelos clássicos lineares que serão abordados neste trabalho são apropriados para processos estacionários ou não estacionários homogêneos, também denominados como processos integráveis, ou ainda, processos não explosivos. Nos processos não estacionários homogêneos há a suposição de que o mecanismo gerador da série produz erros autocorrelacionados e que as séries são não estacionárias em nível (média) e/ou inclinação (tendência). As séries não estacionárias em nível e/ou inclinação oscilam ao redor de um nível e/ou numa certa direção por algum tempo e posteriormente mudam para outro nível e/ou direção (MORETTIN e TOLOI, 2006).

Um importante processo estocástico refere-se ao ruído branco. Um processo $\{\varepsilon(t), t \in T\}$ é especificado como ruído branco se

1. $E[\varepsilon(t)] = 0$, constante, para todo $t \in T$;
2. $Var[\varepsilon(t)] = \sigma_\varepsilon^2$, para todo $t \in T$;
3. $cov[\varepsilon(t), \varepsilon(s)] = 0$, para todo $t \neq s$.

Essas especificações representam uma coleção de variáveis aleatórias não correlacionadas, centradas em zero e variância finita e constante. Usualmente utiliza-se a notação

$\varepsilon(t) \sim \text{RB}(0, \sigma_\varepsilon^2)$ para indicar o particular tipo de processo estocástico.

Um processo estocástico é dito ser um passeio aleatório se a primeira diferença deste processo resulta em um ruído branco. Uma definição mais formal consiste em considerar uma sequência aleatória de variáveis aleatórias $\{\varepsilon_t, t \geq 1\}$ identicamente distribuídas com média μ_ε e variância σ_ε^2 . Um processo $\{Y_t\}$ é chamado de passeio aleatório se

$$Y_t = Y_{t-1} + \varepsilon_t.$$

Na prática existem séries temporais que se comportam como um passeio aleatório, como a difusão gasosa ou líquida, o caminho percorrido por uma molécula e os preços diários de ações (EHLERS, 2009).

2.3.2 Autocovariância e autocorrelação

Uma importante ferramenta para analisar a estrutura de uma série temporal consiste na quantificação da dependência entre as instâncias da série. Essa quantificação é realizada através da autocovariância ou autocorrelação amostral, de forma análoga ao coeficiente de correlação amostral ou covariância entre duas diferentes variáveis. A autocorrelação nada mais é do que a correlação defasada, em um número h de unidades de tempo, de uma série consigo mesma e que pode ser estimada por

$$r_h = \frac{\sum_{t=1}^{n-h} [Y(t) - \bar{Y}(t)][Y(t+h) - \bar{Y}(t)]}{\sum_{t=1}^n [Y(t) - \bar{Y}(t)]^2} \quad \text{ou} \quad r_h = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

em que $\hat{\gamma}(h)$ é a covariância amostral na defasagem h ($h = 0, 1, 2, \dots$); $\hat{\gamma}(0)$ é a variância amostral, ao considerar séries estacionárias; e $\bar{Y}(t)$ é a média da série temporal. A autocorrelação é adimensional e varia de $-1 < r_h < 1$.

Uma forma de representar as autocorrelações é através do gráfico denominado correlograma ou função de autocorrelação amostral. Em tal gráfico esboça-se os primeiros valores de r_h contra os valores não-negativos das defasagens h considerados na análise.

Em linhas gerais, interpreta-se o correlograma buscando a identificação de padrões que retratem características da série temporal. Por exemplo, se uma série temporal é puramente aleatória, isto é, um ruído branco, os seus valores defasados não são correlacionados e espera-se que r_h seja zero; teoricamente, os coeficientes de autocorrelação

amostral se distribuem aproximadamente como uma normal com média zero e variância igual ao inverso do tamanho amostral ($\frac{1}{N}$), sob a suposição de estacionariedade fraca. Já se o correlograma apresenta decaimento muito lento das autocorrelações amostrais há uma indicação de não estacionariedade na série (GUJARATI, 2000).

O correlograma pode ser utilizado também na identificação do modelo paramétrico mais adequado para ajustar aos dados investigados, especialmente o ARIMA, como estimativa da função autocorrelação (fac). A identificação do modelo também pode ser realizada através das funções estimadas de autocovariância (facv) e autocorrelação parcial (facp), assim como são apresentadas na Subseção 2.4.2.

A ideia da função autocorrelação parcial é semelhante a ideia do coeficiente de regressão parcial, já que ambos controlam a influência dos demais fatores considerados na correlação; variáveis, no caso da regressão, e observações no contexto de séries temporais. A autocorrelação parcial quantifica a correlação entre observações que estejam em h períodos afastados (isto é, entre as observações Y_t e Y_{t-h}), após ocorrer o controle das correlações nas defasagens anteriores a h (GUJARATI, 2000; MORETTIN e TOLOI, 2006).

2.3.3 Testes estatísticos para avaliar características em séries temporais

Antes de estimar os componentes ou iniciar o processo de modelagem de uma série recomenda-se analisar as características da série temporal. A seguir serão apresentados alguns testes úteis para este fim.

2.3.3.1 Testes para a detecção de estacionariedade

- Teste com base no correlograma

Assim como mencionado na subseção anterior (2.3.2), se o correlograma apresentar autocorrelações amostrais que diminuem gradualmente, tem-se um indicativo de não estacionariedade na série temporal. Por outro lado, se as autocorrelações forem todas nulas a partir da primeira defasagem, encontra-se o padrão das autocorrelações de um processo estocástico puramente aleatório.

Entre essas duas situações extremas, na prática, pode-se encontrar correlogramas com perfis não tão facilmente identificáveis. Por esse motivo, costuma-se também avaliar a significância estatística das autocorrelações estimadas (r_h). Essa avaliação pode ser realizada

através do erro padrão do estimador do coeficiente de autocorrelação $\left(EP(r_h) = \frac{1}{\sqrt{N}}\right)$, assim como, através do intervalo de confiança para o parâmetro ρ , isto é, o coeficiente de autocorrelação populacional. O intervalo de confiança pode então ser incluído no correlograma. Ao considerar 95% de confiança, por exemplo, o intervalo de confiança para qualquer ρ_h será dado por $IC(\rho_h; 95\%) = \left[0 \mp 1,96 \frac{1}{\sqrt{N}}\right] = \left[-1,96 \frac{1}{\sqrt{N}}; 1,96 \frac{1}{\sqrt{N}}\right]$. Desta forma, há 95% de chance de que cada ρ_h esteja dentro dos limites de confiança. E se um particular r_h estiver incluso no $IC(\rho_h; 95\%)$, a hipótese de que $\rho_h = 0$ não é rejeitada.

Pode-se também testar se todos os coeficientes de autocorrelação são simultaneamente iguais a zero. Para testar a hipótese conjunta, costuma-se utilizar a estatística Q de Box e Pierce ou o teste alternativo de Ljung-Box, que mostra-se mais poderoso em pequenas amostras do que a estatística Q de Box e Pierce (GUJARATI, 2000; BROCKWELL e DAVIS, 2002; EHLERS, 2009).

As estatísticas dos testes Q de Box e Pierce e Ljung-Box encontram-se brevemente descritas na Seção A.1 do Apêndice A.

- Testes para raiz unitária

A estacionariedade em uma série ainda pode ser avaliada através dos testes de raiz unitária. Estes testes verificam se o polinômio (autorregressivo e/ou de médias móveis) analisado possui uma raiz sobre o círculo unitário.

Para efeito de ilustração, seja um modelo que considera a regressão do valor de Y_t sobre o seu valor no instante imediatamente anterior ($t-1$), conhecido como autorregressivo de ordem 1 [AR(1)],

$$Y_t = \phi Y_{t-1} + \varepsilon_t, \text{ sendo } \varepsilon_t \sim RB(0, \sigma_\varepsilon^2), \quad (2.3)$$

em que, o peso ϕ , em valor absoluto, é inferior a 1 ($|\phi| < 1$).

Se a hipótese nula de existência de uma raiz unitária ($\phi = 1$) não for rejeitada, tem-se uma situação de não estacionariedade, conhecida como passeio aleatório (GUJARATI,

2000).

Na literatura especializada, o teste de raiz unitária mais conhecido é o clássico Dickey-Fuller (abreviadamente DF), proposto no trabalho de Dickey e Fuller (1979).

Todavia, vários estudos apresentam discussões sobre o desempenho do teste de Dickey-Fuller. Esses estudos sugerem que testes com hipótese nula de estacionariedade, assim como, hipótese nula de uma raiz unitária sejam utilizados antes de decidir se os dados são estacionários ou integrados (DEJONG *et al.*, 1989; DIEBOLD e RUDEBUSCH, 1990; KWIATKOWSKI *et al.*, 1992). Os autores argumentam que, em geral, os testes de raiz unitária fracassam ao não rejeitar a hipótese nula de existência de uma raiz unitária em muitas séries econômicas, sendo então duvidosa a existência de uma raiz unitária.

Nessa linha de pesquisa, ocorreram melhorias no teste de Dickey-Fuller (como o DF aumentado e testes para mais de uma raiz unitária) e propostas de testes específicos que serviriam para complementar a investigação de estacionariedade. Como é o caso, por exemplo, do teste Kwiatkowski-Phillips-Schmidt-Shin (KPSS), proposto por Kwiatkowski e colaboradores (1992).

A descrição dos testes de Dickey-Fuller e KPSS pode ser encontrada na Seção A.2 do Apêndice A.

2.3.3.2 Testes para a detecção de tendência

A presença do componente de tendência pode ser avaliada através de testes estatísticos, além, é claro, da inspeção gráfica da série temporal.

Na literatura de séries temporais existem algumas propostas de testes que investigam a presença do componente de tendência. Pode-se citar o teste de estabilidade de Mann-Kendall, sugerido por Mann (1945); os testes de aleatoriedade; o teste de Wald-Wolfowitz; o teste de Cox-Stuart; o teste de Spearman para séries temporais, entre outros.

Os testes para tendência comumente utilizados em estudos de séries temporais são os tradicionais testes de Cox-Stuart e de Spearman. Tais testes encontram-se expostos na Seção A.3 do Apêndice A.

2.3.3.3 Testes para sazonalidade

Os testes utilizados para avaliar a presença de sazonalidade são específicos para o caso determinístico, isto é, as séries que possuem sazonalidade determinística podem ser prevista perfeitamente a partir de observações anteriores.

A sazonalidade determinística pode ser testada usando testes usuais ao considerar as suposições dos modelos lineares generalizados, como podem ser visto em Draper e Smith (1998); ou fazendo uso dos testes não paramétricos, como por exemplo, o teste de Friedman (adequado para várias amostras dependentes/emparelhadas); e teste de Kruskal-Wallis (apropriado para várias amostras independentes).

Informações sobre as estatísticas dos testes de Friedman e de Kruskal-Wallis estão sintetizadas na Seção A.4 do Apêndice A.

2.3.3.4 Comentários gerais

Em geral, os testes para tendência e para sazonalidade são adaptações de testes de hipóteses clássicos; os quais possuem suposições específicas, como por exemplo, a independência entre as instâncias, que contradiz a dependência natural entre as observações temporais. Por essa razão, sugere-se utilizá-los de forma consciente; à luz da realidade dos dados, isto é, avaliando o comportamento da série de forma minuciosa.

Caso a série temporal tenha mais de um componente não aleatório, é conveniente testar a existência de um deles (tendência ou sazonalidade) após a eliminação do outro componente (MORETTIN e TOLOI, 2006).

2.4 Estratégias de modelagem e previsão clássica

Nesta seção serão abordados alguns modelos clássicos lineares utilizados na modelagem e previsão de séries temporais.

Em muitos estudos de séries temporais a previsão é, no mínimo, uma etapa necessária da análise. A previsão costuma ser realizada a partir de um modelo que detecte o padrão de comportamento da série temporal fundamentado nas suas observações passadas, distinguindo-o de qualquer ruído ou característica aleatória. Sendo assim, as previsões dos valores futuros da série dependem da escolha do modelo que melhor ajusta-se aos dados históricos.

Na literatura de séries temporais, encontram-se disponíveis duas estratégias clássicas de modelagem. Uma das estratégias é considerada a mais simples e refere-se aos modelos de suavização exponencial e a outra abordagem caracteriza a conhecida metodologia de Box-Jenkins.

2.4.1 Modelos de suavização exponencial

Os métodos de suavização (alisamento ou amortecimento) exponencial são também denominados como *ad hoc*, isto é, técnicas desenvolvidas para um fim específico e que não há fundamentação probabilística. Apesar da ausência de justificação probabilística, os métodos de alisamento exponencial possuem uma grande popularidade atribuída, principalmente, à sua simplicidade; à sua eficiência computacional e à sua razoável precisão (MORETTIN e TOLOI, 2006).

Os métodos de suavização exponencial utilizam a ideia da distribuição de pesos no decorrer dos instantes, almejando considerar ponderadores variantes no tempo.

Nas próximas subseções encontram-se brevemente descritos os modelos de suavização exponencial simples, suavização exponencial de *Holt* e suavização exponencial de *Holt-Winters*.

2.4.1.1 Suavização exponencial simples (SES)

O alisamento exponencial simples ou SES é adequado para séries temporais localmente constantes. Ou seja, séries que não apresentam tendência nem sazonalidade, compostas apenas pelo seu nível e ruído aleatório. Isto é, a série temporal pode ser decomposta como

$$Y_t = N_t + \varepsilon_t ; t = 1, 2, \dots, N,$$

em que $E(\varepsilon_t) = 0$; $Var(\varepsilon_t) = \sigma_\varepsilon^2$; N_t representa o nível da série no instante atual t . O parâmetro N_t é estimado através de uma média ponderada das observações anteriores, cujos pesos crescem exponencialmente no decorrer do tempo, ou seja, pesos relativamente maiores são atribuídos às observações mais recentes, assim como resumido na expressão abaixo

$$N_t = \alpha Y_t + \alpha(1 - \alpha) Y_{t-1} + \alpha(1 - \alpha)^2 Y_{t-2} + \alpha(1 - \alpha)^3 Y_{t-3} + \dots + \alpha(1 - \alpha)^{N-1} Y_1, \quad (2.4)$$

onde $Y_t, Y_{t-1}, Y_{t-2}, Y_{t-3}, \dots, Y_1$ são as N observações de uma série temporal, $t = 1, 2, \dots, N$. Com $0 < \alpha < 1$ representando a constante de alisamento, que deve ser escolhida pelo analista.

Todavia, na prática, utiliza-se uma expressão em função da observação no instante atual e do nível no instante imediatamente anterior; evitando o uso de todas as observações da série, fato este que torna trabalhosa a aplicação da fórmula (2.4).

Ao realizar manipulações simples de matemática na expressão (2.4), esta é transformada na conhecida forma de recorrência do algoritmo, dada por

$$N_t = \alpha Y_t + (1 - \alpha) N_{t-1}. \quad (2.5)$$

Conforme expresso em (2.5), a estimativa atual do nível é uma média ponderada entre a estimativa anterior do nível (N_{t-1}) e a observação atual da série (Y_t).

A previsão de Y_{t+h} a partir do instante t e com h passos à frente de t , é realizada da seguinte forma

$$\hat{Y}_t(h) = N_t; h = 1, 2, 3, \dots; \quad (2.6)$$

sendo que, geralmente, o algoritmo inicializa com a primeira observação ($N_1 = Y_1$).

Na literatura especializada há sugestões para a escolha da constante de alisamento (α). Sugerem-se a escolha de acordo com a experiência do analista; ou através da inspeção visual da série, especificamente com base no comportamento evolutivo dos dados; ou ainda de forma objetiva, utilizando o valor de α que fornece a melhor previsão, ou seja, aquela que minimiza a soma dos quadrados dos erros de previsão um passo à frente. Segundo Morettin e Tolo (2006), as previsões finais serão mais estáveis quanto menor for o valor de α , já que nessa condição os pesos maiores serão associados às observações passadas e desta forma flutuação aleatória, mais recente, exercerá peso menor no cálculo da previsão.

Ao realizar a inspeção visual dos dados, costuma-se utilizar um valor baixo para α se a série for bastante dinâmica, com flutuações rápidas; por outro lado, se a série evolui de forma suave utiliza-se um α alto.

Na sugestão objetiva, considera-se a minimização da soma

$$S_\alpha = \sum_{t=3}^N e_t^2; t = 3, 4, \dots, N, \quad (2.7)$$

em que $e_t = Y - \hat{Y}_{t-1}(1)$. Sendo que, a partir de (2.6) tem-se que $\hat{Y}_{t-1}(1) = N_{t-1}$, e assim

$$e_t = Y_t - N_{t-1}. \quad (2.8)$$

Considerando as equações (2.5) e (2.8) obtém a forma de correção de erros do algoritmo, dado por

$$N_t = N_{t-1} + \alpha e_t. \quad (2.9)$$

A expressão (2.9) é outra forma de obter a estimativa do nível no instante atual t . Observa-se que, agora N_t é estimado em função da estimativa do nível anterior e de um múltiplo do erro de previsão, caso e_t não seja nulo. Sendo assim, a magnitude do ajuste do algoritmo aos dados é controlada através da constante de suavização (α) (GOMES, 2003; MORETTIN e TOLOI, 2006).

2.4.1.2 Suavização exponencial de Holt (SEH)

A suavização exponencial de Holt é semelhante a suavização exponencial simples, contudo a SEH permite a modelagem de séries que apresentam, adicionalmente, tendência linear (crescente ou decrescente). A tendência não precisa ser constante ou global; esta pode também apresentar mudanças, imprevisíveis, no decorrer do tempo.

A série temporal considerada agora pode ser equacionada através da soma de nível, tendência e resíduo aleatório. Assim como apresentado na expressão a seguir

$$Y_t = N_t + m_t + \varepsilon_t; t = 1, 2, \dots, N,$$

em que, assim como em SES, ε_t possui esperança nula e variância constante. Sendo N_t o nível da série e m_t o componente de tendência no instante t .

O nível e a tendência serão estimados através da seguinte forma de recorrência do algoritmo SEH

$$\begin{cases} N_t = \alpha Y_t + (1 - \alpha)(N_{t-1} + m_{t-1}); \\ m_t = \beta(N_t - N_{t-1}) + (1 - \beta)m_{t-1}, \end{cases}$$

em que $0 < \alpha < 1$ e $0 < \beta < 1$ são as duas constantes de suavização do algoritmo.

A previsão de Y_{t+h} , feita no instante t , é dada por

$$\hat{Y}_t(h) = N_t + hm_t; h = 1, 2, 3, \dots \quad (2.10)$$

Tipicamente, o algoritmo inicia com a segunda observação; especificamente $N_2 = Y_2$ e $m_2 = Y_2 - Y_1$.

De maneira análoga ao SES, pode-se obter a forma de correção dos erros do algoritmo de Holt, que é apresentada como

$$\begin{cases} N_t = N_{t-1} + m_{t-1} + \alpha e_t; \\ m_t = m_{t-1} + \alpha\beta e_t. \end{cases}$$

Com relação a escolha das constantes de suavização, esta pode ser realizada de forma semelhante à apresentada para a SES. Contudo, agora será preciso considerar a constante adicional β . Maiores detalhes podem ser encontrados no artigo do Winters (1960). Agora, a magnitude do ajuste é controlada pelas duas constantes de suavização (α e β) (GOMES, 2003; MORETTIN e TOLOI, 2006).

2.4.1.3 Suavização exponencial de Holt-Winters (H-W)

O algoritmo de Holt-Winters é útil para séries temporais com padrão de comportamento mais complexo, que apresentam sazonalidade, além dos outros componentes já contemplados pelos métodos anteriores (SES e SEH). Desta forma, três equações são consideradas, em que cada uma delas corresponde ao tratamento realizado para cada componente observável, a saber: nível, tendência e sazonalidade.

A incorporação do comportamento sazonal no algoritmo HW pode ser realizada através de dois enfoques distintos, dependentes do padrão sazonal identificado na série, são eles: sazonalidade multiplicativa ou sazonalidade aditiva.

- *Série com sazonalidade multiplicativa*

Ao considerar a sazonalidade multiplicativa, a série temporal pode ser representada como

$$Y_t = N_t S_t + m_t + \varepsilon_t; t = 1, 2, \dots, N, \quad (2.11)$$

sendo N_t o nível da série; S_t o fator sazonal; m_t o componente de tendência; e ε_t o erro aleatório no instante t .

A forma de recorrência do método HW, com fator sazonal multiplicativo, é dada pelas seguintes equações envolvendo as três constantes de suavização

$$\begin{cases} N_t = \alpha \frac{y_t}{S_{t-s}} + (1 - \alpha) (N_{t-1} + m_{t-1}) & ; \\ m_t = \beta (N_t - N_{t-1}) + (1 - \beta) m_{t-1} & ; \\ S_t = \gamma \frac{y_t}{N_t} + (1 - \gamma) S_{t-s} & , \end{cases}$$

em que $0 < \alpha < 1$, $0 < \beta < 1$ e $0 < \gamma < 1$ são as constantes de suavização do algoritmo. E s representa o período de sazonalidade da série, onde para séries mensais s indica o número de observações por ano.

As previsões dos valores futuros levam em consideração o período de alcance, ou seja, em cada equação o fator sazonal contabiliza o período correspondente, assim como resumido nas expressões abaixo

$$\hat{Y}_t(h) = (N_t + hm_t) S_{t+h-s}; h = 1, 2, 3, \dots, s$$

$$\hat{Y}_t(h) = (N_t + hm_t) S_{t+h-2s}; h = s + 1, s + 2, s + 3, \dots, 2s$$

$$\hat{Y}_t(h) = (N_t + hm_t) S_{t+h-3s}; h = 2s + 1, 2s + 2, 2s + 3, \dots, 3s$$

$$\vdots$$

Para o presente enfoque, a forma de correção dos erros é dada por

$$\left\{ \begin{array}{l} N_t = N_{t-1} + m_{t-1} + \alpha \frac{e_t}{F_{t-s}} \quad ; \\ m_t = m_{t-1} + \alpha \beta \frac{e_t}{F_{t-s}} \quad ; \\ S_t = S_{t-s} + \gamma(1 - \alpha) \frac{e_t}{N_t} \quad . \end{array} \right.$$

- *Série com sazonalidade aditiva*

O outro enfoque do método de HW é aplicado quando a série possui sazonalidade da forma aditiva.

Tomando como base o fator sazonal aditivo, a série temporal fica representada pela soma de todos os componentes, como pode ser visto a seguir

$$Y_t = N_t + m_t + S_t + \varepsilon_t; t = 1, 2, \dots, N. \quad (2.12)$$

Para o caso de sazonalidade aditiva, a atualização dos dados ocorre através das seguintes equações

$$\left\{ \begin{array}{l} N_t = \alpha (Y_t - S_{t-s}) + (1 - \alpha) (N_{t-1} + m_{t-1}) \quad ; \\ m_t = \beta (N_t - N_{t-1}) + (1 - \beta) m_{t-1} \quad ; \\ S_t = \gamma (Y_t - N_t) + (1 - \gamma) S_{t-s} \quad , \end{array} \right.$$

em que, novamente, $0 < \alpha < 1$; $0 < \beta < 1$; $0 < \gamma < 1$ são as constantes de suavização do algoritmo e s indica a periodicidade da série.

Os valores futuros são previstos considerando as equações a seguir:

$$\widehat{Y}_t(h) = (N_t + hm_t) + S_{t+h-s} ; h = 1, 2, 3, \dots, s$$

$$\widehat{Y}_t(h) = (N_t + hm_t) + S_{t+h-2s} ; h = s + 1, s + 2, s + 3, \dots, 2s$$

$$\widehat{Y}_t(h) = (N_t + hm_t) + S_{t+h-3s} ; h = 2s + 1, 2s + 2, 2s + 3, \dots, 3s$$

⋮

E o procedimento de correção dos erros, para o caso de sazonalidade aditiva, passa a ser denotado por

$$\begin{cases} N_t = N_{t-1} + m_{t-1} + \alpha e_t & ; \\ m_t = m_{t-1} + \alpha \beta e_t & ; \\ S_t = S_{t-s} + \gamma(1 - \alpha)e_t & . \end{cases}$$

2.4.1.4 Comentários gerais

Os algoritmos SES, SEH e HW possuem uma estrutura de autoaprendizagem a partir dos erros de previsão cometidos, isto é, estes algoritmos aprendem com os seus erros. Ao avaliar a estrutura de autoaprendizagem desses algoritmos, o mecanismo pode ser generalizado da seguinte forma:

- * Se $e_t = 0$, então a previsão no instante anterior foi perfeita e a estimativa (nível, tendência e/ou sazonalidade) subsequente à t será mantida.
- * Se $e_t > 0$, então ocorreu subprevisão (subestimação) na última previsão, e assim a próxima previsão ficará acrescida por um múltiplo do último erro de previsão.
- * Se $e_t < 0$, então a última previsão foi maior do que o observado na série (superestimação), e portanto, a previsão posterior será diminuída por um múltiplo do último erro de previsão.

Vale salientar que, o uso adequado do algoritmo de Holt-Winters refere-se à existência de sazonalidade, independentemente da presença da tendência. Assim como acontece com o algoritmo de Holt, onde a presença de tendência linear na série é suficiente para a aplicação de tal técnica.

Em linhas gerais, as técnicas SES, SEH e HW são de fácil entendimento, aplicação e implementação computacional. Contudo, esses algoritmos apresentam dificuldades na determinação do valor mais adequado para as constantes de suavização.

2.4.2 Modelos de Box e Jenkins

A metodologia de Box e Jenkins considera os modelos autorregressivos integrados de médias móveis (ARIMA, sigla oriunda do termo *AutoRegressive Integrate Moving Average*) e suas extensões. Tal metodologia fundamenta-se na construção de modelos ajustados aos dados à luz das suas propriedades probabilísticas. Isto é, o método possibilita realizar a descrição e previsão em séries temporais estacionárias, podendo ter o seu uso estendido para séries não estacionárias homogêneas.

A construção dos modelos na metodologia de Box e Jenkins segue um ciclo iterativo que possibilita identificar o processo estocástico gerador dos dados, assim como os seus parâmetros. O ciclo consiste de quatro estágios, são eles:

1. A especificação dos modelos a serem considerados na análise de acordo com as características existentes na série temporal. Ou seja, uma classe de modelos é proposta, por exemplo, a classe dos modelos autorregressivos (AR);
2. A identificação de um particular modelo e seus parâmetros;
3. A estimação dos parâmetros do modelo identificado na etapa anterior com base nos dados. As variações dos métodos de máxima verossimilhança e dos mínimos quadrados são os métodos de estimação usuais;
4. A checagem ou diagnóstico do modelo ajustado. Estágio em que verifica se o modelo retrata satisfatoriamente a dinâmica dos dados. Usualmente, investiga-se as estimativas dos erros, os resíduos, com o intuito de avaliar se estes são adequados para o objetivo principal da análise. Por exemplo, quando a previsão é o propósito principal da análise espera-se que os resíduos sejam ruídos brancos. Para este fim, pode-se utilizar os testes para autocorrelação residual, semelhantes aos apresentados na Subseção 2.3.3; sendo que agora analisa-se a autocorrelação entre os resíduos ao invés da autocorrelação entre as observações da série temporal.

Caso o modelo selecionado seja adequado à realidade dos dados, este será usado no processo de previsão da série temporal. Por outro lado, se o ajuste realizado não for satisfatório, reinicia-se o ciclo a partir do segundo estágio ou fase, esta é considerada a fase crítica da metodologia. Isso porque, há a possibilidade de que diferentes analistas

identifiquem modelos distintos ao investigar a mesma série temporal (GUJARATI, 2000; MORETTIN e TOLOI, 2006).

De uma forma geral, os analistas utilizam modelos com poucos parâmetros, mas com uma representação satisfatória da estrutura dos dados; prevalecendo na modelagem o princípio da parcimônia, isto é, modelos mais simples costumam ser melhores.

Na fase de identificação, a escolha do modelo é realizada através das autocorrelações e autocorrelações parciais estimadas; as quais devem representar as funções teóricas de autocorrelação e autocorrelação parcial, fac e facp , respectivamente. Além dos métodos alternativos que utilizam, por exemplo, funções que penalizam o ajuste de modelos com muitos parâmetros ou não parcimoniosos, tais como: Critério de Informação de Akaike (AIC) (AKAIKE, 1973, 1974); Critério Bayesiano (BIC) (AKAIKE, 1977); Critério de Hannan e Quinn (HANNAN e QUINN, 1979); entre outros. Para o presente trabalho utilizou-se os critérios AIC e BIC por estes serem comumente aplicados na seleção de modelos paramétricos.

Ao utilizar o critério AIC, Akaike (1973, 1974) indica a seleção de um modelo cujos parâmetros (p e q) minimizam a função

$$AIC = -2\ln(L) + 2(p + q),$$

onde L é a verossimilhança maximizada. De forma semelhante, Akaike (1977) sugere minimizar a função dada por

$$BIC = -2\ln(L) + (p + q)\ln(N),$$

em que, N é o tamanho da série.

A metodologia de Box e Jenkins foi desenvolvida para séries temporais estacionárias ou que tornam estacionárias após transformações na variável, como a aplicação de diferenças nos dados. Tal suposição é justificável, pois qualquer modelo obtido a partir de séries com características constantes no tempo pode ser interpretado como estável ou estacionário, o que valida as observações que servirão como base para as previsões futuras (POKORNY, 1987 *apud*. GUJARATI, 2000).

Antes de iniciar a descrição dos modelos de Box e Jenkins, faz-se necessário apresentar a notação de operadores que será utilizada nas próximas subseções.

Um operador bastante útil para a representação dos modelos refere-se ao operador

translação para o passado, ou simplesmente, operador defasagem. O operador defasagem será denotado por

$$BY_t = Y_{t-1}; B^2Y_t = Y_{t-2}; B^3Y_t = Y_{t-3}; \dots; B^mY_t = Y_{t-m}.$$

O operador diferença, apresentado na Seção 2.3.1, será denotado por

$$\Delta Y_t = Y_t - BY_t = (1 - B)Y_t; \Delta^2 Y_t = (1 - B)^2 Y_t; \dots; \Delta^d Y_t = (1 - B)^d Y_t,$$

em que d indica o número de diferenças realizadas na série temporal original.

A seguir serão apresentados os modelos lineares estacionários, autorregressivos (AR); médias móveis (MA); autorregressivos e de médias móveis (ARMA), e os modelos não estacionários ARIMA e SARIMA.

2.4.2.1 Modelos autorregressivos (AR)

Considere $\{Y_t\}$ uma série temporal com N observações. O modelo autorregressivo de ordem p , AR(p), é expresso como

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t, \quad (2.13)$$

em que ε_t representa o erro aleatório que não pode ser explicado pelo modelo e supõe ser um ruído branco [$\varepsilon_t \sim RB(0, \sigma_\varepsilon^2)$]; $\phi_1, \phi_2, \dots, \phi_p$ são parâmetros autorregressivos do modelo AR; e c é uma constante, que não é a média.

Ao ser aplicado o operador defasagem em (2.13) e evidenciando Y_t , resultará na seguinte expressão resumida do modelo AR

$$\phi(B)Y_t = c + \varepsilon_t, \quad (2.14)$$

em que $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ é o polinômio operador de defasagens ou operador autorregressivo de ordem p .

A esperança e a j -ésima autocovariância de um processo $AR(p)$ são dadas, respectivamente, por

$$E(Y_t) = \frac{c}{1 - \phi_1 - \dots - \phi_p};$$

$$\gamma_j = \phi_1\gamma_{j-1} + \phi_2\gamma_{j-2} + \dots + \phi_p\gamma_{j-p}, \gamma_j = \gamma_{-j}, j > 0.$$

Ao dividir γ_j por $\gamma_0 = Var(Y_t)$, resulta na função autocorrelação (fac)

$$\rho_j = \phi_1\rho_{j-1} + \phi_2\rho_{j-2} + \dots + \phi_p\rho_{j-p}, j > 0; \quad (2.15)$$

em que

$$Var(Y_t) = \frac{\sigma_\varepsilon^2}{1 - \phi_1\rho_1 - \phi_2\rho_2 - \dots - \phi_p\rho_p}.$$

A função autocorrelação (fac) teórica de um processo autorregressivo decai segundo exponenciais e/ou senóides amortecidas, e é infinita em extensão. E as autocorrelações parciais de um $AR(p)$ têm as p primeiras autocorrelações estatisticamente diferentes de zero. Ou seja, a função autocorrelação parcial (facp) de um processo autorregressivo apresenta um corte na correspondente ordem do modelo.

Um processo autorregressivo de ordem p é dito estacionário se todas as raízes do polinômio $\phi(B)$ estiverem fora do círculo unitário, que pode ser denotado por $a^2 + b^2 = 1$. A estacionariedade é determinada pelos valores dos parâmetros autorregressivos. Por exemplo, um $AR(1)$ será estacionário se $-1 < \phi_1 < 1$ (ou $|\phi_1| < 1$); e para o modelo $AR(2)$ a região de estacionariedade corresponde a

$$-1 < \phi_2 < 1, \phi_1 + \phi_2 < 1 \text{ e } \phi_1 - \phi_2 < 1.$$

2.4.2.2 Modelo de média móvel (MA)

Considere o processo $\{Y_t\}$. Tal processo será reconhecido como modelo de média móvel de ordem q , denotado por $MA(q)$, se poder ser expresso da forma

$$Y_t = \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \dots - \theta_q\varepsilon_{t-q}; \quad (2.16)$$

sendo $[\varepsilon_t \sim RB(0, \sigma_\varepsilon^2)]$; μ a média do processo e $\theta_1, \theta_2, \dots, \theta_q$ são os parâmetros de médias móveis.

A equação (2.16) pode ser reestruturada como

$$Y_t = \mu + \theta(B)\varepsilon_t, \quad (2.17)$$

em que $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ é o polinômio operador de médias móveis de ordem q .

Os processos de médias móveis são caracterizados pelas seguintes estatísticas

$$E(Y_t) = \mu;$$

$$E[(Y_t - \mu)(Y_{t-j} - \mu)] = \gamma_j = \begin{cases} 1 + (\theta_1^2 + \theta_2^2 + \dots + \theta_q^2) \sigma_\varepsilon^2, & j = 0; \\ (-\theta_j + \theta_1 \theta_{j+1} + \theta_2 \theta_{j+2} + \dots + \theta_q \theta_{q-j}) \sigma_\varepsilon^2, & j = 1, \dots, q; \\ 0, & j > q; \end{cases}$$

em que $E(Y_t)$ representa a média do processo; γ_j a função de autocovariância (facv), onde γ_0 é a variância de Y_t . Novamente, ao dividir γ_j por γ_0 , a variância da série, tem-se a função de autocorrelação (fac) e esta pode ser representada por

$$\rho_j = \begin{cases} 1, & j = 0; \\ \frac{-\theta_j + \theta_1 \theta_{j+1} + \theta_2 \theta_{j+2} + \dots + \theta_q \theta_{q-j}}{1 + (\theta_1^2 + \theta_2^2 + \dots + \theta_q^2)}, & j = 1, \dots, q; \\ 0, & j > q. \end{cases} \quad (2.18)$$

A ordem de um modelo de médias móveis é indicada através da função de autocorrelação (fac) ao invés da função autocorrelação parcial (facp), como nos modelos autorregressivos. De acordo com a função de autocorrelação do MA, pode-se verificar que até o lag q as autocorrelações são não nulas, indicando que a fac de um MA (q) é finita. Já a facp de um MA(q) comporta-se de forma semelhante à fac de um AR(p).

Para exemplificar o comportamento da função autocorrelação (fac) e autocorrelação parcial (facp) de processos estocásticos puros, simulou-se duas séries temporais, com 200 observações cada, conforme um processo de médias móveis de ordem 1, assim como um modelo autorregressivo de ordem 1. Tais aplicações encontram-se ilustradas no Apêndice B.

Os processos de médias móveis são sempre estacionários, isto é, os modelos MA não possuem restrições nos seus parâmetros. Porém, esses modelos não têm unicidade. Para que a unicidade possa ser garantida, as restrições com relação à invertibilidade do processo devem ser consideradas.

Um processo MA(q) é invertível se todas as raízes do polinômio $\theta(B)$ estiverem fora do círculo unitário. Para exemplificar, considere um MA(2); este processo será invertível se as seguintes condições forem satisfeitas

$$-1 < \theta_2 < 1, \theta_1 + \theta_2 < 1 \text{ e } \theta_1 - \theta_2 < 1,$$

o que equivale às condições de estacionariedade para um AR(2).

2.4.2.3 Modelo autorregressivo de média móvel (ARMA)

Um processo de médias móveis puro é pouco intuitivo para representar o comportamento de uma particular série temporal, enquanto que a aplicação dos modelos autorregressivos é bastante natural em muitas áreas do conhecimento. Na prática, considera-se os termos autorregressivos e de médias móveis simultaneamente, buscando a melhoria no ajuste ao considerar poucos parâmetros. Essa combinação de termos caracteriza os modelos autorregressivos e de médias móveis, denotado por ARMA(p, q).

Um processo ARMA(p, q) apresenta a forma

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}; \quad (2.19)$$

sendo $[\varepsilon_t \sim RB(0, \sigma_\varepsilon^2)]$.

A expressão de um ARMA(p, q) também pode ser reescrita, considerando os operadores autorregressivo e de médias móveis usuais, o que resulta na forma compacta de (2.19)

$$\phi(B)Y_t = c + \theta(B)\varepsilon_t, \quad (2.20)$$

As condições de estacionariedade em um processo ARMA(p, q) são as mesmas de um processo AR(p) e as restrições para que haja invertibilidade permanecem idênticas a de um MA(q).

A função autocovariância (facv) de um processo ARMA(p, q) é a mesma de um AR(p) para defasagens ou ‘lags’ maiores que q , esta costuma ser representada por

$$\gamma_j = \phi_1\gamma_{j-1} + \phi_2\gamma_{j-2} + \cdots + \phi_p\gamma_{j-p}, \quad j > q. \quad (2.21)$$

E a partir de (2.21) obtém a função autocorrelação (fac) de um ARMA(p, q), que é dada por

$$\rho_j = \phi_1\rho_{j-1} + \phi_2\rho_{j-2} + \cdots + \phi_p\rho_{j-p}, \quad j > q. \quad (2.22)$$

Para “lags” menores ou iguais a q , deduz-se que as autocorrelações são afetadas pelos parâmetros de médias móveis. É possível verificar que, se o processo ARMA for formado por mais parâmetros autorregressivos do que de médias móveis ($q < p$), a fac comporta-se como exponenciais e/ou senóides amortecidas após o lag $q - p$; todavia, se $q = p$, as autocorrelações iniciais $\rho_0, \rho_1, \dots, \rho_{q-p}$ não seguirão este padrão (MORETTIN e TOLOI, 2006).

2.4.2.4 Modelo autorregressivo integrado de média móvel (ARIMA)

A classe dos modelos autorregressivos integrados de médias móveis, ARIMA(p, d, q), é uma generalização da classe de modelos ARMA. Os modelos ARIMA permitem trabalhar com séries temporais não estacionárias, porém considera-se que o processo gerador dos dados seja homogêneo ou não explosivo.

Seja d um inteiro não negativo e Z_t um processo estacionário, então $\{Y_t\}$ é um processo ARIMA (p, d, q) se

$$Z_t = \Delta^d Y_t = (1 - B)^d Y_t$$

para um processo ARMA(p, q). Em outras palavras, os processos ARIMA (p, d, q) reduzem a processos ARMA(p, q), estacionários e invertíveis, ao serem diferenciados um número finito de vezes. O modelo ARIMA (p, d, q) é também conhecido como um processo integrado de ordem d , e neste contexto, um processo $\{Y_t\}$ é estacionário se e somente se $d = 0$.

Costuma-se definir um processo ARIMA de ordem (p, d, q) através do operador autorregressivo, operador diferença e o operador de médias móveis, assim como é apresentado a seguir:

$$\phi(B) \left[(1 - B)^d Y_t - \mu \right] = \theta(B) \varepsilon_t, \quad (2.23)$$

em que μ representa o valor esperado da série com d diferenças; p , d e q as ordens dos operadores $\phi(B)$, $\Delta^d Y_t$ e $\theta(B)$, respectivamente.

Uma expressão equivalente a (2.23) pode ser obtida ao utilizar um operador autorregressivo não estacionário $\varphi(B)$, assim como apresentado abaixo

$$\varphi(B) Y_t = \theta(B) \varepsilon_t, \quad (2.24)$$

em que $\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_{p+d} B^{p+d}$ é um polinômio de ordem $p + d$, com d raízes iguais a 1 e as demais p raízes fora do círculo unitário (MORETTIN e TOLOI, 2006).

A forma extensa do modelo ARIMA, útil para a obtenção das previsões, é representada pela equação

$$Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_{p+d} Y_{t-p-d} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}, \quad (2.25)$$

sendo $[\varepsilon_t \sim RB(0, \sigma_\varepsilon^2)]$ e $\varphi(B)$ o operador autorregressivo não estacionário.

2.4.2.5 Modelo autorregressivo integrado de média móvel sazonal (SARIMA)

Em muitos fenômenos físicos, biológicos e econômicos a variabilidade natural dos dados coincide com as flutuações sazonais. Já que, em geral, nesses processos observa-se uma forte dependência entre as observações em defasagens múltiplas do período sazonal s , também conhecido como *lags* sazonais ou defasagens sazonais. Por exemplo, os fenômenos meteorológicos tomados trimestralmente apresentam correlações mais elevadas em “*lags*” múltiplos de $s = 4$, o que corresponde às estações do ano. Já em dados econômicos mensais, ocorrendo repetição no período anual de $s = 12$, espera-se uma dependência maior entre as instâncias em múltiplos de 12, devido às atividades relacionadas ao ano civil.

Diante da situação exposta no parágrafo anterior, é adequado considerar uma sazonalidade estocástica para explicar o comportamento sazonal da série. Isto é, recomenda-se trabalhar com uma das variações do modelo ARIMA, nomeado como ARIMA sazonal multiplicativo ou simplesmente SARIMA.

O modelo geral sazonal multiplicativo, denotado por ARIMA $(p, d, q) \times (P, D, Q)$, pode ser representado como

$$\phi(B) \Phi_P(B^s) [\Delta^d \Delta_s^D - \mu] Y_t = \theta(B) \Theta_Q(B^s) \varepsilon_t; \quad (2.26)$$

sendo $[\varepsilon_t \sim RB(0, \sigma_\varepsilon^2)]$. Onde são utilizados os polinômios usuais definidos nas três últimas subseções $\phi(B)$, Δ^d e $\theta(B)$; o polinômio estacionário autoregressivo sazonal de ordem P

$$\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps};$$

o polinômio invertível de médias móveis sazonal de ordem Q

$$\Theta_Q(B^s) = 1 - \theta_1 B^s - \theta_2 B^{2s} - \dots - \theta_Q B^{Qs};$$

e o operador diferença sazonal de ordem D

$$\Delta_s^D = (1 - B^s)^D Y_t;$$

em que, em geral, a primeira diferenciação sazonal $\Delta_s = (1 - B^s) Y_t = Y_t - Y_{t-s}$ consegue eliminar a sazonalidade da série temporal.

Um clássico modelo SARIMA refere-se ao *airline model* ou SARIMA $(0, 1, 1) \times (0, 1, 1)$. Tal modelo foi inicialmente aplicado por Box e Jenkins (1970), que modelaram o logaritmo da quantidade mensal de passageiros em companhias aéreas. Posteriormente, observou-se que este modelo apresentava potencial para explicar o comportamento de outras séries sazonais (GOMES, 2003; MORETTIN e TOLOI, 2006).

O processo de modelagem em modelos SARIMA ocorre de forma análoga ao já apresentado. Contudo, as fases de especificação e identificação do ciclo da metodologia Box e Jenkins ficam mais extensas, já que agora será preciso também indicar o número de diferenças sazonais (D); e inspecionar a *fac* e *facp* nas defasagens sazonais para a obtenção de P e Q, além da inspeção nos demais “*lags*” para a identificação de p e q .

Detalhes sobre os modelos sazonais podem ser encontrados em Brockwell e Davis (2002); Morettin e Tolo (2006); e Shumway e Stoffer (2006).

2.4.2.6 Previsão

As previsões em modelos ARIMA podem ser obtidas usando a equação do correspondente modelo identificado. De posse da equação do modelo identificado e estimado, obtém-se a predição Y_{t+h} , com $h = 1$, substituindo os valores futuros da série

$$Y_{N+1}, Y_{N+2}, Y_{N+3}, \dots$$

pela sua esperança condicionada aos valores passados, ou seja,

$$Y_{t+h} = E(Y_{t+h} | Y_t, Y_{t-1}, \dots, Y_1);$$

trocando os valores futuros dos erros por zero; e os valores passados de Y e dos erros pelos seus valores observados.

Comumente, utiliza-se a conveniente notação $\hat{Y}_t(h)$ para representar a previsão de origem t e horizonte h . Toda a teoria desenvolvida para a previsão supõe que a série temporal ou o processo integrado seja estacionário e invertível.

