

UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação
ISSN 0103-2569

**DESCRIÇÃO DE MODELOS ESTATÍSTICOS E DE APRENDIZADO
DE MÁQUINA PARA PREDIÇÃO DE SÉRIES TEMPORAIS**

ANTONIO RAFAEL SABINO PARMEZAN
GUSTAVO ENRIQUE DE ALMEIDA PRADO ALVES BATISTA

Nº 412

RELATÓRIOS TÉCNICOS



São Carlos – SP
Ago/2016

Instituto de Ciências Matemáticas e de Computação

ISSN - 0103-2569

**Descrição de Modelos Estatísticos e de Aprendizado de Máquina para
Predição de Séries Temporais**

**Antonio Rafael Sabino Parmezan
Gustavo Enrique de Almeida Prado Alves Batista**

Nº 412

RELATÓRIOS TÉCNICOS DO ICMC

São Carlos
Agosto/2016

**Descrição de Modelos Estatísticos e de Aprendizado de Máquina para
Predição de Séries Temporais* †**

**Antonio Rafael Sabino Parmezan
Gustavo Enrique de Almeida Prado Alves Batista**

Universidade de São Paulo
Instituto de Ciências Matemáticas e de Computação
Departamento de Ciências de Computação
Laboratório de Inteligência Computacional
Caixa Postal 668, 13560-970 - São Carlos, SP, Brasil
e-mail: {antoniop, gbatista}@icmc.usp.br

Agosto, 2016

*Trabalho desenvolvido com o apoio da FAPESP (Processo 2013/10978-8).

†Disponível *on-line* em: [Relatórios Técnicos do ICMC](#).

Resumo

A escolha do algoritmo mais promissor para explicar e/ou prever um determinado fenômeno reside em uma das atividades mais críticas do processo de Mineração de Dados Temporais. Como diferentes modelos são capazes de descrever uma mesma sequência de dados, argumenta-se que nenhum método de modelagem pode ser considerado o melhor independentemente do problema em questão. Todavia, são poucos os trabalhos científicos que expõem, de maneira circunstanciada, o funcionamento dos algoritmos mais populares para predição de Séries Temporais e há uma carência do registro do desempenho desses métodos quando executados sobre sequências de dados complexas e altamente não-lineares. Neste trabalho, atendendo majoritariamente aos interesses da comunidade de Aprendizado de Máquina, são detalhadas as abordagens para predição de valores futuros com seus respectivos algoritmos usuais e algumas técnicas automáticas que oportunizam suas parametrizações. No intuito de evidenciar em que situações um método de predição supera o outro e quais aspectos dos dados tem maior influência no desempenho desses algoritmos, uma avaliação experimental foi realizada empregando 40 conjuntos de dados sintéticos, os quais foram identificados por meio de uma revisão sistemática da literatura.

Palavras-Chave: Estimação de Parâmetros, Construção de Modelos Preditivos, Processamento de Séries Temporais, Mineração de Dados

Este documento foi preparado com o formatador de textos \LaTeX . O sistema de citações de referências bibliográficas utiliza o padrão *Apalike* do sistema $\text{bib}\text{\TeX}$.