Em modelos autorregressivos de ordem p [AR(p)] a função de previsão pode ser representada como

$$\begin{aligned}
\widehat{Y}_t(1) &= \alpha_1 Y_t + \alpha_2 Y_{t-1} + \cdots + \alpha_p Y_{t-p+1}; \\
\widehat{Y}_t(2) &= \alpha_1 \widehat{Y}_t(1) + \alpha_2 Y_t + \cdots + \alpha_p Y_{t-p+2}; \\
\widehat{Y}_t(3) &= \alpha_1 \widehat{Y}_t(2) + \alpha_2 \widehat{Y}_t(1) + \cdots + \alpha_p Y_{t-p+3}; \\
&\vdots \\
\widehat{Y}_t(p+1) &= \alpha_1 \widehat{Y}_t(p) + \alpha_2 \widehat{Y}_t(p-1) + \cdots + \alpha_p \widehat{Y}_t(1),
\end{aligned} \tag{2.27}$$

em que $\alpha_1, \alpha_2, \dots, \alpha_p$ são constantes reais. Assim como pode ser visto na última equação de (2.27), caso o horizonte seja superior a ordem p , as previsões levarão em consideração apenas as previsões nos passos anteriores. Situação que deve ser evitada para que não haja acúmulo dos erros de previsão.

Nos modelos de médias móveis de ordem q pode-se representar as expressões recorrentes como

$$\begin{aligned}
\widehat{Y}_t(1) &= \beta_1 \varepsilon_t + \beta_2 \varepsilon_{t-1} + \cdots + \beta_q \varepsilon_{t-q+1}; \\
\widehat{Y}_t(2) &= \beta_2 \varepsilon_t + \beta_3 \varepsilon_{t-1} + \cdots + \beta_q \varepsilon_{t-q+2}; \\
\widehat{Y}_t(3) &= \beta_3 \varepsilon_t + \beta_4 \varepsilon_{t-1} + \cdots + \beta_q \varepsilon_{t-q+3}; \\
&\vdots \\
\widehat{Y}_t(q) &= \beta_q \varepsilon_t; \\
\widehat{Y}_t(q+j) &= 0, \quad j=1.
\end{aligned} \tag{2.28}$$

Sendo assim, a previsão é nula para qualquer horizonte maior do que a ordem q do modelo de médias móveis.

Extensões das funções de previsão podem ser realizadas para os modelos autorregressivos integrados e de médias móveis, assim como para os modelos sazonais.

Vale salientar que, as previsões de curto prazo (ou horizonte pequeno), são mais precisas do que às de longo prazo; pois ao considerar σ_ε^2 fixo, a variância do erro de

previsão cresce conforme aumenta o horizonte de previsão h , isto é, ao passo que ocorre o afastamento da origem t (MORETTIN e TOLOI, 2006; EHLERS, 2009).

Com relação à qualidade das previsões oriundas de métodos distintos, Newbold e Granger (1974) *apud*. Gomes (2003) comentam que as previsões através dos modelos ARIMA são mais precisas do que as obtidas pelos algoritmos de alisamento exponencial, quando se considera poucos passos à frente. Para previsões mais distantes, observou-se equivalência nos métodos.

Maiores detalhes sobre a previsão na metodologia de Box e Jenkins podem ser encontrados nos livros de Brockwell e Davis (2002); Morettin e ToloI (2006); e Shumway e Stoffer (2006).

2.4.2.7 Comentários gerais

Em suma, os modelos ARIMA consideram que o valor atual da série temporal é a combinação de $p-1$ observações defasadas, $q-1$ termos aleatórios defasados e um impacto aleatório atual (ε_t), modelando as partes autorregressivo e de médias móveis em séries não estacionárias homogêneas. Em séries estacionárias, o ajuste é realizado através da classe dos modelos AR, MA ou ARMA.

A utilização do método de Box e Jenkins exige conhecimento especializado sobre as propriedades das classes de modelos, além da experiência do analista, principalmente, no que tange a identificação do modelo mais apropriado aos dados. Fatos que podem ser considerados desvantagens da metodologia de Box e Jenkins.

2.4.3 Medidas de erro

A estimação dos valores futuros está susceptível a ocorrência de erros, independente da estratégia de previsão adotada. Como os verdadeiros valores futuros são desconhecidos, não há como quantificar a precisão dos valores previstos. Contudo, costuma-se avaliar o melhor método de previsão ao retirar as n últimas observações da série temporal e compará-las, através de uma medida de erro, com os correspondentes valores preditos a partir do modelo estimado.

A comparação entre as estratégias de previsão abordadas neste trabalho foi realizada utilizando as conhecidas medidas de erro: erro quadrático médio (*MSE*, do inglês *Mean Square Error*) e erro percentual total (*TPE*, do inglês *Total Percentage Error*).

O erro quadrático médio leva em consideração a quantidade de observações na série completa (T), além dos erros cometidos na estimação, assim como apresentado na expressão abaixo:

$$MSE = \frac{1}{n} \sum_{j=k+1}^T e_j^2 \quad (2.29)$$

em que $e_j = Y_j - \widehat{Y}_j$ são os erros de previsão; \widehat{Y}_j a previsão de Y_j ; n é a quantidade de observações reservadas para a avaliação da capacidade preditiva, comumente identificado como conjunto de teste, e $k = T - n$.

No erro percentual total, acrescenta-se os valores observados Y_j com o intuito de obter uma medida relativa do erro. O TPE pode ser representado como

$$TPE = \left(\frac{\sum_{j=k+1}^T e_j}{\sum_{j=k+1}^T Y_j} \right) \times 100. \quad (2.30)$$

O capítulo seguinte apresenta a descrição da técnica *Singular Spectrum Analysis* (SSA). As definições básicas da Álgebra linear e da Estatística Multivariada necessárias para um bom entendimento da metodologia da SSA encontram-se descritas no Apêndice C.

Singular Spectrum Analysis

3.1 Introdução

A Análise Espectral Singular (SSA, do inglês *singular spectrum analysis*) é mencionada como uma técnica moderna e poderosa para análise de séries temporais, que incorpora as fundamentações teóricas da análise clássica de série temporal, da análise multivariada, da dinâmica de sistemas e do processamento de sinal (ELSNER e TSONIS, 1996; GOLYANDINA, NEKRUTKIN e ZHIGLJAVSKY, 2001).

O termo *singular spectrum analysis* é pouco intuitivo, uma vez que, este termo não transmite a ideia das múltiplas técnicas incorporadas a SSA. Além de não caracterizar exclusivamente a técnica, já que, segundo Elsner e Tsonis (1996), a decomposição utilizada em dados multivariados também representa uma análise do espectro singular. Esses autores citam um nome mais sugestivo para a técnica, apesar de longo, a saber: *análise de séries temporais utilizando o espectro singular*.

Na literatura sobre a ferramenta de análise SSA são utilizadas outras terminologias. Dentre essas terminologias inclui-se: *Caterpillar*, *singular systems analysis*, *singular spectrum approach*, expansão de *Karhunen-Loève* e *empirical orthogonal function* (EOF). Contudo, para o presente trabalho a terminologia mais usual é considerada; doravante *singular spectrum analysis* ou, simplesmente, SSA.

Dois dos principais autores da literatura especializada, Elsner e Tsonis (1996), apresentam uma breve discussão sobre a etimologia do nome da técnica SSA. O termo *singular spectrum* tem sua origem na decomposição espectral de uma matriz \mathbf{X} em seu conjunto de autovalores λ , conjunto este identificado por *spectrum*. Enquanto que, o termo *singular* remete a matriz singular $(\mathbf{X} - \lambda\mathbf{I})$.

A ideia da SSA fundamenta-se na decomposição da série temporal em seus componentes constituintes, que são independentes (ou separáveis) e interpretáveis. Como por exemplo, a decomposição aditiva envolvendo o componente de tendência, componentes periódicos e estruturas de ruído ou resíduo. O termo *separável* associa-se com uma das propriedades da SSA, a separabilidade, que reporta à capacidade de separação dos componentes da série.

A técnica SSA dispensa o conhecimento sobre o modelo paramétrico da série temporal, sendo considerada também como um método de estimação espectral não paramétrico. Na

aplicação e investigação das propriedades da técnica, não são exigidas suposições estatísticas relativas ao sinal ou ruído. Em outras palavras, qualquer série, simples ou complexa, com alguma potencial estrutura pode ser analisada através da SSA. A SSA também pode ser aplicada em séries com uma quantidade relativamente pequena de observações (GOLYANDINA, NEKRUTKIN e ZHIGLJAVSKY, 2001; HASSANI, 2007).

A Análise Espectral Singular possui basicamente duas finalidades, são elas:

- Modelagem ou ajuste dos dados: destina-se a modelar a dependência entre as observações da série temporal. Para este objetivo da SSA, o propósito principal é o entendimento do comportamento da série, ou seja, é parte responsável pela identificação ou extração de padrões geradores da série temporal.
- Previsão: objetiva realizar previsões de valores futuros da série temporal; após o processo de modelagem dos dados, assim como ocorre nos métodos clássicos de previsão.

Em linhas gerais, o algoritmo da SSA particiona uma série temporal em séries defasadas e posteriormente, ao fazer uso de conceitos e manipulações matemáticas, obtém-se uma aproximação da série inicial. Nas seções subsequentes, as metodologias relativas à SSA básica e à previsão são descritas; assim como os conceitos importantes relacionados a cada passo da técnica.

3.1.1 Estado da arte

A Análise Espectral Singular tem sua origem associada com os trabalhos de Broomhead e King (1986a, 1986b) e Broomhead e colaboradores (1987), os quais mostravam que a decomposição do valor singular é eficaz na redução de ruídos (Myung, 2009). Enquanto isso, as ideias do SSA se desenvolviam de forma independente na Rússia (St. Petersburg, Moscou), com o nome de *Caterpillar*; e em vários grupos do Reino Unido (UK) e dos EUA. Contudo, as documentações com aspectos metodológicos e aplicações da técnica foram consolidadas a partir de 1996 com a publicação do livro *Singular Spectral Analysis. A New Tool in Time Series Analysis* de autoria de Elsner e Tsonis. Em termos de *papers* incluindo aspectos metodológicos e práticos, são conhecidas mundialmente, no estado da arte, aplicações da SSA nas áreas de meteorologia e geofísica, como as encontradas em Allen e Smith (1996); Ghil e Taricco (1997); Danilov e Zhigljavsky (1997); Yiou, Sornette e Ghil (2000) (GOLYANDINA, NEKRUTKIN e ZHIGLJAVSKY, 2001; HASSANI, 2007).

Posteriormente, em 2001, publicou-se o primeiro livro sobre a SSA com enfoque Estatístico, intitulado *Analysis of time series structure: SSA and related techniques*, tendo

Nina Golyandina como autora principal. Tal livro resume os resultados publicados sobre a técnica SSA nos 15 anos anteriores à sua publicação, sendo uma fonte que contém a teoria matemática formal descrita de forma clara e concisa (GOLYANDINA, NEKRUTKIN e ZHIGLJAVSKY, 2001).

Atualmente, a aplicação da SSA tem sido bem sucedida em muitas áreas do conhecimento; como por exemplo, na economia e na pesquisa de mercado. Por outro lado, esta ferramenta já tornou-se padrão na análise de séries temporais meteorológicas e geofísicas; e como esperado as revistas internacionais *Journal of Climate*, *Journal of the Atmospheric Sciences* e *Journal of Geophysical Research* são as principais fontes de publicação de trabalhos relacionados ao SSA (GOLYANDINA, NEKRUTKIN e ZHIGLJAVSKY, 2001).

Apesar da vasta literatura e do grande número de aplicações da SSA, essa técnica ainda é pouco conhecida pela comunidade científica brasileira. Em termos de documentos técnicos científicos publicados no Brasil, aplicou-se a técnica em apenas duas teses e duas dissertações, oriundas do Departamento de Engenharia Biomédica da Universidade Federal do Rio de Janeiro (UFRJ), e em um trabalho desenvolvido no Departamento de Engenharia Elétrica da Universidade Federal de Minas Gerais (UFMG). As teses e dissertações apresentadas ao Programa de Pós-graduação em Engenharia Biomédica utilizaram aplicações com extração de sinais ultra-sônicos (MACIEL, 2000; SAITO, 2003; BAKITA, 2005; MACHADO, 2005), e a dissertação oriunda da UFMG discute aplicações sobre o consumo de energia elétrica, utilizando a SSA para a decomposição das séries (JESUS, 2008). Essas informações foram obtidas através de um levantamento no Banco de Teses do Portal da Capes ao utilizar a palavra-chave “Análise Espectral Singular” e “*Singular Spectrum Analysis*” para todos os anos, anteriores ao corrente ano de 2012, disponíveis no sistema. Além dos artigos publicados em revistas que foram derivados do grupo de pesquisa do Departamento de Engenharia Biomédica da UFRJ como, por exemplo, Pereira e Maciel (2001); Pereira e colaboradores (2002); Pereira e colaboradores (2004); Machado e Pereira (2006); dentre outros.

3.2 Descrição da SSA básica

A técnica SSA básica fundamenta-se em dois estágios complementares: decomposição e reconstrução da série temporal. Cada estágio é composto por dois passos, constituindo os quatro passos da técnica, são eles: *Embedding*, *SVD*, *Grouping* e *Diagonal averaging*. Estes passos serão descritos nas próximas subseções.

3.2.1 Decomposição

No estágio da decomposição, a série temporal inicial é decomposta em uma soma de poucas subséries, de modo que cada subsérie possa ser identificada e interpretada como os componentes constitutivos.

3.2.1.1 Embedding

Considere uma série temporal unidimensional real e não nula, com pelo menos um valor diferente de zero, $Y_t = Y_1, \dots, Y_N$; $t = 1, 2, \dots, N$, sendo N o comprimento da série ou a quantidade de observações ao longo do intervalo de tempo investigado.

Inicialmente, a série Y_t unidimensional é representada como uma série multidimensional com dimensão L denominada *window length* ou comprimento da janela.

O comprimento da janela L é o único parâmetro do estágio da decomposição e representa a quantidade de componentes em que a série original é decomposta. Tal parâmetro deve ser um valor inteiro entre $2 \leq L \leq N - 1$. Contudo, segundo resultados teóricos o tamanho de L deve ser suficientemente grande, mas não superior a $\frac{N}{2}$ (GOLYANDINA, NEKRUTKIN e ZHIGLJAVSKY, 2001; HASSANI, 2007; MYUNG, 2009; GOLYANDINA, 2010).

A série temporal multidimensional, que é uma sequência de vetores constituídos por elementos da série Y_t , forma a matriz apresentada na expressão (3.1), denominada como matriz trajetória. A matriz trajetória é o resultado deste primeiro passo (GOLYANDINA, NEKRUTKIN e ZHIGLJAVSKY, 2001; HASSANI, 2007; HASSANI, HERAVI e ZHIGLJAVSKY, 2009).

$$\mathbf{X}_{L \times K} = \begin{bmatrix} \mathbf{X}_1 & \dots & \mathbf{X}_K \end{bmatrix} = \begin{bmatrix} Y_1 & Y_2 & \dots & Y_K \\ Y_2 & Y_3 & \dots & Y_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ Y_L & Y_{L+1} & \dots & Y_N \end{bmatrix}; \quad (3.1)$$

em que, K representa o número de vetores ou subséries defasadas no tempo. Sendo que, K é especificado de acordo com a completa alocação das observações da série original Y_t nos extremos da matriz \mathbf{X} , sequência formada pelos valores da primeira coluna e última linha em (3.1); o que corresponde a $K = N - L + 1$.

A matriz trajetória \mathbf{X} tem a estrutura semelhante a de uma matriz Hankel. Uma matriz Hankel é simétrica e suas entradas $\{a_{ij}\}$ são constantes ao longo das diagonais paralelas à diagonal secundária.

Nota-se que, a matriz transposta \mathbf{X}^T também é a matriz trajetória da série Y_t , contudo o comprimento da janela fica igual a K , em vez de L ; assim como expresso em

$$\mathbf{X}^T = \begin{bmatrix} Y_1 & Y_2 & \cdots & Y_L \\ Y_2 & Y_3 & \cdots & Y_{L+1} \\ \vdots & \vdots & \ddots & \vdots \\ Y_K & Y_{K+1} & \cdots & Y_N \end{bmatrix}. \quad (3.2)$$

Em suma, o passo do *embedding* é considerado como um mapeamento que transforma uma série unidimensional Y_t em uma série multidimensional \mathbf{X} formada pelas subséries $\mathbf{X}_1, \dots, \mathbf{X}_K$; em que $\mathbf{X}_i = [Y_i, Y_{i+1}, \dots, Y_{i+L-1}]^T$, sendo $1 \leq i \leq K$.

3.2.1.2 SVD

No passo da decomposição do valor singular (SVD, em inglês) é realizada a decomposição da matriz trajetória \mathbf{X} em uma soma de matrizes elementares.

Seja \mathbf{S} o produto entre a matriz trajetória e a sua transposta, $\mathbf{S} = \mathbf{X}\mathbf{X}^T$. Ao realizar a SVD da matriz \mathbf{S} , obtém-se os seus autovalores, que podem ser ordenados de acordo com as suas magnitudes ($\lambda_1 \geq \dots \geq \lambda_L \geq 0$), e os correspondentes autovetores U_1, \dots, U_L ; ortogonais e normalizados.

Ao ser considerada a transformação dada por $V_i = \frac{\mathbf{X}^T U_i}{\sqrt{\lambda_i}}$, ($i = 1, \dots, d$), a SVD da matriz trajetória pode ser escrita como:

$$\mathbf{X} = E_1 + \cdots + E_d = \sum_{i=1}^d E_i, \quad (3.3)$$

em que $d = \max\{i, \text{tal que } \lambda_i > 0\} = \text{posto}(\mathbf{X})$, ou seja, d denota o número de autovalores diferentes de zero da matriz \mathbf{S} ($d \leq L$); e $E_i = \sqrt{\lambda_i} U_i V_i^T$ representa uma matriz de posto unitário, que é comumente nomeada como matriz elementar.

Na literatura da SSA, muitas terminologias da SVD são preservadas ao abordar o

segundo passo da técnica. A coleção ou tripla $(\sqrt{\lambda_i}, U_i, V_i)$ é chamada de i -ésimo autotriple (*eigen-triple*) da matriz \mathbf{X} ; $\sqrt{\lambda_i}$ é conhecido como valor singular da matriz \mathbf{X} ; e o conjunto $\{\sqrt{\lambda_i}\}$ representa o espectro de \mathbf{X} (GOLYANDINA, NEKRUTKIN e ZHIGLJAVSKY, 2001).

Os vetores U_i e V_i são identificados como vetores singulares da matriz \mathbf{X} , todavia outras nomenclaturas mais específicas também aparecem na literatura da SSA. Nomea-se U_i como vetor singular de esquerda (*left singular vector*), ou autovetor principal, ou ainda, *factor empirical orthogonal function* (EOF). Enquanto que, o V_i é nomeado das seguintes formas: vetor singular de direita (*right singular vector*), *factor vector* e componente principal.

Os principais textos metodológicos da literatura trazem que a SVD em (3.3) é ótima, já que entre todas as matrizes $E^{(r)}$ de posto inferior a d ($r < d$), a matriz dada por $\left(\sum_{i=1}^r E_i\right)$ fornece a melhor aproximação para a matriz trajetória \mathbf{X} ; de forma que, o mínimo é alcançado através da norma $\|\mathbf{X} - E^{(r)}\|$.

De forma semelhante à análise de componentes principais (ACP), a contribuição de cada matriz elementar E_i , na expansão (3.3), também pode ser considerada. Essa contribuição é computada como

$$\frac{\lambda_i}{\sum_{i=1}^d \lambda_i}; i = 1, \dots, d; \quad (3.4)$$

em que, $\sum_{i=1}^d \lambda_i = \|\mathbf{X}\|^2$ e o autovalor $\lambda_i = \|E_i\|^2$.

Herança também recebida da ACP, considera-se as contribuições acumuladas transmitidas pelas r primeiras matrizes elementares; o que equivale a fração de explicabilidade associada à aproximação ótima da matriz trajetória pela matriz de posto r . Essa proporção pode ser expressa como segue

$$\frac{\sum_{i=1}^r \lambda_i}{\sum_{i=1}^d \lambda_i}; i = 1, \dots, d. \quad (3.5)$$

3.2.2 Reconstrução

No presente estágio, busca-se os grupos de matrizes mais semelhantes para a formação dos componentes e posterior reconstrução da série temporal.

3.2.2.1 Grouping

O procedimento *grouping* tem como principal objetivo a distinção dos componentes aditivos da série temporal em termos de matrizes separáveis. Em outras palavras, neste passo, identifica-se as componentes mais correlacionadas entre si, com o intuito de organizá-las em um mesmo grupo.

Matematicamente, o passo *grouping* particiona o conjunto de índices das matrizes elementares da expressão (3.3), $\{1, \dots, d\}$, em subconjuntos disjuntos I_1, \dots, I_m , o que corresponde a representação

$$\mathbf{X} = E_{I_1} + \dots + E_{I_m} = \sum_{p=1}^m E_{I_p}; \quad (3.6)$$

onde E_{I_1}, \dots, E_{I_m} são conhecidas como matrizes resultantes. E cada matriz resultante é obtida a partir da soma de matrizes elementares em um particular conjunto índice I_p , isto é,

$$E_{I_p} = \sum_{i \in I_p} E_i; p = 1, \dots, m. \quad (3.7)$$

Na expressão (3.6) tem-se uma nova decomposição de matrizes, esta é denominada como decomposição agrupada.

A organização dos índices $\{1, \dots, d\}$ em grupos I_1, \dots, I_m ocorre de tal forma que, as suas matrizes satisfaçam (3.6) e sejam próximas de matrizes Hankel, para que assim, possam transformar-se em matrizes trajetórias de séries que são separáveis pela expansão (3.3).

O procedimento de escolher os conjuntos I_1, \dots, I_m é chamado *eigentriple grouping*, este fornece o último parâmetro da técnica SSA. A escolha deste parâmetro utiliza o conceito de separabilidade que é apresentado na Seção 3.4.

Em linhas gerais, no terceiro passo ocorre a divisão das matrizes elementares E_i em vários grupos e a soma das matrizes dentro de cada grupo. E assim, tem-se o resultado deste passo: representação da matriz trajetória como uma soma de matrizes resultantes.

3.2.2.2 Diagonal averaging

A operação realizada no último passo, conhecida como *diagonal averaging*, transforma cada matriz resultante em um componente aditivo da série original. Ou seja, cada matriz da decomposição agrupada (3.6) é convertida em uma nova série de tamanho N . O que possibilita a obtenção de uma série unidimensional, considerada como uma aproximação da série inicial.

A transformação das matrizes resultantes em séries ocorre ao aplicar o operador linear de Hankelização (\mathcal{H}). Tal operador age em uma matriz arbitrária de forma que transforme-a em uma matriz Hankel e, portanto, em uma matriz trajetória e, por fim em uma série. No contexto do quarto passo, o operador \mathcal{H} contabiliza as médias ao longo das linhas paralelas as diagonais secundárias das matrizes E_{I_p} , para $p = 1, \dots, m$.

Considere uma matriz $\mathbf{Z}_{L \times K} = (z_{ij})$; $1 \leq i \leq L$ e $1 \leq j \leq K$, em que $L \leq K$. O resultado da aplicação do operador \mathcal{H} na matriz \mathbf{Z} é a matriz Hankel $\mathcal{H}\mathbf{Z}$, que é a matriz trajetória da série obtida como um resultado da *diagonal averaging*.

Com o intuito de exemplificação, denote a matriz $\mathcal{H}\mathbf{Z}$ como \mathbf{H} . Uma matriz Hankel pode ser definida como uma matriz cujas diagonais perpendiculares à diagonal principal são valores constantes (FREITAS, 2006). As entradas dessa matriz satisfazem a seguinte lei de formação:

$$a_{i,j} = a_{i-1,j+1}; \forall i, j \in N, j > 0.$$

Isto é, uma matriz Hankel possui a forma representada por

$$\mathbf{H} = \begin{bmatrix} a_0 & a_1 & a_2 & a_3 & a_4 & a_5 & \cdots \\ a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & \cdots \\ a_2 & a_3 & a_4 & a_5 & a_6 & a_7 & \cdots \\ a_3 & a_4 & a_5 & a_6 & a_7 & a_8 & \cdots \\ a_4 & a_5 & a_6 & a_7 & a_8 & a_9 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Voltando ao contexto da matriz $\mathbf{Z}_{L \times K}$, z_{ij} representa um elemento de tal matriz e, então cada termo da série “resultante” (oriunda da operação de Hankelização) G_t ($t = 1, \dots, N$) é obtido da seguinte maneira:

$$G_t = \begin{cases} \frac{1}{s-1} \sum_{q=1}^{s-1} z_{(q,s-q)}, & 2 \leq s \leq L-1 \\ \frac{1}{L} \sum_{q=1}^L z_{(q,s-q)}, & \leq s \leq K+1 \\ \frac{1}{K+L-s+1} \sum_{q=s-K}^L z_{(q,s-q)}, & K+2 \leq s \leq K+L \end{cases} \quad (3.8)$$

Em que, s é a soma dos índices da matriz $\mathbf{Z}_{L \times K}$ ($s=i+j$) e $N = L + K - 1$.

A expressão (3.8) corresponde a média dos elementos z_{ij} sobre as diagonais paralelas à diagonal secundária de \mathbf{Z} . Por exemplo, se $s = 2$ tem-se

$$G_1 = z_{11};$$

para $s = 3$,

$$G_2 = \frac{1}{2} \sum_{q=1}^2 z_{(q,3-q)} = \frac{z_{12} + z_{21}}{2};$$

para $s = 4$,

$$G_3 = \frac{1}{3} \sum_{q=1}^3 z_{(q,4-q)} = \frac{z_{13} + z_{22} + z_{31}}{3};$$

e assim sucessivamente até a obtenção do último termo (G_N) de uma série “resultante”.

Caso a matriz \mathbf{Z} tenha mais linhas do que colunas ($L > K$), o processo de obtenção dos elementos da matriz $\mathcal{H}\mathbf{Z}$ é análogo ao caso anterior ($L \leq K$). As alterações referem-se a substituição $L \leftrightarrow K$ e a transposição dos elementos de \mathbf{Z} ($z_{ij} \rightarrow z_{ji}$).

Na prática, a matriz trajetória \mathbf{X} da série original possui um número de linhas inferior ao de colunas ($L < K$). Isso porque, como N é fixo e existem limites para L ($2 \leq L \leq \frac{N}{2}$), o valor mínimo de K fica determinado por $K_{min} = N - (L_{max}) + 1 = \frac{N}{2} + 1$; onde o L_{max}

denota o valor máximo de L . Desta forma, tem-se que $L_{max} < K_{min}$.

A série “resultante” $\tilde{Y}_t^{(p)} = \tilde{Y}_1^{(p)}, \dots, \tilde{Y}_N^{(p)}$ é produzida ao aplicar o procedimento de Hankelização em cada matriz resultante E_{I_p} ; $p = 1, \dots, m$. Ou seja, se o operador \mathcal{H} é aplicado a todos os componentes da matriz (3.6), obtém-se a última expansão da técnica SSA, a saber:

$$\mathbf{X} = \tilde{E}_{I_1} + \dots + \tilde{E}_{I_m} = \sum_{p=1}^m \tilde{E}_{I_p}; \quad (3.9)$$

em que, $\tilde{E}_{I_p} = \mathcal{H}E_{I_p}$.

Detalhes teóricos sobre o operador \mathcal{H} podem ser encontrados na seção 6.2 do livro da Golyandina, Nekrutkin e Zhigljavsky (2001).

Assim como citado na Subseção anterior (3.2.2.1), um agrupamento adequado leva à decomposição (3.6), em que as matrizes resultantes E_{I_p} são quase Hankel; o que corresponde a separabilidade aproximada (Seção 3.4). E sob a suposição de que cada matriz \tilde{E}_{I_p} da expansão (3.9) é uma matriz Hankel, cada uma dessas matrizes determina unicamente a série $\tilde{Y}_t^{(p)}$; e portanto, a série inicial Y_t é decomposta na soma de m séries, resultando em uma série temporal aproximada \tilde{Y}_t , assim como segue:

$$\tilde{Y}_t = \tilde{Y}_t^{(1)} + \dots + \tilde{Y}_t^{(m)} = \sum_{p=1}^m \tilde{Y}_t^{(p)}; \quad (3.10)$$

em que, $t = 1, \dots, N$ e para cada p a série $\tilde{Y}_t^{(p)}$ é o resultado do processo de Hankelização da matriz E_{I_p} . Na decomposição dada pela equação (3.10) tem-se a soma de m componentes separáveis, já expressas na forma de séries temporais.

Como exemplificação, considere $\tilde{Y}_1, \dots, \tilde{Y}_N$ como sendo as observações da série temporal aproximada. A observação \tilde{Y}_1 é obtida somando o primeiro termo de cada uma das m séries “resultantes”; assim como, obtém-se a observação \tilde{Y}_N ao somar o N -ésimo termo de cada uma das m séries. Ou seja,

$$\tilde{Y}_1 = \tilde{Y}_1^{(1)} + \dots + \tilde{Y}_1^{(m)};$$

⋮

$$\tilde{Y}_N = \tilde{Y}_N^{(1)} + \cdots + \tilde{Y}_N^{(m)}.$$

3.3 Informações adicionais

A escolha dos parâmetros da técnica SSA depende do objetivo da análise e das informações preliminares sobre a série temporal. Na literatura especializada há algumas informações complementares, as quais auxiliam na escolha do parâmetro L para atingir uma boa separabilidade dos componentes; assim como na maneira de agrupamento, que possibilita a identificação adequada dos autotriples mais importantes para a extração dos componentes (tendência, periodicidade e ruído). Nas próximas subseções encontram-se descrições de informações adicionais e algumas ferramentas úteis para separar o sinal do ruído. Maiores detalhes podem ser verificados em Golyandina, Nekrutkin e Zhigljavsky (2001) e Hassani (2007).

3.3.1 Separabilidade

Os procedimentos SVD e *grouping* apoiam-se na propriedade denominada separabilidade. Esta propriedade utiliza conceitos que possibilitam a separação de diferentes componentes da série temporal. Sendo assim, a decomposição e reconstrução da série Y_t tem êxito se os seus componentes aditivos são separáveis um do outro.

Considere que a série original Y_t é dada pela soma de duas subséries ($Y_t = Y_t^{(1)} + Y_t^{(2)}$). Neste caso, a separabilidade das subséries $Y_t^{(1)}$ e $Y_t^{(2)}$ significa que os termos da SVD da matriz trajetória \mathbf{X} podem ser divididos em dois diferentes grupos, de forma que as somas dos termos dentro de cada grupo origina as matrizes trajetórias $\mathbf{X}^{(1)}$ e $\mathbf{X}^{(2)}$ das subséries $Y_t^{(1)}$ e $Y_t^{(2)}$, respectivamente.

Em termos teóricos, a separabilidade implica que cada linha da matriz trajetória $\mathbf{X}^{(1)}$ é ortogonal a cada linha da matriz trajetória $\mathbf{X}^{(2)}$, e o mesmo vale para as colunas. Diante desta suposição de ortogonalidade, pode-se dizer que as subséries $Y_t^{(1)}$ e $Y_t^{(2)}$ são fracamente separáveis. Existe também uma separabilidade mais restrita, esta é conhecida como separabilidade forte. Para este caso, além da suposição de separabilidade fraca, considera-se que os conjuntos formados pelos valores singulares das matrizes $\mathbf{X}^{(1)}$ e $\mathbf{X}^{(2)}$ são disjuntos.

Conforme apresentado por Golyandina, Nekrutkin e Zhigljavsky (2001), na prática, não ocorre separabilidade exata. Nas situações práticas, fala-se somente de separabilidade aproximada. Comumente, avalia-se a qualidade da separabilidade através de uma medida

natural de dependência entre as subséries. Essa medida é denominada como “correlação ponderada” ou “w-correlação”.

A correlação ponderada entre duas subséries $Y_t^{(1)}$ e $Y_t^{(2)}$ pode ser expressa como segue:

$$\rho_{12}^{(w)} = \frac{\langle Y_t^{(1)}, Y_t^{(2)} \rangle_w}{\|Y_t^{(1)}\|_w \|Y_t^{(2)}\|_w}; \quad (3.11)$$

em que a norma da i -ésima subsérie é dada por

$$\|Y_t^{(i)}\|_w = \sqrt{\langle Y_t^{(i)}, Y_t^{(i)} \rangle_w};$$

o produto interno entre um par de subséries é

$$\langle Y_t^{(i)}, Y_t^{(j)} \rangle = \sum_{c=1}^N w_c Y_c^{(i)} Y_c^{(j)}; i, j = 1, 2;$$

sendo que, os ponderadores w_c são representados por $w_c = \min\{c, L, N - c\}$, e assume-se que $L \leq \frac{N}{2}$.

Caso o valor absoluto da correlação ponderada seja pequeno, tem-se que as duas séries são quase ortogonais. Desta forma, uma correlação ponderada entre dois componentes reconstruídos igual a zero ($\rho_{12}^{(w)} = 0$) significa que estes componentes são separáveis. Por outro lado, se o valor absoluto é alto, então as séries não são tão bem separáveis. Ou seja, valores para $\rho_{12}^{(w)} \approx 1$ indicam que os componentes devem ser reunidos em um mesmo grupo, correspondendo ao mesmo componente na decomposição SSA.

Na Subseção 3.3.3.2 é apresentada uma ferramenta de agrupamento que utiliza os valores absolutos da correlação ponderada.

3.3.2 Comprimento da janela (L)

Assim como apresentado na Subseção 3.2.1, o primeiro passo da técnica SSA exige a entrada de um valor para o parâmetro L . Tal parâmetro é considerado o principal do SSA básico, pois a sua escolha inadequada dificulta um bom agrupamento, e conseqüentemente, diminui a precisão na aproximação da série.

Segundo estudos de simulação, valores entre $[2; \frac{N}{2}]$ são suficientes para a realização da decomposição da série temporal. Contudo, a decomposição da série temporal torna-se mais detalhada na medida em que o comprimento da janela aumenta. Assim, pode-se notar que o comprimento ótimo de L , com relação a proximidade entre a série original e a reconstruída, coincide com o seu tamanho máximo ($L_{max} = \frac{N}{2}$). Desta forma, sugere-se usar o valor ótimo de L se a análise exigir uma alta precisão na reconstrução da série (ESQUIVEL, SENNA e GOMES, 2011a). O resumo estendido do trabalho desenvolvido por Esquivel, Senna e Gomes (2011a) encontra-se no Apêndice D.

Todavia, a literatura indica a necessidade de atenção na escolha de L ao trabalhar com séries temporais periódicas. Para alcançar separabilidade suficiente dos componentes sugere-se utilizar um comprimento L proporcional ao período de sazonalidade dos dados (GOLYANDINA, NEKRUTKIN e ZHIGLJAVSKY, 2001; HASSANI, 2007).

De qualquer forma, indica-se que o L seja suficientemente grande para que cada vetor $L - de\ fasado$, discutido em 3.2.1.1, incorpore essencialmente o comportamento da série inicial Y_t .

Ao tratar do efeito do comprimento da janela, Golyandina, Nekrutkin e Zhigljavsky (2001) mostram a influência de L através de um exemplo, exibindo o resultado da decomposição de uma série caracterizada como complexa. Tal série refere a dados diários no decorrer de 14 anos, tendo $N = 5113$ pontos no tempo. Nesta série existem dois comportamentos periódicos, periodicidade semanal (período $T_1 = 7$) e anual ($T_2 = 365$). Desde que T_2 seja muito maior do que T_1 , é natural que considere um L múltiplo de T_2 .

A escolha de $L = T_2$ garantiu a extração simultânea da periodicidade anual e semanal. Além disso, este valor de L permitiu a extração da tendência da série usando um único autotriple principal. Enquanto que, se o L for aumentado em três vezes ($L = 3 \times T_2 = 1095$), o componente de tendência fica misturado com os componentes de periodicidade. Então, em geral, valores altos para L conduz a violação da suposição de separabilidade forte.

Todavia, vale salientar que, se não é importante o problema da separação da tendência e periodicidade, então valores de L superiores a T_2 também podem ser utilizados; neste caso, tem-se a vantagem de uma decomposição mais detalhada.

Voltando ao contexto do exemplo, caso utilize $L = 365$, a estrutura geral da série é obtida com apenas 5 autotriples, enquanto que, ao considerar $L = 1095$ precisa-se de dezenas de autotriples. Por outro lado, se for considerado um L pequeno (em relação ao período anual, por exemplo, $L = 28$), então o comportamento global da série deve ser descrito por apenas um autotriple principal. Isto indica que, em situações semelhantes,

ao aumentar o comprimento da janela, há a necessidade de considerar mais autotriples na etapa de reconstrução da série, para que estes possam captar melhor as características das componentes.

O trabalho de Myung (2009) contém uma seção que apresenta brevemente os efeitos do comprimento da janela. Uma de suas discussões aponta o efeito de L na tendência principal da série. É possível notar que, diante da escolha adequada do comprimento L , a tendência da série comporta-se segundo um padrão suave; já que não incorpora características de outros componentes, como o componente sazonal.

Em suma, a escolha adequada do comprimento da janela depende, de modo fundamental, do propósito da análise. Deve-se ter uma maior atenção na escolha de L quando o interesse principal é a extração de componentes específicos (como tendência e periodicidades). Para que deste modo, a suposição de separabilidade possa ser garantida.

3.3.3 Escolha dos autotriples

O segundo parâmetro da técnica SSA é estrutural, isto é, refere-se a forma de agrupamento dos autotriples. Diante do terceiro passo da SSA, *grouping*, precisa-se selecionar os autotriples oriundos da SVD da matriz \mathbf{X} para que estes formem os grupos do correspondente passo.

O processo de formação dos grupos também tem sua importância, pois, espera-se que, um agrupamento adequado leve a uma boa separação (aproximada) dos componentes da série temporal. Desta forma, o problema envolve a escolha dos autotriples de interesse entre todos aqueles autotriples gerados pela série temporal. Em outras palavras, precisa-se identificar os autotriples correspondentes aos principais componentes constituintes da série inicial. Para isso, costuma-se avaliar os valores singulares oriundos do segundo passo da SSA.

O agrupamento dos autotriples também depende do propósito da análise. Golyandina, Nekrutkin e Zhigljavsky (2001), apresentam o efeito do agrupamento para diferentes objetivos, tais como: agrupamento para extração de tendência; para suavização da série; agrupamento para extração de componentes periódicos e agrupamento para obtenção de uma decomposição refinada da série. Contudo, independente do objetivo, os autotriples são organizados em dois grandes blocos: o sinal e o ruído. O sinal é formado por todos os componentes de interesse e os demais componentes são alocados para o outro bloco.

Por exemplo, se o interesse é encontrar uma estrutura refinada da série, assim como no contexto do presente trabalho; então os autotriples correspondentes aos componentes de

tendência e oscilações são tratados como *signal*; e os componentes restantes são associados ao ruído. Ou seja, o ruído, assim como os resíduos da análise clássica de séries temporais, não deve ter uma estrutura de tendência e/ou periódica.

O comportamento dos vetores singulares, especialmente os *factor vectors*, e dos valores singulares da SVD da matriz trajetória auxilia no agrupamento adequado dos autotriples. Algumas dicas de agrupamento são apresentadas a seguir .

3.3.3.1 Valores singulares

O comportamento dos valores singulares podem ser observados através de um gráfico similar ao *scree plot*, em que os seus L valores são apresentados de forma decrescente de magnitude ($\sqrt{\lambda_1} \geq \sqrt{\lambda_2} \geq \dots \geq \sqrt{\lambda_L}$), exibindo o espectro de \mathbf{X} .

O primeiro valor singular, aquele com maior valor absoluto, sempre está associado ao componente de tendência. Assim como a tendência de uma série temporal pode ser descrita através de uma função com variação lenta e estável no decorrer de um período longo de tempo, o comportamento esperado do primeiro *factor vector* ou componente principal (V_1) deve apresentar também uma variação lenta e sem oscilações.

Vautard e Ghil (1989), *apud*. Pereira e Maciel (2001), mostrou que a presença de autovalores com valores próximos o suficiente para ser identificado como um “par” pode ser associado a um componente periódico do sinal. Sendo assim, pode-se identificar pares de valores singulares quase iguais no gráfico do espectro de \mathbf{X} , retratando os componentes periódicos da série temporal.

Teoricamente, uma série puramente residual produz uma sequência lentamente decrescente de valores singulares. E se um ruído é adicionado a um sinal, composto por eigentriples com valores singulares altos, então pode-se observar uma quebra no espectro de \mathbf{X} . Desta forma, pode-se distinguir os eigentriples relacionados ao sinal daqueles associados ao ruído.

Com o propósito de exemplificação, considere a clássica série sobre os dados mensais de mortes acidentais nos EUA entre 1973 e 1978 ($N = 72$), também analisada por Hassani (2007). A tendência linear e uma sazonalidade aditiva são as principais características da referida série (Figura 3.1).

Observa-se através do gráfico (Figura 3.1), que há uma periodicidade anual. Sendo assim, para obter uma melhor separabilidade deste componente periódico sugere-se considerar um comprimento da janela proporcional ao período $T = 12$. Hassani (2007) escolheu

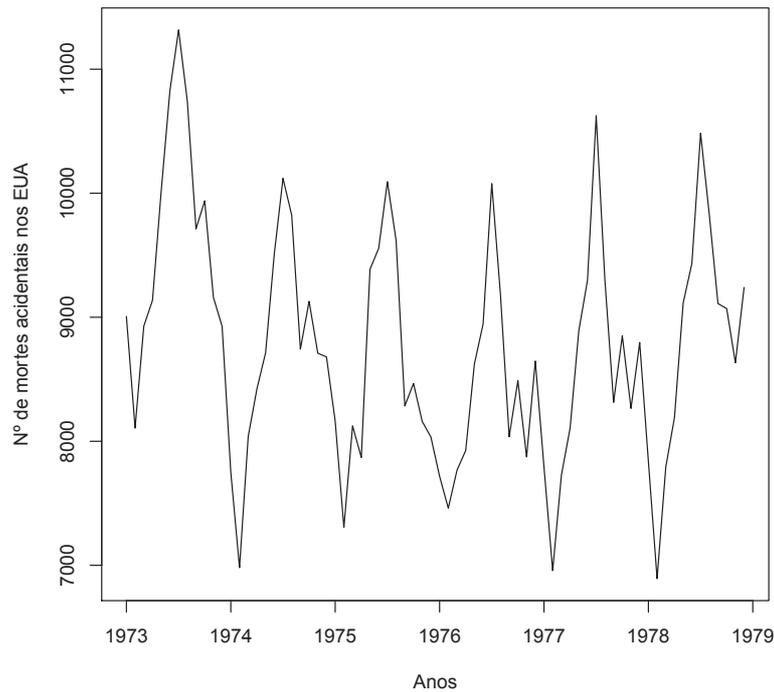


Figura 3.1: Número mensal de mortes acidentais nos EUA entre 1973 e 1978

$L = 24$, isto é, duas vezes a periodicidade da série ($L = 2 \times T$).

Ao considerar $L = 24$, os autovalores resultantes da decomposição da série original podem ser visualizados através do próximo gráfico (Figura 3.2).

Segundo a Figura 3.2, pode-se identificar naturalmente o componente de tendência, representado no gráfico pelo autovalor de maior magnitude (posto 1).

Posteriormente, pode-se buscar os pares de autovalores. Assim como mencionado anteriormente, estes pares formam autotriples que correspondem aos componentes periódicos da série temporal. Neste exemplo, o padrão de sazonalidade da série original começa a ser identificado a partir do segundo componente. De acordo com a Figura 3.2, pode-se visualizar a formação de três autotriples, a saber: (2-3); (4-5) e (11-12). Apesar da não observância trivial de mais pares de autotriples, os autovalores de postos intermediários (6 a 10) podem ser considerados também na formação do comportamento periódico da série.

Por fim, os autotriples restantes podem ser identificados como o ruído da série. Pode-se observar que entre a sequência decrescente de forma lenta (a partir do posto 15) e os autovalores com maiores magnitudes, tem-se uma quebra no gráfico (autotriples de ordem 13 e 14); o que representa um limite entre o sinal e o ruído. Fato este que concorda

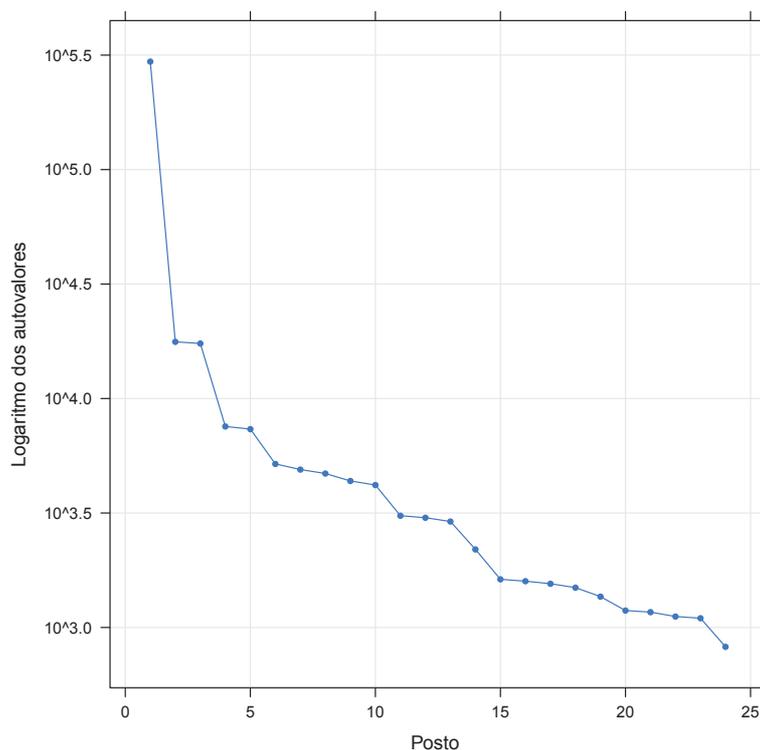


Figura 3.2: Autovalores resultantes da decomposição da série em 24 componentes

com os resultados teóricos disponíveis na literatura.

3.3.3.2 Correlação ponderada

Assim como apresentado na Seção 3.3.1, uma condição necessária para a separabilidade (aproximada) de duas séries é a existência de correlação ponderada nula (ou próxima de zero) dos componentes reconstruídos. Em outras palavras, os componentes altamente correlacionados se tornam pertencentes a um mesmo grupo de autotriples.

Uma ferramenta complementar utilizada para o processo de agrupamento é conhecida como “matriz w -correlação”. Tal ferramenta consiste na análise gráfica da matriz das correlações ponderadas (Seção 3.3.1), que indica as correlações entre as componentes da SVD através de uma escala de cores variando do branco ao preto, correspondendo aos valores absolutos das correlações (0 a 1).

A forma da matriz w -correlação fornece um indicativo de como realizar um agrupamento adequado, a saber: espera-se uma divisão clara dos componentes em duas partes. A primeira parte é constituída por componentes altamente correlacionados (tonalidade mais escura), o que caracteriza o grupo de autotriples relacionado ao sinal da série. Por

outro lado, a segunda parte retrata o ruído, exibindo muitos componentes com correlações baixas (tonalidades claras).

A Figura 3.3 mostra a aplicação da ferramenta “matriz w-correlação” para a série exemplificada na subseção anterior.

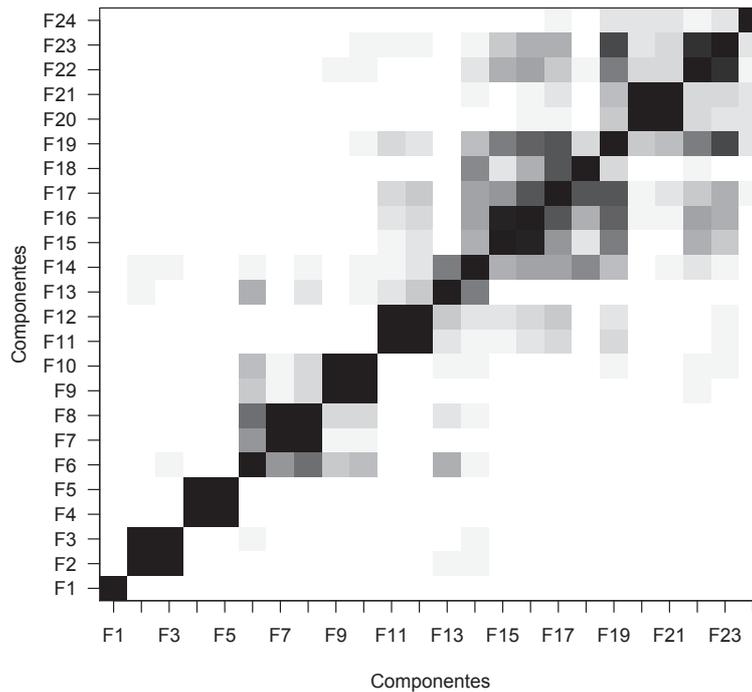


Figura 3.3: Matriz w-correlação (correlações ponderadas)

Nota-se, através da Figura 3.3, que o sinal da série pode ser formado pelos doze autotriples principais, indicando que os doze primeiros componentes são suficientes para reconstruir a série original. Por outro lado, os demais componentes formam o ruído, ou seja, a partir do décimo terceiro componente observa-se a existência de mais componentes com correlações maiores, representadas através das tonalidades mais fortes.

3.4 Previsão via SSA

De acordo com os aspectos metodológicos da previsão SSA, pode-se obter previsões em série temporal que satisfazem (aproximadamente) as fórmulas recorrentes linear (LRFs, do inglês *linear recurrent formulae*). Diz-se que uma série Y_t satisfaz uma LRF de ordem ou dimensão $L - 1$ se existirem coeficientes a_1, \dots, a_{L-1} , tal que

$$Y_{N-j} = \sum_{p=1}^{L-1} a_p Y_{N-j-p}; \quad (3.12)$$

em que $0 \leq j \leq N - L$.

Conforme apresentado por Hassani (2007), os coeficientes a_1, \dots, a_{L-1} podem ser encontrados através dos autovetores obtidos a partir do passo SVD. Para maiores informações pode-se consultar Golyandina, Nekrutkin e Zhigljavsky (2001); capítulos 2 e 5.

A classe de séries temporais regidas por LRFs é bastante ampla, tal classe inclui polinômios, termos harmônicos e exponenciais. Uma série que satisfaz alguma LRF pode ser representada como uma soma de produtos de polinômios, exponenciais e harmônicos; assim como segue:

$$Y_t = \sum_{p=1}^q a_p(t) e^{t\mu_p} \sin(2\pi t\omega_p + \varphi_p); \quad (3.13)$$

onde $a_p(t)$ representam polinômios; μ_p , ω_p e φ_p são constantes arbitrárias. E q denota o número de termos linearmente independentes.

Golyandina, Nekrutkin e Zhigljavsky (2001) apresentam a ideia das fórmulas recorrentes linear expressando séries governadas por uma LRF de ordem mínima, cuja dimensão é representada por d . Segundo características da decomposição da SSA (Subseção 3.2.1.2), existem no máximo d valores singulares não nulos na SVD da matriz \mathbf{S} ; e portanto, mesmo que o L e K sejam maiores do que d , precisa-se de no máximo d matrizes elementares E_i para reconstruir a série.

Contudo, na prática, dificilmente espera-se ter uma série que seja regida por uma LRF de dimensão pequena. Diante de uma série governada por uma LRF (3.12), pode-se dar continuidade para um número arbitrário de passos à frente, ao usar a mesma LRF. Na verdade, não é necessário buscar uma dimensão mínima da LRF, pois qualquer LRF que governe uma determinada série temporal produz a mesma “continuação” da série original Y_t . Esta “continuação” é obtida ao aplicar uma LRF nos termos passados da série Y_t .

A previsão SSA recebeu uma contribuição importante do Danilov (1997a, 1997b). Este autor obteve o seguinte resultado: se o espaço linear \mathcal{L}_r , também conhecido como espaço trajetória, gerado pelas colunas da matriz trajetória tem dimensão (r) menor do que L e este espaço é um espaço “não-vertical”, então a série satisfaz, naturalmente, uma LRF de dimensão $L - 1$. Em linhas gerais, diz-se que \mathcal{L}_r é um espaço “não-vertical” se a base

$e_L \notin \mathcal{L}_r$, em que $e_L = (0, 0, \dots, 0, 1)^T \in R^L$.

Diante da suposição de espaço “não-vertical”, o espaço trajetória \mathcal{L}_r produz LRFs adequadas. Tais LRFs podem ser usadas para a previsão dos componentes da série, correspondendo às matrizes de posto unitário selecionadas no terceiro passo da técnica SSA; o que representa a seleção de um espaço r – dimensional ($\mathcal{L}_r \in R^L$) gerado pelos correspondentes *left singular vector*. Este processo de previsão é conhecido como *recurrent forecasting algorithm*.

Para o presente trabalho foi utilizado o algoritmo recorrente de previsão baseado no SSA básico, que encontra-se descrito na próxima subseção, isto é, Seção 3.4.1.

Assim como no SSA básico, a característica da separabilidade ajuda na seleção do espaço trajetória. Segundo Golyandina, Nekrutkin e Zhigljavsky (2001), a separabilidade relaciona-se diretamente com as LRFs, ou seja, caso duas séries sejam separáveis (conforme apresentado na Subseção 3.3.1); estas certamente serão regidas por certas LRFs.

Geralmente, objetiva-se realizar a previsão do sinal na presença do ruído. Isto é, supõe-se que a série original Y_t é formada pelo sinal $Y_t^{(1)}$ e pelo ruído $Y_t^{(2)}$; componentes aditivos e separáveis entre si ($Y_t = Y_t^{(1)} + Y_t^{(2)}$). Em outras palavras, o interesse concentra-se em dar continuidade ao componente $Y_t^{(1)}$ da série temporal Y_t ; sendo assim, assume-se que $Y_t^{(1)}$ satisfaz alguma LRF e é separável da série residual $[Y_t^{(2)} = Y_t - Y_t^{(1)}]$ para o valor selecionado de L (discutido na Subseção 3.3.2). Como mencionado anteriormente, se duas séries são fortemente separáveis, então cada uma delas deve satisfazer alguma LRF .

A teoria utilizada na previsão SSA pode ser encontrada no capítulo 5 da obra da Golyandina, Nekrutkin e Zhigljavsky (2001). E maiores detalhes sobre as LRFs podem ser encontrados na seção 5.2 da obra ora citada.

3.4.1 Algoritmo recorrente de previsão

Na presente seção o algoritmo recorrente de previsão, baseado no SSA básico, é brevemente apresentado.

Antes de apresentar o algoritmo propriamente dito, faz-se necessário introduzir algumas notações e comentários úteis para o bom entendimento do algoritmo recorrente de previsão SSA.

Seja $Y_t = (Y_1; \dots; Y_N)$ a série original e $\tilde{Y}_t = \tilde{Y}_1 + \dots + \tilde{Y}_N$ a série reconstruída,

conforme o segundo estágio da SSA; apresentado na Seção 3.2.2.

Defina um vetor $P \in R^L$ e denote o vetor dos $L - 1$ primeiros termos do vetor P como $P^\nabla \in R^{L-1}$. Considere P_1, \dots, P_r uma base ortonormal do espaço linear \mathcal{L}_r .

Pode-se provar que o último termo Z_L de qualquer vetor $Z = (Z_1, \dots, Z_L)^T$ é uma combinação linear dos primeiros componentes (Z_1, \dots, Z_{L-1}) . Isto é, $Z_L = a_1 Z_{L-1} + a_2 Z_{L-2} + \dots + a_{L-1} Z_1$ (GOLYANDINA, NEKRUTKIN e ZHIGLJAVSKY, 2001; HASSANI, HERAVI e ZHIGLJAVSKY, 2009).

A previsão dos valores futuros \widehat{Y}_{N+h} , com $h = 1, \dots, M$, é obtida a partir da seguinte expressão recorrente:

$$\widehat{Y}_j = \sum_{p=1}^{L-1} a_p \widehat{Y}_{j-p}; j = N + 1, \dots, N + M. \quad (3.14)$$

Em que, a_p são os coeficientes da combinação linear entre os $L - 1$ últimos termos da série considerada, ou seja,

$$\widehat{Y}_{N+h} = \sum_{p=1}^{L-1} a_p \widehat{Y}_{(N+h)-p}; h = 1, \dots, M.$$

Os coeficientes a_p podem ser representados na forma de vetor $A = (a_1, \dots, a_{L-1})$. Tal vetor pode ser expresso como

$$A = \frac{1}{1 - \nu^2} \sum_{i=1}^r \pi_i P_i^\nabla; \quad (3.15)$$

em que, $\nu^2 = \pi_1^2 + \dots + \pi_r^2$, sendo π_i o último componente do vetor P_i ; e, como citado anteriormente, o vetor P_i^∇ consiste dos primeiros $L - 1$ componentes do vetor P . O vetor dos coeficientes A não depende da escolha da base P_1, \dots, P_r no espaço linear \mathcal{L}_r .

Na literatura especializada, como ν^2 representa o cosseno ao quadrado do ângulo entre o vetor e_L e o espaço linear \mathcal{L}_r , este pode ser chamado como “coeficiente de verticalidade” de \mathcal{L}_r .

A equação (3.14) é a continuação da série aproximada pelo SSA básico \widetilde{Y}_j para $j =$

$1, \dots, N$, ou equivalentemente ao apresentado na equação (3.10) ($\tilde{Y}_t, t = 1, \dots, N$).

Vale salientar que, quanto maior o número de passos à frente (h), as previsões tornam-se mais dependentes das predições anteriores. Para que a predição considere pelo menos um valor aproximado (\tilde{Y}_j) da série original, recomenda-se utilizar um horizonte de no máximo $M = L - 1$.

Em síntese, os números $\hat{Y}_{N+1}, \dots, \hat{Y}_{N+M}$ formam os M termos preditos pelo algoritmo recorrente de previsão fundamentado no SSA básico. No presente trabalho, por brevidade, este algoritmo é referido como algoritmo recorrente.

Detalhes sobre o algoritmo de previsão recorrente estão disponíveis no capítulo 2 da Golyandina, Nekrutkin e Zhigljavsky (2001).

3.4.2 Exemplos de aplicações bem sucedidas

Com o intuito de exemplificação, nesta subseção apresenta-se algumas aplicações bem sucedidas da literatura sobre a previsão SSA.

Os principais resultados das previsões e conclusões discutidas nos artigos do Hassani (2007) e Hassani, Heravic e Zhigljavskya (2009) são abordados.

No primeiro *paper*, Hassani confronta as previsões SSA com quatro técnicas tradicionais, analisadas por Brockwell e Davis (2002), aplicadas na clássica série “mortes acidentais nos EUA” (Subseção 3.3.3.1).

Os autores Brockwell e Davis utilizaram dois dos modelos de Box-Jenkins (médias móveis e ARIMA sazonal), um algoritmo conhecido como ARAR; que é uma transformação para encurtamento de memória, e o algoritmo sazonal de Holt-Winters (BROCKWELL e DAVIS, 2002).

Brockwell e Davis (2002) ajustaram à série “mortes acidentais” um modelo MA(13) incompleto, especificamente, eles consideraram os seguintes parâmetros de médias móveis: $\theta_1, \theta_6, \theta_{12}$ e θ_{13} . O conjunto de previsões oriundas desse modelo foi rotulado por Hassani (2007) como Modelo II.

Os outros três conjuntos de previsões discutidos em Brockwell e Davis (2002) foram obtidos a partir de um SARIMA(0,1,1)(0,1,1) com período sazonal de 12 meses; e a partir dos algoritmos ARAR e Holt-Winters aditivo; os quais foram rotulados como Modelo I, ARAR e HWS, respectivamente.

Segundo Hassani (2007), considerou-se os doze primeiros autotriples (1-12) para a obtenção das previsões via SSA.

Para o cálculo da precisão, os autores ora citados fizeram uso de duas medidas de erro, a saber: o erro absoluto médio (MAE, do inglês *Mean Absolute Error*) e o erro absoluto médio relativo (MRAE, do inglês *Mean Relative Absolute Error*).

A Tabela 3.1 mostra os resultados dos diferentes métodos de previsão para os seis próximos meses do ano de 1979 ($h = 1, 2, 3, 4, 5, 6$).

Tabela 3.1: Previsões na clássica série “mortes acidentais”.

Horizonte	Valor real	Modelo I	Modelo II	HWS	ARAR	SSA
1	7798	8441	8345	8039	8168	7782
2	7406	7704	7619	7077	7196	7428
3	8363	8549	8356	7750	7982	7804
4	8460	8885	8742	7941	8284	8081
5	9217	9843	9795	8824	9144	9302
6	9316	10279	10179	9329	9465	9333
MAE	-	524	415	351	227	180
MRAE (%)	-	6	5	4	3	2

Nota. Fonte: Adaptado de Hassani (2007).

Através da Tabela 3.1 observa-se que as predições oriundas da técnica SSA são bem próximas aos dados reais (observados). Os valores das medidas MAE e MRAE indicam o ótimo desempenho da previsão SSA.

Nota-se que as previsões via SSA resultaram nas menores medidas de erro indicando melhor performance em relação aos demais métodos considerados na comparação.

A Tabela 3.1 mostra, por exemplo, que o MAE ou MRAE para a técnica SSA é três vezes menor do que o Modelo I e duas vezes menor do que o HWS.

Desta forma, para a série *mortes acidentais*, as comparações dos resultados revelam que a SSA gera previsões mais precisas do que aquelas oriundas dos métodos clássicos considerados.

No segundo artigo citado, os autores Hassani, Heravic e Zhigljavskya compararam a SSA, o SARIMA e o Holt-Winters para a previsão de oito indicadores da produção industrial em três países da Europa. Sendo assim, esses autores trabalharam com um total de 24 séries não ajustadas sazonalmente.

Os índices mensais da produção industrial foram nomeados como: *Food product*, *Chemicals*, *Basic metals*, *Fabricated metal*, *Machinery*, *Electrical machinery*, *Vehicles* e *Electricity and gas*. Cada um desses indicadores foram analisados para os seguintes países: Alemanha, França e Reino Unido.

As principais medidas de erro utilizadas foram as seguintes: raiz quadrada do erro quadrático médio (*RMSE*, do inglês *Root Mean Squared Error*); a razão entre o *RMSE* do SSA e *RMSE* do modelo concorrente (SARIMA ou Holt-Winters); e um escore que contabiliza o número de vezes em que a previsão SSA resulta no menor *RMSE* da comparação para cada horizonte considerado ($h = 1, 3, 6, 12$).

A Tabela 3.2 mostra uma parte da Tabela 3 do artigo Hassani, Heravic e Zhigljavskya (2009) para o país Alemanha. Na intenção de sintetizar os resultados apresentados no artigo ora citado considerou-se quatro dos oito índices mensais da produção industrial.

Tabela 3.2: Precisão das previsões para algumas séries da Alemanha.

Horizonte	$RRMSE_{(SSA/SARIMA)}$	$RRMSE_{(SSA/Holt-Winters)}$
Série <i>Food product</i>		
1	0.82	0.83
3	0.69	0.63
6	0.66	0.61
12	0.56	0.70
Série <i>Basic metals</i>		
1	1.24	1.13
3	1.01	0.95
6	0.73	0.58
12	0.44	0.37
Série <i>Vehicles</i>		
1	0.84	0.52
3	0.79	0.44
6	0.53	0.29
12	0.83	0.43
Série <i>Electricity and gas</i>		
1	0.94	0.92
3	0.71	0.71
6	0.67	0.68
12	0.61	0.61

Nota. Fonte: Adaptado de Hassani, Heravic e Zhigljavskya (2009).

Os autores mostraram que, de uma forma geral, a SSA superou o ARIMA e Holt-Winters na predição dos valores futuros das séries analisadas, de acordo com os critérios considerados, especialmente para os horizontes mais longos ($h = 3, 6$ e 12 meses); assim como exemplificado na Tabela 3.2.

Tabela 3.3: Escores para os países.

Horizonte	Reino Unido	Alemanhã	França
SSA/SARIMA			
1	5	5	7
3	5	5	6
6	7	8	7
12	8	8	7
SSA/Holt-Winters			
1	5	6	6
3	6	7	7
6	7	8	7
12	7	8	7

Nota. Fonte: Adaptado de Hassani, Heravic e Zhigljavskya (2009).

Assim como sintetizado na Tabela 3.3, nota-se que para qualquer horizonte a *RMSE* da SSA foi inferior a *RMSE* do modelo concorrente em mais da metade das séries analisadas. O menor escore observado foi igual a 5, retratando que a previsão SSA foi melhor em no mínimo 5 das oito séries da produção industrial.

Os resultados completos para os três países podem ser consultados na obra original dos autores Hassani, Heravic e Zhigljavskya (2009).

O próximo capítulo, intitulado “Desenvolvimento da Pesquisa”, apresenta a metodologia da presente dissertação.

Desenvolvimento da pesquisa

O desenvolvimento do trabalho para alcançar o objetivo geral da pesquisa é discutido neste capítulo. A descrição e características das séries temporais investigadas são apresentadas, isto é, os materiais adotados no trabalho; assim como o relato dos métodos considerados nas análises apresentadas no Capítulo 5.

Em suma, nesse capítulo encontram-se descritos a metodologia utilizada na presente dissertação.

4.1 *Materiais e métodos*

4.1.1 *Cenários- Descrição dos dados*

Com o propósito de comparar alguns métodos propostos na literatura para a análise de séries temporais definiu-se alguns cenários, para os quais modelou-se e estimou-se valores futuros de dois tipos de séries temporais reais (estudo empírico) e séries artificiais (estudo simulado).

Para o estudo empírico considerou-se séries temporais de duas distintas áreas do conhecimento, a saber: meteorologia e setor financeiro. A motivação para a escolha das séries contextualizadas em áreas diferentes relaciona-se a possibilidade de comparação de previsões para séries com dinâmicas variadas.

Adicionalmente, utilizou-se séries simuladas objetivando a avaliação do desempenho preditivo da técnica SSA em cenários já bem definidos na literatura especializada.

4.1.1.1 *Estudo empírico: Meteorologia*

Uma das séries analisadas no presente trabalho refere-se a dados meteorológicos, especificamente, a velocidade média do vento. Essa variável é a principal característica para avaliar o potencial de energia eólica de um certo local.

O potencial de energia eólica das principais regiões geográficas do Brasil encontra-se

descrito em Atlas Eólicos, como por exemplo, o Atlas do Potencial Eólico brasileiro e Atlas do Potencial Eólico da Bahia. Segundo informações disponíveis no sítio do Centro de Energia Eólica da Pontifícia Universidade Católica do Rio Grande do Sul (PUC/RS), em geral, grande parte do litoral brasileiro apresenta velocidades de vento adequadas para o aproveitamento de energia eólica em larga escala.

A Figura 4.1 mostra a distribuição da velocidade média anual na Bahia para estações meteorológicas com quatro diferentes patamares de altura; 10 metros, 30 metros, 50 metros e as maiores com, no máximo, 70 metros.

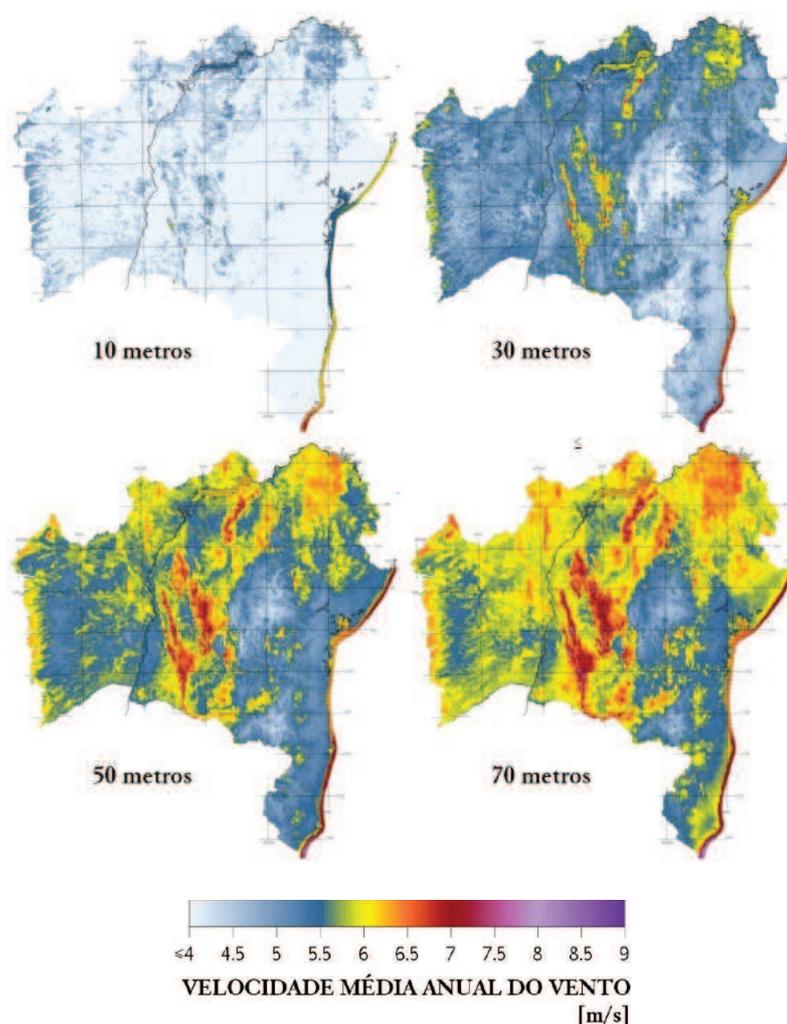


Figura 4.1: Potencial Eólico (Fonte: Atlas do Potencial Eólico do Estado da Bahia)

Normalmente, aplicações em larga escala requerem uma velocidade média de, no mínimo, 6,5 m/s a 7,5 m/s. Por outro lado, a utilização em sistemas pequenos exige uma velocidade média de 3,5 m/s a 4,5 m/s.

No contexto da presente pesquisa, utilizou-se a série temporal disponibilizada pelo setor responsável por uma mini estação meteorológica. Tal mini estação fica localizada na unidade CIMATEC do SENAI-BA e encontra-se instalada há uma altura de 36 metros

em relação ao nível do solo.

Os dados trabalhados representam os registros das médias das velocidades do vento, as quais foram armazenadas a cada intervalo de 30 minutos. A série considerada apresenta um histórico que teve início no dia 03 de setembro de 2010 às 12:00 e término às 11:30 do dia 12 de setembro de 2011. O que corresponde a 17892 observações no tempo.

As comparações das previsões foram realizadas em um conjunto de teste constituído pelas observações posteriores ao conjunto de treinamento citado no parágrafo anterior (03/09/10 às 12:00 - 12/09/2011 às 11:30). Especificamente, o conjunto de teste abrange o período de 12/09/11 às 12:00 até 07/11/2011 às 12:00; correspondendo a 2688 instâncias no tempo.

A série histórica meteorológica será rotulada como *VENTO* na correspondente aplicação.

4.1.1.2 *Estudo empírico: Setor Financeiro*

A outra série temporal é representada pelos índices diários da Bolsa de Valores de São Paulo (Bovespa). Historicamente, a Bovespa e as demais bolsas brasileiras eram entidades vinculada às secretarias de finanças da época (até meados da década de 60). Já na década de 2000, houve a integração de todas as bolsas de valores brasileiras e a Bovespa passou a ser a bolsa oficial do Brasil. Posteriormente, no ano de 2008, a Bovespa une-se à Bolsa de Mercadorias e Futuros (BM&F), surgindo uma nova companhia de capital brasileiro, denominada como BM&FBOVESPA.

Atualmente, a BM&FBOVESPA é a principal companhia brasileira de intermediação para operações do mercado de capitais. A sede dessa instituição encontra-se localizada no centro da cidade de São Paulo. O principal índice econômico da companhia é o conhecido Índice Bovespa (Ibovespa), que retrata o valor em moeda corrente de uma carteira teórica de ações estabelecida em 02/01/1968 (valor-base: 100 pontos). Em linhas gerais, o Ibovespa representa um indicador do desempenho médio das cotações do mercado de ações no Brasil, com base em uma aplicação hipotética; retratando a quantidade de investimentos em ações (compra ou venda).

No ano de 2011, empresas famosas como Abril Educação, Technos, Magazine Luiza e Arezzo deram início a negociação de ações na BM&FBOVESPA. A título de curiosidade, informações adicionais sobre a BM&FBOVESPA e a metodologia de construção do Índice Bovespa podem ser encontradas na *webpage* da companhia.

A série do Ibovespa utilizada neste trabalho abrange o período de 01 de fevereiro de 1996 a 03 de janeiro de 2011, mensalmente, perfazendo um total de 180 observações ao longo do período especificado. Considerou-se os valores brutos do índice, isto é, o número de pontos no final do primeiro dia útil de cada mês. Esses dados foram coletados na *homepage* do *Yahoo Finanças*.

A série temporal financeira aplicada neste trabalho será identificada como *IBOVESPA*. A escolha do período de tempo analisado não foi influenciada por indicações econômicas. Optou-se por considerar o histórico dos últimos 16 anos visando trabalhar com uma série mensal com poucos instantes de tempo, contrastando com a série temporal *VENTO*. Sendo que, as doze últimas observações (01 de fevereiro de 2011 a 02 de janeiro de 2012) não foram analisadas, estas formaram o conjunto de teste; útil para avaliar a precisão das estimativas.

4.1.1.3 Estudo simulado: Séries artificiais

O estudo simulado dividiu-se em dois contextos. No primeiro contexto, as séries temporais foram geradas de acordo com quatro diferentes processos estocásticos. Para o segundo contexto escolheu-se quatro séries artificiais disponibilizadas na *webpage* de um repositório de séries temporais intitulado *Time Series Data Library*, criado por Rob Hyndman. Em tal sítio pode-se encontrar as 3003 séries temporais utilizadas na clássica obra *M3-Competition*, de autoria do Makridakis e Hibon (2000).

Em ambos os contextos trabalhou-se com séries simuladas segundo os seguintes processos estocásticos: AR, MA, ARMA e ARIMA.

Assim, para o atual enfoque as séries temporais foram obtidas a partir de clássicos processos estocásticos, sendo assim dispensada a realização da análise exploratória completa em cada série simulada.

- Simulação 1

As simulações nesse cenário foram realizadas considerando a função `arima.sim` do software R, a qual gera séries da classe dos modelos ARIMA. A escolha dos parâmetros dos processos estocásticos fundamentou-se nos exercícios do capítulo 5 do livro do Morettin e Toloi (2006).

Para esse cenário considerou-se os processos estocásticos AR ($p=1$), MA ($q=2$), ARMA ($p = 1, q = 2$) e ARIMA ($p=1, d=1, q=2$). O único parâmetro da parte regressiva de

cada processo foi fixado em $\phi_1 = 0,4$ e os valores para os parâmetros de médias móveis foram: $\theta_1 = -0,3$ e $\theta_2 = 0,8$.

Antes de iniciar o processo de simulação fixou-se, através da função `set.seed` do R, o gerador de números aleatórios no valor 1986. De acordo com as configurações ora especificadas simulou-se 212 observações para cada um dos quatro processos, sendo que, as doze últimas instâncias foram alocadas para o conjunto de teste.

- Simulação 2

O segundo conjunto de séries artificiais foi considerado com o intuito de ter uma base de dados clássica e acessível. O autor do repositório *Time Series Data Library* disponibiliza várias séries artificiais com as respectivas obras de referência, contudo não discrimina todos os valores dos parâmetros utilizados na simulação.

Dentre as séries artificiais do sítio escolheu-se quatro oriundas de exemplos do Brockwell e Davis (1991), rotuladas como E921, E1042, E923 e E951; correspondendo aos processos estocásticos AR ($p=2$), MA ($q=1$), ARMA ($p=2, q=1$) e ARIMA ($p=1, d=2, q=1$), respectivamente.

As séries advindas dos processos AR, ARMA e ARIMA possuem 200 instâncias, enquanto que a série oriunda do processo MA contém 160 observações. Para esses quatro processos considerou-se também os 12 últimos instantes de tempo na formação do conjunto de teste.

4.1.2 Métodos

Com o intuito de avaliar as capacidades preditivas dos distintos métodos considerou-se diferentes horizontes de previsão. Para a extensa série meteorológica, os seguintes passos foram adotados: 1, 24, 48, 96, 192, 336, 672, 1008, 1344, 2016, 2688. Enquanto que as predições na série IBOVESPA e simuladas são sintetizadas considerando no máximo 12 passos à frente.

As predições são comparadas mediante as medidas de precisão descritas na Subseção 2.4.3, a saber: erro quadrático médio (*MSE*) e o erro percentual total (*TPE*); além do erro cometido na estimação para cada horizonte.

- Clássicos

Os algoritmos aditivo e multiplicativo de Holt-Winters foram considerados na fase de modelagem e previsão das séries temporais empíricas. A escolha dos valores dos parâmetros de alisamento foi realizada de acordo com o procedimento objetivo descrito na Seção 2.4.1, isto é, utilizando a minimização da soma dos quadrados dos erros de previsão a um passo à frente.

A referida minimização dos erros de previsão ocorre de forma automática através da função `HoltWinters` do `R`. Todos os parâmetros que alimentam essa função foram considerados no seu formato padrão. Todavia o parâmetro `seasonal` de tal função foi alternado para contemplar as duas formas do algoritmo (aditivo e multiplicativo).

Salienta-se que, a forma multiplicativa do algoritmo de Holt-Winters não foi possível ser ajustada à série meteorológica, porque a correspondente série temporal possui valores nulos.

A modelagem clássica nas séries artificiais foi realizada a partir da Suavização exponencial de Holt (SEH), ao considerar a ausência do parâmetro γ na função `HoltWinters`, e pelo correspondente modelo da classe ARIMA.

Nas análises do estudo simulado utilizou-se a SEH ao invés do H-W, porque as séries simuladas não apresentam comportamentos sazonais, assim como será apresentado na Seção 5.3 do Capítulo 5.

Cada modelo de Box-Jenkins foi considerado como um `padrão ouro`, isto é, cada modelo da classe ARIMA servirá como técnica de comparação por parte dos outros métodos adotados. Pois, espera-se que a predição realizada a partir dos dados oriundos do processo estocástico puro tenha maior precisão do que as demais técnicas.

A modelagem através do enfoque Box-Jenkins ocorre de acordo com um ciclo de quatro estágios: especificação, identificação, estimação e diagnóstico de um modelo; assim como explicado na Seção 2.4.2. A especificação dos modelos fundamentou-se na análise exploratória das séries.

Conforme será apresentado na Seção 5.1, as séries temporais VENTO e IBOVESPA não são estacionárias, estas exibem tendência e sazonalidade. Desta forma, o método de Box-Jenkins mais adequado para modelar as fontes de variações nessas séries será o SARIMA. Assim como será apresentado na análise exploratória (Subseção 5.1), ambos estudos empíricos precisarão de apenas uma diferença para tornarem processos integrados $[I(1)]$, desta forma, especificou-se a classe do modelo SARIMA $(p,1,q) \times (P,D,Q)$.

A fase de identificação do modelo foi realizada à luz dos critérios AIC e BIC. Nessa fase ocorreu a seleção dos modelos mais adequados para representar o processo gerador das séries.

Para o processo de seleção dos modelos utilizou-se uma função externa ao *software* R que minimiza o valor do AIC ou BIC para vários modelos diferentes. Esta função seleciona automaticamente as ordens p e q dos modelos, mediante a fixação das demais ordens do SARIMA, isto é, d , P , D e Q . Detalhes sobre essa função de seleção podem ser encontradas em Frery e Cribari-Neto (2005).

Para a seleção dos modelos nas análises do presente trabalho considerou-se que p e q assumiria no máximo valor 3. Seguindo o princípio da parcimônia, o processo de seleção iniciou-se com o modelo mais simples, a saber: SARIMA($p,1,q$) $\times(0,0,0)$; que corresponde ao modelo ARIMA($p,1,q$).

- SSA

Objetivando a comparação entre o enfoque SSA e a modelagem das séries através das técnicas clássicas adotadas no trabalho, realizou-se a decomposição e a reconstrução dos componentes constituintes das séries. Tais análises foram sucedidas ao comparar o erro quadrático médio dos resíduos gerados por cada um dos métodos clássicos em relação ao ruído classificado após a decomposição SSA.

Assim como discutido no Capítulo teórico 3, a modelagem via SSA ocorre mediante a escolha dos seus dois únicos parâmetros: o comprimento da janela L , no estágio de decomposição, e o processo estrutural de agrupamento, na reconstrução da série temporal.

No presente trabalho de dissertação a seleção dos parâmetros foi realizada à luz das informações preliminares sobre a série temporal, obtidas na fase da análise exploratória, e das informações adicionais apresentadas na Seção 3.3.

Na análise via SSA considerou-se o comprimento ótimo da janela para os estudos empírico e simulado, com exceção da série temporal VENTO. Pois o L_{timo} para essa última série ficou computacionalmente intensivo.

A decomposição SSA na série VENTO levou em consideração um comprimento da janela de tamanho razoável, em relação ao comprimento da série (N). A escolha do valor para L ocorreu à luz do período sazonal da série. Isto é, escolheu-se um comprimento múltiplo da sazonalidade. O L máximo processado em um computador pessoal foi de $L = 48 \times 84 = 4032$.

A separação entre o sinal e o ruído na fase de reconstrução das séries foi avaliada de acordo com os valores singulares (subseções 3.3.3 e 3.3.3.1) e/ou a matriz de correlação ponderada (Subseção 3.3.3.2).

Além dos critérios ora citados, investigou-se também o percentual acumulado de explicabilidade dos possíveis componentes formadores do sinal, assim como o comportamento dos pares de autovetores; visando facilitar o processo de agrupamento e alcançar a propriedade da separabilidade.

5.1 Estudos empíricos: Análise exploratória

5.1.1 VENTO

As características da série temporal *VENTO* podem não ser identificadas facilmente através do seu gráfico temporal. Assim como apresentado na Figura 5.1, nota-se uma grande massa de dados ao longo dos instantes de tempo.

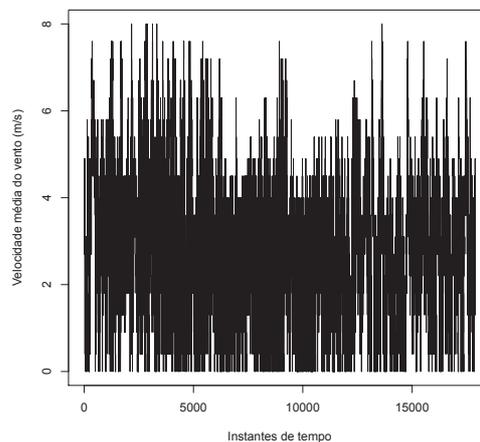


Figura 5.1: Velocidade média do vento ao longo do tempo.

Para o período considerado, a série *VENTO* apresentou uma média de 2,87 m/s e desvio padrão de 1,65 m/s. Observou-se também que a série possui um coeficiente de variação de 57,49%; um valor máximo igual a 8,00 m/s; e velocidade média mínima de 0,00 m/s.

Informações adicionais sobre a distribuição de frequência da série em análise podem ser obtidas através da Figura 5.2.

A partir do histograma nota-se que a faixa de velocidade média mais frequente corresponde a classe modal 3,5 - 4,0 m/s; além disso, pode-se observar que as maiores frequências correspondem às menores velocidades médias. O valor mediano para a velocidade média do vento foi de 2,70 m/s. As principais medidas descritivas são exibidas no *box-plot* da Figura 5.2.

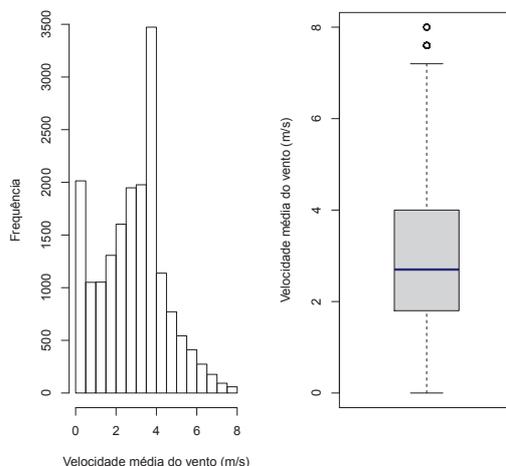


Figura 5.2: Distribuição da velocidade média do vento (histograma à esquerda e *box-plot* à direita).

De acordo com o *box-plot* (grafico à direita da Figura 5.2), pode-se observar que as duas maiores velocidades do vento (8,00 m/s e 7,60 m/s) são apontadas como *outliers*. Ao analisar esses valores ditos “atípicos” notou-se um total de 59 distintos instantes no tempo, dos quais 45 (76,27%) referem-se à diferentes dias da estação primavera; sendo que a maioria dos instantes correspondem a novembro de 2010 (66,67%). Os outros 14 instantes de tempo condizem a 12 dias do inverno e 2 dias do verão.

Ao avaliar a suposição de estacionariedade o teste KPSS não indicou nível estacionário nem tendência estacionária com 95 % de confiança (p -valor $< 0,01$). Contudo, o teste de Dickey-Fuller refuta a hipótese de raiz unitária, isto é, a hipótese de não estacionariedade (p -valor $< 0,01$).

Com o intuito de obter mais investigações sobre a estacionariedade avaliou-se o comportamento da função de autocorrelação amostral na série VENTO. O primeiro gráfico da Figura 5.3 apresenta o correlograma da série para uma quantidade de 20 defasagens no tempo e o segundo gráfico mostra uma ampliação do gráfico anterior, exibindo as autocorrelações estimadas até o *lag* de ordem 3.

Através da Figura 5.3 verifica-se que o correlograma apresenta oscilações ao longo das defasagens. Em um primeiro momento, as autocorrelações estimadas decrescem lentamente e em seguida observa-se um crescimento gradual.

Em complemento à análise visual do correlograma, aplicou-se os testes de Box-Pierce e de Ljung-Box. Ambos os testes apontaram a existência de autocorrelação conjunta estatisticamente significativa, ou seja, os coeficientes de autocorrelação não são simultaneamente iguais a zero ($p < 2,2 \times 10^{-16}$).

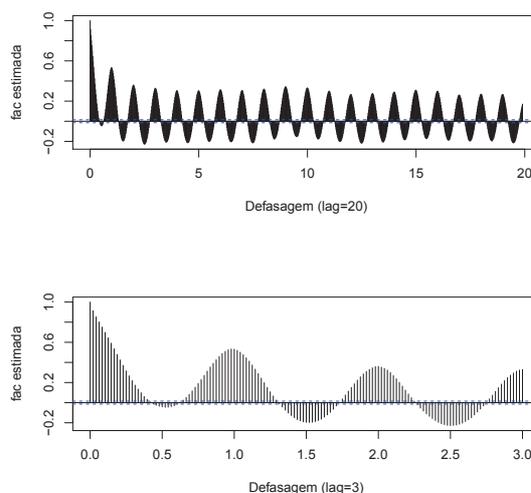


Figura 5.3: Correlograma (lag= 20 e lag= 3).

De acordo com o comportamento do correlograma e dos testes para autocorrelação conjunta tem-se indícios de que a série *VENTO* não é estacionária.

Dando prosseguimento à análise exploratória, ao investigar a decomposição clássica na série *VENTO*, nota-se que um modelo multiplicativo apresentou menores resíduos, em geral, do que um modelo aditivo (Figura 5.4). Observou-se que uma decomposição a partir do modelo multiplicativo gera um erro quadrático médio (*MSE*) ligeiramente menor do que o resultante do modelo aditivo (1,22 e 1,25, respectivamente). Sendo que, o erro mínimo observado na forma multiplicativa foi 0,00 m/s e o máximo de 6,47 m/s; em contraste, a decomposição aditiva resultou em um valor mínimo e máximo de $-4,65$ m/s e 5,26, respectivamente.

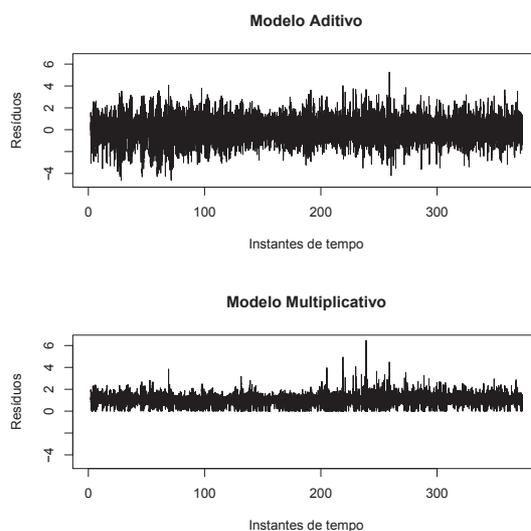


Figura 5.4: Resíduos oriundos da decomposição clássica (modelo aditivo e multiplicativo).

A Figura 5.5 apresenta a decomposição da série VENTO através de uma técnica de alisamento conhecida como STL, do inglês *Seasonal Decomposition of Time Series by Loess* (*Local Polynomial Regression Fitting*); ver Cleveland e colaboradores (1990). Nesta figura observa-se a estimação da sazonalidade e da tendência; além dos resíduos que são obtidos pela diferença entre a série original e a combinação das estimativas dos dois componentes citados inicialmente (sazonalidade e tendência).

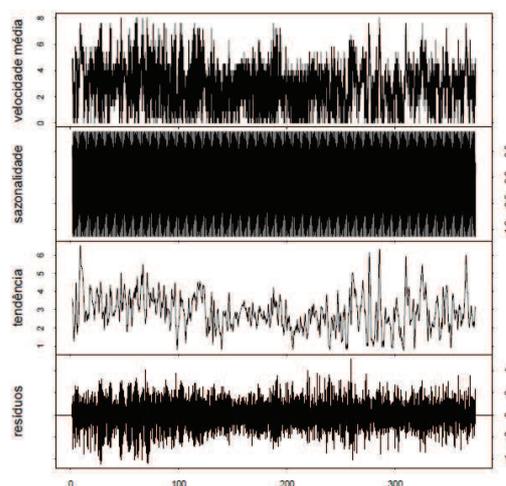


Figura 5.5: Decomposição da série VENTO ao longo do tempo.

Através da decomposição realizada na série VENTO, verifica-se que há um padrão sazonal e uma tendência linear com comportamentos simples (Figura 5.5).

Com o intuito de facilitar a visualização da periodicidade da série, considerou-se um recorte da sazonalidade da série temporal VENTO. A Figura 5.6 apresenta duas ampliações do componente sazonal, em que os comportamentos do componente periódico para os 336 e 1344 instantes de tempo iniciais são mostrados.

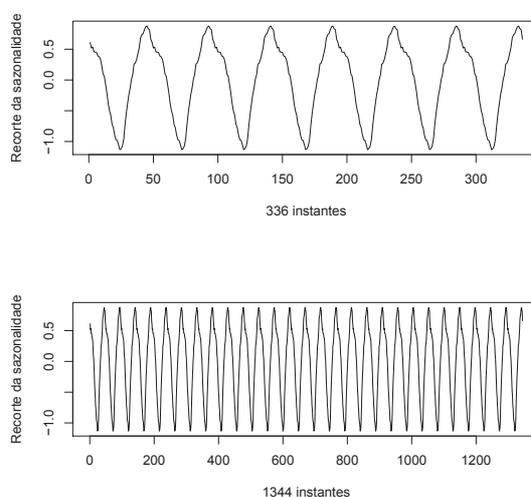


Figura 5.6: Sazonalidade da série VENTO para os primeiros 336 e 1344 instantes de tempo.

Através da Figura 5.6 é possível notar a periodicidade da série VENTO. Verifica-se uma repetição no comportamento sazonal da série, esta repetição tem uma amplitude igual a 48; correspondendo a um período de 48 instantes de tempo por dia.

Com relação aos resíduos gerados na decomposição STL, foi possível observar um MSE de 1,13, resíduo mínimo de $-4,41$ e máximo igual a 5,07.

De acordo com os testes Cox-Stuart e Spearman, a série VENTO apresenta tendência ($p < 2,2 \times 10^{-16}$ para ambos os testes). Os testes para sazonalidade considerados (Kruskal-Wallis e de Friedman) indicaram a presença da componente sazonal ($p < 2,2 \times 10^{-16}$ para os dois testes).

A fim de uma possível identificação de um modelo estocástico analisou-se o comportamento da função autocorrelação (fac) amostral e função autocorrelação parcial (facp) amostral. Na Figura 5.7 encontram-se os gráficos da fac e facp estimadas para a série VENTO.

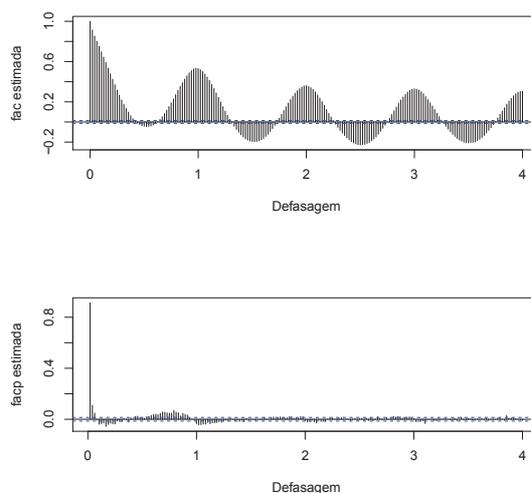


Figura 5.7: Função autocorrelação (fac) e função autocorrelação parcial (facp) da série VENTO.

Como apresentado na Seção 2.4.2, as ordens de um modelo de médias móveis $[MA(q)]$ e de um processo $AR(p)$ são definidas pela sua fac e facp, respectivamente. Sendo assim, é necessário buscar um corte no ponto q da fac amostral, para um modelo de médias móveis, ou no ponto p da facp (para um modelo autoregressivo).

Nota-se que nenhuma das funções de autocorrelação (fac ou facp) indicam um corte na sua ordem. Ou seja, não é observada a permanência de todas as autocorrelações amostrais dentro do intervalo assintótico $\left[-1, 96 \frac{1}{\sqrt{N}}; 1, 96 \frac{1}{\sqrt{N}}\right]$ a partir de algum ponto, (Figura 5.7).

Desta forma, um processo autoregressivo ou de médias móveis puro não é suficiente para explicar o processo gerador da série VENTO.

A Figura 5.8 mostra a aplicação de uma e duas diferenças na série VENTO. Graficamente, não é possível notar grandes distinções entre as séries diferenciadas com $lag=1$ e $lag=2$.

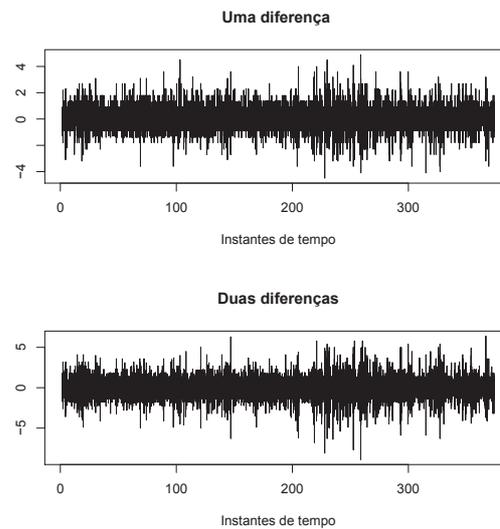


Figura 5.8: Diferenciação na série VENTO.

Ao investigar a quantidade de diferenças necessárias para tornar a série temporal das velocidades médias do vento estacionária, observou-se que uma diferenciação foi suficiente (Figura 5.8).

Os testes para a avaliação de estacionariedade resultaram nos seguintes p -valores: $p = 0,1$ para ambas as formas do KPSS (nível e tendência) e $p = 0,01$ para o teste de Dickey-Fuller. Esses resultados indicam que a série VENTO pode ser considerada um processo $I(1)$, isto é, integrável de ordem 1.

5.1.2 IBOVESPA

As principais características da série do Ibovespa considerada apresenta uma estrutura com tendência global não-linear e sazonalidade complexa. Essas características podem ser visualizadas através da Figura 5.9. Tal figura mostra claramente que a série histórica não tem um comportamento estacionário.

A série IBOVESPA possui uma média e desvio padrão amostral de 27807 e 20725 pontos, respectivamente. E o coeficiente de variação da série fica em torno de 74,5%, o que representa uma dispersão da série relativamente alta.

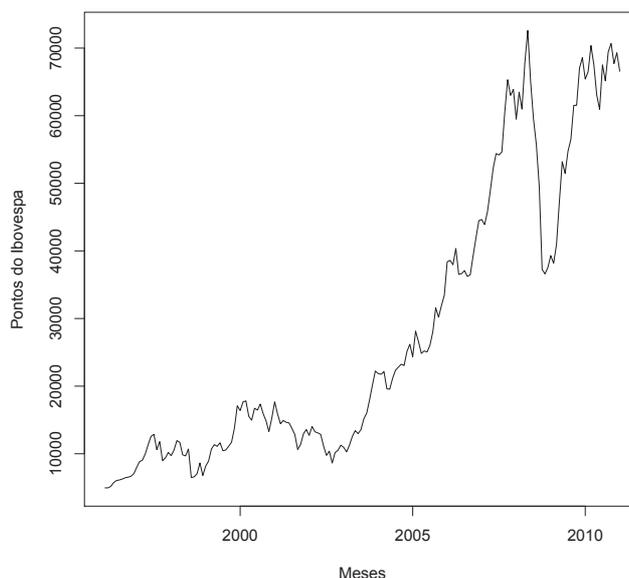


Figura 5.9: Número de pontos do Ibovespa no final do primeiro dia útil de cada mês ao longo do tempo (01 de fevereiro de 1996 a 03 de março de 2011).

Ao longo do período considerado observou-se que a maior quantidade de pontos do índice corresponde ao mês de maio de 2008 (72593 pontos), e o menor valor refere-se ao mês de março de 1996 (4955 pontos). Sendo que, a partir do maior pico (valor máximo) da série nota-se uma queda brusca nos pontos mensais estendendo até novembro de 2008; fato associado ao fenômeno histórico da crise econômica de 2008.

Adicionalmente, apesar do gráfico temporal (Figura 5.9) não mostrar estacionariedade na série, pode-se avaliar essa clássica suposição através dos testes estatísticos citados na Subseção 2.3.3.1. Ao aplicar as duas formas de avaliação do teste KPSS, com 95 % de confiança, rejeita-se a hipótese de nível estacionário, assim como a hipótese de tendência estacionária (p -valor $< 0,01$). Conclusão similar é obtida ao aplicar o teste de Dickey-Fuller, a partir do qual não há evidências para rejeitar a hipótese de raiz unitária (p -valor = 0,36). Ou seja, como esperado, a série histórica não apresenta estacionariedade.

Ao aplicar a decomposição clássica por Média Móveis na série de dados; ver Kendall e Stuart (1983), pode-se verificar que um modelo multiplicativo mostra-se mais adequado aos dados do que um aditivo. Pois, os resíduos provenientes da decomposição de um modelo multiplicativo são menores do que aqueles oriundos de um modelo aditivo (Figura 5.10). Ao sintetizar os resíduos, observa-se um MSE igual a 1,00 para um modelo multiplicativo, apresentando um erro mínimo e máximo de 0,75 e 1,23, respectivamente; enquanto que, uma decomposição aditiva gera um MSE de $7,395818 \times 10^6$ (mínimo = -10624,06 e máximo=12942,86).

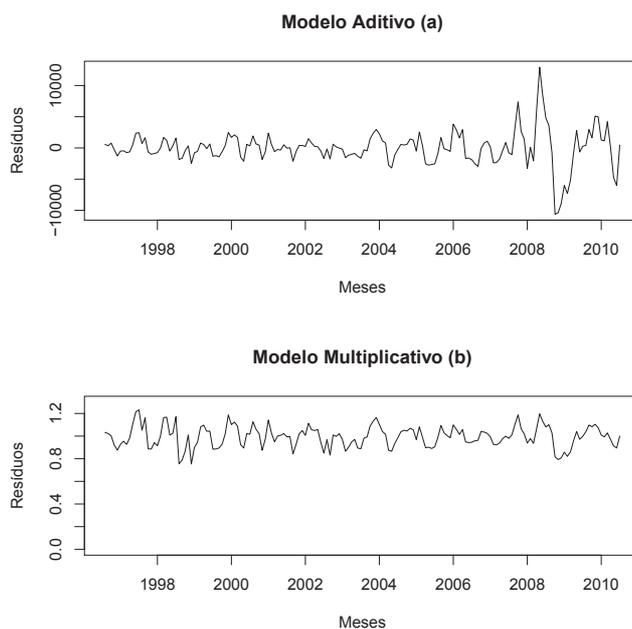


Figura 5.10: Resíduos para os modelos aditivo e multiplicativo.

De acordo com as medidas descritivas e o gráfico dos resíduos provenientes da forma aditiva (a), observa-se que a sua distribuição não é homogênea ao longo do tempo, principalmente nos últimos instantes de tempo considerados. Por outro lado, o gráfico dos resíduos do modelo multiplicativo (b) mostra pequenas flutuações em torno da média, que é igual a 0,9951.

Com o intuito de facilitar a visualização do comportamento dos resíduos oriundos do modelo aditivo, separou-se tais resíduos em duas partes. A segunda parte foi formada pelas últimas instâncias, representando um grupo de resíduos com altas variações em relação a um valor central. Já para a primeira parte, considerou-se a parcela mais concentrada dos resíduos (Figura 5.11).

O primeiro gráfico (a) da Figura 5.11 representa os resíduos do modelo aditivo correspondente ao período de agosto de 1996 até o mês de setembro 2007. E o segundo gráfico (b) da mesma figura retrata os resíduos mensais de outubro de 2007 à julho de 2010.

Segundo os gráficos da Figura 5.11, observa-se facilmente que há uma tendência de crescimento na amplitude dos resíduos no decorrer do tempo.

A decomposição da série IBOVESPA através do método STL é apresentada na Figura 5.12. As estimativas da sazonalidade, da tendência e dos resíduos são exibidas nesta sequência.

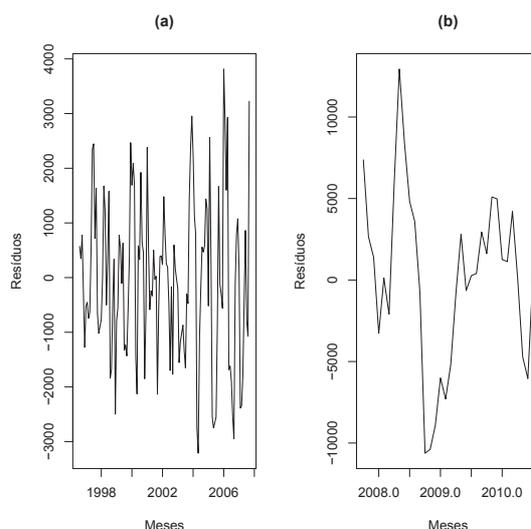


Figura 5.11: Separação dos resíduos provenientes do modelo aditivo.

Os resíduos resultantes da decomposição STL apresentam as seguintes medidas: $MSE=6,37912 \times 10^5$; Mínimo= $-1,041 \times 10^4$ e Máximo= $1,249 \times 10^4$.

Nota-se, através dos gráficos e medidas de sínteses, semelhanças entre os resíduos da decomposição aditiva e os resíduos provenientes do STL. Desta forma, nota-se que a decomposição envolvendo um modelo multiplicativo conseguiu captar melhor as variações da série IBOVESPA.

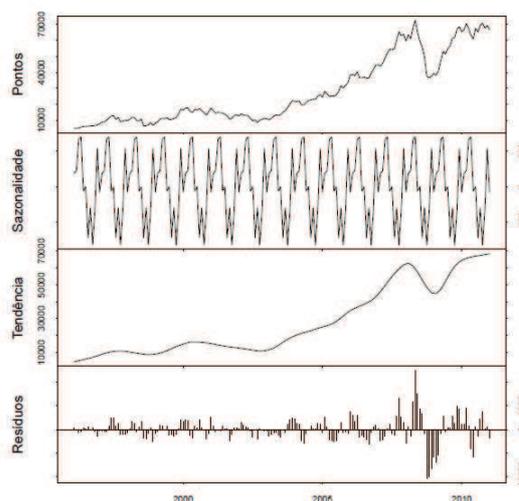


Figura 5.12: Decomposição da série IBOVESPA ao longo do tempo.

Prosseguido com a análise exploratória, aplicou-se testes estatísticos para avaliar a existência de tendência e sazonalidade. Ambos os testes para a detecção de tendência descritos na equação 2.3.3.2 foram aplicados na série IBOVESPA sem a parte sazonal; assim como a presença de sazonalidade foi testada na série livre de tendência. Os testes de Cox-Stuart e de Spearman indicaram a existência de tendência ($p < 2,2 \times 10^{-16}$

para os dois testes). Apesar dos métodos de decomposição apontarem estimativas para a componente sazonal, os testes de Kruskal-Wallis e de Friedman não conseguiram validar a presença de sazonalidade (p -valor = 0,216 e p -valor = 0,262, respectivamente).

A Figura 5.13 mostra o correlograma ou função de autocorrelação amostral (fac estimada) e a função de autocorrelação parcial amostral (facp estimada). Assim como comentado na Seção 2.4.2, o correlograma e facp estimada pode ser utilizados para auxiliar na identificação do processo estocástico gerador dos dados.

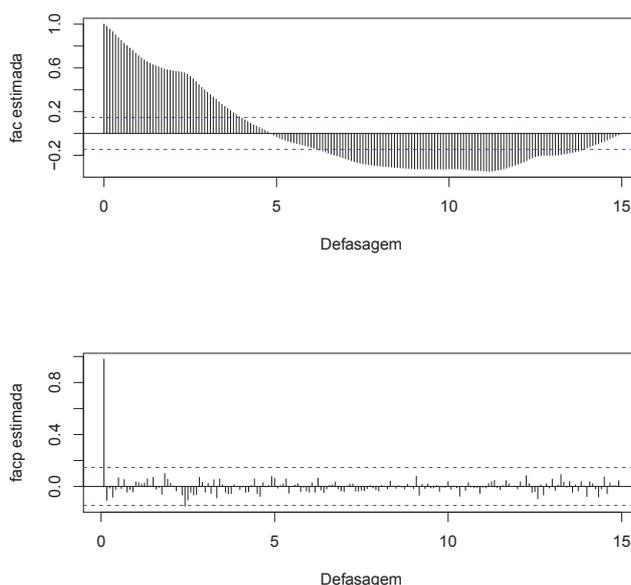


Figura 5.13: Correlograma (fac estimada) e a função de autocorrelação parcial amostral (facp estimada) da série IBOVESPA.

A função de autocorrelação amostral da série IBOVESPA apresenta um comportamento semelhante a de uma senóide. E a facp estimada revela um corte no ponto $p = 1$, ou seja, observa-se que apenas a primeira autocorrelação é estatisticamente diferente de zero, ao considerar 95% confiança. Sendo assim, a fac e facp estimadas exibem padrões similares a de um processo autoregressivo de primeira ordem [AR(1)] (Figura 5.13).

Como a série IBOVESPA não é estacionária, precisa-se investigar o comportamento das séries defasadas no tempo. Na Figura 5.14 são apresentadas a primeira e a segunda diferença da série IBOVESPA.

Verifica-se, através da Figura 5.14, que com apenas uma diferença a série IBOVESPA já aparenta ser estável. Os testes Dickey-Fuller e KPSS validaram que a série defasada com uma diferença é estacionária ($p < 0, 1$ e p -valor = 0,01, respectivamente).

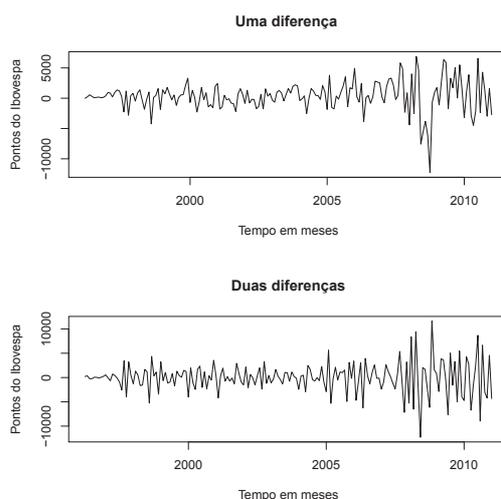


Figura 5.14: Primeira e segunda diferença da série IBOVESPA.

5.2 Estudos empíricos: Modelagem e previsão

Na atual seção são apresentados os resultados do processo de modelagem e previsão das séries temporais investigadas nesse trabalho.

As previsões são comparadas mediante as duas medidas de precisão descritas na Subseção 2.4.3, a saber: erro quadrático médio (MSE) e o erro percentual total (TPE).

5.2.1 Clássica

Na presente subseção serão estudados alguns modelos clássicos candidatos a representar a dinâmica das séries analisadas. E a partir desses modelos as previsões serão obtidas.

5.2.1.1 Alisamento Exponencial de Holt-Winters

- VENTO

As previsões geradas pelo método de Holt-Winters, na sua forma aditiva, os valores reais do conjunto de teste e os erros de previsão para os horizontes considerados são apresentadas na Tabela 5.1.

Tabela 5.1: Previsões provenientes da forma aditiva do Holt-Winters (VENTO).

h	Valor real	HW Aditivo	Erro
1	3,100	3,512	-0,412
24	3,600	3,041	0,559
48	2,200	3,590	-1,390
96	3,600	3,596	0,004
192	5,800	3,607	2,193
336	5,800	3,625	2,175
672	4,900	3,665	1,235
1008	6,700	3,705	2,995
1344	4,000	3,745	0,255
2016	3,600	3,825	-0,225
2688	4,500	3,905	0,595

De acordo com a Tabela 5.1, nota-se que quase todas as previsões advindas do algoritmo sazonal aditivo subestimaram os valores reais da série, com exceção das previsões referentes aos horizontes $h = 1$, $h = 48$ e $h = 2016$.

A distribuição das previsões oriundas do algoritmo de Holt-Winters aditivo apresentou uma predição mínima de 1,99 m/s e máxima de 4,49 m/s. A média e mediana da referida distribuição foram bem próximas, a saber: 3,18 m/s e 3,14 m/s, respectivamente. As principais medidas descritivas estão resumidas na Subseção 5.2.3 (Tabela 5.8).

Na Figura 5.15 pode ser visualizada a distribuição do conjunto de teste, nomeada como “Valor real”, e das previsões oriundas da forma aditiva do método de Holt-Winters (H-W).

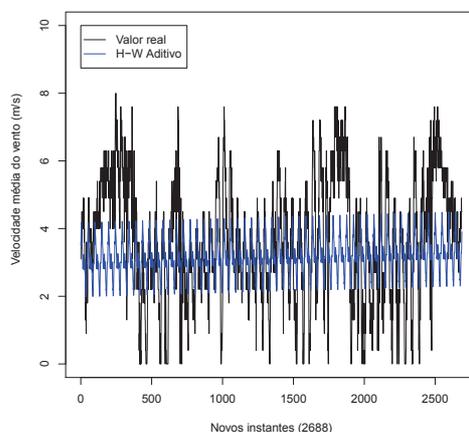


Figura 5.15: Velocidade média observada e prevista pela forma aditiva do algoritmo de H-W (VENTO).

Segundo a Figura 5.15, observa-se que as previsões via Holt-Winters aditivo comportam-se em torno da sua média e com menores flutuações, em relação ao conjunto de teste, ao longo dos 2688 instantes de tempo.

Sendo assim, pode-se notar que as predições do algoritmo de Holt-Winters encontram-se limitadas em uma pequena faixa de variação. Desta forma, observa-se uma clara limitação no poder preditivo desse algoritmo para a série VENTO, haja visto que os verdadeiros valores futuros são bem dinâmicos.

- IBOVESPA

As previsões resultantes do algoritmo de Holt-Winters aditivo e multiplicativo encontram-se comparadas aos valores observados da série IBOVESPA na Tabela 5.2. Nesta tabela são exibidas as previsões de fevereiro de 2011 a janeiro de 2012, contabilizando 12 passos à frente.

Através da Tabela 5.2, nota-se que apenas para os dois primeiros horizontes ($h = 1$ e $h = 2$) há subestimação nas predições geradas pelo método sazonal (aditivo ou multiplicativo). Assim como esperado, os erros absolutos das predições crescem com o aumento do horizonte de previsão.

Tabela 5.2: Previsões obtidas por Holt-Winters aditivo e multiplicativo (IBOVESPA).

h	Valor real	HW aditivo	Erro 1	HW multiplicativo	Erro 2
1	67383	67370,06	12,94	66923,98	459,02
2	68587	67574,69	1012,31	68197,87	389,13
3	66133	68581,05	-2448,05	69922,31	-3789,31
4	64620	70124,64	-5501,64	70422,35	-5802,35
5	62404	71492,87	-9088,87	69361,66	-6957,66
6	58823	71985,68	-13162,68	73489,21	-14666,21
7	56495	69820,60	-13325,60	73850,75	-17355,75
8	52324	70069,92	-17745,92	75642,14	-23318,14
9	58338	70178,58	-11840,58	71226,20	-12888,20
10	56875	70270,18	-13395,18	71374,23	-14499,23
11	56754	70531,58	-13777,58	73236,19	-16482,19
12	59147	71296,84	-12149,84	70940,04	-11793,04

Ainda na Tabela 5.2, verifica-se facilmente que quase todos os erros absolutos provenientes do método sazonal aditivo foram inferiores aos erros gerados pelo correspondente

algoritmo multiplicativo, com exceção dos erros referentes aos meses de março e junho de 2011, além do mês mais atual - janeiro de 2012.

Na Figura 5.16 tem-se um gráfico comparativo incluindo as previsões advindas dos métodos Holt-Winters sazonal aditivo e multiplicativo, além dos valores reais da série IBOVESPA.

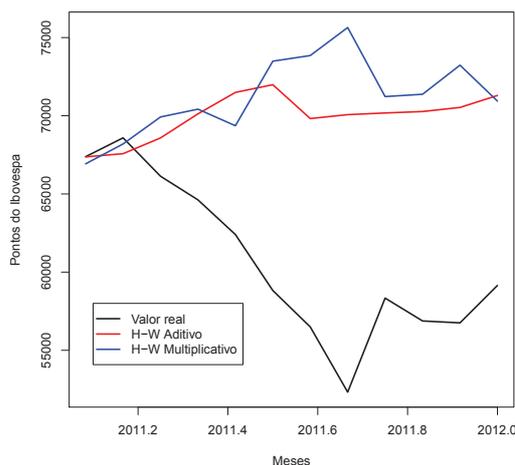


Figura 5.16: Previsões via Holt-Winters comparadas com os valores reais (IBOVESPA).

De forma geral, as previsões oriundas do método H-W sazonal aditivo aproximam-se mais dos valores reais da série IBOVESPA; apesar de não conseguir detectar satisfatoriamente o padrão de variação dessa série.

5.2.1.2 Método de Box-Jenkins

- VENTO

Em virtude da grande quantidade de observações na série meteorológica, a função de seleção, externa ao *software* R, não conseguiu ser executada. Desta forma, considerou-se apenas os modelos ajustados aos dados através da função padrão, interna ao R, que utiliza o AIC como critério de seleção.

Os dois modelos de Box-Jenkins que apresentaram menores AIC foram os seguintes: SARIMA(2,1,2)x(1,1,1); com AIC de 35394,05 e SARIMA(3,1,3)x(1,1,1); com AIC de 35389,52.

Com o intuito de facilitar a representação tabular e gráfica, os dois melhores modelos foram rotulados como: SARIMA 1 e SARIMA 2 para SARIMA(2,1,2)x(1,1,1) e

SARIMA(3,1,3) \times (1,1,1), respectivamente.

As previsões para os diferentes horizontes considerados são mostradas na Tabela 5.3.

Tabela 5.3: Previsões de Box e Jenkins (VENTO).

h	Valor real	SARIMA 1	Erro 1	SARIMA 2	Erro 2
1	3,100	3,470	-0,370	3,480	-0,380
24	3,600	2,690	0,910	2,715	0,885
48	2,200	3,703	-1,503	3,745	-1,545
96	3,600	3,708	-0,108	3,757	-0,157
192	5,800	3,707	2,093	3,759	2,041
336	5,800	3,704	2,096	3,762	2,038
672	4,900	3,699	1,201	3,769	1,131
1008	6,700	3,693	3,007	3,777	2,923
1344	4,000	3,687	0,313	3,784	0,216
2016	3,600	3,675	-0,075	3,798	-0,198
2688	4,500	3,664	0,836	3,812	0,688

A partir da Tabela 5.3, observa-se que os horizontes 1, 48, 96 e 2016 apresentam previsões superestimadas pelos dois modelos selecionados (SARIMA 1 e SARIMA 2). Nota-se também que para esses mesmos passos o SARIMA 1 exibe menores erros absolutos; indicando que, na maioria dos horizontes considerados, o SARIMA 2 apresenta melhor desempenho.

A performance global de cada modelo SARIMA, em relação ao conjunto de teste, pode ser observada na Figura 5.17.

A Figura 5.17 exibe concomitantemente todas as 2688 previsões advindas dos modelos SARIMA 1 e SARIMA 2. Observa-se que esses modelos apresentam poucas flutuações ao longo dos 2688 instantes de tempo, os quais não atingem toda a amplitude da real distribuição das velocidades médias do vento.

De forma geral, nota-se que os modelos SARIMA 1 e SARIMA 2 possuem comportamentos bem similares. Tal interpretação pode ser ratificada através das principais medidas descritivas mostradas na Subseção 5.2.3 (Tabela 5.8).

O diagnóstico do ajuste de ambos os modelos selecionados foi realizado através da avaliação dos seus resíduos e encontra-se sintetizado na Figura 5.18.

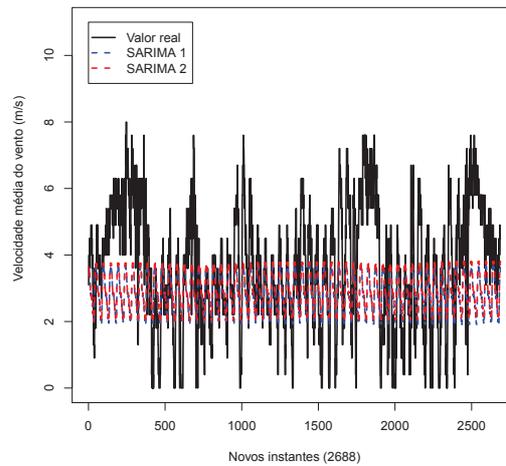


Figura 5.17: Previsões do SARIMA 1 e SARIMA 2 comparadas com os valores reais (VENTO).

A Figura 5.18 mostra dois conjuntos de gráficos, um para cada modelo SARIMA. Em cada conjunto, o primeiro gráfico expõe a distribuição dos resíduos ao longo do tempo (Resíduos padronizados); o segundo mostra o correlograma dos resíduos e o último gráfico exhibe os p -valores do teste de Ljung-Box para 10 diferentes $lags$.

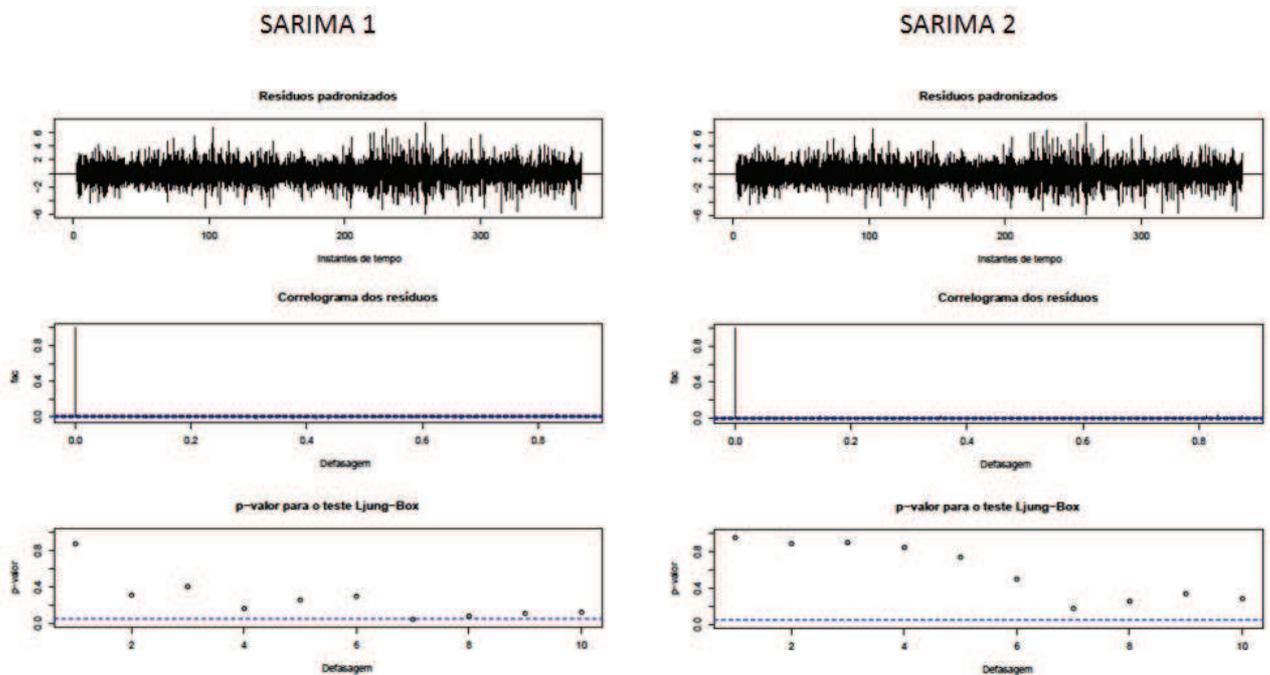


Figura 5.18: Diagnóstico dos modelos SARIMA 1 e SARIMA 2 (VENTO).

Os gráficos da Figura 5.18 mostram que, de forma geral, ambos os modelos apresentam resíduos com pequenas amplitudes. De acordo com o correlograma e teste de Ljung-Box,

observa-se que os resíduos não apresentam grandes problemas de autocorrelação residual.

- IBOVESPA

Para a série IBOVESPA, de acordo com o critério AIC, os dois melhores modelos foram: SARIMA(3,1,2)×(1,1,1); com AIC=3115,23 e SARIMA(2,1,3)×(1,1,2); com AIC=3119,70. Por outro lado, a função de seleção, alimentada pelo critério BIC, indicou os seguintes modelos: SARIMA(0,1,0)×(1,1,1) e SARIMA(1,1,0)×(1,1,2); com BIC=3125,75 e BIC=3130,62, respectivamente.

Diante das quatro indicações objetivas escolheu-se dois modelos considerando o bom senso da analista. Inicialmente decidiu-se pela adoção de um modelo indicado por cada critério. Os modelos foram: SARIMA(2,1,3)×(1,1,2) e SARIMA(1,1,0)×(1,1,2); sugerido pelo AIC e BIC, respectivamente.

Posteriormente, ao realizar o diagnóstico dos modelos (análise dos resíduos), verificou-se que o SARIMA(3,1,2)×(1,1,1) superou, em termos do ajuste aos dados, o modelo preditivo indicado pelo BIC (Figura 5.19). Desta forma, como modelos finais, adotou-se os dois indicados pelo critério AIC.

As Figuras 5.19 e 5.20 apresentam os principais resultados da análise dos resíduos. Tais figuras possibilitaram avaliar a adequação dos modelos propostos. Ou seja, através dos gráficos dispostos nessas figuras, avaliou-se se os modelos estimados detectaram a dinâmica da série de forma satisfatória.

Os gráficos da Figura 5.19 exibem uma comparação entre os resíduos resultantes do SARIMA(1,1,0)×(1,1,2), em (a), e do SARIMA(3,1,2)×(1,1,1), em (b).

Através dos gráficos Resíduos padronizados na Figura 5.19, nota-se que, de uma forma geral, os resíduos são pequenos. Os correlogramas não mostram problemas de correlação serial nos resíduos. E o teste de Ljung-Box indica a não rejeição da hipótese nula de ausência de autocorrelação conjunta. Contudo, no geral, nota-se que os resíduos gerados pelo SARIMA(3,1,2)×(1,1,1) apresentam autocorrelações menores no decorrer das defasagens.

O SARIMA(2,1,3)×(1,1,2) mostrou-se satisfatório para modelar as fontes de variações da série, assim como mostrado através da Figura 5.20.

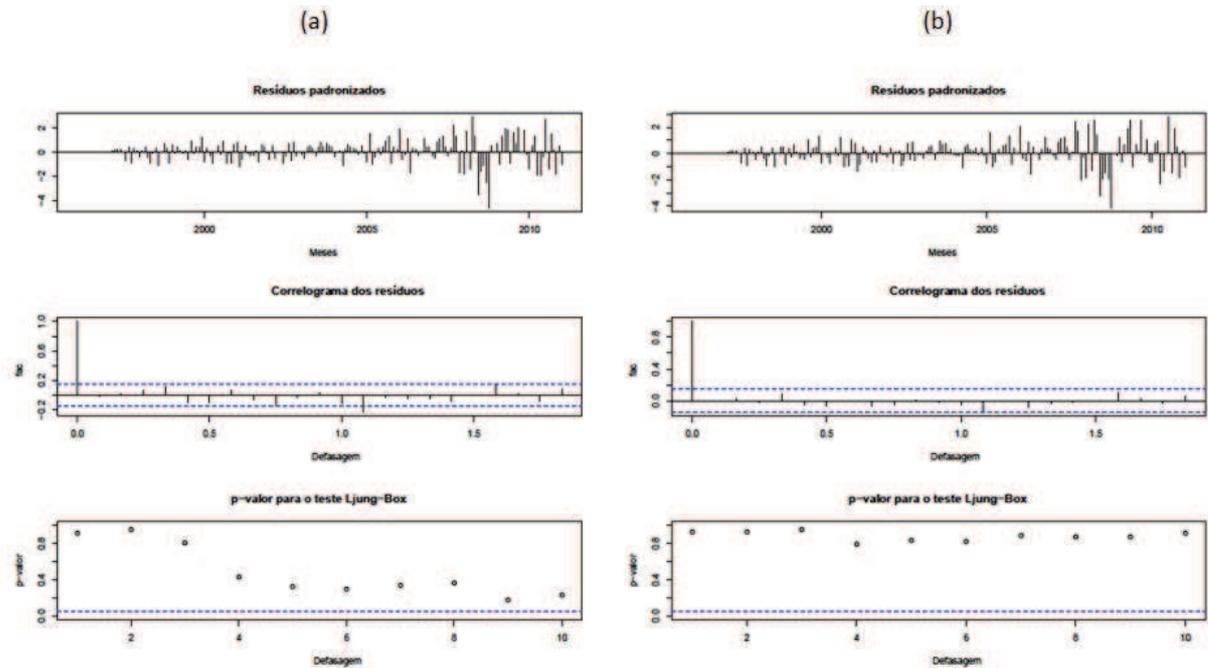


Figura 5.19: Diagnóstico do SARIMA(1,1,0)×(1,1,2) (a) e SARIMA(3,1,2)×(1,1,1) (b) (IBOVESPA).

De forma global, a distribuição dos erros estimados pelo modelo SARIMA(2,1,3)×(1,1,2) comporta-se como o esperado. Os resíduos não são marcados por tendências, os quais mostram pequenas autocorrelações e a ausência de evidências contra a hipótese de não autocorrelação conjunta.

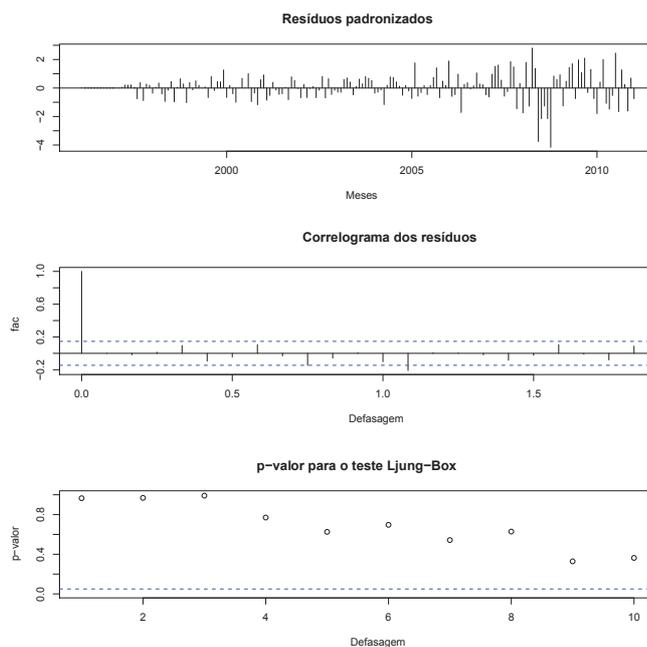


Figura 5.20: Diagnóstico do modelo SARIMA(2,1,3)×(1,1,2) (IBOVESPA).

A Tabela 5.4 apresenta os resultados das previsões advindas dos dois modelos finais. Identificou-se o modelo SARIMA(2,1,3) \times (1,1,2) como SARIMA 1, enquanto que o rótulo SARIMA 2 representa o modelo SARIMA(3,1,2) \times (1,1,1).

A partir da Tabela 5.4 pode-se observar que ambos os modelos obtiveram previsões subestimadas apenas nos dois primeiros horizontes, assim como ocorreu para os resultados da modelagem via Holt-Winters.

Ao comparar os erros absolutos provenientes dos modelos SARIMA 1 e SARIMA 2, nota-se que para os horizontes intermediários ($h= 5, 6, 7, 8$) e para $h= 9$ o SARIMA 2 proporcionou menores valores. Desta forma, o SARIMA 1 apresentou melhor performance do que o SARIMA 2 na maioria dos horizontes considerados (Tabela 5.4).

Tabela 5.4: Previsões obtidas através do SARIMA (IBOVESPA).

h	Valor real	SARIMA 1	Erro 1	SARIMA 2	Erro 2
1	67383	66290,53	1092,47	66041,27	1341,73
2	68587	66513,22	2073,78	66249,73	2337,27
3	66133	67664,96	-1531,96	68007,28	-1874,28
4	64620	68666,07	-4046,07	68713,10	-4093,10
5	62404	68873,62	-6469,62	67664,47	-5260,47
6	58823	69534,84	-10711,84	67863,79	-9040,79
7	56495	68878,63	-12383,63	67985,48	-11490,48
8	52324	69008,87	-16684,87	68673,25	-16349,25
9	58338	68135,12	-9797,12	67856,94	-9518,94
10	56875	68572,69	-11697,69	68933,20	-12058,20
11	56754	69825,54	-13071,54	70555,27	-13801,27
12	59147	69999,77	-10852,77	70250,05	-11103,05

A Figura 5.21 mostra um gráfico comparativo envolvendo os valores observados na série IBOVESPA e as previsões dos modelos preditivos selecionados, na classe SARIMA.

Assim como exibido no gráfico da Figura 5.21, o SARIMA 2 aparenta-se aproximar mais da série real de dados do que o SARIMA 1, apesar do SARIMA 2 apresentar erros absolutos mais elevados em grande parte dos horizontes ($h= 1, 2, 3, 4, 10, 11$ e 12).

Pode-se verificar que os modelos SARIMA(2,1,3) \times (1,1,2) e SARIMA(3,1,2) \times (1,1,1) geram comportamentos preditivos muito semelhantes, com exceção das previsões a médio prazo; onde o SARIMA 2 apresenta uma ligeira melhora na habilidade preditiva.

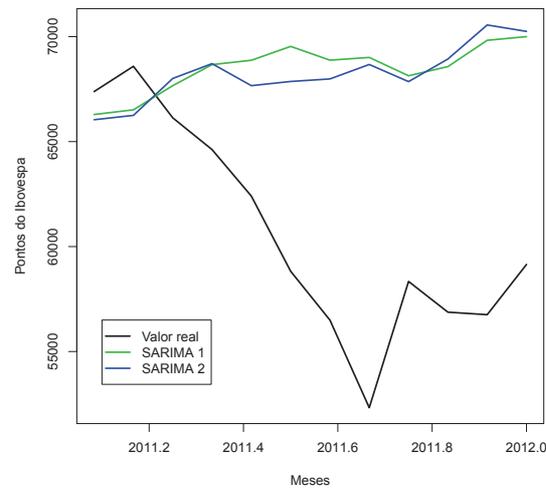


Figura 5.21: Previsões advindas dos modelos SARIMA (IBOVESPA).

Todavia, os modelos preditivos de Box-Jenkins ajustados à série IBOVESPA não conseguiram reproduzir bem as variações encontradas na série real. Tais modelos resultaram em previsões bastante distantes da realidade observada.

5.2.2 Enfoque SSA

- VENTO

Assim como citado na Seção 4.1.2, adotou-se um comprimento de janela igual a $L = 4032$ para a série VENTO. A Figura 5.22 apresenta o comportamento do logaritmo dos autovalores.

Através da Figura 5.22 é possível identificar claramente o autovalor de maior magnitude, haja visto que este destaca-se dos demais. Assim como discutido na Seção 3.3.3.1, tal autovalor representa o componente de tendência.

Adicionalmente, pode ser de interesse identificar os componentes periódicos da série. De acordo com a Figura 5.22, devido ao elevado comprimento da janela, não é possível visualizar a formação desses pares.

Como alternativa, costuma-se avaliar os pares de autovetores. A Figura 5.23 mostra um conjunto de *scatterplots*, isto é, gráficos de dispersão para os primeiros pares de autovetores.

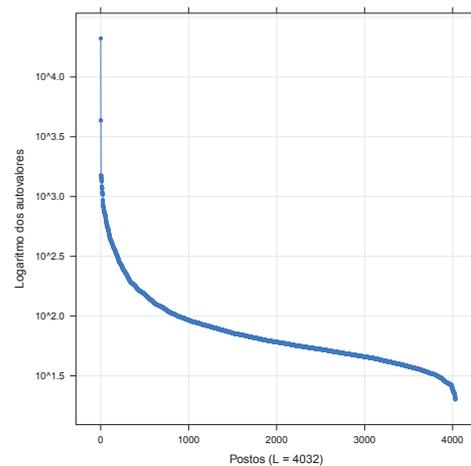


Figura 5.22: Logaritmo dos autovalores (VENTO).

Ainda através da Figura 5.23, pode-se notar a fração de explicabilidade advinda de cada autotriple. Nota-se que o primeiro autovalor, a tendência da série VENTO, contribui com uma fração de 74,96%. Individualmente, os dois próximos componentes explicam um pouco mais de 3,18% e a partir do quarto autotriple, estes detêm conjuntamente apenas 18,65% de toda fonte de variação da série.

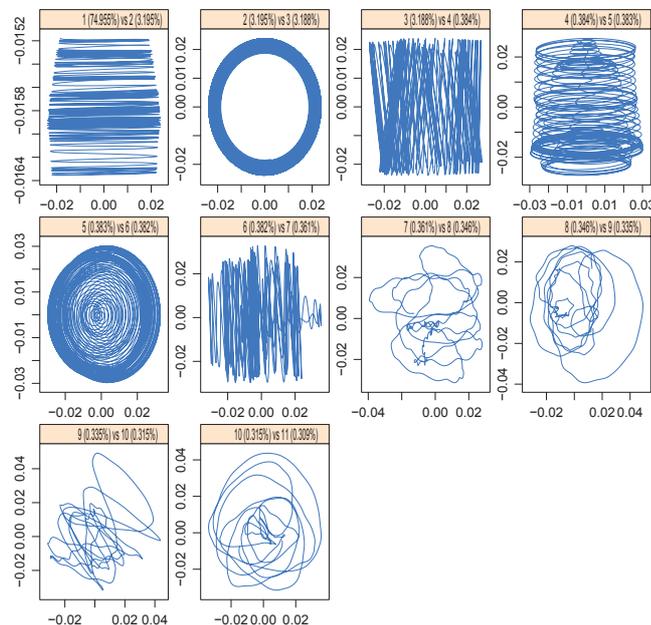


Figura 5.23: Primeiros pares de autovetores (VENTO).

Posteriormente, avaliou-se a ferramenta *matriz w-correlação* com o intuito de encontrar um limiar de separação entre o sinal e o ruído. Em virtude da grande massa de dados e conseqüente exigência de processamento computacional, não foi possível estimar as correlações ponderadas entre todos os 4032 componentes da decomposição; a quantidade máxima processada não superou 1000 componentes.

A representação gráfica dessa ferramenta é exibida na Figura 5.24. Nessa Figura também é mostrado o gráfico com as correlações ponderadas entre os 336 primeiros componentes da série VENTO, gráfico (b).

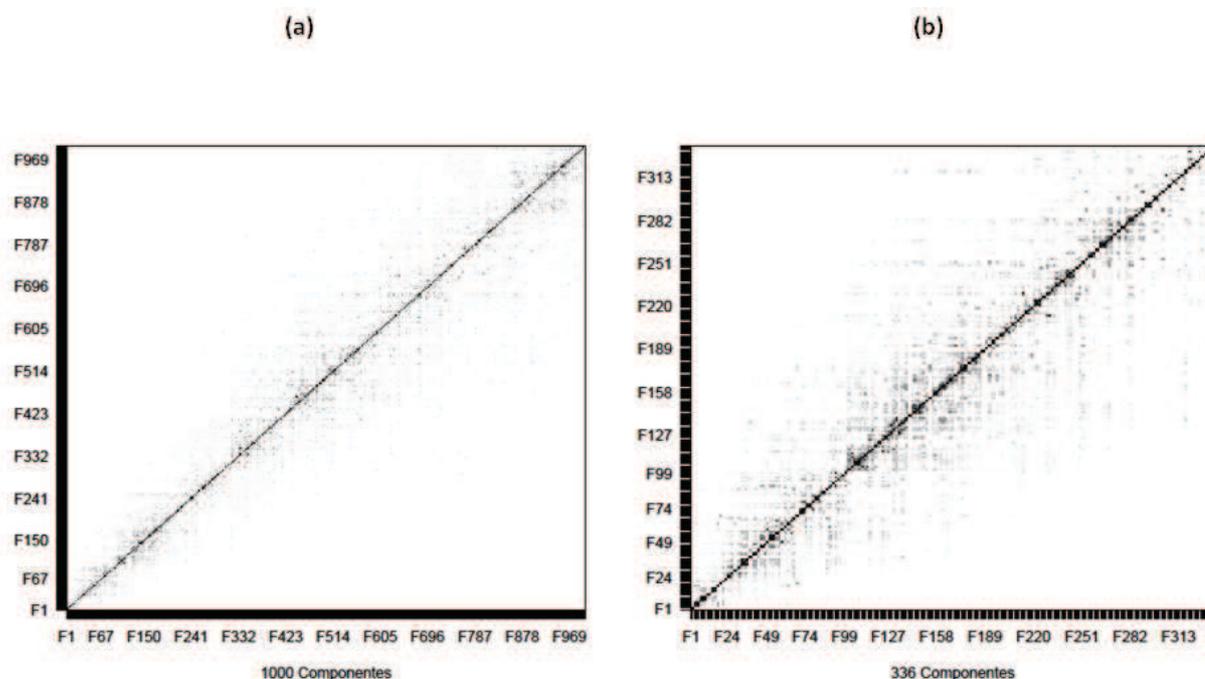


Figura 5.24: Matriz w-correlação (VENTO).

Assim como era de esperar, haja visto o elevado número de componentes considerados na decomposição, a ferramenta matriz w-correlação não apresentou uma separação trivial entre o sinal e ruído.

Todavia, ao analisar tal ferramenta com 336 componentes, gráfico (b), pode-se observar que a partir do componente de posto em torno de 200 há muitos componentes com cores de tonalidades mais claras (correlações baixas), os quais estariam representando o grupo do ruído.

Após realizar a separação entre o sinal e ruído, analisou-se a reconstrução de algumas subséries, tais como: a tendência estimada, uma estimativa do componente periódico e do ruído da série VENTO.

A Figura 5.25 exibe os três principais componentes na forma de subséries, além do gráfico com a série VENTO original.

O gráfico (a) da Figura 5.25 apresenta a série original. A tendência foi reconstruída considerando o primeiro componente principal (CP1), gráfico (b), já a parte periódica foi formada pelos pares de autovalores de postos (2,3) e (4,5), gráfico (c). E formou-se o ruído a partir do componente de ordem 201, isto é, CP201; gráfico (d).

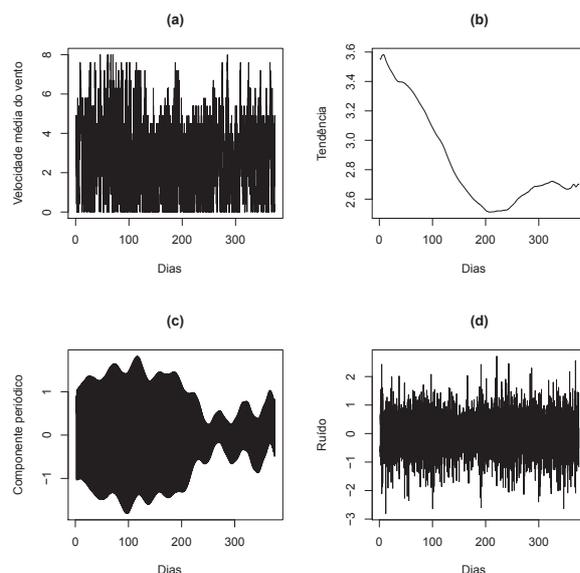


Figura 5.25: Decomposição da série via SSA (VENTO).

Ao comparar a estimativa do componente residual advinda do método clássico de decomposição STL com a gerada pela SSA, observa-se uma diferença de 0,8143 no erro quadrático médio ($MSE = 1,1288$ e $MSE = 0,3145$, respectivamente). Sendo assim, pode-se notar que o método clássico de decomposição estimou os resíduos de forma menos acurada do que a técnica SSA.

A Figura 5.26 mostra a reconstrução da série VENTO ao considerar dois conjuntos de agrupamento: (1-1000) e (1-4032), além de apresentar concomitantemente o componente de tendência (em (a)). A reconstrução a partir do agrupamento (1-4032) foi rotulada como **reconstrução completa**, haja visto que, todos os $L = 4032$ componentes contribuíram para a formação da série aproximada.

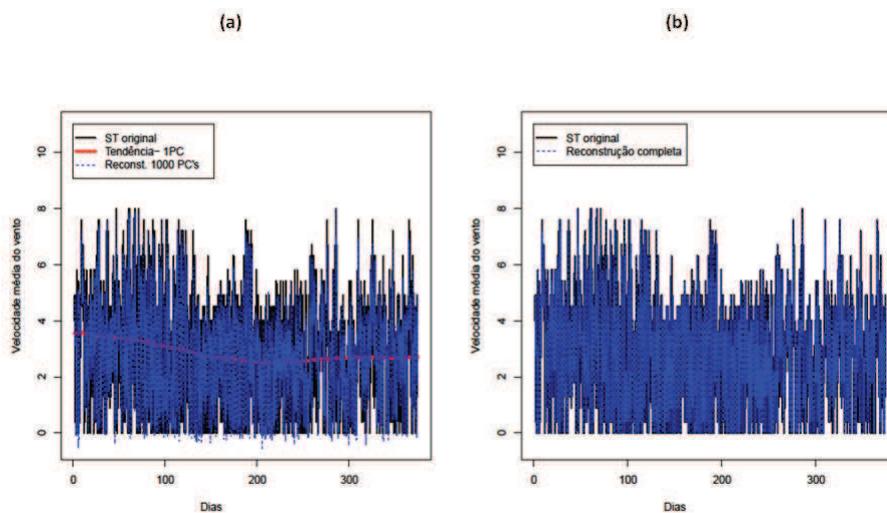


Figura 5.26: Reconstrução da série via SSA (VENTO).

Segundo a Figura 5.26, de modo geral, ambos conjuntos de agrupamentos retratam bem as características originais da série VENTO. Contudo, na aproximação que considera (1-4032), gráfico (b), observa-se que a série original, em preto, quase não aparece no gráfico. Tal fato retrata a capacidade da SSA ao realizar a reconstrução com todos os autotriples da decomposição.

Nota-se que, as séries aproximadas resultantes da técnica SSA gerou séries quase iguais à série VENTO original. O erro quadrático médio de cada série reconstruída foi o seguinte: $MSE = 4,924 \times 10^{-10}$, para o agrupamento (1-1000) e $MSE = 4,884 \times 10^{-36}$, levando em consideração todos os componentes da SVD.

Ao finalizar a análise dos estágios da SSA (decomposição e reconstrução), aplicou-se o algoritmo recorrente de previsão SSA na série reconstruída a partir dos 1000 primeiros componentes da SVD.

A Figura 5.27 exhibe simultaneamente os comportamentos das previsões e dos valores futuros. Através de tal figura é possível observar que a predição SSA não se afastou da realidade dos valores reais da série VENTO, para os 2688 instantes futuros.

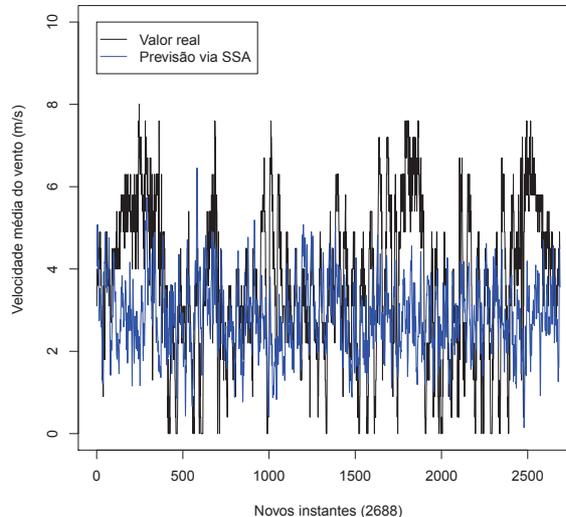


Figura 5.27: Previsão da série via SSA (VENTO).

Nota-se que os valores preditos via SSA, em azul, não representam tão bem as reais velocidades médias do vento mais elevadas (a partir de 5 m/s). Por outro lado, o algoritmo de previsão SSA representa melhor as velocidades médias reais menores; conseguindo captar os extremos inferiores dos novos instantes avaliados (Figura 5.27).

Os valores pontuais preditos pela SSA e os respectivos erros para os onze horizontes considerados encontram-se descritos na Tabela 5.5.

Tabela 5.5: Previsões via SSA (VENTO).

Horizonte (h)	Valor real	SSA	Erro
1	3,100	4,063	-0,963
24	3,600	3,055	0,545
48	2,200	4,457	-2,257
96	3,600	3,518	0,082
192	5,800	3,973	1,827
336	5,800	4,072	1,728
672	4,900	3,767	1,133
1008	6,700	3,297	3,403
1344	4,000	3,557	0,443
2016	3,600	3,166	0,434
2688	4,500	3,560	0,940

Observa-se que o algoritmo de previsão SSA superestimou em apenas dois dos onze horizontes, a saber: $h=1$ e $h=48$. De uma maneira geral, os valores reais se distanciam pouco dos valores previstos pela SSA (Tabela 5.5). Por exemplo, o menor erro observado, entre os horizontes considerados, corresponde ao horizonte $h = 96$ (0,082 m/s). Enquanto que, o maior erro ficou em torno de 3,4 m/s, correspondendo ao horizonte $h = 1008$ (Tabela 5.5).

- IBOVESPA

A decomposição da série via SSA encontra-se sumarizada na Figura 5.28.

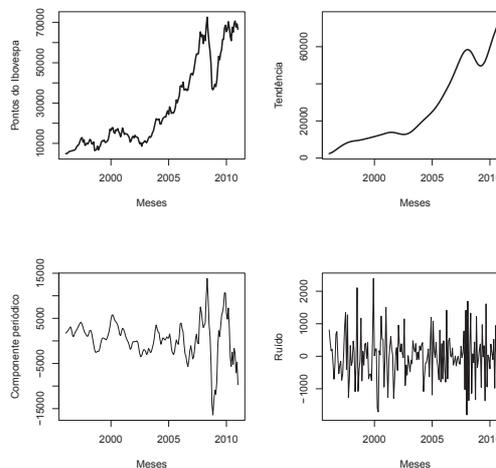


Figura 5.28: Decomposição via SSA (IBOVESPA).

Na decomposição SSA considerou-se o L ótimo para o comprimento da série IBOVESPA, isto é, $L = 90$. Através do gráfico “Tendência” (em 5.28), observa-se que a forma da tendência extraída pela técnica SSA assemelha-se com as estimativas dadas pelos métodos STL e Médias Móveis.

A distribuição da série classificada como “Ruído” apresentou um MSE de $5,143374 \times 10^5$; sendo que o mínimo observado foi de $-1,80352 \times 10^3$ e o máximo igual a $2,39133 \times 10^3$.

Após a decomposição da série IBOVESPA observou-se que o maior autovalor destaca-se dos demais. Este autovalor indica a tendência principal da série, sendo representada pelo primeiro componente do espectro da matriz trajetória (Figura 5.29).

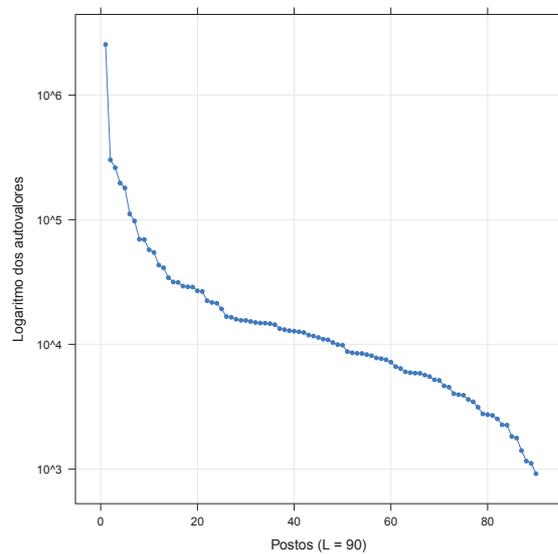


Figura 5.29: Logaritmo dos autovalores da SVD (IBOVESPA).

De acordo com a Figura 5.29, pode-se identificar um par de autotriples constituído pelos autovalores de ordem 8 e 9. Esse par representa um componente periódico da série IBOVESPA. Aparentemente existem outros conjuntos de autovalores com valores próximos, contudo o gráfico do espectro dos autovalores da série IBOVESPA não permite a fácil identificação dos demais pares de autovalores.

Para auxiliar na identificação dos pares de autovalores investigou-se os comportamentos dos autovetores relacionados aos primeiros autotriples. A Figura 5.30 mostra gráficos de dispersão de dez pares de autovetores.

Ao analisar os primeiros pares de autotriples da Figura 5.30 observa-se que o comportamento oscilatório não é tão bem definido como na série VENTO. Todavia, pode-se extrair alguma característica periódica na série IBOVESPA através dos seguintes pares: (4,5), (6,7), (8,9) e (10,11).

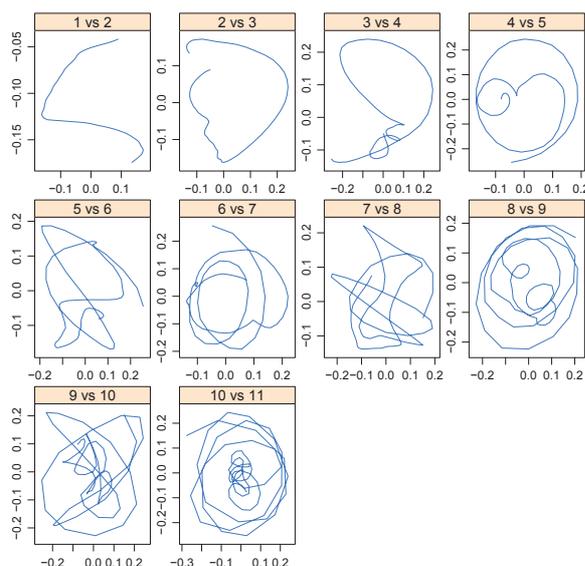


Figura 5.30: Primeiros pares de autovetores (IBOVESPA).

A Figura 5.31 exibe nove componentes principais, na forma de séries temporais, relacionados aos correspondentes autotriples da decomposição.

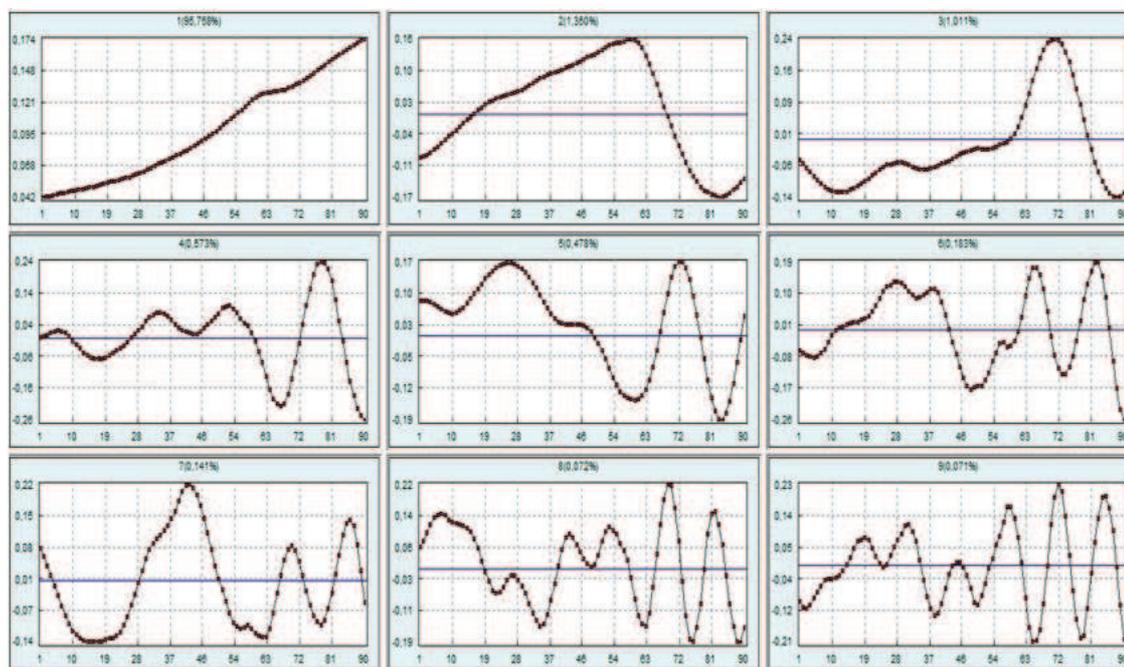


Figura 5.31: Primeiros autovetores (IBOVESPA).

Ao analisar a decomposição da série IBOVESPA observou-se que, apenas o primeiro componente explica cerca de 95,8% da variação da série. Os demais componentes contribuem com frações relativamente pequenas, como por exemplo, o segundo autotriple é responsável por explicar apenas 1,35% de toda variação da série; já o terceiro autotriple contribui com 1,01% (Figura 5.31).

Ainda através da Figura 5.31, verifica-se que o primeiro componente da decomposição indica a tendência principal da série e a partir do segundo componente o padrão de sazonalidade começa a ser identificado.

Com relação a reconstrução da série, nota-se que ao combinar os dois primeiros componentes principais tem-se uma estimativa refinada da tendência da série temporal. Ambos componentes são responsáveis por 97,1% da variação da série (Figura 5.32).

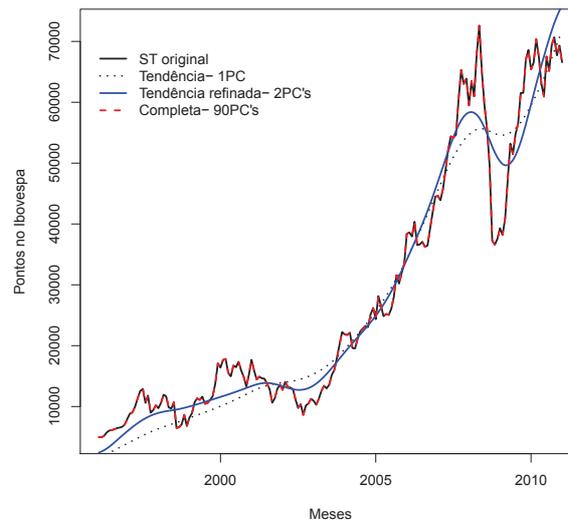


Figura 5.32: Reconstrução da série IBOVESPA.

O estágio de reconstrução da série IBOVESPA ao considerar 90 componentes produziu uma série aproximada bem semelhante a série original (Figura 5.32). O erro quadrático médio da diferença entre a série original e a série aproximada pela SSA é baixíssimo, em torno de $2,54 \times 10^{-22}$. A diferença mínima observada é $-5,821 \times 10^{-11}$ e a máxima não ultrapassa $3,64 \times 10^{-11}$.

Na Figura 5.33 tem-se a matriz w-correlação formada pelos 90 componentes considerados no estágio de decomposição da série IBOVESPA.

Assim como discutido na Subseção 3.3.3.2, a matriz w-correlação indica como realizar um agrupamento adequado. De acordo com a Figura 5.33, pode-se verificar dois grandes blocos aparentemente separáveis. O primeiro bloco identifica o sinal da série, caracterizado por um grupo de autotriples altamente correlacionados (do primeiro até o componente com ordem em torno de 28). A segunda parte da figura apresenta muitos componentes com correlações baixas (tonalidades mais claras) e podem ser agrupados em um mesmo bloco, retratando o ruído da série IBOVESPA.

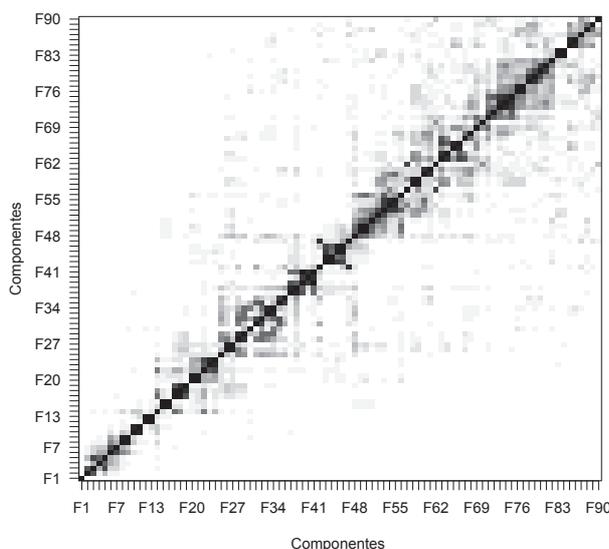


Figura 5.33: Matriz de correlações ponderadas (IBOVESPA).

Ao considerar os autotriples de 1 a 28 na formação do *signal*, observou-se que o valor absoluto da correlação ponderada entre o *signal* e *ruído* foi bem baixo ($\rho_{12}^{(w)} = 0,0068$). Fato este que pode ser considerado como uma confirmação de que a separação entre ambos grupos é bastante razoável, isto é, os componentes ou subséries possuem a propriedade da separabilidade (ver Seção 3.3.1).

Na Figura 5.34 encontra-se a representação gráfica da matriz correlação ponderada entre o *signal* e *ruído* da série IBOVESPA.

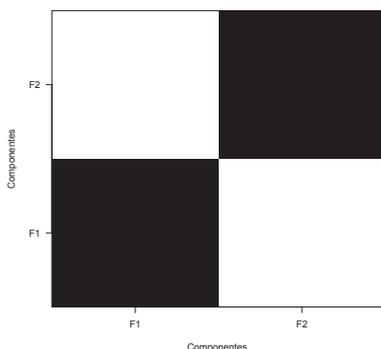


Figura 5.34: Sinal e ruído: Correlação ponderada (IBOVESPA).

Visualmente nota-se que as duas subséries consideradas (*signal* e *ruído*) são separáveis entre si, já que é possível verificar a formação de apenas dois grandes blocos de componentes (Figura 5.34). O primeiro bloco, nomeado como F1 no gráfico, é formado pelos 28 primeiros componentes correlacionados. E os componentes remanescentes (29-90) foram agrupados no bloco seguinte.

Portanto, de forma geral, o sinal da série (tendência e componentes periódicos) pode ser formado pelos 28 primeiros componentes e os demais componentes ficam identificados como ruído. Para esse agrupamento verificou-se que o sinal responde por 99,93% de toda variabilidade da série IBOVESPA.

Após a identificação do sinal e ruído realizou-se a previsão utilizando o algoritmo recorrente.

Inicialmente considerou-se o espaço trajetória com os vinte e oito primeiros componentes. Em virtude do distanciamento das predições em relação aos valores reais, em especial nas previsões a longo prazo, considerou-se também dois outros espaços trajetória, a saber: agrupamento dos dez primeiros autotriples (1-10), o qual contempla 99,69% da variação da série, e o agrupamento dos sete primeiros componentes (1-7), que explica 99,49%.

Para a previsão via SSA na série IBOVESPA foi levado em consideração a mesma quantidade de horizontes, $h = 12$ pontos à frente. A título de comparação, a previsão foi realizada a partir da série reconstruída, de acordo com as três formas de agrupamentos citados, e a partir da própria série observada (série inicial).

A Tabela 5.6 exhibe as predições e os correspondentes erros advindos do algoritmo recorrente ao considerar as três distintas estruturas de agrupamento [(1-28), (1-10) e (1-7)]. As correspondentes previsões fundamentadas na série inicial são mostradas na forma de gráficos e resumidas através das medidas de precisão na Subseção 5.2.3.

Tabela 5.6: Previsões obtidas através da SSA (IBOVESPA).

h	Valor real	(1-28)	Erro 1	(1-10)	Erro 2	(1-7)	Erro 3
1	67383	62170,65	5212,35	66652,03	730,97	72125,71	-4742,71
2	68587	53443,52	15143,48	66227,49	2359,52	74804,11	-6217,11
3	66133	48405,38	17727,62	66298,44	-165,44	77572,53	-11439,53
4	64620	54824,15	9795,85	67347,72	-2727,72	80301,65	-15681,65
5	62404	64544,54	-2140,54	69743,74	-7339,75	82861,22	-20457,22
6	58823	76904,82	-18081,82	73608,46	-14785,46	85121,96	-26298,96
7	56495	90344,69	-33849,69	78741,17	-22246,17	86963,36	-30468,36
8	52324	100348,18	-48024,18	84633,78	-32309,78	88282,62	-35958,63
9	58338	105364,79	-47026,79	90562,37	-32224,37	89005,92	-30667,92
10	56875	109052,14	-52177,14	95733,77	-38858,77	89094,82	-32219,82
11	56754	105666,32	-48912,33	99454,27	-42700,27	88550,96	-31796,96
12	59147	88820,66	-29673,66	101261,21	-42114,21	87421,50	-28274,50

Através da Tabela 5.6 nota-se que na maioria dos horizontes considerados, a partir de $h = 5$, o algoritmo recorrente SSA superestimou os reais pontos observados da série IBOVESPA ($h = 5, \dots, 12$), para o agrupamento (1-28).

Ao observar os erros absolutos das previsões fundamentadas no *signal* formado pelos 10 autotripes iniciais (1-10), pode-se constatar que a SSA superestimou apenas em horizontes menores ($h= 1$ e $h=2$). Em paralelo, pode-se notar que todas as previsões advindas do agrupamento (1-7) ficaram subestimadas em relação aos valores reais observados (Tabela 5.6).

Ainda a partir dos resultados da Tabela 5.6, ao realizar um comparativo entre as previsões SSA oriundas de cada um dos três agrupamentos, nota-se que o algoritmo recorrente com agrupamento (1-10) gerou menores erros absolutos para os oito primeiros horizontes analisados. Por outro lado, as previsões a longo prazo ($h= 9, 10, 11, 12$) foram mais acuradas ao considerar o agrupamento (1-7).

Com o intuito de simplificação, para a realização das comparações entre os métodos (Subseção 5.2.3) considerou-se o algoritmo recorrente SSA com agrupamento (1-10), já que este gerou erros inferiores na maioria dos horizontes investigados.

5.2.3 Comparação dos métodos preditivos

- VENTO

A Tabela 5.7 apresenta um comparativo geral dos resultados de previsão de cada método analisado. Essa tabela resume a acurácia das técnicas apresentando o erro quadrático médio (*MSE*) e o erro percentual total (*TPE*) para predições 1, 24, 48, 96, 192, 336, 672, 1008, 1344, 2016 e 2688 passos à frente para a série VENTO.

Tabela 5.7: Comparação entre as técnicas consideradas (VENTO).

h	HW aditivo		SARIMA 1		SARIMA 2		SSA	
	<i>MSE</i>	<i>TPE</i>	<i>MSE</i>	<i>TPE</i>	<i>MSE</i>	<i>TPE</i>	<i>MSE</i>	<i>TPE</i>
1	0,170	-13,285	0,137	-11,948	0,144	-12,250	0,927	-31,065
24	1,242	18,404	1,629	24,884	1,626	25,088	1,117	6,597
48	0,944	6,917	1,312	12,250	1,323	11,763	1,301	-3,244
96	0,819	13,203	1,150	18,211	1,112	17,269	1,208	3,976
192	2,693	28,372	3,200	32,619	3,064	31,584	3,296	26,668
336	4,608	36,790	5,361	40,719	5,146	39,669	5,201	35,471

672	3,529	21,609	3,874	27,027	3,760	25,500	4,053	23,109
1008	2,889	15,713	3,103	22,126	3,012	20,293	3,262	19,798
1344	2,441	12,424	2,603	19,690	2,522	17,603	3,100	17,988
2016	2,891	14,081	3,222	22,373	3,091	19,984	3,721	22,009
2688	2,924	13,842	3,333	23,295	3,168	20,573	3,881	22,748

Observa-se que, de acordo com o MSE , o método de Holt-Winters sazonal aditivo foi superior aos demais, pois para quase todos os horizontes, os MSE 's para o método H-W foram inferiores ao SARIMA 1, SARIMA 2 e SSA, com exceção do passo $h = 24$; o qual apontou performance melhor para a previsão SSA (Tabela 5.7).

Se a avaliação recair sobre a medida TPE , o algoritmo H-W aditivo e SSA vencem em igual quantidade de horizontes. O algoritmo aditivo apresentou melhor desempenho em horizontes maiores ($h = 672, 1008, 1344, 2016, 2688$); enquanto que para previsões à curto e médio prazo (24, 48, 96, 192 e 336 passos à frente) a previsão SSA apresentou uma melhor performance.

A Tabela 5.8 exibe as principais medidas descritivas da distribuição do conjunto de teste (**valor real**), assim como as medidas para as previsões oriundas de cada técnica investigada.

De acordo com as descritivas mostradas na Tabela 5.8, pode ser notado que a distribuição das previsões oriundas da SSA assemelha-se mais com a distribuição do conjunto teste no que tange as medidas de dispersão [desvio padrão (0,892 m/s) e coeficiente de variação (31,28%)] e os valores extremos [mínimo (0,144 m/s) e máximo (6,454 m/s)].

Com relação às medidas de tendência central (média e mediana), a distribuição das previsões via H-W aditivo é mais semelhante ao conjunto de teste [$mdia_{H-W} = 3,181$ m/s; $mdia_{Real} = 3,692$ m/s e $mediana_{H-W} = 3,139$ m/s; $mediana_{Real} = 3,600$ m/s] (Tabela 5.8).

Tabela 5.8: Medidas descritivas das distribuições das previsões (VENTO).

Descritivas	Conjunto teste	HW aditivo	SARIMA 1	SARIMA 2	SSA
Mínimo	0,000	1,986	1,906	1,970	0,144
Quartil 1	2,700	2,895	2,317	2,422	2,235
Mediana	3,600	3,139	2,789	2,885	2,797
Média	3,692	3,181	2,832	2,932	2,852
Quartil 3	4,900	3,513	3,464	3,559	3,449
Máximo	8,000	4,493	3,708	3,812	6,454

Desvio padrão	1,758	0,553	0,571	0,568	0,892
Coef. de variação (%)	47,611	17,398	20,169	19,364	31,280

Sendo assim, pode-se constatar que a distribuição das previsões geradas pela SSA retrata bem o conjunto de teste formado pelos 2688 valores futuros da série VENTO.

Avaliou-se também o segmento das 2688 previsões advindas de cada método preditivo. A Figura 5.35 apresenta uma comparação gráfica envolvendo as previsões advindas das técnicas preditivas e as verdadeiras velocidades médias do vento (*valor real*). Em paralelo, visando facilitar a interpretação gráfica, realizou-se uma amostragem sequencial dos instantes de tempo. Analisou-se as previsões para 135 distintos horizontes, a saber: $h = 1, 20, 40, \dots, 2640, 2660, 2680$ (Figura 5.36).

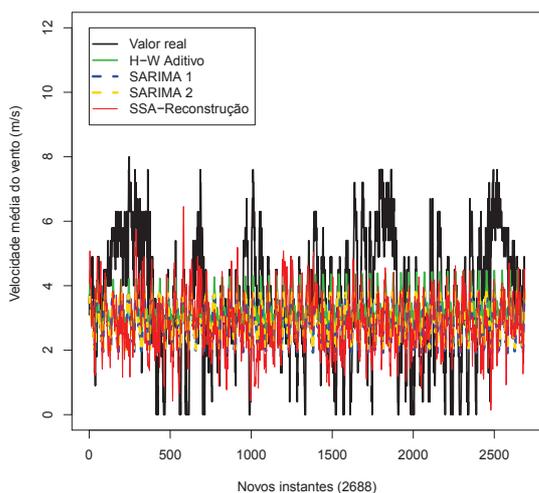


Figura 5.35: Comparação geral (VENTO)

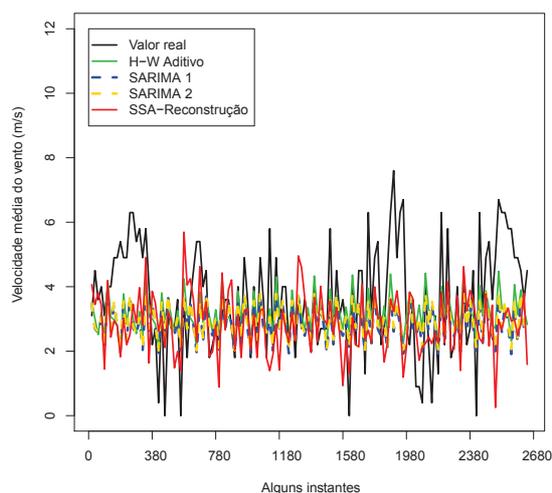


Figura 5.36: Comparação para alguns instantes

Observa-se que o algoritmo recorrente de previsão SSA conseguiu detectar melhor as flutuações das velocidades médias do vento futuras (*valor real*) em comparação ao H-W aditivo e os modelos ARIMA sazonais (Figura 5.35 e Figura 5.36).

Em segundo plano, através da figuras 5.35 e 5.36, nota-se também que o algoritmo H-W aditivo resultou em previsões melhores do que os modelos ARIMA sazonais, isso porque o segmento correspondente ao H-W aditivo (em verde) possui uma amplitude um pouco maior do que as observadas nos segmentos em azul [SARIMA $(2,1,2) \times (1,1,1)$] e em amarelo [SARIMA $(3,1,3) \times (1,1,1)$].

Esses achados podem ser confirmados ao avaliar a amplitude total (AT), diferença entre o máximo e o mínimo, de cada distribuição na Tabela 5.8. Verificou-se as seguintes me-

didadas para cada distribuição: $AT_{Real} = 8,000$ m/s; $AT_{H-W} = 2,507$ m/s; $AT_{SARIMA1} = 1,802$ m/s; $AT_{SARIMA2} = 1,842$ m/s e $AT_{SSA} = 6,310$ m/s.

Se o interesse tiver em avaliar o ranqueamento entre as técnicas preditivas consideradas na presente dissertação, segundo os resultados de previsão obtidos na série VENTO, pode-se obter a seguinte ordenação de desempenho:

1. SSA (1-1000);
2. Holt-Winters aditivo;
3. SARIMA (3,1,3)×(1,1,1);
4. SARIMA (2,1,2)×(1,1,1).

Globalmente falando, a previsão SSA apresentou um comportamento mais fidedigno à realidade da série VENTO, o que ratifica todas as análises preditivas realizadas na série VENTO.

- IBOVESPA

A Tabela 5.9 apresenta as medidas de precisão consideradas no trabalho: o erro quadrático médio (MSE) e o erro percentual total (TPE) para os horizontes 1, 3, 6 e 12 meses; para ambos enfoques do algoritmo de Holt-Winters, dois modelos da classe SARIMA e o algoritmo recorrente de previsão SSA com os dez primeiros autotriples (1-10).

Tabela 5.9: Comparação entre as técnicas clássicas (IBOVESPA).

h	HW aditivo		HW multiplicativo		SARIMA 1		SARIMA 2	
	MSE	$TPE(\%)$	MSE	$TPE(\%)$	MSE	$TPE(\%)$	MSE	$TPE(\%)$
1	$1,67 \times 10^2$	0,02	$2,11 \times 10^5$	0,68	$1,19 \times 10^6$	1,62	$1,80 \times 10^6$	1,99
3	$2,34 \times 10^6$	-0,70	$4,91 \times 10^6$	-1,46	$2,61 \times 10^6$	0,81	$3,59 \times 10^6$	0,89
6	$4,89 \times 10^7$	-7,52	$5,20 \times 10^7$	-7,83	$3,01 \times 10^7$	-5,05	$2,28 \times 10^7$	-4,28
9	$1,03 \times 10^8$	-12,99	$1,47 \times 10^8$	-15,12	$7,87 \times 10^7$	-10,53	$6,97 \times 10^7$	-9,72
12	$1,20 \times 10^8$	-15,31	$1,62 \times 10^8$	-17,41	$9,45 \times 10^7$	-12,93	$9,05 \times 10^7$	-12,49

De acordo com a Tabela 5.9, observa-se que o algoritmo de Holt-Winters sazonal aditivo apresenta um desempenho global superior ao método multiplicativo, pois as medidas MSE

e *TPE* oriundas da forma aditiva do método foram menores do que as medidas advindas do algoritmo sazonal multiplicativo.

Ao confrontar as técnicas clássicas adotadas e os diferentes enfoques, observa-se que a forma aditiva do modelo de Holt-Winters apresentou melhor desempenho à curto prazo, ou seja, nos horizontes 1 e 3. Já o modelo SARIMA 2, isto é, o SARIMA $(3,1,2) \times (1,1,1)$ mostrou-se superior para as previsões à médio e longo prazo (6, 9 e 12 passos à frente).

A Figura 5.37 apresenta um comparativo de todas as previsões oriundas de cada método considerado no estudo.

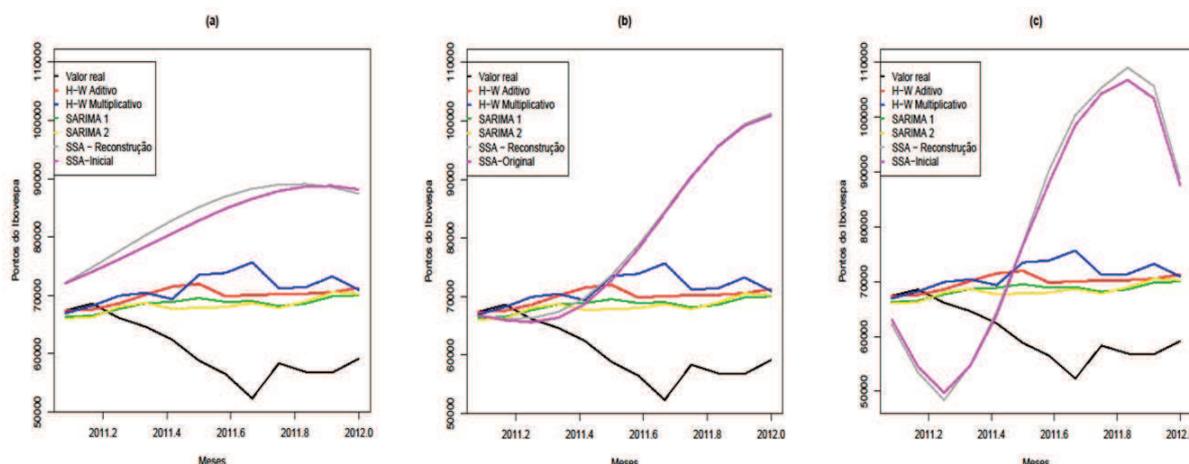


Figura 5.37: Comparação geral das previsões [Agrupamento SSA: (a) 1-7; (b) 1-10 e (c) 1-28 (IBOVESPA)].

Os gráficos da Figura 5.37 mostram que os quatro modelos clássicos analisados, Holt-Winters sazonal aditivo, Holt-Winters sazonal multiplicativo, SARIMA 1 e SARIMA 2 apresentam habilidades preditivas semelhantes; ainda que distantes do conjunto de teste ou série observada (valor real).

Com relação às previsões oriundas do algoritmo de previsão SSA, nota-se que a configuração com menos autotriples no espaço trajetória resulta, em geral, em uma melhor capacidade preditiva, assim como apresentado na Figura 5.37 e Tabela 5.10.

Tabela 5.10: Comparação entre os diferentes agrupamentos da SSA (IBOVESPA).

h	SSA (1-28)		SSA (1-10)		SSA (1-7)	
	MSE	$TPE(\%)$	MSE	$TPE(\%)$	MSE	$TPE(\%)$
1	$2,72 \times 10^7$	7,74	$5,34 \times 10^5$	1,08	$2,25 \times 10^7$	-7,04
3	$1,90 \times 10^8$	18,84	$2,04 \times 10^6$	1,45	$6,40 \times 10^7$	-11,08
6	$1,66 \times 10^8$	7,13	$4,77 \times 10^7$	-5,65	$2,58 \times 10^8$	-21,87
9	$7,40 \times 10^8$	-18,24	$3,18 \times 10^8$	-19,58	$5,23 \times 10^8$	-32,77
12	$1,05 \times 10^9$	-31,87	$6,64 \times 10^8$	-31,93	$6,30 \times 10^8$	-37,67

Outro resultado interessante refere-se a observação de uma tendência de melhora no desempenho da SSA ao passo que diminui a quantidade de autotriples no espaço trajetória para a série IBOVESPA.

E ao comparar as previsões SSA advindas das duas séries de base (reconstruída e inicial) observou-se que as diferenças foram bem sutis. A maior discrepância observada refere-se ao agrupamento (1-7), assim como apresentado no gráfico (a) da Figura 5.37.

De acordo com os resultados das previsões na série IBOVESPA, pode-se ordenar o desempenho geral dos modelos preditivos analisados, da seguinte forma:

1. SARIMA(3,1,2)×(1,1,1);
2. SARIMA(2,1,3)×(1,1,2);
3. Holt-Winters aditivo;
4. Holt-Winters multiplicativo;
5. SSA (1-10);
6. SSA (1-7);
7. SSA (1-28).

Em linhas gerais, pode-se afirmar que, apesar das suposições clássicas e exigências da SSA terem sido satisfeitas, nenhum método conseguiu reproduzir satisfatoriamente a dinâmica da série ao longo de todo o horizonte considerado.

5.3 Estudos de simulação

Para o atual contexto analisou-se o processo de modelagem nas séries apenas através da SSA, já que o algoritmo de previsão SSA exige a inserção das informações de agrupamento para a definição do espaço trajetória (Seção 3.4).

5.3.1 Simulação 1

As estruturas das séries simuladas estão apresentadas na Figura 5.38.

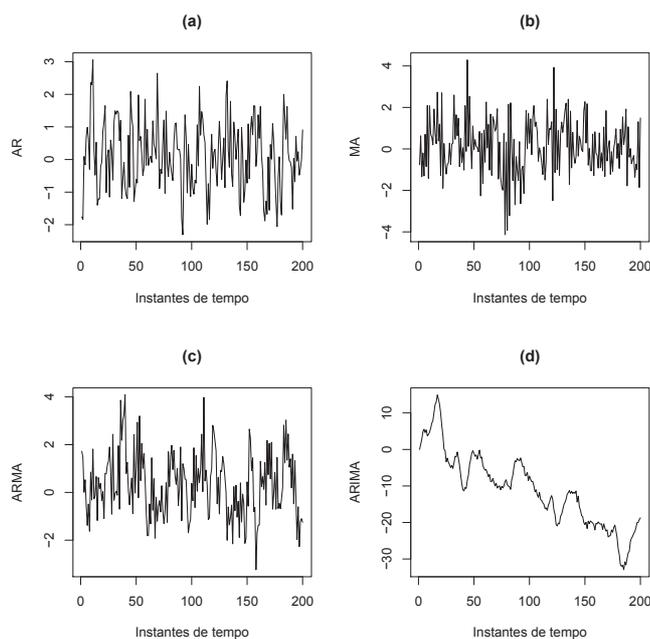


Figura 5.38: Processos AR (a), MA (b), ARMA (c) e ARIMA (d) (Simulação 1).

Em cada gráfico da Figura 5.38 apresenta-se o comportamento de cada processo estocástico teórico simulado: AR (a), MA (b), ARMA (c) e ARIMA (d). De acordo com os gráficos (a), (b) e (c), as respectivas séries geradas indicam um desenvolvimento em torno de um nível com a presença de tendências locais em poucos pontos do tempo. Enquanto que o processo ARIMA simulado, gráfico (d), indica uma clara tendência decrescente com algumas flutuações, mas sem caracterizá-las como sazonais.

Após realizar a decomposição via SSA nas séries simuladas avaliou-se os espectros das correspondentes matrizes trajetórias através da Figura 5.39.

Através do gráfico (a), pode-se observar a formação de três principais pares de autovalores, são eles: (1,2); (5,6) e (7,8). Outro aspecto investigado refere-se a avaliação das quebras observadas em tal gráfico. Notou-se que o gráfico (a) exhibe duas grandes quebras ao longo dos 100 postos. Ao investigar a primeira quebra, isto é, considerando o agrupando (1-8) como sinal e os demais componentes como ruído; observou uma correlação ponderada entre o sinal e o ruído de $\rho_{12}^{(w)} = 0,0644$, sendo que tal sinal é responsável por explicar 36,24% da variação da correspondente série. Enquanto que, ao considerar a próxima quebra, ou seja, a formação do sinal até o componente de ordem 78 (CP78); verificou-se que a correlação ponderada aumentou ($\rho_{12}^{(w)} = 0,0965$), nesse caso, o grupo até o CP78 detém 98,55% de todas as variações da série original.

Para a série advinda de um processo de média móvel não observou-se a formação de pares de autotriples. O agrupamento dos autotriples para constituir o sinal fundamentou-se na primeira grande quebra do gráfico (b) na Figura 5.39 (posto 40). Em outras palavras,

formou-se o ruído da série a partir do quadragésimo primeiro autotriple da SVD; o que resultou em uma fração de variabilidade desprezada em torno de 12,03%. Tal separação entre o sinal e o ruído proporcionou uma correlação de $\rho_{12}^{(w)} = 0,0799$. Adicionalmente avaliou-se o agrupamento a partir da segunda grande quebra: autotriples até CP52. Para essa segunda opção notou-se que quase não houve redução na correlação ponderada ($\rho_{12}^{(w)} = 0,0784$), sendo que 5,50% da variabilidade da série foi associada ao ruído.

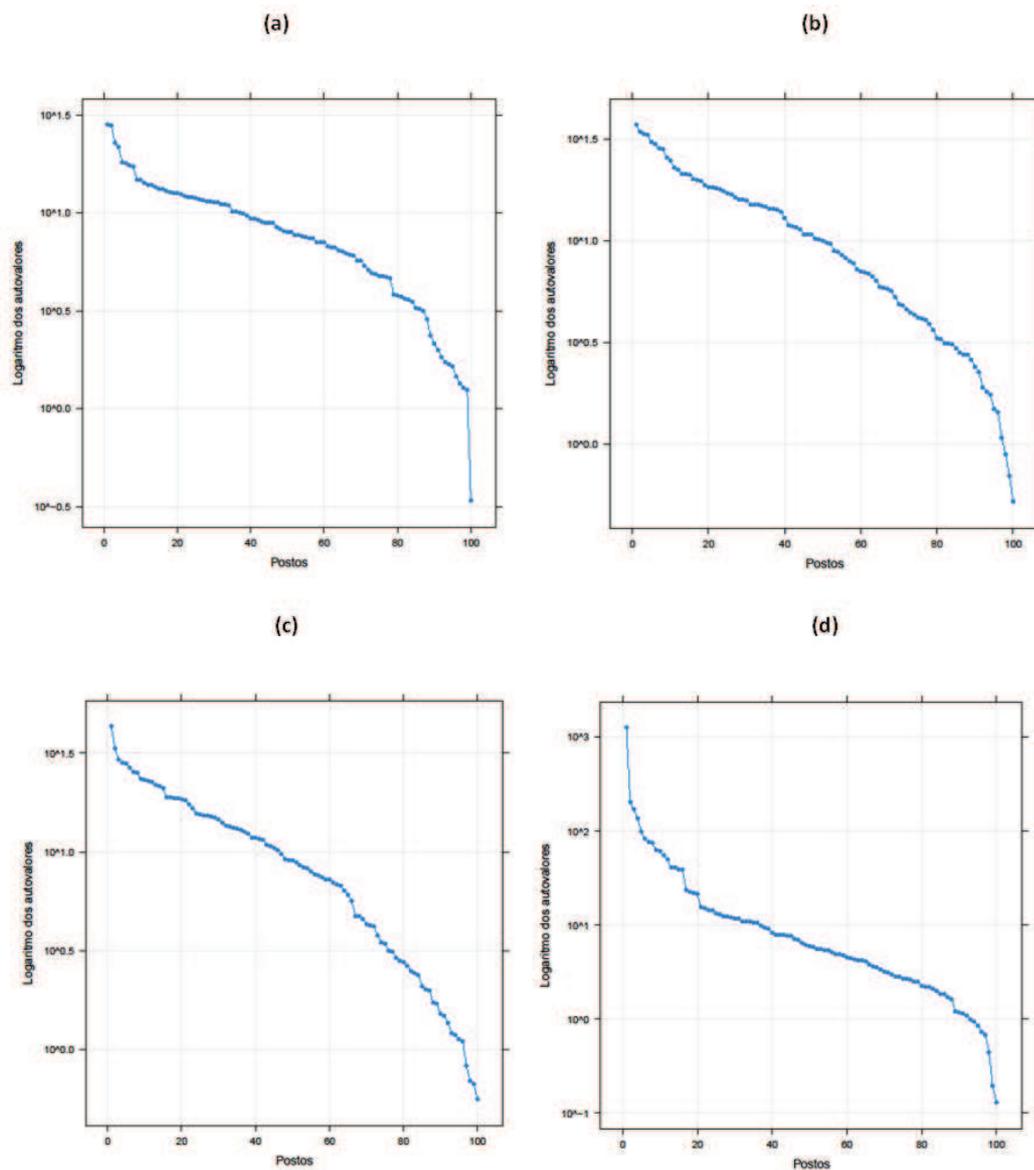


Figura 5.39: Logaritmo dos autovalores: AR (a), MA (b), ARMA (c) e ARIMA (d) (Simulação 1).

A decomposição SSA na série artificial oriunda de um processo ARMA não aponta a formação nítida de pares de autovalores. Através da Figura 5.39, gráfico (c), nota-se a presença de uma tendência geral, retratada pelo primeiro componente da SVD.

Com relação às quebras do espectro no gráfico (c) da Figura 5.39, avaliou-se as suas três primeiras quebras. Para a primeira quebra, até o décimo quinto componente (CP15), verificou-se uma w -correlação de $\rho_{12}^{(w)} = 0,1207$; em que o sinal explica 57,33% da série

original. Analisando a segunda quebra, até CP47, observou-se uma correlação de $\rho_{12}^{(w)} = 0,1163$; tendo uma fração de explicabilidade de 92,73% por parte do sinal. Por outro lado, ao levar em consideração a terceira quebra, até CP66, verifica-se que o sinal explica 98,65% da série e que a correlação ponderada foi igual a $\rho_{12}^{(w)} = 0,0418$.

Por fim, de acordo com o gráfico (d) da Figura 5.39, a decomposição na série proveniente de um processo ARIMA indica claramente o componente de tendência, representado pelo primeiro autovalor em destaque. Ainda através do gráfico (d), analisou-se as três primeiras quebras do gráfico ora citado. As frações de explicabilidade e as correlações entre o sinal e o ruído ao considerar a primeira quebra (até CP16), a segunda quebra (até CP20) e a terceira quebra (até CP88) foram as seguintes: 99,66% e $\rho_{12}^{(w)} = 0,0069$; 99,77% e $\rho_{12}^{(w)} = 0,0039$; 99,99% e $\rho_{12}^{(w)} = 0,0017$, respectivamente.

De posse de todas as opções de agrupamento pré-selecionadas aplicou-se o algoritmo recorrente de previsão nas séries provenientes de cada processo teórico, confrontando as previsões geradas a partir dos diferentes espaços trajetória, em cada série simulada.

Para a série originária do processo AR(1), em geral, o melhor resultado preditivo observado corresponde ao sinal formado pelos oito primeiro autotriples. Na série advinda do processo MA(2) verificou-se que o agrupamento até CP40 (1-40) para o sinal gerou previsões com menores erros. Enquanto que, para o processo ARMA(1,2) os melhores resultados foram obtidos ao considerar todos os autotriples a partir do quadragésimo oitavo como ruído. Já para a série proveniente do processo ARIMA(1,1,2) o sinal constituído pelos 88 componentes principais proporcionou menores erros de previsão.

Na Figura 5.40 são mostradas as correlações ponderadas finais, após a avaliação dos agrupamentos pré-selecionados, para cada série simulada.

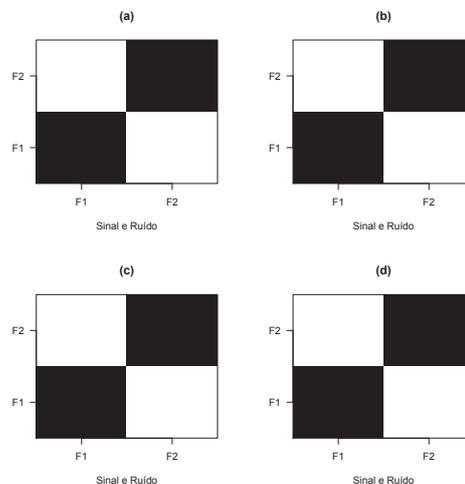


Figura 5.40: Correlações entre o sinal e o ruído: AR, MA, ARMA e ARIMA (Simulação 1).

Tabela 5.11: Comparação entre as previsões (Processo AR - Simulação 1).

h	Real	SEH			AR(1)			SSA		
		Erro	MSE	TPE	Erro	MSE	TPE	Erro	MSE	TPE
1	0,053	-0,254	0,064	-479,801	-0,330	0,109	-623,523	0,541	0,293	1021,986
2	-0,341	-0,637	0,235	308,772	-0,545	0,203	303,145	0,359	0,211	-311,794
3	0,481	0,197	0,170	-360,541	0,339	0,173	-278,221	0,303	0,171	624,861
4	-1,078	-1,350	0,583	230,903	-1,198	0,489	195,876	-1,633	0,795	48,596
5	0,199	-0,061	0,467	307,053	0,086	0,393	240,353	-0,018	0,636	65,289
6	-0,660	-0,910	0,527	224,041	-0,771	0,426	179,781	-1,368	0,842	134,927
7	1,278	1,040	0,606	2907,948	1,167	0,560	1843,200	-0,032	0,722	2720,662
8	0,010	-0,217	0,536	3760,526	-0,101	0,491	2320,150	-0,769	0,705	4490,029
9	-1,324	-1,540	0,740	269,872	-1,434	0,665	201,553	-1,791	0,984	318,845
10	-0,244	-0,448	0,686	256,860	-0,355	0,611	193,055	-1,221	1,034	346,003
11	0,083	-0,110	0,625	277,681	-0,027	0,556	205,149	-0,646	0,978	406,343
12	-0,987	-1,168	0,686	215,530	-1,097	0,610	168,491	-1,001	0,980	287,439

Através da Figura 5.40 nota-se que os agrupamentos considerados na formação de cada espaço trajetória possuem a importante propriedade da separabilidade (ver Subseção 3.3.1).

As Tabelas 5.11, 5.12, 5.13 e 5.14 apresentam os valores futuros ou conjunto de teste (rotulado nas tabelas como “Real”), a diferença entre o valor original do conjunto de teste e o previsto por cada técnica (nomeado como “Erro”), além das medidas MSE e TPE ; para o primeiro estudo de simulação.

A Tabela 5.11 contém os resultados para a série oriunda do processo estocástico AR(1). Observa-se que para SEH houve subestimação apenas nos horizontes 3 e 7. Para o padrão ouro AR(1) os horizontes 3, 5 e 7 resultaram em previsões subestimadas. Já as previsões via SSA ficaram superestimadas em quase todos os horizontes (4 a 12).

Ao comparar os resultados das três técnicas consideradas, de acordo com a medida erro quadrático médio (MSE), verifica-se que o padrão ouro ganhou a competição, ou seja, apresentou menores valores para os horizontes $h = 2, 4, 5, \dots, 12$, enquanto que para esses mesmos horizontes a técnica SEH ficou em segunda posição e para 1 e 3 passos à frente a SEH venceu a competição. Já a previsão SSA mostrou-se a segunda melhor técnica preditiva para os horizontes 2 e 3 (Tabela 5.11).

Ao considerar o erro percentual total (TPE), observa-se que o padrão ouro exibe menores percentuais para os horizontes $h = 2, 3, 7, \dots, 12$. A SEH mostrou-se melhor para $h = 1$ e ficou em segundo lugar para os horizontes $h = 2, 9, \dots, 12$. Enquanto que para os passos 4, 5 e 6 a SSA liderou a competição e para o horizonte 7 essa técnica apresentou o segundo menor percentual (Tabela 5.11).

Tabela 5.12: Comparação entre as previsões (Processo MA - Simulação 1).

h	Real	SEH			MA(2)			SSA		
		Erro	MSE	TPE	Erro	MSE	TPE	Erro	MSE	TPE
1	-0,289	-0,201	0,041	69,611	0,167	0,028	-57,613	2,008	4,031	-694,268
2	-0,202	-0,191	0,039	79,860	-0,283	0,054	23,666	-1,740	3,530	-54,412
3	-0,899	-0,964	0,336	97,603	-1,037	0,395	82,964	-0,048	2,354	-15,797
4	-0,040	-0,182	0,260	107,627	-0,178	0,304	93,113	-1,996	2,762	124,189
5	-3,163	-3,382	2,496	107,147	-3,302	2,423	100,867	-3,206	4,265	108,468
6	1,602	1,306	2,364	120,867	1,463	2,376	105,956	-0,194	3,561	173,033
7	-1,151	-1,524	2,358	124,068	-1,289	2,274	107,641	0,010	3,052	124,718
8	-1,523	-1,973	2,550	125,530	-1,662	2,335	108,028	-2,462	3,428	134,644
9	2,354	1,828	2,638	159,562	2,215	2,621	117,910	3,138	4,142	135,594
10	-2,845	-3,448	3,563	141,837	-2,983	3,249	111,882	-3,886	5,237	136,056
11	1,063	0,383	3,252	163,911	0,925	3,031	117,077	0,546	4,788	153,727
12	-1,500	-2,256	3,406	160,849	-1,638	3,002	115,291	-2,399	4,869	155,140

Os resultados da previsão para a série simulada a partir do processo MA(2) encontram-se resumidos na Tabela 5.12. Nota-se que ocorreu subestimação nos horizontes 6, 9 e 11 tanto para a SEH quanto para a SSA. E o padrão ouro mostrou previsões subestimadas para os passos 6 e 11.

A medida MSE indicou melhor desempenho para o padrão ouro nos horizontes $h = 1, 5, 7, \dots, 12$, não havendo a segunda colocação para tal técnica. Sendo que, para esses mesmos horizontes a SEH apresentou o segundo menor valor e tal método venceu a competição nos passos 2, 3, 4 e 6. A performance da previsão SSA, segundo a MSE , mostrou-se inferior aos outros dois métodos considerados (Tabela 5.12).

Ao avaliar o TPE verifica-se que o padrão ouro exibiu menores valores para $h = 1, 2, 8, \dots, 12$. Para os horizontes $h = 1, 4, \dots, 8$ a SEH apresentou-se como a segunda coloca na competição. A previsão SSA liderou no horizonte 3 e mostrou-se vice-líder nos horizontes $h = 2, 10, \dots, 12$ (Tabela 5.12).

Para a série simulada proveniente do processo ARMA(1,2) os resultados preditivos são mostrados na Tabela 5.13. É possível notar que para os horizontes 1, 2, 4, 6 e 9 a SEH, assim como o padrão ouro e a SSA apresentaram previsões superestimadas.

Conforme os resultados do MSE apresentado na Tabela 5.13, o padrão ouro mostrou-se melhor para todos os horizontes, exceto para $h = 2$, o qual foi responsável pela segunda colocação e a SEH assumiu a liderança em tal horizonte.

De acordo com o TPE , o padrão ouro apresentou os menores percentuais em 9 dos horizontes considerados ($h = 1, 3, 5, 7, \dots, 12$). A SEH exibiu um segundo melhor desempenho para 1 e 4 passos à frente; e venceu no horizonte 2. Já a SSA mostrou um segundo melhor desempenho em 7 dos 12 horizontes ($h = 5, 7, \dots, 12$); a qual liderou no passo 6.

Tabela 5.13: Comparação entre as previsões (Processo ARMA - Simulação 1).

h	Real	SEH			ARMA(1,2)			SSA		
		Erro	MSE	TPE	Erro	MSE	TPE	Erro	MSE	TPE
1	-0,817	0,499	0,249	-60,994	-0,212	0,045	25,925	-2,574	6,627	314,968
2	-1,378	-0,023	0,125	-21,680	-0,640	0,227	38,806	-3,072	8,033	257,176
3	0,300	1,696	1,041	-114,548	0,443	0,217	21,565	0,450	5,423	274,125
4	-0,708	0,728	0,914	-111,384	-0,824	0,333	47,361	-1,136	4,389	243,242
5	1,453	2,929	2,446	-506,836	1,223	0,565	0,854	4,513	7,586	158,175
6	-1,334	0,181	2,044	-241,938	-1,614	0,905	65,366	0,756	6,417	42,795
7	1,207	2,763	2,843	-687,215	0,906	0,893	56,227	4,109	7,912	-238,645
8	0,555	2,151	3,065	-1514,319	0,244	0,789	65,649	1,698	7,284	-657,770
9	-0,018	1,617	3,015	-1695,700	-0,333	0,714	109,110	0,082	6,475	-652,642
10	1,045	2,721	3,454	4988,343	0,728	0,695	-25,629	-0,085	5,829	1550,014
11	0,848	2,563	3,737	1544,628	0,530	0,658	39,156	-1,276	5,447	300,361
12	1,531	3,285	4,325	786,346	1,212	0,725	61,997	1,513	5,183	185,454

Tabela 5.14: Comparação entre as previsões (Processo ARIMA - Simulação 1).

h	Real	SEH			ARIMA(1,1,2)			SSA		
		Erro	MSE	TPE	Erro	MSE	TPE	Erro	MSE	TPE
1	-18,579	-0,321	0,103	1,729	0,564	0,318	-3,035	-1,788	3,196	9,623
2	-18,821	-1,016	0,568	3,576	-0,438	0,255	-0,337	1,575	2,838	0,569
3	-18,377	-1,026	0,729	4,237	-0,294	0,199	0,301	-2,555	4,069	4,963
4	-20,244	-3,345	3,344	7,509	-2,279	1,447	3,218	5,462	10,510	-3,543
5	-22,230	-5,784	9,366	11,697	-4,311	4,874	6,878	0,724	8,513	-3,478
6	-24,278	-8,285	19,245	16,141	-6,377	10,840	10,720	3,980	9,734	-6,037
7	-26,532	-10,992	33,757	20,642	-8,639	19,953	14,607	2,105	8,976	-6,375
8	-28,318	-13,230	51,417	24,805	-10,427	31,049	18,153	5,552	11,708	-8,487
9	-29,625	-14,990	70,671	28,497	-11,735	42,901	21,225	-5,121	13,320	-4,799
10	-30,217	-16,035	89,316	31,626	-12,327	53,807	23,718	0,687	12,035	-4,477
11	-28,670	-14,941	101,490	33,835	-10,781	59,482	25,215	-8,083	16,880	-0,955
12	-29,909	-16,633	116,087	36,037	-12,020	66,565	26,729	-8,574	21,600	2,040

Desta forma, segundo a Tabela 5.13, a previsão SSA apresentou-se melhor do que a sua concorrente SEH.

Os principais resultados preditivos para a última série simulada, advinda do processo ARIMA(1,1,2), encontram-se na Tabela 5.14. Pode-se observar que houve superestimação das previsões em todos horizontes e métodos avaliados.

Segundo a Tabela 5.14, ao analisar o MSE nota-se que a liderança do padrão ouro permaneceu em poucos horizontes, são eles: $h = 2, 3, 4, 5$; e observou-se a sua vice-liderança nos passos $h = 6, \dots, 12$. Enquanto que para esses últimos horizontes a SSA mostrou-se líder e assumiu a vice-liderança no passo 5.

Ao avaliar a medida relativa, pode-se verificar desempenhos semelhantes ao observado pelo MSE . Nota-se que o padrão ouro apresenta a melhor performance em 3 dos hori-

zontes ($h = 2, 3, 4$); com a segunda posição para os passos $h = 1, 5, \dots, 12$. Contrastando com a SSA que mostrou-se melhor em 8 dos horizontes ($h = 5, \dots, 12$); cuja vice-liderança foi observada nos passos 2 e 4.

Com o intuito de sintetizar os resultados, pode-se considerar um indicador ou escore de desempenho que contabilize o número de vezes em que o algoritmo preditivo SSA resulte em menores erros, no que tange o *MSE* ou *TPE*, ao longo dos horizontes analisados.

A Tabela 5.15 apresenta os escores de desempenho do algoritmo SSA para cada processo estocástico investigado.

Tabela 5.15: Escores de desempenho preditivo SSA (Simulação 1).

Processos	Primeiro lugar		Segundo lugar	
	<i>MSE</i>	<i>TPE</i>	<i>MSE</i>	<i>TPE</i>
AR	0	3	2	1
MA	0	1	0	4
ARMA	0	1	0	7
ARIMA	7	8	1	2

Em resumo, observou-se que a capacidade preditiva da SSA foi destacada ao analisar os processos estocásticos menos simples, como o ARIMA. Nota-se que para o processo ARIMA (1,1,2) a previsão SSA resultou nos escores 7 e 8, segundo o *MSE* e *TPE*, respectivamente. Isso significa que, a SSA apresentou as menores medidas de precisão em uma maior quantidade de horizontes, superando o **padrão ouro** e a SEH. Com relação a segunda posição, verifica-se que a SSA apresentou desempenho levemente superior à SEH; escores iguais a 1 e 2, de acordo com o *MSE* e *TPE*, respectivamente (Tabela 5.15).

Os gráficos das Figuras 5.41 e 5.42 mostram uma comparação geral entre todos os métodos preditivos e o conjunto de teste para cada processo estocástico analisado. De acordo com os gráfico (a) e (b) da Figura 5.41, nota-se que as previsões oriundas do **padrão ouro**, em azul, e da SEH apresentam comportamentos próximos, todavia estas previsões estão distantes da realidade do conjunto teste, haja visto que ambas técnicas estimaram previsões em torno de um único nível.

Por outro lado, as previsões via SSA, em verde, contornam satisfatoriamente os valores do conjunto de teste, principalmente para a série originária do processo MA, gráfico (b) da Figura 5.41. Observa-se que a SSA apresentou algumas previsões quase iguais aos valores reais, por exemplo: para os horizontes 5 (erro de -0,018) e 7 (erro de -0,032) na série AR, gráfico (a); e para os passos 3 (erro igual a -0,048), 6 (erro de 0,01) e 7 (erro de -0,194) no processo MA, gráfico (b).

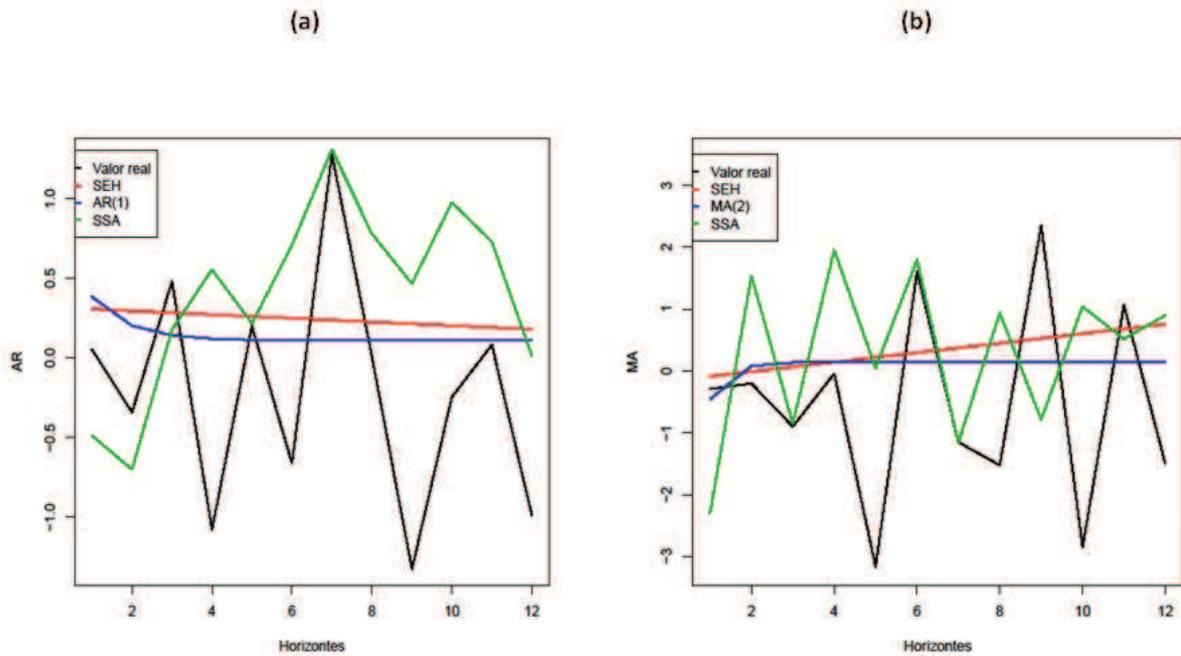


Figura 5.41: Comparação das previsões (AR (a) e MA (b) - Simulação 1).

A Figura 5.42 exibe o comparativo para as séries advindas de um ARMA, gráfico (a), e ARIMA, gráfico (b).

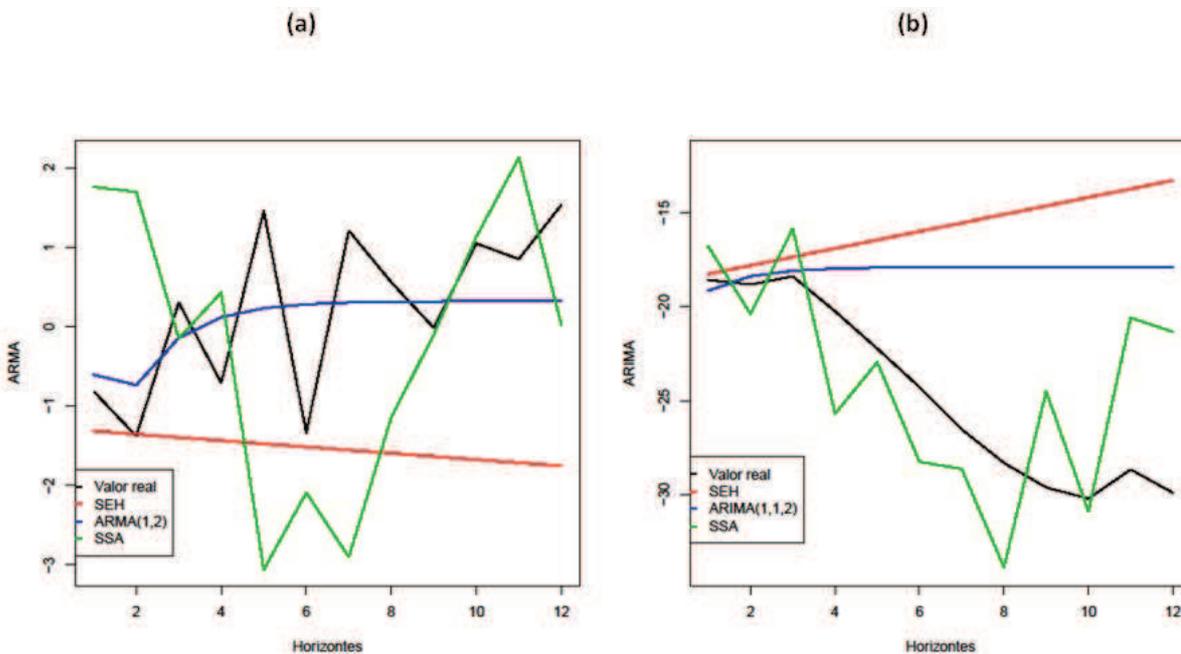


Figura 5.42: Comparação das previsões (ARMA (a) e ARIMA (b) - Simulação 1).

Em (a), observa-se graficamente que o padrão ouro gerou uma curva preditiva que simplifica o comportamento real dos valores futuros em torno de dois patamares: um nível para o intervalo formado entre os passos 1 e 2; e outro nível a partir do terceiro horizonte. Já a SEH apresentou uma curva estimada bastante simples, reta decrescente; o que dis-

tância da realidade observada. Enquanto que, em (b) nota-se que tanto a SEH quanto o padrão ouro estimaram uma reta, sendo que para esse último método as previsões foram quase constantes.

Ainda através da Figura 5.42, conforme apresentado em (a), há um trecho do gráfico em que a previsão SSA apresenta-se bastante precisa, ou seja, a SSA reproduziu bem o conjunto de teste entre os horizontes 9 e 10, apresentando erros de 0,082 e $-0,085$, respectivamente. Ao comparar os resultados dados pela SSA com os valores reais, é perceptível que a curva resultante do algoritmo recorrente de previsão SSA acompanha a curva do conjunto de teste do início ao fim, apesar da SSA exibir mais picos e variações do que a curva real.

De forma geral, para as quatro séries analisadas no primeiro estudo de simulação, a previsão via SSA manifestou melhor as peculiaridades dos reais valores futuros, pois as curvas preditivas geradas pela SSA são mais semelhantes ao conjunto de teste do que as curvas geradas pelos demais métodos avaliados, principalmente para as séries provenientes dos processos AR(1) e MA(2).

5.3.2 Simulação 2

As séries artificiais analisadas nessa etapa do estudo encontram-se representadas nos gráficos da Figura 5.43.

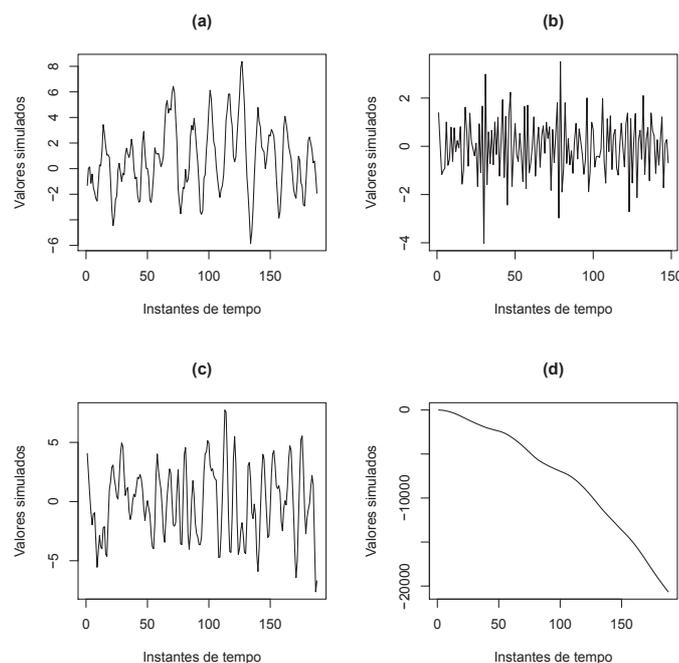


Figura 5.43: Processos AR (a), MA (b), ARMA (c) e ARIMA (d) (Simulação 2).

Observa-se que o gráfico (a) apresenta três padrões de comportamento, são eles: nos instantes iniciais, até $t = 50$, tem-se um padrão simples com pequenas flutuações; entre os instantes 51 e 150 a série apresenta uma tendência de crescimento marcada por grandes amplitudes; já nos instantes finais a série retorna ao seu patamar inicial. Os processos mostrados nos gráficos (b) e (d) retratam séries monótonas, sendo que em (d) há uma evidente tendência de decrescimento. Já a série representada no gráfico (c) não é portadora de uma tendência bem marcada.

A decomposição via SSA foi realizada em cada uma das séries simuladas conforme descrito na Subseção 4.1.2 do Capítulo 4, usando o comprimento da janela otimizado. Os espectros das matrizes trajetórias de cada processo teórico gerado encontra-se apresentado na Figura 5.44.

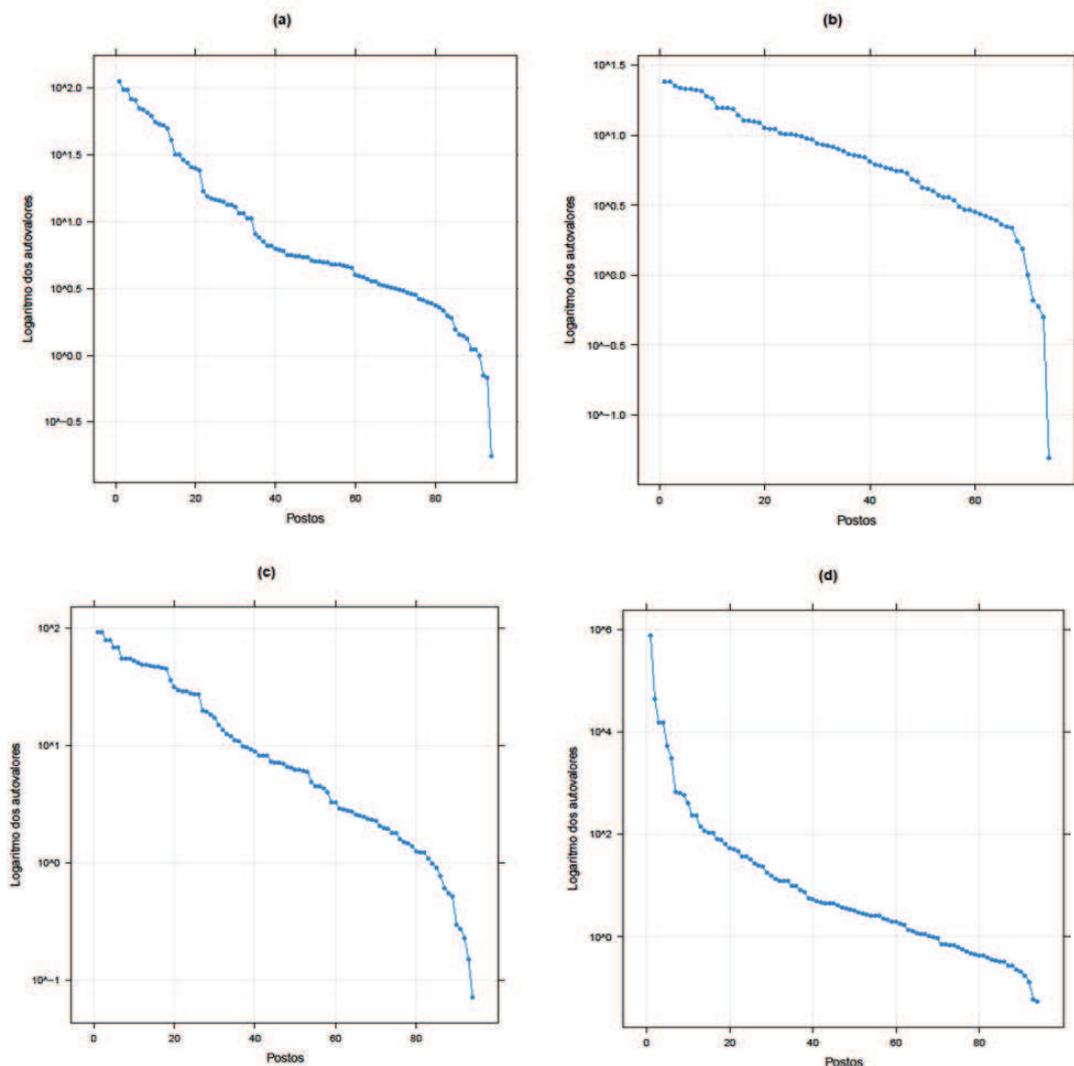


Figura 5.44: Logaritmo dos autovalores: AR (a), MA (b), ARMA (c) e ARIMA (d) (Simulação 2).

A partir da análise no gráfico (a) pode-se notar a formação de três principais pares de autovalores, a saber: (2,3); (4,5) e (6,7). Ao avaliar a separação entre o sinal e o ruído,

na perspectiva das quebras representadas no gráfico (a), observa-se os seguintes possíveis sinais: o agrupamento (1-13), com explicabilidade de 87,51% e correlação ponderada de $\rho_{12}^{(w)} = 0,0633$; os autotriples (1-21), responsáveis por explicar 95,94% da variação da série e cuja w-correlação resultou em $\rho_{12}^{(w)} = 0,0474$; e o grupo (1-34), com uma fração de 98,75% de explicabilidade e correlação ponderada de $\rho_{12}^{(w)} = 0,0131$.

A decomposição SSA na série temporal advinda do particular processo de média móvel, gráfico (b) da Figura 5.44, não indicou a formação evidente de pares de autovalores, com exceção dos dois primeiros, apesar do gráfico apresentar muitos conjuntos de autovalores próximos. Com relação às quebras do espectro, investigou-se cinco quebras: até CP8, CP10, CP19, CP30 e CP67. Para a primeira quebra, até o oitavo componente, verificou-se uma correlação de $\rho_{12}^{(w)} = 0,1404$; em que o sinal reproduz 45,80% da série original. A segunda quebra, até CP10, produziu uma correlação de $\rho_{12}^{(w)} = 0,0992$ e uma fração de explicabilidade de 53,82% por parte do sinal. Na queda seguinte, até CP19, observou-se uma correlação ponderada de $\rho_{12}^{(w)} = 0,1140$; tendo uma fração de explicabilidade de 74,66% por parte do sinal. As duas últimas quebras ora citadas resultaram em uma explicabilidade de 87,87% e 99,91%, respectivamente; e exibiram uma correlação ponderada de $\rho_{12}^{(w)} = 0,1330$ e $\rho_{12}^{(w)} = 0,0194$, respectivamente.

Ao avaliar o gráfico (c), nota-se a presença de três principais pares de autovalores: (1,2); (3,4) e (5,6). Ainda, segundo a análise gráfica, observou-se também as quebras no espectro da matriz trajetória da particular série. Para o processo ARMA(2,1) considerado pode-se notar a presença de quatro importantes quebras ao longo dos 94 postos. Ao investigar a primeira quebra, isto é, considerando o grupo formado por CP1 a CP6 como sinal e os demais componentes como ruído; verifica-se uma correlação ponderada igual a $\rho_{12}^{(w)} = 0,0672$, sendo que esse sinal explica 49,20% da variação da específica série artificial. Ao considerar a próxima quebra, isto é, a formação do sinal até o componente de ordem 18; verificou-se que a correlação ponderada aumentou para $\rho_{12}^{(w)} = 0,1181$, nesse caso, o sinal detém 86,61% de todas as variações da série original. A quebra seguinte, até CP26, retrata um sinal com 95,52% de explicabilidade e uma w-correlação de $\rho_{12}^{(w)} = 0,0483$. Enquanto que, o agrupamento formado até CP53, acumula 99,72% de explicabilidade da série; apresentando uma correlação ponderada de $\rho_{12}^{(w)} = 0,4150$.

Finalmente, ao avaliar o gráfico (d) da Figura 5.44, pode-se observar a existência dos seguintes pares de autovalores: (3,4), (7-8) e (11,12). Analisou-se também as quatro primeiras quebras de tal gráfico, são elas: até CP4, CP6, CP12 e CP38. Ao considerar a primeira quebra, até CP4, nota-se que a w-correlação resulta em $\rho_{12}^{(w)} = 0,0039$; sendo que o particular sinal já consegue explicar 99,9935% da variabilidade do específico processo simulado [ARIMA(1,2,1)]. As correlações ponderadas entre o sinal e o ruído, e as frações de explicabilidade ao considerar as demais quebras foram as seguintes: $\rho_{12}^{(w)} = 0,00018$ e 99,9998%; $\rho_{12}^{(w)} = 5,1283 \times 10^{-5}$ e 99,9999%; $\rho_{12}^{(w)} = 4,9312 \times 10^{-6}$ e quase 100%,

respectivamente.

Ainda através do gráfico (d), Figura 5.44, nota-se que o espectro expressa nitidamente a presença da tendência da série temporal, a qual é representada pelo primeiro autovalor e contribui individualmente com 99,58% para explicar as variações na série original.

De forma análoga ao realizado no primeiro estudo de simulação, avaliou-se todas as opções de agrupamento descritas nos parágrafos anteriores à luz dos erros de previsão provenientes dos diferentes espaços trajetória, para cada série simulada.

No contexto geral, para o processo estocástico AR(2), o sinal formado pelos vinte e um primeiros autotriples apresentou melhores resultados preditivos. Para a série artificial advinda do processo MA(1) os melhores resultados foram obtidos ao considerar os autotriples a partir do trigésimo primeiro como o ruído. Já para o processo ARMA(2,1) observou-se que o agrupamento (1-26) gerou previsões com erros menores. Enquanto que, para o processo teórico ARIMA(1,2,1) o sinal composto pelos 12 componentes principais proporcionou menores erros de previsão.

A Figura 5.45 mostra as correlações ponderadas para os agrupamento finais, para cada série simulada analisada.

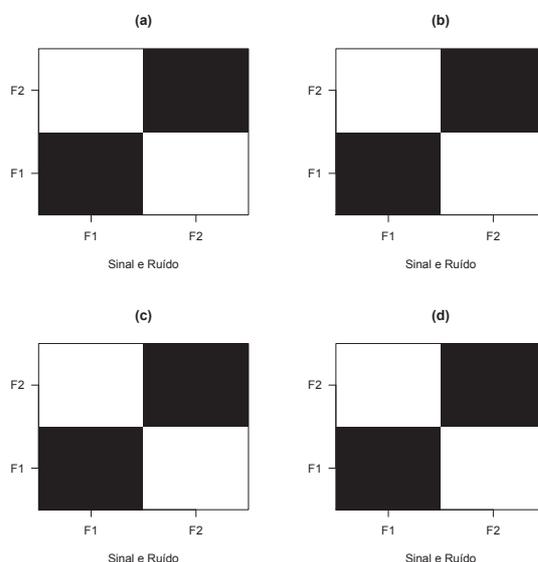


Figura 5.45: Correlação ponderada entre o Sinal e o Ruído: AR, MA, ARMA e ARIMA (Simulação 2).

Os resultados das medidas de precisão avaliadas encontram-se apresentadas nas Tabelas 5.16, 5.17, 5.18 e 5.19.

Ao avaliar os erros cometidos na estimação, nota-se que a SEH, o padrão ouro e a SSA subestimaram as previsões em todos os horizontes considerados (Tabela 5.16).

Tabela 5.16: Comparação entre as previsões (Processo AR - Simulação 2).

h	Real	SEH			AR(2)			SSA		
		Erro	MSE	TPE	Erro	MSE	TPE	Erro	MSE	TPE
1	-1,342	1,924	3,700	-143,310	1,007	1,014	-75,020	6,756	45,640	-503,311
2	-1,327	3,284	7,243	-195,079	0,723	0,768	-64,796	10,012	72,936	-628,081
3	-0,580	5,378	14,469	-325,791	0,748	0,699	-76,262	13,052	105,413	-917,752
4	0,623	7,926	26,556	-704,760	1,087	0,820	-135,740	14,931	134,790	-1703,718
5	0,744	9,392	38,889	-1481,784	0,419	0,691	-211,602	13,799	145,912	-3109,125
6	1,750	11,744	55,396	-29712,483	0,838	0,693	-3613,949	11,859	145,031	-52763,459
7	1,881	13,221	72,454	3025,721	0,627	0,650	311,892	8,108	133,703	4493,421
8	2,710	15,397	93,029	1531,437	1,344	0,795	152,413	5,178	120,342	1877,525
9	5,429	19,461	124,772	887,363	4,123	2,595	110,424	5,269	110,055	899,858
10	6,464	21,842	160,003	670,126	5,321	5,167	99,313	5,435	102,004	577,337
11	6,148	22,872	193,014	588,655	5,205	7,160	95,307	6,096	96,110	446,661
12	5,854	23,923	224,623	551,496	5,095	8,727	93,600	8,193	93,694	383,338

Tabela 5.17: Comparação entre as previsões (Processo MA - Simulação 2).

h	Real	SEH			MA(1)			SSA		
		Erro	MSE	TPE	Erro	MSE	TPE	Erro	MSE	TPE
1	1,117	1,532	2,348	137,173	0,703	0,494	62,910	0,462	0,213	41,361
2	0,368	0,818	1,508	158,262	0,394	0,325	73,871	1,784	1,698	151,246
3	-0,386	0,099	1,009	222,828	-0,360	0,260	67,086	-0,642	1,269	145,951
4	-0,041	0,478	0,814	276,682	-0,015	0,195	68,295	-1,077	1,242	49,811
5	-1,377	-0,823	0,786	-659,704	-1,351	0,521	196,915	-1,380	1,375	267,398
6	1,339	1,928	1,275	395,328	1,365	0,745	72,267	2,309	2,034	142,745
7	-0,386	0,238	1,101	673,495	-0,360	0,657	59,529	0,551	1,787	316,562
8	1,279	1,937	1,432	324,480	1,305	0,788	87,961	-1,298	1,774	37,062
9	-0,040	0,653	1,321	366,278	-0,014	0,700	89,107	0,508	1,606	64,976
10	0,829	1,557	1,431	311,517	0,855	0,703	93,422	2,521	2,081	138,342
11	-0,486	0,277	1,308	392,317	-0,460	0,658	93,166	-0,742	1,942	135,199
12	-1,895	-1,098	1,299	2366,362	-1,869	0,895	61,008	-3,407	2,747	-128,037

Conforme descrito na Tabela 5.16, as medidas de precisão consideradas indicaram o mesmo achado. Ou seja, o MSE e o TPE apontaram que o padrão ouro apresentou menores valores para os doze horizontes; ganhando a competição em todos os passos. A SEH ficou na segunda colocação para as previsões à curto e médio prazo. Por outro lado, a SSA assumiu tal colocação nos horizontes finais ($h = 10, 11, 12$).

A Tabela 5.17 mostra os resultados na série temporal advinda do processo MA(1). Quase todas as previsões geradas pela SEH ficaram superestimadas, com exceção dos passos 5 e 12. O padrão ouro superestimou as predições nos horizontes 3, 4, 5, 7, 9, 11 e 12. Enquanto que, a SSA subestimou nos horizontes $h = 1, 2, 6, 7, 9$ e 10.

Ao analisar o MSE verifica-se que, o padrão ouro produziu os menores valores para quase todos os horizontes, com exceção do primeiro passo. Para esses mesmos horizontes a SEH foi a técnica preditiva vice-líder. Sendo assim, a SSA apresentou menor MSE no passo $h = 1$ (Tabela 5.17).

Tabela 5.18: Comparação entre as previsões (Processo ARMA - Simulação 2).

h	Real	SEH			ARMA(2,1)			SSA		
		Erro	MSE	TPE	Erro	MSE	TPE	Erro	MSE	TPE
1	-4,628	1,137	1,293	-24,572	-2,285	5,220	49,368	-3,580	12,816	77,355
2	-2,189	2,645	4,144	-55,480	-3,704	9,470	87,849	-2,129	8,675	83,747
3	0,788	4,691	10,098	-140,538	-2,506	8,407	140,904	2,747	8,299	49,129
4	2,950	5,921	16,339	-467,473	0,018	6,305	275,314	7,918	21,897	-160,946
5	3,860	5,900	20,034	2598,700	2,494	6,288	-766,157	9,835	36,864	1893,946
6	3,083	4,192	19,624	633,781	3,337	7,096	-68,500	6,257	37,244	544,771
7	1,430	1,608	17,189	492,966	2,599	7,047	-0,899	-1,030	32,075	378,163
8	0,737	-0,016	15,041	432,445	1,952	6,643	31,586	-7,173	34,497	212,995
9	0,472	-1,213	13,533	382,409	1,157	6,053	47,084	-9,769	41,268	47,295
10	0,243	-2,372	12,742	333,443	0,271	5,455	49,409	-8,132	43,753	-74,957
11	-1,011	-4,558	13,473	312,760	-1,421	5,143	33,345	-4,397	41,534	-164,864
12	-2,212	-6,690	16,080	319,222	-2,719	5,330	-22,901	0,341	38,082	-258,699

Segundo os resultados da medida TPE , os menores percentuais foram provenientes do padrão ouro e da SSA. Esta última técnica apresentou-se líder nos seguintes horizontes: $h = 1, 4, 8$ e 9 .

A Tabela 5.18 exibe os resultados da série temporal originária do processo ARMA(2,1).

Com relação ao erro de previsão cometido na estimação, verifica-se que a SEH obteve previsões superestimadas nos horizontes finais (8 a 12). O padrão ouro apresentou previsões superestimadas nos passos 1, 2, 3, 11 e 12. E a SSA resultou em previsões subprevisadas para os horizontes 3, 4, 5, 6 e 12 (Tabela 5.18).

Para a série artificial gerada a partir de um ARMA(2,1) o padrão ouro mostrou os menores MSE 's para os horizontes 4 a 12. A SEH e a SSA lideraram nos passos $h = 1, 2$ e $h = 3$, respectivamente. Os menores erros percentuais relativos correspondeu ao padrão ouro, para os passos 5 a 12; à SEH, para os dois primeiros horizontes; e à SSA, para os passos 3 e 4.

Os resultados para a série artificial advinda do processo ARIMA (1,2,1) são mostrados na Tabela 5.19. A SEH subestimou e a SSA supertimou em todos horizontes; enquanto que o específico padrão ouro subpreveu nos horizontes 4, 5 e 6. Nota-se que o MSE e TPE apontaram resultados semelhantes. Para ambas medidas o padrão ouro apresentou os menores valores absolutos em todos os horizontes. A técnica SSA foi responsável pela segunda colocação, nos passos $h = 10, 11, 12$, segundo o MSE . E de acordo com o TPE , a SSA apresentou melhores resultados para os dois últimos horizontes (Tabela 5.19).

Os escores de desempenho da previsão SSA para os processos teóricos considerados são mostrados na Tabela 5.20.

Tabela 5.19: Comparação entre as previsões (Processo ARIMA - Simulação 2).

h	Real	SEH			ARIMA(1,2,1)			SSA		
		Erro	MSE	TPE	Erro	MSE	TPE	Erro	MSE	TPE
1	-20827,690	5,750	33,063	-0,028	-0,510	0,260	0,002	-36,700	1346,890	0,176
2	-20993,730	17,410	168,085	-0,055	-0,607	0,314	0,003	-50,874	1967,527	0,209
3	-21154,390	34,450	507,658	-0,091	-0,151	0,217	0,002	-65,832	2756,302	0,244
4	-21310,900	55,640	1154,696	-0,134	0,217	0,174	0,001	-82,141	3754,012	0,279
5	-21463,650	80,590	2222,706	-0,183	0,622	0,217	0,000	-99,352	4977,374	0,317
6	-21613,980	107,960	3794,815	-0,237	0,178	0,186	0,000	-117,863	6463,093	0,355
7	-21762,120	137,520	5954,378	-0,295	-0,946	0,287	0,001	-136,938	8218,653	0,395
8	-21908,270	169,070	8783,163	-0,356	-2,601	1,097	0,002	-155,862	10227,942	0,436
9	-22053,190	201,850	12334,303	-0,420	-5,239	4,025	0,005	-174,576	12477,813	0,477
10	-22195,570	237,170	16725,834	-0,487	-7,281	8,924	0,008	-191,083	14881,303	0,516
11	-22335,070	275,370	22098,816	-0,557	-8,150	14,150	0,010	-204,504	17330,447	0,554
12	-22474,410	313,730	28459,457	-0,629	-10,357	21,910	0,013	-217,163	19816,224	0,589

Tabela 5.20: Escores de desempenho preditivo SSA (Simulação 2).

Processos	Primeiro lugar		Segundo lugar	
	MSE	TPE	MSE	TPE
AR	0	0	3	3
MA	1	4	0	8
ARMA	1	2	1	9
ARIMA	0	0	3	2

Ao analisar os escores criados para sintetizar os resultados, nota-se que para o processo teórico MA(1), conforme o TPE , a SSA apresentou o segundo melhor desempenho em 8 dos 12 horizontes considerados. Ainda de acordo com a medida relativa TPE , a SSA assumiu a segunda posição em 9 horizontes para o processo ARMA (2,1). Em ambos os casos o algoritmo de previsão SSA superou o modelo de SEH (Tabela 5.20).

A Figura 5.46 mostra o comportamento geral dos três métodos preditivos ao longo dos horizontes considerados, para cada processo teórico investigado: AR (a), MA (b), ARMA (c) e ARIMA (d).

Através do gráfico (a), Figura 5.46, observa-se que a técnica considerada como **padrão ouro**, no caso AR(2), apresentou-se mais próxima da realidade dos dados reservados para o conjunto de teste. Ao comparar os resultado gerados pela SEH e SSA, nota-se claramente que o método SSA expressa melhor as flutuações existentes no conjunto real, já que a SEH estimou uma reta decrescente.

O gráfico (b) da Figura 5.46 demonstra que, em geral, a técnica SSA apresenta melhor desempenho, pois a curva resultante das predições reflete bem os picos e variações do particular processo de média móvel simulado [MA(1)]. Como pode ser visto no gráfico ora citado, o **padrão ouro** e a SEH estimaram uma reta.

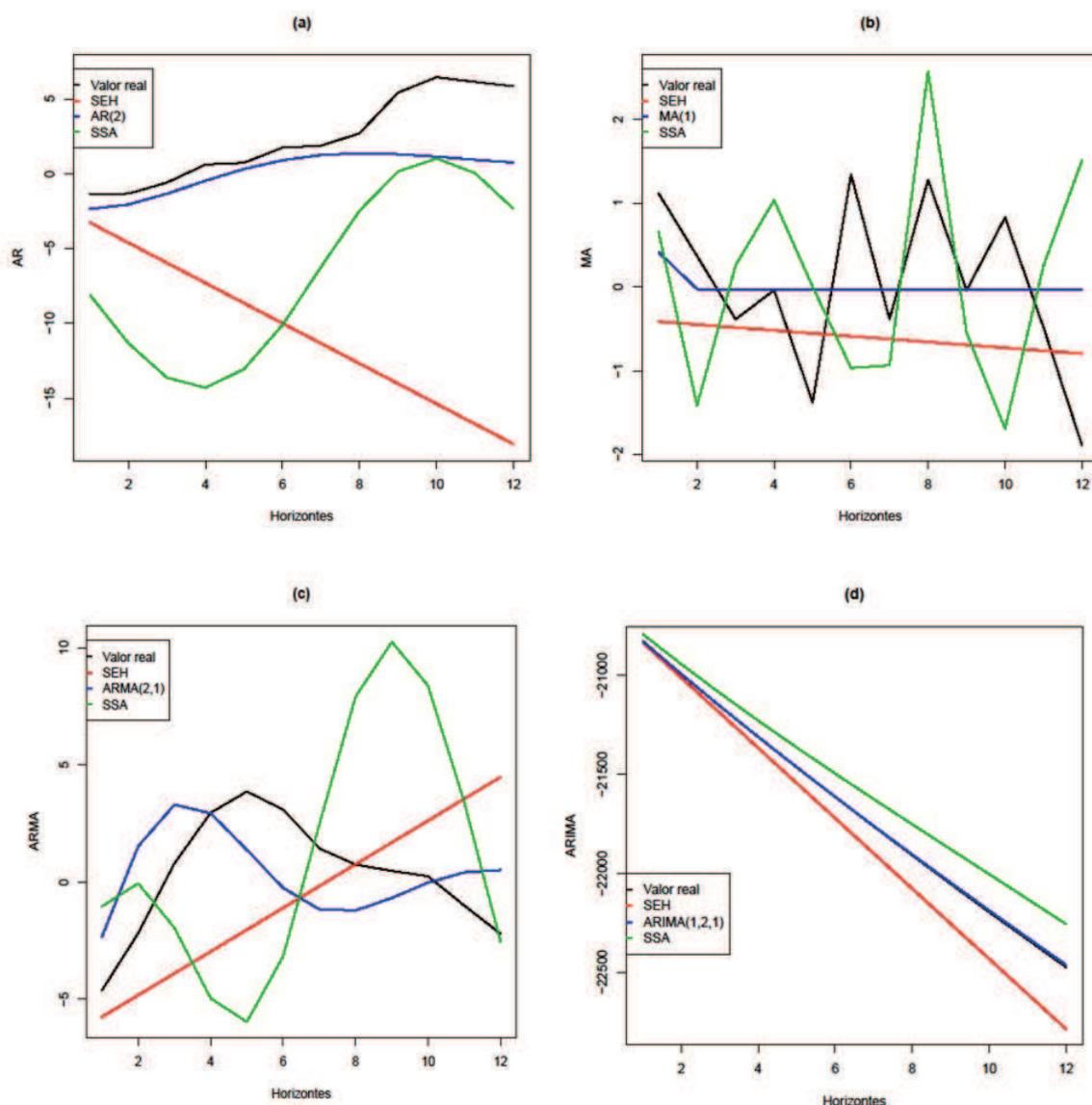


Figura 5.46: Comparação gráfica (AR (a), MA (b), ARMA (c) e ARIMA (d) - Simulação 2).

Ainda de acordo com a Figura 5.46, o gráfico (c) mostra que a curva oriunda do padrão ouro é a que mais assemelha-se à curva real, em preto. Por outro lado, a SSA exibiu estimativas mais voláteis, embora as previsões não tenham sido precisas.

Tendo como base a série obtida a partir de um processo ARIMA, gráfico (d) da Figura 5.46, verificou-se que as previsões oriundas dos três métodos resultaram em uma reta decrescente, assim como a representação gráfica do correspondente conjunto de teste. Isto é, todos os métodos conseguiram reproduzir satisfatoriamente bem a realidade dos valores futuros da série. Todavia, como era de se esperar, o padrão ouro mostrou-se mais preciso do que os demais métodos, embora a diferença entre os métodos seja mais acentuada nas previsões à médio e longo prazo. Pode-se notar que à longo prazo a SSA expôs previsões mais precisas do que a SEH.

Considerações finais

O presente trabalho científico teve como objetivo central avaliar o processo de modelagem e a capacidade preditiva da técnica SSA em séries univariadas empíricas e simuladas, confrontando-a com os principais métodos clássicos. Aplicou-se a família de algoritmos de alisamento exponencial, especificamente, a suavização exponencial de Holt (SEH) e a suavização exponencial de Holt-Winters (H-W); assim como a classe de modelos de Box e Jenkins como estratégias padrões de modelagem e previsão.

De acordo com os resultados apresentados no Capítulo 5, observou-se que a Análise Espectral Singular é uma técnica bastante poderosa no processo de modelagem das séries temporais, principalmente na identificação dos seus componentes constitutivos, o que levou a aproximações com erros quase nulos; superando as estimativas geradas pelas decomposições clássicas. No que tange a capacidade preditiva, constatou-se que, no geral, o algoritmo recorrente de previsão SSA produziu resultados tão bons ou superiores aos gerados pelos métodos clássicos considerados na pesquisa.

6.1 Conclusões

As principais conclusões das análises mostradas no capítulo anterior são dissertadas nessa seção.

No estudo empírico, verificou-se que, quando analisado ambos contextos do algoritmo de suavização exponencial, a estratégia aditiva de Holt-Winters mostrou-se mais precisa do que sua contrapartida multiplicativa. No que tange a metodologia de Box e Jenkins, observou-se que o modelo SARIMA com mais parâmetros apresentou-se mais eficiente, no que diz respeito a capacidade preditiva, para ambas séries históricas.

Para os estudos empírico e simulado, observou-se que geralmente as medidas de erro MSE e TPE não indicaram os mesmos métodos como vencedores das comparações investigadas. Em geral, a SSA apresentou melhor capacidade preditiva quando o interesse estava na produção de previsões globalmente precisas, assim como aferido pelo TPE , já que esta medida quantifica o acúmulo dos erros em termos relativos. Por outro lado, de uma forma geral, caso o interesse estivesse na obtenção de previsões pontualmente precisas, como as mensuradas através do MSE , as técnicas clássicas indicaram menores erros.

No contexto do estudo empírico, verificou-se que para a série VENTO, série histórica com uma estrutura bem definida, a SSA mostrou-se melhor nas previsões à curto ou médio prazo, ao considerar a medida relativa *TPE*; e ao avaliar o *MSE*, o algoritmo de Holt-Winters apresentou, em geral, melhor desempenho preditivo. Já para a série IBOVESPA, série tradicionalmente imprevisível e de difícil modelagem, o algoritmo recorrente de previsão SSA não representou bem os valores futuros, apesar de apresentar erros razoáveis nos primeiros horizontes. Todavia, as demais estratégias de previsão avaliadas também não conseguiram reproduzir satisfatoriamente bem o conjunto de teste.

Outro resultado interessante, com relação a previsão SSA, na série financeira refere-se a observação de equivalência ao utilizar a série original como a série resultante, ao invés da série aproximada através do processo de reconstrução SSA, pois não houve grandes diferenças nas predições de ambas as estratégias.

No âmbito do estudo simulado, observou-se que para os processos teóricos com estruturas mais complexas o algoritmo de previsão SSA tende a apresentar melhor capacidade preditiva do que a SEH, e para alguns horizontes, a SSA mostrou-se equivalente ou superior ao desempenho do correspondente **padrão ouro**. Esse resultado foi observado nos seguintes processos estocásticos: ARMA (1,2) e ARIMA (1,1,2), para a simulação 1; MA(1) e ARMA (2,1), para a simulação 2. Vale ressaltar que, tanto pelo *MSE* quanto pelo *TPE*, a SSA exibiu predições mais precisas para o processo ARIMA (1,1,2); processo teórico com comportamento bastante comum na realidade.

De uma forma geral, para todas as séries analisadas no trabalho de dissertação, notou-se que o algoritmo recorrente de previsão SSA consegue representar melhor as variações existentes nos dados, como flutuações sazonais e picos, características tradicionais encontradas na prática, isto é, em muitas séries temporais reais. Portanto, a previsão SSA apresentou um comportamento global mais fidedigno à realidade dos valores futuros das séries. Além disso, observou-se também que o desempenho preditivo da SSA diminui conforme aumenta-se a quantidade de autotriples no espaço trajetória.

A grande vantagem em aplicar a SSA em detrimento aos algoritmos de suavização exponencial ou modelos de Box-Jenkins refere-se a sua simplificação no entendimento das séries temporais e conseqüente diminuição da intervenção do analista, pois o processo de modelagem e previsão via SSA já leva em consideração uma fundamentação teórica que envolve a decomposição e reconstrução dos componentes constitutivos da série temporal. Em outras palavras, os métodos clássicos ora citados exigem uma detalhada análise exploratória visando definir os parâmetros necessários para o funcionamento dessas técnicas; em contrapartida, a SSA exige a definição de apenas dois parâmetros: o comprimento da janela e a forma de agrupamento.

Contudo, o procedimento de agrupamento dos autotriples, parâmetro do primeiro passo da reconstrução SSA, pode ser considerado um ponto crítico da análise via SSA; pois, por ser um parâmetro estrutural, a sua definição baseia-se em poucos critérios objetivos. Este ponto crítico seria equivalente a uma das desvantagens das clássicas técnicas consideradas no trabalho, a saber: a identificação do modelo mais adequado aos dados; fato este que possibilita a definição de múltiplos modelos para uma mesma série temporal.

6.2 Contribuições

O presente trabalho contribuiu cientificamente com o estudo do processo de modelagem e previsão através da técnica Análise Espectral Singular, pois espera-se que ocorra a difusão de uma alternativa metodológica, que tem se mostrado poderosa na identificação ou extração dos componentes constituintes da série temporal e na sua capacidade de previsão.

Ao longo do período de investigação da presente pesquisa científica, tivemos a oportunidade de apresentar, em formato pôster, e publicar trabalhos nos anais dos seguintes eventos científicos: XIV Escola de Séries Temporais e Econometria (ESTE), realizada no Centro de Eventos da FAURGS em Gramado- RS, no período de 01 a 05 de agosto de 2011; e no XLIII Simpósio Brasileiro de Pesquisa Operacional (SBPO) realizado no período de 15 a 18 de agosto de 2011 em Ubatuba- SP (ESQUIVEL, SENNA e GOMES, 2011a; ESQUIVEL, SENNA e GOMES, 2011b).

6.3 Atividades futuras de pesquisa

Naturalmente, assim como apontado na Seção 1.3, uma proposta de continuação deste trabalho consiste na implementação da metodologia de previsão SSA no R; *software* estatístico de código-fonte aberto e com a possibilidade de trabalhar com interface amigável. Outro ponto de extensão interessante e desafiador refere-se ao desenvolvimento de um sistema automatizado que contemple todos os passos da modelagem e previsão via SSA, além de incorporar a otimização do seu parâmetro estrutural, isto é, otimizar a escolha dos autotriples no processo de agrupamento.

Acredita-se que essas promissoras propostas irão disponibilizar aos planejadores e pesquisadores uma importante ferramenta de trabalho, já que essa ferramenta facilitará a utilização por parte dos usuários da Estatística nas diversas áreas do conhecimento. Além de contribuir para a disseminação da Análise Espectral Singular entre os estatísticos, visto que o R é um *software* bastante utilizado por esse público.

Com relação a trabalhos científicos futuros, a atual pesquisa renderá mais alguns produtos. Um resumo estendido foi submetido para apresentação, em formato pôster, na XX edição do Simpósio Nacional de Probabilidade e Estatística (SINAPE), evento que será realizado em João Pessoa-PB entre os dias 30 de julho e 3 de agosto de 2012. Além da submissão de um artigo para apresentação, na categoria oral, no evento CLAIO/SBPO; que será realizado no Rio de Janeiro, de 24 a 28 de setembro do corrente ano. A título de curiosidade, a CLAIO é a conferência bianual organizada pela ALIO [do espanhol, *Asociación Latino-Iberoamericana de Investigación Operativa*, uma associação regional Latino-Americana da IFORS (do inglês, *International Federation of Operational Research Societies*)]. Outra publicação bastante almejada refere-se a submissão de um artigo científico, em uma revista de grande impacto na área de Estatística, como, por exemplo, o *Journal of Forecasting*.

Testes estatísticos em séries temporais

A.1 Testes de Box-Pierce e Ljung-Box

A estatística de Box e Pierce pode ser definida da seguinte forma:

$$Q = N \sum_{k=1}^m \hat{r}_k^2, \quad (\text{A.1})$$

onde N é tamanho da amostra e m a duração da defasagem. Em grandes amostras, a estatística Q se distribui segundo uma distribuição qui-quadrado com m graus de liberdade (gl).

A estatística de Ljung-Box pode ser expressa como:

$$LB = N(N+2) \sum_{k=1}^m \left(\frac{\hat{r}_k^2}{N-k} \right), \quad (\text{A.2})$$

A distribuição assintótica da estatística LB também é qui-quadrado com m graus de liberdade (gl).

Para ambos os testes, a hipótese nula refere-se a não existência de autocorrelação conjunta. A hipótese nula será rejeitada se o valor da estatística obtida através da série exceder o valor crítico da distribuição qui-quadrado para certo nível de significância α e grau de liberdade $gl = m$.

A.2 Testes de Dickey-Fuller e KPSS

- Teste de Dickey-Fuller

A estatística do clássico teste Dickey-Fuller (DF) não segue uma distribuição tradicional ou conhecida, esta foi tabulada considerando simulações de Monte Carlo (DICKEY e FULLER, 1979; GUJARATI, 2000).

Ao considerar o modelo simples, apresentado na expressão (2.3), pode-se subtrair Y_{t-1} de ambos os lados e obter a seguinte expressão equivalente:

$$\Delta Y_t = \phi^* Y_{t-1} + \varepsilon_t, \quad (\text{A.3})$$

onde $\phi^* = 1 - \phi$. Desta maneira, a hipótese nula passa a ser representada por $H_0: \phi^* = 0$, ao invés da $H_0: \phi = 1$.

A partir da expressão (A.3), costuma-se desenvolver a estatística de teste de DF ou estatística *tau* para o modelo AR(1), que pode ser expressa da seguinte maneira:

$$\hat{\tau} = \frac{\frac{\sum_{t=2}^N Y_{t-1} \varepsilon_t}{N}}{\sqrt{\sigma_\varepsilon^2 \times \left(\frac{\sum_{t=2}^N Y_{t-1}^2}{N^2} \right)}}. \quad (\text{A.4})$$

- Teste KPSS

O teste KPSS considera duas formas de avaliação da estacionariedade. O KPSS testa a hipótese nula de nível estacionário, assim como, a hipótese de tendência estacionária contra a alternativa de uma raiz unitária. Para tal teste, assume-se que a série pode ser decomposta na soma de uma tendência determinística (ξd), um passeio aleatório (W_t), e um erro estacionário (ε_t). Em outras palavras, o teste fundamenta-se no modelo

$$Y_t = \xi d + W_t + \varepsilon_t, \quad (\text{A.5})$$

$$W_t = W_{t-1} + u_t$$

onde u_t representa as flutuações do passeio aleatório, que são identicamente distribuídos (iid) com média zero e variância σ_u^2 e possui a suposição adicional de normalidade [$u_t \stackrel{iid}{\sim} N(0, \sigma_u^2)$]; ε_t é um ruído branco normal [$\varepsilon_t \stackrel{iid}{\sim} \text{RB } N(0, \sigma_\varepsilon^2)$].

A hipótese de estacionariedade pode ser formulada como $H_0: \sigma_u^2 = 0$. Se $\xi = 0$, sob a H_0 , tem-se o caso especial do modelo (A.5), em que Y_t é estacionária ao redor de um nível

(W_0), ao invés de estacionária ao redor de uma tendência (KWIATKOWSKI *et al.*,1992; LIMA, 2008)

A estatística de teste do KPSS é dada por

$$KPSS = \sum_{t=1}^N \frac{S_t^2}{\sigma_\varepsilon^2}, \quad (\text{A.6})$$

em que $S_t = \sum_{i=1}^t e_i$; $t = 1, 2, \dots, N$. Sendo e_i os erros estimados de ε_t ou os resíduos e σ_ε^2 a variância do erro ε_t .

A.3 Testes de Cox-Stuart e Spearman

- Teste de Cox-Stuart

O teste de Cox-Stuart é uma adaptação do teste dos sinais para análise de tendência, aplicável a uma ampla variedade de situações, dentro do contexto de séries temporais é aplicado para conhecer a evolução dos dados.

Para tal teste é necessário dividir a série temporal em duas partes, formando assim duas amostras ou grupos de observações dependentes (Grupo 1 e Grupo 2). Desta forma, as observações de cada grupo são pareadas da seguinte maneira:

$$(Y_1; Y_{1+c}), (Y_2; Y_{2+c}), (Y_3; Y_{3+c}), \dots, (Y_{(N-1)-c}; Y_{N-1}), (Y_{N-c}; Y_N),$$

$$\text{em que } c = \begin{cases} \frac{N}{2}, & N \text{ par} ; \\ \frac{N+1}{2}, & N \text{ mpar.} \end{cases}$$

Em seguida são atribuídos os sinais. Se um particular valor do Grupo 1 (Y_1, Y_2, \dots, Y_{N-c}) for menor do que um valor do Grupo 2 ($Y_{1+c}, Y_{2+c}, \dots, Y_N$), o sinal '+' é associado ao correspondente par. E o sinal '-', se a observação no Grupo1 exceder o valor no Grupo 2. Os empates são descartados do procedimento.

A hipótese nula de ausência de tendência é testada através de uma estatística que quantifica o número de pares com sinais positivos. Esta estatística distribui-se segunda uma Binomial ($n, \frac{1}{2}$), se $n \leq 20$, sendo n o número de pares sem empate; e se $n > 20$,

pode-se utilizar a aproximação normal. E a hipótese nula será rejeitada se o módulo do valor obtido através da estatística do teste for maior ou igual ao quantil da distribuição Binomial ou Normal.

A partir do resultado da estatística de teste pode-se ter um indicativo de qual sinal é mais provável, ou seja, a direção da tendência pode ser avaliada ao considerar um teste unilateral.

- Spearman para séries temporais

O teste de Spearman para séries temporais fundamenta-se no coeficiente de correlação de Spearman. Em tal teste avalia-se a existência de tendência crescente ou decrescente entre as observações e o tempo. Segundo Morettin e Tolo (2006), geralmente o teste de Spearman é mais poderoso que o teste de Cox-Stuart.

Considere R_t como sendo os postos associados às observações da série Y_t . A expressão abaixo representa a estatística do teste de Spearman para séries temporais:

$$T_{\text{Spearman}} = \sum_{t=1}^N [R_t - t]^2 \quad (\text{A.7})$$

em que, t identifica os instantes de tempo ($t = 1, 2, \dots, N$).

Os quantis da distribuição da estatística T_{Spearman} encontra-se tabelada e disponível em vários livros de Estatística.

A hipótese da ausência de tendência na série será rejeitada se o valor obtido a partir da estatística T_{Spearman} for menor que o quantil $Q_{\alpha/2}$ ou maior do que o quantil $Q_{1-\alpha/2}$.

Usando o nível descritivo ou p -valor, rejeita-se a hipótese nula se o p -valor for no máximo igual ao nível de significância adotado.

A.4 Testes de *Kruskal-Wallis* e *Friedman*

- Teste de *Kruskal-Wallis*

Ao aplicar o teste de *Kruskal-Wallis* em uma série temporal mensal, cada mês da série é suposto ser uma das k amostras de uma população dentro de cada um dos p

anos. Considere Y_{ij} ($i= 1, \dots, n_j ; j= 1, \dots, k$) as observações da série. Sejam R_{ij} os postos obtidos ao ordenar todas as N observações $\left(N = \sum_{j=1}^k n_j\right)$ e R_j a soma dos postos associados à j -ésima amostra $\left(R_j = \sum_{i=1}^{n_j} R_{ij}\right)$.

A estatística do teste de *Kruskal-Wallis* será expressa por

$$T_{KW} = \frac{12}{N(N+1)} \sum_{j=1}^k \frac{(R_j)^2}{n_j} - 3(N+1). \quad (\text{A.8})$$

Sob H_o , a distribuição assintótica de T_{KW} é uma qui-quadrado com $k-1$ graus de liberdade. A hipótese nula de ausência de sazonalidade é rejeitada se o valor calculado a partir de T_{KW} for maior ou igual ao valor crítico da distribuição da estatística T_{KW} .

- Teste de *Friedman*

Para o teste de *Friedman*, considera-se os “tratamentos” ou “colunas” como sendo os meses e os anos são considerados “blocos” ou “linhas”. A ideia deste teste é semelhante a do teste de *Kruskal-Wallis*, porém no teste de *Friedman* as ordenações das observações são feitas por bloco em vez da ordenação geral (todas as N observações da série).

A estatística de teste é a seguinte:

$$T_F = \frac{12}{pk(k+1)} \sum_{j=1}^k (R_j^*)^2 - 3p(k+1), \quad (\text{A.9})$$

em que j indexa as colunas ou meses da série ($j = 1, \dots, k$); R_j^* representa a soma dos postos na coluna/mês j ; e $\sum_{j=1}^k (R_j^*)^2$ denota o somatório dos quadrados das somas dos postos R_j^* sobre as k colunas.

A estatística de *Friedman* tem distribuição aproximadamente qui-quadrado com $k-1$ graus de liberdade, para p (número de linhas) e k (número de colunas) não muito pequenos. Caso tenha amostras pequenas, a distribuição da estatística T_F é exata e encontra-se tabelada (SIEGEL, 1975; CAMPOS, 1983).

Como regra de decisão do teste tem-se que, se T_F calculado exceder o valor crítico da distribuição, a hipótese nula será rejeitada.

Exemplos simulados- comportamento da fac e facp

A Figura B.1 apresenta o gráfico de uma série temporal simulada com 200 observações segundo um modelo autoregressivo de ordem 1 [AR(1)] e parâmetro autorregressivo igual a $\phi_1=0,7$ (a); assim como os gráficos da sua fac (b) e facp (c).

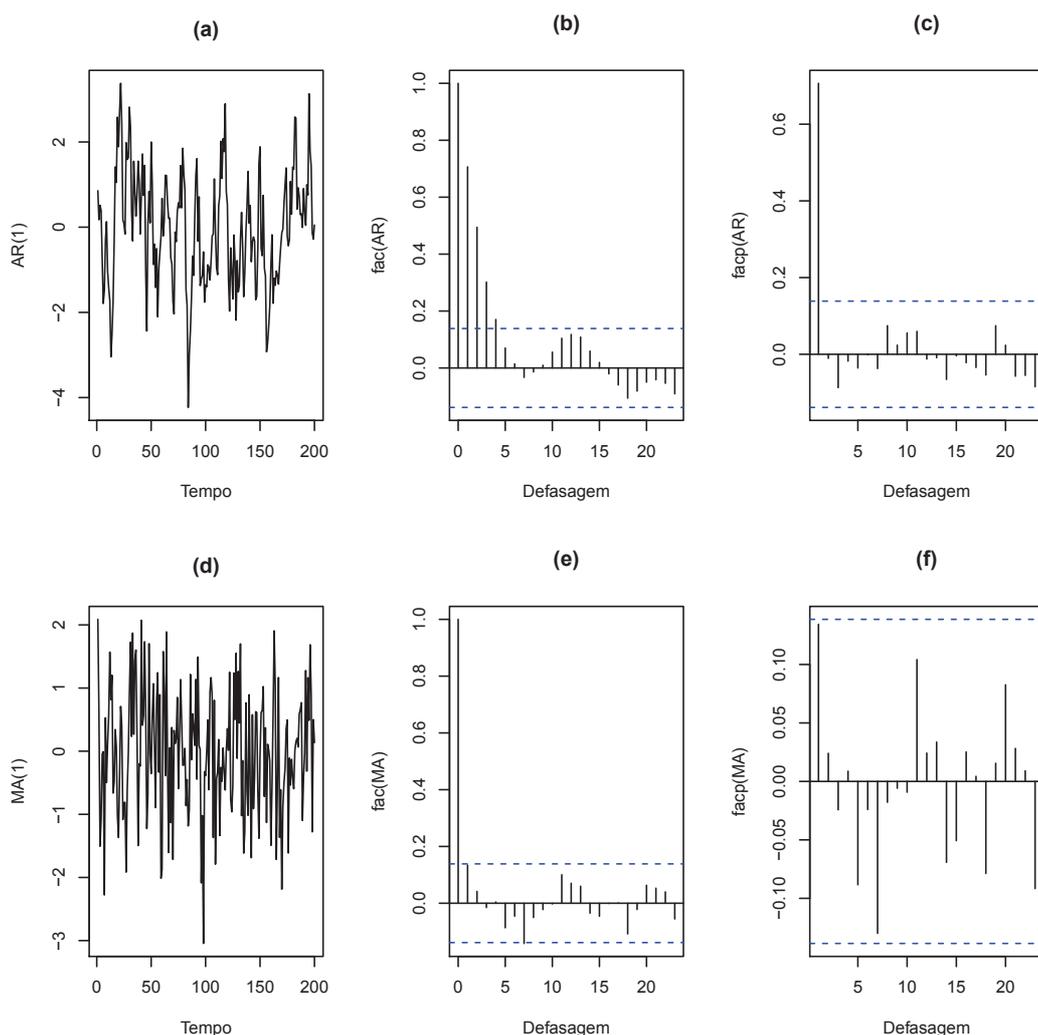


Figura B.1: Gráficos das séries simuladas, modelo AR(1) e $\phi_1=0,7$ (a); fac do AR (b) e facp do processo AR(1) (c); modelo MA(1) e $\theta_1 = 0,1$ (d); fac do MA (e) e facp do MA (f).

Como pode ser observado, a fac estimada (b) para os dados simulados apresenta um decaimento exponencial rápido para zero; e o gráfico da facp (c) mostra que apenas a primeira autocorrelação parcial é estatisticamente diferente zero, com 95% de confiança, indicando que o processo gerador da série é um autoregressivo de ordem 1.

A segunda parte da Figura B.1 mostra uma outra série temporal simulada, também com 200 observações, conforme um processo de médias móveis de ordem 1 [MA(1)] e $\theta_1 = 0,1$ (d); a função autocorrelação (e) e autocorrelação parcial (f).

Assim como apresentado no gráfico da fac estimada do MA (e), apenas a primeira autocorrelação difere de zero, indicado que o processo gerador dos dados refere-se a um MA de ordem 1.

Álgebra de matrizes e Estatística Multivariada

O presente apêndice expõe as definições básicas de Álgebra linear matricial e Estatística Multivariada, em um contexto introdutório para auxiliar no entendimento da metodologia da técnica SSA.

C.1 Autovalores e autovetores

Os autovalores e seus correspondentes autovetores são úteis em muitas situações práticas. De forma geral, os autovalores têm o importante papel de simplificar a matriz de dados. E os autovetores são comumente interpretados como os eixos de um sistema de coordenadas diferentes dos originais.

A seguir as definições de autovalores e autovetores são apresentadas.

Dada uma transformação linear de um espaço vetorial ($T : V \rightarrow V$), o interesse encontra-se na identificação de um vetor (\mathbf{v}) e um escalar (λ), tais que

$$T(\mathbf{v}) = \lambda\mathbf{v}. \quad (\text{C.1})$$

Se existirem $\mathbf{v} \in V$, com $\mathbf{v} \neq 0$, tais que (C.1) é satisfeita; λ é chamado *autovalor* ou *valor característico* de T e \mathbf{v} é um *autovetor* ou *vetor característico* associado a λ .

C.1.1 Autovalores e autovetores de uma matriz

No contexto de matrizes, um autovalor λ e um autovetor \mathbf{v} de uma matriz $\mathbf{A}_{n \times n}$ são soluções da equação

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}. \quad (\text{C.2})$$

Comumente, utiliza-se um procedimento prático para encontrar os autovalores e autovetores de uma matriz ao invés de trabalhar com a definição dada em (C.1) ou (C.2).

Em tal procedimento é aplicado um polinômio conhecido como *polinômio característico* ou *equação característica*.

Dada uma matriz, $\mathbf{A}_{n \times n}$, os seus autovalores e autovetores satisfazem a equação (C.2) ou de forma prática satisfazem a expressão $\mathbf{A}\mathbf{v} = (\lambda\mathbf{I})\mathbf{v}$, ou ainda

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = 0; \quad (\text{C.3})$$

em que \mathbf{I} representa a matriz identidade de ordem n .

Fazendo uso das propriedades de sistemas de equações lineares homogêneas, a única forma de encontrar autovetores \mathbf{v} (soluções não nulas) é resolver a equação característica da matriz \mathbf{A} (BOLDRINI *et.al.*, 1980; ELSNER e TSONIS, 1996), ou seja,

$$P(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}) = 0. \quad (\text{C.4})$$

Sendo $P(\lambda)$ um polinômio de grau n e as suas raízes os autovalores da matriz \mathbf{A} .

Elsner e Tsonis (1996), também definem os autovalores como sendo os λ 's que tornam a matriz $\mathbf{A} - \lambda\mathbf{I}$ singular.

C.2 Estatística Multivariada

O propósito desta seção é apresentar os principais conceitos utilizados na estatística multivariada. Os conceitos sobre vetor aleatório, matrizes de covariâncias e de correlação, bem como algumas propriedades importantes da análise de dados multivariados são tratados a seguir.

C.2.1 Vetor e matriz aleatória

Ao considerar dados multivariados, com p variáveis aleatórias e uma quantidade n de observações para cada variável, os dados são organizados em uma matriz formada por p vetores aleatórios; assim como mostrado abaixo

$$\mathbf{X}_{n \times p} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1} & x_{n1} & \cdots & x_{np} \end{bmatrix} = \begin{bmatrix} X_1^T \\ X_2^T \\ \vdots \\ X_p^T \end{bmatrix}; \quad (\text{C.5})$$

em que x_{ij} indica a i -ésima observação da j -ésima variável .

O vetor aleatório $\mathbf{X} = \begin{bmatrix} X_1^T \\ X_2^T \\ \vdots \\ X_p^T \end{bmatrix}$ é frequentemente denotado por

$$X^T = \begin{bmatrix} X_1 & \cdots & X_p \end{bmatrix} \text{ ou } X = \begin{bmatrix} X_1 & \cdots & X_p \end{bmatrix}^T,$$

o que representa o vetor transposto de \mathbf{X} .

C.2.1.1 Médias, covariâncias e correlações

As médias do vetor aleatório \mathbf{X} são definidos por

$$E(\mathbf{X}) = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \mu; \quad (\text{C.6})$$

em que $E(X_i) = \mu_i$ denota a esperança ou média da variável aleatória X_i , $i = 1, \dots, p$.

Tradicionalmente, o grau de dependência entre as variáveis é apresentado na matriz de variâncias e covariâncias do vetor aleatório \mathbf{X} , esta pode ser denotada como

$$Cov(\mathbf{X}) = \Sigma_{p \times p} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}; \quad (\text{C.7})$$

onde , $\sigma_{ij} = Cov(X_i, X_j) = E[(X_i - \mu_i)(X_j - \mu_j)]$.

A matriz de covariâncias é uma matriz real e simétrica , ou seja , $\sigma_{ij} = \sigma_{ji}, \forall i \neq j$. Além disso, esta matriz pode ser não negativa definida, isto é , $C_{1 \times p} \Sigma_{p \times p} C_{p \times 1} \geq 0$; em que $C_{p \times 1} \in \mathbb{R}^p$ é um vetor de constantes , assim como o seu transposto $C_{1 \times p}$. Condição esta que implica em autovalores não negativos ($\lambda_i \geq 0, \forall i = 1, \dots, p$). Ou ainda, a matriz de covariâncias pode ser positiva definida, ou seja, $C_{1 \times p} \Sigma_{p \times p} C_{p \times 1} > 0$, para todo vetor de constantes não nulo. Para este último caso, os autovalores de $\Sigma_{p \times p}$ são todos positivos ($\lambda_i > 0$) e a matriz de covariâncias é dita não singular, e assim sendo, o seu determinante será maior que zero e admitirá inversa (MINGOTI, 2005).

Na diagonal principal da matriz (C.7) , para $i=j$, tem-se a variância de cada variável X_i . Ou seja,

$$Cov(X_i, X_i) = Var(X_i).$$

Outra medida de variabilidade muito útil na análise de dados multivariados é conhecida como *variância total*. A variância total sumariza a variância global da distribuição multivariada através do traço da matriz de covariâncias $\Sigma_{p \times p}$, ou seja, considera a soma das variâncias das p variáveis envolvidas na análise. Esta medida pode ser expressa como

$$traço(\Sigma_{p \times p}) = tr(\Sigma_{p \times p}) = \sigma_{11} + \sigma_{22} + \dots + \sigma_{pp}. \quad (C.8)$$

Onde, um valor alto para (C.8) indica que há grande dispersão global das variáveis $X_i, i = 1, \dots, p$.

Ao considerar a análise multivariada, a matriz de dados é investigada conjuntamente, já que pode ter relacionamentos interessantes entre as p -variáveis . Os relacionamentos lineares são identificados através dos coeficientes de correlação entre as variáveis, que costumam ser sintetizados na matriz de correlações.

A matriz de correlações populacionais é uma matriz também real e simétrica, $p \times p$, que pode ser denotada como

$$\rho_{p \times p} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{bmatrix} \quad (C.9)$$

em que, $-1 \leq \rho_{ij} \leq 1$; $i, j = 1, 2, \dots, p$. O coeficiente de correlação populacional entre a i -ésima e j -ésima variável do vetor \mathbf{X} é definido por

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}}.$$

E nos casos em que $i = j$, tem-se $\rho_{ii} = 1$; o que representa o coeficiente de correlação entre uma mesma variável aleatória.

O coeficiente de correlação é adimensional, já que este é a covariância padronizada pelos desvios padrão. E sendo assim, esta medida não sofre influências das diferenças de escalas de medidas entre variáveis. Além disso, como o coeficiente de correlação é limitado entre os valores de referência -1 e 1 , este pode identificar a magnitude da correlação linear entre as variáveis; por exemplo, uma correlação próxima de zero indica um não relacionamento linear entre duas das variáveis em questão X_i e X_j .

Na prática, como frequentemente trabalha-se com amostras ao invés de populações, as estimativas das médias, das covariâncias e das correlações são consideradas na análise dos dados.

C.2.2 Decomposição espectral

Em estatística multivariada um teorema da álgebra de matrizes muito importante é intitulado como *teorema da decomposição espectral*. Este teorema relaciona a matriz de dados com seus autovalores e autovetores. Em outras palavras, a decomposição espectral é um meio de expansão de matrizes simétricas.

Considere $\mathbf{A}_{p \times p}$ uma matriz simétrica e $\mathbf{O}_{p \times p}$ uma matriz ortogonal, ou seja, $\mathbf{O}^T \mathbf{O} = \mathbf{O} \mathbf{O}^T = \mathbf{I}_{p \times p}$, tal que

$$\mathbf{O}^T \mathbf{A} \mathbf{O} = \Lambda = \begin{bmatrix} \lambda_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \lambda_p \end{bmatrix}. \quad (\text{C.10})$$

Em que $\lambda_1 \geq \dots \geq \lambda_p$ são os autovalores de \mathbf{A} . E a matriz diagonal Λ é dita ser similar à matriz simétrica $\mathbf{A}_{p \times p}$; o que implica em

1. $\det(\mathbf{A}) = \det(\Lambda) = \prod_{i=1}^p \lambda_i$
2. $\text{tr}(\mathbf{A}) = \text{tr}(\Lambda) = \sum_{i=1}^p \lambda_i$

A matriz $\mathbf{O}_{p \times p}$ é constituída por p autovetores ortonormais \mathbf{e}'_i s correspondentes aos autovalores λ'_i s da matriz $\mathbf{A}_{p \times p}$. Ou seja, vetores ortogonais desde que o produto interno $\langle \mathbf{e}_i^T, \mathbf{e}_j \rangle = 0$ ou $\mathbf{e}_i^T \mathbf{e}_j = 0$, para $i \neq j$; e normalizados, pois $\mathbf{e}_i^T \mathbf{e}_i = 1$, para $\forall i$; sendo $i, j = 1, \dots, p$.

A forma matricial de $\mathbf{O}_{p \times p}$ é representada como

$$\mathbf{O}_{p \times p} = \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1p} \\ e_{21} & e_{22} & \cdots & e_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ e_{p1} & e_{p2} & \cdots & e_{pp} \end{bmatrix}. \quad (\text{C.11})$$

A título de ilustração, considere que \mathbf{v}_i é um dos p autovetores da matriz \mathbf{A} . O particular autovetor normalizado é obtido por

$$\mathbf{e}_i = \frac{\mathbf{v}_i}{\|\mathbf{v}_i\|},$$

onde $\|\mathbf{v}_i\|$ denota a norma ou comprimento do autovetor \mathbf{v}_i e é calculada como

$$\|\mathbf{v}_i\| = \sqrt{\mathbf{v}_{i1}^2 + \mathbf{v}_{i2}^2 + \cdots + \mathbf{v}_{ip}^2} = \sqrt{\langle \mathbf{v}_i, \mathbf{v}_i \rangle},$$

em que $\mathbf{v}_{i1}^2, \mathbf{v}_{i2}^2, \dots, \mathbf{v}_{ip}^2$ representam os componentes do autovetor \mathbf{v}_i .

Após o processo de normalização o autovetor \mathbf{v}_i transforma-se em

$$\mathbf{e}_i = \begin{bmatrix} e_{i1} \\ e_{i2} \\ \vdots \\ e_{ip} \end{bmatrix}.$$

Como uma matriz real e simétrica, por exemplo \mathbf{A} , pode ser diagonalizada por uma matriz ortogonal \mathbf{O} , então $\mathbf{O}^T \mathbf{O} = \mathbf{I}$, e portanto, $\mathbf{O}^T = \mathbf{O}^{-1}$. E assim, a expressão (C.10) é equivalente a

$$\mathbf{O}^{-1} \mathbf{A} \mathbf{O} = \Lambda. \quad (\text{C.12})$$

então $\mathbf{A} = \mathbf{O} \Lambda \mathbf{O}^T$. O que equivale a expressão geral da decomposição espectral da matriz \mathbf{A} , a saber:

$$\mathbf{A} = \lambda_1 \mathbf{e}_1 \mathbf{e}_1^T + \lambda_2 \mathbf{e}_2 \mathbf{e}_2^T + \cdots + \lambda_p \mathbf{e}_p \mathbf{e}_p^T = \sum_{i=1}^p \lambda_i \mathbf{e}_i \mathbf{e}_i^T. \quad (\text{C.13})$$

Detalhes sobre a álgebra de matrizes aplicadas à análise multivariada podem ser encontrados em Johnson e Wichern (2002) ou em Mingoti (2005), assim como nos três primeiros capítulos do livro do Elsner e Tsonis (1997).

C.2.3 *Análise de componentes principais*

A técnica intitulada por Análise de Componentes Principais (ACP) foi introduzida no início do século 20 por Karl Pearson e posteriormente fundamentada no artigo de Hotelling (1933).

O principal propósito da ACP é explicar a estrutura de covariância de um vetor composto por p variáveis aleatórias, através de combinações lineares das variáveis originais. Tais combinações lineares são denominadas como componentes principais (CP's) e não são correlacionadas entre si. Geometricamente, as CP's representam um novo sistema de coordenadas obtidas por rotação do sistema original.

Caso a distribuição de probabilidade da matriz de dados seja normal p -variada, os componentes principais também são independentes e têm distribuição normal. Contudo, a suposição de normalidade não é necessária para que a técnica possa ser utilizada (JOHNSON e WICHERN, 2002; MINGOTI, 2005).

Na ACP cada variável em análise gera uma componente principal e, em geral, busca-se obter redução no número de variáveis e interpretação das componentes principais. Ou seja, procura-se sumarizar a informação das variáveis originais de uma forma simples. Sendo assim, na ACP considera-se um número k de componentes principais ($k < p$) que

será uma aproximação do sistema completo com p dimensões. A qualidade da aproximação dependerá da quantidade de componentes preservadas no sistema, assim, as k componentes principais poderão substituir as p variáveis.

O processo de obtenção das componentes principais envolve a decomposição da matriz de covariâncias ou da matriz de correlações. A extração das componentes através da matriz de correlações é indicada nos casos em que as variáveis possuem variabilidades discrepantes ou escalas distintas.

C.2.3.1 Componentes principais exatas

Seja o vetor aleatório $\mathbf{X}_{n \times p}$ apresentado em (C.5), a matriz de covariâncias $\Sigma_{p \times p}$ em (C.7), os autovalores da matriz ($\Sigma_{p \times p}$) e os respectivos autovetores normalizados $\mathbf{e}_i = [e_{i1}, e_{i2}, \dots, e_{ip}]^T$. Considere o vetor aleatório $\mathbf{Y} = \mathbf{O}^T \mathbf{X}$; em que \mathbf{Y} é um vetor constituído por p combinações lineares das variáveis originais do vetor \mathbf{X} ; e $\mathbf{O}_{p \times p}$ é uma matriz ortonormal formada pelos vetores \mathbf{e}_i s (C.11).

Cada uma das p combinações lineares ou componentes principais também pode ser definida usando a forma extensa, a saber :

$$Y_i = e_{i1}X_1 + e_{i2}X_2 + \dots + e_{ip}X_p = \mathbf{e}_i \mathbf{X}; i = 1, \dots, p. \quad (\text{C.14})$$

Sob a suposição de que , $e_i^T e_i = 1$ e que em cada combinação linear ($\mathbf{e}_i \mathbf{X}$) tem-se variância máxima.

A variância de uma particular componente Y_i é definida como

$$\text{Var}[Y_i] = e_i^T \Sigma_{p \times p} e_i = \lambda_i; i = 1, \dots, p.$$

E a covariância entre qualquer duas componentes é nula, isto é,

$$\text{cov}(Y_i, Y_u) = e_i^T \Sigma_{p \times p} e_u = 0,$$

para $i \neq u$.

Como as variáveis aleatórias do vetor \mathbf{Y} não são correlacionadas entre si, utiliza-se as

combinações lineares no lugar do vetor \mathbf{X} . Pois, assim como descrito nas expressões da equação (C.15), os vetores \mathbf{Y} e \mathbf{X} possuem a mesma variância total.

Na ACP, cada um dos autovalores λ_i , oriundos da matriz de covariâncias ou da matriz de correlações, representa a variabilidade de uma componente principal Y_i . Assim, as componentes principais são ordenadas de acordo com o valor dos correspondentes autovalores. Ou seja, a primeira componente tem maior variabilidade e possui um maior nível de importância, em comparação aos demais componentes. Alguns resultados importantes da ACP são apresentados a seguir.

Segundo o teorema da decomposição espectral, subseção C.2.2, a variância total de \mathbf{X} pode ser expressa a partir da variância total de \mathbf{Y} , pois

$$\sum_{i=1}^p \text{Var}[X_i] = \sum_{i=1}^p \sigma_{ii} = \text{tr}(\Sigma_{p \times p}) = \text{tr}(\Lambda_{p \times p}) = \sum_{i=1}^p \lambda_i = \sum_{i=1}^p \text{Var}[Y_i] \quad (\text{C.15})$$

onde, σ_{ii} denota a variância de uma particular variável original, isto é, $\sigma_{ii} = \text{Var}[X_i]$, $i = 1, \dots, p$. E conseqüentemente, a proporção da variância total de \mathbf{X} explicada pela i -ésima componente principal é

$$\frac{\lambda_i}{\sum_{i=1}^p \lambda_i}; i = 1, \dots, p. \quad (\text{C.16})$$

Se a proporção da variância total devido aos primeiros componentes, para p grande, for alta (por exemplo, 90 %); então as p -variáveis originais podem ser substituídas por essas primeiros componentes com pouca perda de informação sobre a estrutura de variâncias e covariâncias de \mathbf{X} .

Ao restringir o foco da análise às k primeiras componentes principais, à luz do teorema da decomposição espectral, tem-se a seguinte aproximação da matriz de covariâncias:

$$\Sigma_{p \times p} \approx \sum_{i=1}^k \lambda_i \mathbf{e}_i \mathbf{e}_i^T \quad (\text{C.17})$$

De acordo com (C.17), a variabilidade de \mathbf{X} é aproximada pela soma de k matrizes. Caso k seja igual a p , a matriz de covariâncias é reproduzida exatamente por

$$\Sigma_{p \times p} = \sum_{i=1}^p \lambda_i \mathbf{e}_i \mathbf{e}_i^T.$$

Todos os resultados apresentados anteriormente refere-se a ACP populacional, contudo para informações amostrais esses resultados continuam válidos, sendo que, neste caso, é considerada a estimação das componentes principais. Detalhes podem ser consultados em Johnson e Wichern (2002) ou em Mingoti (2005).

C.2.3.2 Determinação do número de componentes principais

Na literatura especializada, existem sugestões de critérios para determinar o número de componentes principais que devem permanecer no sistema. Contudo, como não há uma unificação, geralmente avalia-se mais de um critério. Para esse fim, examina-se a quantidade da variância amostral explicada, as variâncias dos componentes principais amostrais e o uso prático das componentes selecionadas (interpretação das componentes). Além disso, também analisa-se os valores dos autovalores, já que componentes oriundos de autovalores nulos são considerado sem importância e podem ser eliminados do sistema.

Os critérios mais utilizados nas aplicações envolvendo a análise de componentes principais são apresentados a seguir:

- Quantidade da variância amostral explicada

Neste critério a proposta é buscar a quantidade de componentes k , de tal forma que

$$\gamma = \frac{\sum_{i=1}^k \hat{\lambda}_i}{\sum_{i=1}^p \hat{\lambda}_i} \quad (\text{C.18})$$

em que $\hat{\lambda}_i$ é o autovalor obtido a partir da matriz de covariâncias ou matriz de correlações amostrais. Sendo $0 < \gamma < 1$, contudo não existe um patamar ou ponto de corte definido e a sua escolha é realizada de maneira subjetiva, de forma que, as k primeiras componentes representem conjuntamente grande parte da variância total de \mathbf{X} .

- Critério de Kaiser

Na escolha das componentes principais mais importantes indica-se manter no sistema as componentes cujos autovalores são maiores ou iguais à média de todos os autovalores considerados; que pode ser definida por

$$\hat{\lambda}_m = \frac{\sum_{i=1}^p \hat{\lambda}_i}{p}; i = 1, \dots, p. \quad (\text{C.19})$$

Medida também conhecida como *variância média* das variáveis originais.

Nos casos em que as componentes principais são extraídas de uma matriz de correlações, a variância total estimada $\left(\sum_{i=1}^p \hat{\lambda}_i\right)$ iguala-se ao número de variáveis originais (p). Desta forma, o critério seleciona as componentes relacionadas aos autovalores maiores ou iguais a 1 ($\hat{\lambda}_i \geq 1$).

- *Scree plot*

O *scree plot* é um conhecido gráfico da ACP que esboça os autovalores em ordem decrescente de magnitude. Este critério é um auxílio visual na escolha de um número adequado para k . Para determinar a quantidade apropriada de CP's , procura-se um desnível ou 'dobra' no *scree plot*; o que corresponde ao ponto em que os autovalores restantes são relativamente pequenos e próximos entre si (região de estabilidade). Em geral, observa-se a estabilidade dos autovalores no ponto a partir do qual estes se aproximam de zero.

Com o propósito de exemplificação, considerou-se o *scree plot* de uma situação real envolvendo 18 indicadores da qualidade de gestão dos governos de 59 países industrializados ou em desenvolvimento. As variáveis examinadas e muitas outras podem ser obtidas na *homepage* da *Central Intelligence Agency* (CIA).

A partir de uma ACP via matriz de correlação obteve-se o *scree plot* apresentado na Figura C.1.

Através da Figura C.1 , pode-se observar que a partir do décimo quarto componente os autovalores tendem a se estabilizar em torno do zero, formando componentes principais com baixa contribuição para explicar a variabilidade total da matriz de dados.

A linha tracejada na Figura C.1 corresponde ao autovalor 1, e assim, utilizando o critério de Kaiser, o *scree plot* sugere a retenção dos quatro primeiros componentes.

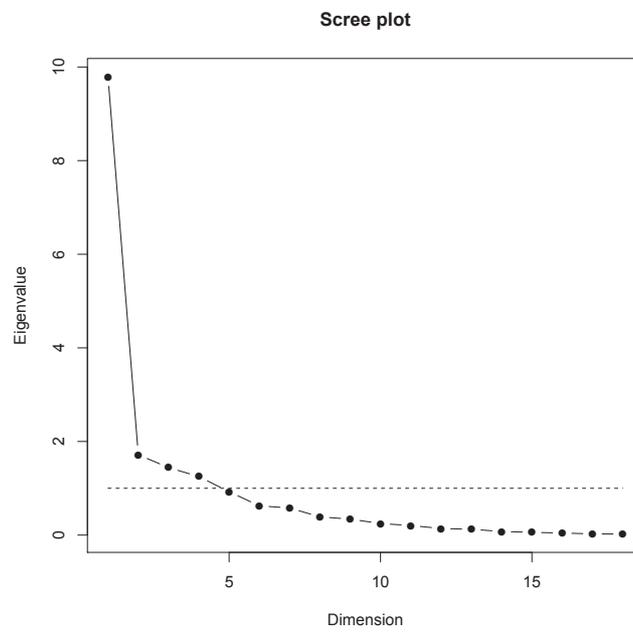


Figura C.1: *Scree plot* ($\hat{\lambda}_i$ versus a i -ésima componente principal).

Modelagem de séries temporais via *Singular Spectrum Analysis*- Uma otimização do *window length*

Renata de M. Esquivel¹; Valter de Senna²; Gecynalda S. da S. Gomes³.

D.1 Introdução

O desenvolvimento de métodos estatísticos para análise de dados obtidos em situações em que as observações são dependentes tem apresentado crescimento relevante nas últimas décadas e, em particular, na análise de dados provenientes de séries temporais.

Os métodos estatísticos para análise de séries temporais encontram-se amplamente desenvolvidos na literatura e vários modelos clássicos estão implementados em programas genéricos (SPPlus, MINITAB, SPSS e R) ou específicos (SCA e STAMP) (MORETTIN e TOLOI, 2006). Contudo, cada modelo clássico possui suposições, como por exemplo, a suavização exponencial de Holt que pode ser aplicada quando a série apresenta tendência, e se além dessa componente, a série possuir sazonalidade esta poderá ser modelada através da suavização de Holt-Winters (MORETTIN e TOLOI, 2006). O uso adequado dos modelos exigirá verificações das suposições, que podem demandar muitos esforços na etapa de identificação do padrão de comportamento da série temporal. Como alternativa, pode-se utilizar a Análise Espectral Singular (SSA, do inglês *Singular Spectrum Analysis*).

A SSA é mencionada como uma técnica moderna e poderosa para análise de séries temporais, que incorpora as fundamentações teóricas da análise clássica de série temporal, da análise multivariada, da dinâmica de sistemas e do processamento de sinal (ELSNER e TSONIS, 1996; GOLYANDINA, NEKRUTKIN e ZHIGLJAVSKY, 2001). Segundo HASSANI (2007), a SSA realiza uma decomposição da série temporal em poucos componentes independentes e interpretáveis. Como por exemplo, a identificação de componentes de tendência, componentes sazonais e estruturas de ruído. A técnica SSA não exige o conhecimento sobre o modelo paramétrico da série temporal, sendo considerado também como um método de estimação espectral não paramétrico (HASSANI, 2007; GOLYANDINA, 2001). Qualquer série aparentemente complexa com alguma potencial estrutura pode ser analisada através da SSA (GOLYANDINA, NEKRUTKIN e ZHIGLJAVSKY,

¹Mestranda do PPGMCTI (SENAI CIMATEC)

²Professor do SENAI CIMATEC

³Professora do Departamento de Estatística da UFBA

2001).

Vale salientar que, a SSA não se detém a identificar e extrair padrões geradores da série, mas a técnica também tem a finalidade de realizar previsões de valores futuros (HASSANI, 2007; GOLYANDINA, NEKRUTKIN e ZHIGLJAVSKY, 2001). O trabalho objetiva apresentar a potencialidade da SSA na identificação de componentes de uma série, utilizando uma otimização do parâmetro *window length*. Com esta finalidade, analisou-se a série do IBOVESPA ao longo de 1314 dias e os dados de companhias aéreas no decorrer de 144 meses (clássica série *AirPassengers*).

D.2 Materiais e Métodos

D.2.1 Breve descrição da SSA básica

A técnica SSA se fundamenta em dois estágios complementares: decomposição e reconstrução da série temporal. Cada estágio é composto por dois passos, constituindo os quatro passos da técnica, a saber: *Embedding*, *SVD*, *Grouping* e *Diagonal averaging* (HASSANI, 2007; GOLYANDINA, NEKRUTKIN e ZHIGLJAVSKY, 2001).

Considere a série temporal unidimensional com valores reais diferentes de zero $Y_t = Y_1, \dots, Y_N$; $t = 1, 2, \dots, N$, em que N é o comprimento da série ou a quantidade de observações ao longo do intervalo de tempo.

D.2.1.1 Decomposição

Na presente etapa, a série temporal original é decomposta em uma soma de poucas subséries, de modo que cada subsérie possa ser identificada como um componente interpretável (GOLYANDINA, NEKRUTKIN e ZHIGLJAVSKY, 2001).

- Primeiro passo: *Embedding*

Inicialmente, a série original unidimensional é representada como uma série multidimensional com dimensão L denominada *window length*. O *window length* é o único parâmetro deste passo e representa a quantidade de partes ou componentes em que a série é decomposta. Tal parâmetro deve ser um valor inteiro entre $2 \leq L \leq N - 1$, e segundo resultados teóricos o tamanho de L deve ser suficientemente grande, mas não superior a $\frac{N}{2}$ (HASSANI, 2007).

A série temporal multidimensional, que é uma sequência de vetores constituídos por elementos da série Y_t , forma a matriz apresentada na expressão D.1, denominada como *trajectory matrix* ou matriz trajetória e é o resultado do passo *embedding* (HASSANI, 2007; HASSANI, HERAVIC e ZHIGLJAVSKYA, 2009; GOLYANDINA, NEKRUTKIN e ZHIGLJAVSKY, 2001).

$$\mathbf{X}_{L \times K} = \begin{bmatrix} Y_1 & Y_2 & \cdots & Y_K \\ Y_2 & Y_3 & \cdots & Y_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ Y_L & Y_{L+1} & \cdots & Y_N \end{bmatrix}; \quad (\text{D.1})$$

em que, $K = N - L + 1$ é o número de vetores deslocados no tempo (*lagged vectors*).

- Segundo passo: *SVD*

No passo da decomposição do valor singular (*SVD*, em inglês) é realizada a decomposição da matriz trajetória \mathbf{X} em uma soma de matrizes elementares.

Seja $\mathbf{S} = \mathbf{X}\mathbf{X}^T$. Ao realizar a SVD da matriz \mathbf{S} , obtém-se os autovalores de \mathbf{S} ($\lambda_1 \geq \cdots \geq \lambda_L \geq 0$) e os seus correspondentes autovetores U_1, \dots, U_L , os quais formam um sistema ortonormal.

Se denotar os componentes principais da matriz trajetória como $V_i = \frac{\mathbf{X}^T U_i}{\sqrt{\lambda_i}}$, ($i = 1, \dots, d$), a SVD da matriz trajetória pode ser escrita da seguinte forma:

$$\mathbf{X} = E_1 + \cdots + E_d = \sum_{i=1}^d E_i, \quad (\text{D.2})$$

em que d denota o número de autovalores diferentes de zero da matriz \mathbf{S} ($d \leq L$); e $E_i = \sqrt{\lambda_i} U_i V_i^T$ (matrizes elementares).

O conjunto $(\sqrt{\lambda_i}, U_i, V_i)$ é conhecido como o i -ésimo *eigen triple* da matriz trajetória \mathbf{X} (HASSANI, 2007; HASSANI, HERAVIC e ZHIGLJAVSKYA, 2009; GOLYANDINA, NEKRUTKIN e ZHIGLJAVSKY, 2001).

D.2.1.2 Reconstrução

- Terceiro passo: *Grouping*

No terceiro passo ocorre a divisão das matrizes elementares E_i em vários grupos e a soma das matrizes dentro de cada grupo. O procedimento *grouping* particiona o conjunto de índices da expressão D.2 ($1, \dots, d$) em subconjuntos disjuntos I_1, \dots, I_m , podendo ser representado através da expressão

$$\mathbf{X} = E_{I_1} + \dots + E_{I_m} = \sum_{p=1}^m E_{I_p}; \quad (\text{D.3})$$

Em suma, o resultado deste passo é a representação da matriz trajetória como uma soma de várias matrizes nomeadas como *resultant matrices* ou matrizes resultantes ($E_{I_1} + \dots + E_{I_m}$).

- Quarto passo: *Diagonal averaging*

A operação realizada no último passo transforma cada *resultant matrices* em um componente da série original, ou seja, converte cada matriz da decomposição agrupada (D.3) em uma nova série de tamanho N . O que possibilitará a obtenção de uma série unidimensional que pode ser considerado como uma aproximação da série original.

D.2.2 Otimização do *window length*

Realizou-se a otimização do parâmetro *window length* (L) considerando o *default* da função `optimize` do software estatístico R. Esta função procura otimizar o intervalo de valores, iniciado no seu limite inferior, segundo a condição especificada como argumento da função interna `optimize`. O método *default* utiliza uma combinação de buscas, que fundamenta-se na proporção áurea e na interpolação parabólica, sendo projetado para funções contínuas. Detalhes sobre a função `optimize` podem ser consultados através do *Help* do R - versão 2.13.

O parâmetro L foi otimizado de forma que minimizasse o erro quadrático médio $\left[EQM = \frac{\sum(Y_t - \hat{Y}_t)^2}{N}\right]$ cometido na reconstrução da série através da SSA, fundamentando-se na condição $2 \leq L \leq \frac{N}{2}$.

Em que \hat{Y}_t representa a série reconstruída com todos os L componentes.

D.2.3 Aplicações: Características das séries temporais

Uma das séries temporais utilizada é representada pelos índices diários do IBOVESPA (Índice da Bolsa de Valores de São Paulo) entre 01 de janeiro de 2006 e 30 de abril de 2011, que totaliza 1314 observações ao longo do período especificado. Esses dados foram coletados na webpágina do "Yahoo!Finanças". A série histórica do IBOVESPA considerada apresenta uma estrutura com tendência não-linear e sazonalidade complexa. Para a série do IBOVESPA o L ótimo foi de 657.

A outra aplicação foi realizada na clássica série de companhias aéreas apresentada por Box e Jenkins. Tal série se refere aos números mensais, em milhares, de passageiros das companhias aéreas internacionais do período de 1949 a 1960, correspondendo a 144 observações no tempo. Esta série temporal pode ser encontrada no *datasets* do software R nomeada por `AirPassengers`. Tal série apresenta uma visível tendência linear e sazonalidade multiplicativa, uma vez que a sua amplitude cresce no decorrer do tempo. Na segunda aplicação o *window length* otimizado resultou em 72.

D.3 Resultados

Para a série do IBOVESPA decomposta em 657 componentes foi possível verificar que o primeiro componente indica a captura da tendência da série e o padrão de sazonalidade da série original começa a ser identificado a partir do segundo componente. A estrutura geral da série do IBOVESPA pode ser representada através dos dois primeiros componentes que têm os maiores valores singulares. Comparando a série original com as reconstruídas nota-se o seguinte erro quadrático médio: $3,128 \times 10^{-22}$.

Após realizar a decomposição da série `AirPassengers` em 72 componentes observou-se que o maior autovalor se destaca dos demais representando claramente a tendência principal da série pelo primeiro componente. A partir do terceiro componente da decomposição os autovalores tendem a se estabilizarem em torno de zero, e assim os 71 componentes restantes são gerados por autovalores "insignificantes" os quais pouco contribuem para a explicação da variabilidade existente na série. Assim, o sinal da série pode ser formado pelos dois primeiros componentes e os demais componentes seriam considerados como ruído da série. Nota-se que o estágio de reconstrução da série com os 72 *eigen triples* produziu uma série muito semelhante a original (erro quadrático médio igual a $2,121 \times 10^{-26}$).

Ao analisar a decomposição das séries observou-se que, apenas o primeiro componente explica mais de 96% da variação de cada série (96,58% para IBOVESPA e 97,98% para AirPassengers).

De forma geral, o sinal da série é formado pelos dois primeiros componentes e os demais componentes são identificados como ruído.

Como a decomposição de ambas as séries foi realizada com muitos componentes, ($L_{timo} = \frac{N}{2}$), observou-se que a SSA discrimina satisfatoriamente bem o sinal e o ruído ao considerar 30 e 10 componentes na reconstrução das séries IBOVESPA e AirPassengers, respectivamente.

D.4 Considerações finais

Como a técnica SSA não pressupõe nenhum modelo paramétrico da série temporal, foi possível analisar as séries de dados com as suas estruturas originais. Desta forma, é notório observar a diminuição dos esforços na etapa de identificação do padrão de comportamento das séries, uma grande vantagem da SSA.

Como esperado, a SSA apresentou um bom desempenho ao modelar séries não estacionárias com tendências e periodicidades não triviais. Ou seja, a modelagem das séries analisadas discriminou satisfatoriamente os principais componentes constituintes, demonstrando que com poucos componentes é possível recompor a estrutura principal das séries analisadas. Os resultados apontam para evidências de que o aumento do número de observações leva a uma ligeira piora na aproximação da série tornando-a menos semelhante à série original.

O uso do *window length* ótimo demonstrou excelentes precisões na etapa de reconstrução das séries temporais. Assim como disposto na literatura, a escolha do *window length* (L) dependerá do objetivo da análise. De acordo com o presente estudo, sugere-se usar o valor ótimo para o parâmetro L se for exigido uma alta precisão na reconstrução da série temporal.

Referências

ELSNER, J.B.; TSONIS, A.A. *Singular spectrum analysis: a new tool in time series analysis*. New York: Plenum Press, 1996.

GOLYANDINA, N.; NEKRUTKIN, V.; ZHIGLJAVSKY, A. *Analysis of time series structure: SSA and related techniques*. New York: Chapman Hall/CRC, 2001.

HASSANI, H. Singular Spectrum Analysis: Methodology and Comparison. *Journal of Data Science*, 5, 239-257, 2007.

HASSANI, H.; HERAVIC, S.; ZHIGLJAVSKYA, A. Forecasting European industrial production with SSA. *International Journal of Forecasting*, 25, 103-118, 2009.

MORETTIN, P.A. e TOLOI, C.M.C. *Análise de séries temporais*. São Paulo: Edgard Blucher, 2006.

Referências Bibliográficas

- AKAIKE, H. Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika*, 60, 255-265, 1973.
- _____. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, AC-19, 716-723, 1974.
- _____. On entropy maximization principle. In: *Applications of Statistics* (P.R. Krishnaiah, Ed.), 27-41. Amsterdam: North-Holland, 1977.
- ALLEN, M.R. and SMITH, L.A. Monte Carlo SSA: Detecting irregular oscillations in the presence of coloured noise. *Journal of Climate*, 9, 3373-3404, 1996.
- ATLAS DO POTENCIAL EÓLICO DO ESTADO DA BAHIA. Projeto de Pesquisa e Desenvolvimento COELBA-ANEEL. Escala 1: 8000000.
- BAKITA, A.L. *Avaliação da técnica de autocorrelação espectral na estimação do espaço médio intertrabecular em Sinais Ultra-Sônicos de osso calcâneo*. Dissertação (Mestrado)-Programa de Pós-graduação em Engenharia Biomédica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 51p., 2005.
- BEVERIDGE, W.H. Wheat prices and rainfall in western. *Europe. J. Royal Statistical Soc*, 85, 412-459, 1922.
- BOLDRINI, J.L., COSTA, S.I.R., FIGUEIREDO, V.L. e WETZLER, H.G. Álgebra Linear. 3.ed. São Paulo: Harper & Row do Brasil, 1980.
- BOX, G.E.P and COX, D.R. An analysis of transformation. *Journal of the Royal Statistical Society* , Series B, 26, 211-243, 1964.
- BOX, G.E.P. and JENKINS, G.M. *Time Series Analysis: Forecasting and Control*. São Francisco: Holden-Day, 1970.
- BOX, G.E.P.; JENKINS, G.M. and REINSEL, G.C. *Time Series Analysis: Forecasting and Control*. 3rd ed. Englewood Cliffs: Prentice Hall, 1994.
- BOX, G.E.P. and PIERCE, D.A. Distribution of residual correlations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 65, 1509-1526, 1970.
- BRILLINGER, D.R. *Time Series: Data analysis and Theory*. Classics in Applied Mathematics. SIAM, 2001.
- BROCKWELL, P.J. and DAVIS, R.A. *Time Series: Theory and Methods*. 2nd ed. New York: Springer Series in Statistics, 1991.

- . *Introduction to Time Series and Forecasting*. 2nd ed. New York: Springer Texts in Statistics, 2002.
- BROOMHEAD, D. S. and KING, G.P. Extracting qualitative dynamics from experimental data. *Physica D*, 20, 217–236, 1986a.
- . On the qualitative analysis of experimental dynamical systems. *Nonlinear Phenomena and Chaos*, Sarkar S (Ed.), Adam Hilger, Bristol, 113–144, 1986b.
- BROOMHEAD, D.S.; JONES, R. and KING, G.P. Addenda and correction. *J. Phys. A.*, 20, L563–L569. 1987.
- BROWN, R.G. *Statistical Forecasting for Inventory Control*. New York: McGraw-Hill, 1959.
- . *Smoothing, Forecasting and Prediction of Discrete Time Series*. Englewood Cliffs: Prentice Hall, 1963.
- CAMPOS, H. *Estatística Experimental Não-Paramétrica*. 4.ed. Piracicaba: Departamento de Matemática e Estatística/ESALQ, 1983.
- CHAVES, A.N. *Bootstrap em Séries Temporais*. Tese (Doutorado) - Programa de Engenharia Elétrica da Universidade Federal do Rio de Janeiro, Rio de Janeiro, 208p., 1991.
- CONOVER, W.J. *Practical Nonparametric Statistics*. New York: John Wiley & Sons, 1971.
- CLEVELAND, R.B.; CLEVELAND, W.S.; MCRAE, J.E. and TERPENNING, I. STL: A Seasonal-Trend Decomposition Procedure Based on Loess. *Journal of Official Statistics*, 6, 3-73, 1990.
- COOLEY, J.W. and TUKEY, J.W. An Algorithm for the Machine Calculation of Complex Fourier Series. *Mathematics of Computation*, 19, 297-301, 1965.
- DANILOV, D. Principal components in time series forecast. *Journal of Computational and Graphical Statistics*, 6, 112-121, 1997a.
- . *The ‘Caterpillar’ method for time series forecasting*. In: D. Danilov and A. Zhigljavsky (Eds.), *Principal Components of Time Series: the ‘Caterpillar’ method*, 73-104. University of St. Petersburg Press, St. Petersburg, Russian, 1997b.
- DANILOV, D. and ZHIGLJAVSKY, A. (Eds.). *Principal Components of Time Series: the ‘Caterpillar’ method*. University of St. Petersburg Press, St. Petersburg, Russian, 1997.
- DEJONG, D.N; NANKERVIS, J.C.; SAVIN, N.E. and WHITEMAN, C.H. *Integration versus trend stationarity in macroeconomic time series*. Working paper no. 89-99. Department of Economics, University of Iowa, Iowa City, IA, 1989.

DICKEY, D.A. and FULLER, W.A. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74, 427-431, 1979.

DIEBOLD, F.X. and RUDEBUSCH, G.D. On the power of Dickey-Fuller tests against fractional alternatives. *Economics Letters*, 35, 155-160, 1990.

EHLERS, R.S. *Análise de Séries Temporais*. Curitiba: Laboratório de Estatística e Geoinformação, Universidade Federal do Paraná, 2009.

ELSNER, J.B. and TSONIS, A.A. *Singular spectrum analysis: a new tool in time series analysis*. New York: Plenum Press, 1996.

ESQUIVEL, R.M., SENNA, V. e GOMES, G.S.S. *Modelagem de séries econômicas via Singular Spectrum Analysis (SSA)*. ESTE 2011: XIV Escola de Séries Temporais e Econometria, Fundação de Apoio da Universidade Federal do Rio Grande do Sul (FAURGS), Gramado-RS, 2011a.

_____. *Modelagem de séries temporais via Singular Spectrum Analysis- Uma otimização do window length*. SBPO 2011: XLIII Simpósio Brasileiro de Pesquisa Operacional, Hotel Recanto das Toninhas, Ubatuba-SP, 2011b.

FREITAS, D.S. *Operadores de Hankel*. Dissertação (Mestrado) - Programa de Pós-graduação em Matemática e Computação Científica, Universidade Federal de Santa Catarina, Florianópolis, 57p., 2006.

FRERY, A.C. e CRIBARI NETO, F. *Elementos de Estatística Computacional Usando Plataformas de Software Livre/Gratuito*. Rio de Janeiro: Instituto de Matemática Pura e Aplicada, 2005.

GARDNER, E.S. Exponential smoothing: The state of the art-Part II. *International Journal of Forecasting*, 22, 637– 666, 2006. Disponível em: <<http://www.forecasters.org/ijf/journal-issue/297/article/6110>> . Acesso em: 02 set. de 2011.

GHIL, M. and TARICCO, C. Advanced spectral analysis methods. In: *Past and present Variability of the Solar-Terrestrial system: Measurement, Data Analysis and Theoretical Model* (Edited by Castagnoli, G.C. and Provenzale, A.), 137-159, 1997.

GNEDENKO, B.V. *A Teoria da Probabilidade*. Tradução da série de textos clássicos da American Mathematical Society. Rio de Janeiro: Ciência Moderna, 2008.

GOLYANDINA, N.; NEKRUTKIN, V.; ZHIGLJAVSKY, A. *Analysis of time series structure: SSA and related techniques*. New York: Chapman & Hall/CRC, 2001.

_____. On the choice of parameters in Singular Spectrum Analysis and related subspace-based methods. *Statistics and Its Interface*, 3, 259-279, 2010.

- GOMES A.S. *Modelagem e previsão da arrecadação do imposto de renda no Brasil*. Dissertação (Mestrado) - Programa de Pós-graduação em Estatística da Universidade Federal de Pernambuco, Recife, 133p., 2003.
- GRANGER, C.W.J. and NEWBOLD, J. P. Forecasting transformed series. *Journal of the Royal Statistical Society, Series B*, 38, 189-203, 1976.
- GUJARATI, D.N. *Econometria Básica*. São Paulo: Pearson Makron Books, 2000.
- HANNAN, E.J. and QUINN, B.G. The determination of the order of an autoregression. *Journal of the Royal Statistical Society, Series B*, 41, 190-195, 1979.
- HASSANI, H. Singular Spectrum Analysis: Methodology and Comparison. *Journal of Data Science*, 5, 239-257, 2007.
- HASSANI, H.; HERAVIC, S.; ZHIGLJAVSKYA, A. Forecasting European industrial production with SSA. *International Journal of Forecasting*, 25, 103–118, 2009.
- HOLT, C.C. *Forecasting seasonals and trends by exponentially weighted moving averages*. Available from the Engineering Library (ONR Memorandum). Pittsburgh, PA: Carnegie Institute of Technology, University of Texas at Austin, 1957.
- _____. Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, 20, 5-10, 2004a.
- _____. Author's retrospective on Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, 20, 11-13, 2004b.
- HOLT, C.C.; MODIGLIANI, F.; MUTH, J.F. and SIMON, H.A. *Planning Production, Inventories and Work Force*. Englewood Cliffs: Prentice Hall, 1960.
- HOTELLING, H. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24, 498-520, 1933.
- HYNDMAN, R.J. *Time Series Data Library*. Disponível em: <<http://robjhyndman.com/TSDL>>. Acesso em: 21 abr. de 2012.
- JESUS, T.A. *Detecção de Determinismo e Modelagem Preditiva de Séries Temporais de Consumo de Energia Elétrica*. Dissertação (Mestrado) - Programa de Pós-graduação em Engenharia Elétrica, Universidade Federal de Minas Gerais, Belo Horizonte, 94p., 2008.
- JOHNSON, R.A. and WICHERN, D.W. *Applied Multivariate Statistical Analysis*. Englewood Cliffs: Prentice Hall, 2002.
- KABOUDAN , M. Genetic programming prediction of stock prices. *Computational Economics*, 16, 207-236, 2000.
- KALMAN, R.E. and BUCY, R. New Results in Linear Filtering and Prediction Theory, Transactions of ASME. *Journal of Basic Engineering*, 83D, 95-108, 1961.

- KLEIN, J.L. *Statistical Visions in Time: A History of Time Series Analysis, 1662-1938*. Cambridge: Cambridge University Press, 1997.
- KUNST, R.M. *Applied Time Series Analysis- Part I*. University of Vienna and Institute for Advanced Studies Vienna, 2009.
- KWIATKOWSKI, D.; PHILLIPS, P.C.B.; SCHMIDT, P. and SHIN, Y. Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root. *Journal of Econometrics*, 54, 159–178, 1992.
- LAURITZEN, S.L. Time series analysis in 1880: a discussion of contributions made by T. N. Thiele. *Internat. Statist. Review*, 49, 319-331, 1981.
- LIMA, A.B. *Contribuições à modelagem de teletráfego fractal*. Tese (Doutorado) – Escola Politécnica, Universidade de São Paulo, São Paulo, 215p., 2008.
- LJUNG, G.M. and BOX, G.E.P. On a measure of lack of fit in time series models. *Biometrika*, 65, 297–303, 1978.
- MACHADO, C.B. *Caracterização da periodicidade de fígado humano sadio e patológico usando Sinais de Ultra-Som*. Dissertação (Mestrado) - Programa de Pós-graduação em Engenharia Biomédica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 144p., 2005.
- MACHADO, C.B and PEREIRA W.C.A. Characterization of in vitro healthy and pathological human liver tissue periodicity using backscattered ultrasound signals. *Ultrasound in Medicine & Biology*, 32, 5, 649–657, 2006.
- MACIEL, C.D. *Análise de Espectro Singular Aplicada à Sinais Ultra-Sônicos*. Tese (Doutorado) - Programa de Pós-graduação em Engenharia Biomédica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 150p., 2000.
- MAKRIDAKIS, S and HIBON, M. The M3-Competition: results, conclusion and implications. *International Journal of Forecasting*, 16, 451–476, 2000.
- MANN, H.B. Nonparametric tests against trend. *Econometrica*, 13, 245-259, 1945.
- MINGOTI, S.A. *Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada*. Belo Horizonte: Editora UFMG, 2005.
- MORETTIN, P.A. e TOLOI, C. M. C. *Análise de séries temporais*. 2.ed. São Paulo: Edgard Blücher/ ABE- Projeto Fisher, 2006.
- MYUNG, N.K. *Singular Spectrum Analysis*. Thesis submitted for the degree Master of Science in Statistics, University of California, Los Angeles, 32p., 2009.
- NEWBOLD, P. and GRANGER, C.W.J. Experience with forecasting univariate time series and the combination of forecasts. *Journal of the Royal Statistical Society A*, 137, 131–146, 1974.

- PEREIRA W.C.A. and MACIEL C.D. Performance of ultrasound echo decomposition using Singular Spectrum Analysis. *Ultrasound in Medicine & Biology*, 27, 9, 1231–1238, 2001.
- PEREIRA W.C.A.; ABDELWAHAB, A.; BRIDAL, S.L. and LAUGIER, P. Singular spectrum analysis applied to 20 MHz backscattered ultrasound signals from periodic and quasi-periodic phantoms. *Acoustical Imaging*, 26, 239-246, 2002.
- PEREIRA, W.C.A.; BRIDAL, S.L.; CORON, A. and LAUGIER, P. Singular Spectrum Analysis Applied to Backscattered Ultrasound Signals from in Vitro Human Cancellous Bone Specimens, *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, 51, 3, 302 - 312, 2004.
- PLOSSER, C.I. Short-term forecasting and seasonal adjustment. *Journal of the American Statistical Association*, 74, 365, 15-24, 1979.
- POKORNY, M. *An Introduction to Econometrics*. Nova York: Basil Blackwell, 1987.
- QUENOUILLE, M. H. *The Analysis of Multiple Time Series*. London: Griffi, 1957.
- SAITO, A.T.K. *Avaliação da detecção automática do espaçamento médio de espalhadores utilizando Sinais Ultra-Sônicos*. Tese (Doutorado) - Programa de Pós-graduação em Engenharia Biomédica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 172p., 2003.
- SHUMWAY, R.H. and STOFFER, D.S. *Time Series Analysis and Its Applications With R Examples*. 2nd ed. New York: Springer Science, 2006.
- SIEGEL, S. *Estatística Não-Paramétrica para as Ciências do Comportamento*. São Paulo: McGraw-Hill do Brasil, 1975.
- TUFTE, E.R. *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press, 1983.
- TSAY, R.S. Time Series and Forecasting: Brief History and Future Research. *Journal of the American Statistical Association*, 95, 450, 638-643, 2000.
- WINTERS, P.R. Forecasting sales by exponentially weighted moving averages. *Management Science*, 6, 324-342, 1960.
- VAUTARD, R. and GHIL, M. Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series, *Physica D*, 35, 395-424, 1989.
- YIOU, P.; SORNETTE, D. and GHIL, M. Data-adaptive wavelets and multi-scale singular spectrum analysis. *Physica D*, 142, 254-290, 2000.
- YULE, G.U. On a Method of Investigating Periodicities in Disturbed Series With Special Reference to Wolfer's Sunspot Numbers. *Philosophical Transactions of the Royal Society London*, Series A, 226, 267-298, 1927.

ZIVOT, E. and WANG, J. *Modeling Financial Time Series With S-PLUS*. New York: Springer Science, 2006.

Análise Espectral Singular: Modelagem de séries temporais através de estudos comparativos usando diferentes estratégias de previsão.

Renata de Miranda Esquivel

Salvador, maio de 2012.