Lista de Figuras

1	Expressão de busca adotada no protocolo de revisão sistemática .	6
2	Distribuição do número de trabalhos por ano de publicação	7
3	Porcentagem do número de publicações que contemplaram uma determinada ferramenta computacional	9
4	Produção mensal de chocolate na Austrália	13
5	Preço anual de ingressos para o <i>Super Bowl</i>	15
6	Tendência da série de preços de ingressos para o <i>Super Bowl</i> . . .	17
7	Exemplo de obtenção da tendência usando <i>MA</i> com $r = 3$	17
8	Tendência da série de produção trimestral de cerveja nos EUA . .	18
9	Tendência da série de produção mensal de chocolate na Austrália	20
10	Técnica do <i>scatter plot</i> para a série de produção trimestral de cerveja nos EUA	22
11	Técnica do <i>scatter plot</i> para a série de produção mensal de chocolate na Austrália	23
12	Sazonalidade da série de produção trimestral de cerveja nos EUA	25
13	Sazonalidade e o efeito da desestacionalização da série de produção mensal de chocolate na Austrália	26
14	Resíduo da série de produção trimestral de cerveja nos EUA . . .	29
15	Resíduo da série de produção mensal de chocolate na Austrália .	29
16	Processo de Mineração de Dados Temporais	30
17	Processo de predição de valores em ST	32
18	Predições obtidas pelo modelo de <i>MA</i> com parâmetro $r = 1, 3$ e 5 .	35
19	Predições computadas pelo modelo de <i>SES</i> com parâmetro $\alpha = 0,1, 0,3$ e $0,9$	38
20	Predições obtidas pelo modelo de <i>HES</i> com parâmetros $\alpha = 0,3$ e $\beta = 0,9$	39
21	Representação dos tipos de variação sazonal	40
22	Predições computadas pelos modelos aditivo e multiplicativo de <i>HW</i> , ambos com parâmetros $\alpha = 0,3, \beta = 0,5, \gamma = 0,7$ e $s = 4$	42
23	Diagrama de atividades para o fluxo de construção de um modelo <i>ARIMA</i> ou <i>SARIMA</i>	46
24	Exemplificação do processo de predição de ST segundo a abordagem global	48
25	Representação da estratégia de predição multi-etapa à frente para a abordagem global	49
26	Estrutura do <i>Perceptron</i>	50
27	Estrutura de uma rede <i>MLP</i> com camada oculta única	52

28	Hiperplano de separação ótima e seus hiperplanos de suporte. Os eixos ordenados x_1 e x_2 representam as dimensões das amostras no espaço 2D	54
29	Mapeamento de dados para um espaço de características de mais alta dimensão utilizando como artifício a função <i>kernel</i>	54
30	Exemplo da aplicação do método baseado em similaridade para predição de ST com parâmetros $k = 3$ e $l = 25$	57
31	Exemplo de variância à amplitude e deslocamento	58
32	Exemplo de variância à complexidade	58
33	Representação da estimativa de complexidade adotada pela <i>CID</i> .	60
34	Exemplo de casamentos triviais	60
35	Séries com sazonalidade aditiva derivadas da composição de Fourier	69
36	Séries com dependência sazonal	70
37	Série com sazonalidade multiplicativa	70
38	Série de alta frequência com sazonalidade multiplicativa	71
39	Padrões temporais provenientes da categoria GCA	72
40	Padrões temporais provenientes da categoria GCB	73
41	Séries com dependência sazonal e ruído	73
42	Mapa Logístico	74
43	Mapa de Hénon	75
44	Sistema de Mackey-Glass	76
45	Sistema de Lorenz	77
46	Sistema de Rössler	78
47	Sinais Caóticos	78
48	Série de valores ECGSYN	79
49	Configuração experimental	79
50	Diagramas de distância crítica para os valores dos índices <i>MSE</i> , <i>TU</i> e <i>POCID</i> provenientes dos métodos de predição sobre ST determinísticas	84
51	Diagramas de distância crítica para os valores dos índices <i>MSE</i> , <i>TU</i> e <i>POCID</i> provenientes dos métodos de predição sobre ST estocásticas	85
52	Diagramas de distância crítica para os valores dos índices <i>MSE</i> , <i>TU</i> e <i>POCID</i> provenientes dos métodos de predição sobre ST caóticas	86
53	Diagramas de distância crítica para os valores dos índices <i>MSE</i> , <i>TU</i> e <i>POCID</i> provenientes dos métodos de predição sobre ST sintéticas	87

Lista de Tabelas

1	Questões de pesquisa que nortearam a revisão sistemática	5
2	Exemplo de obtenção da tendência por <i>MA</i> centrada com $r = 4$. .	19
3	Etapas do método da relação entre <i>MA</i>	23
4	Exemplo de obtenção dos índices sazonais pelo método da relação entre <i>MA</i>	24
5	Exemplo de obtenção das medidas de síntese dos índices para cada período da variação sazonal utilizando o método da relação entre <i>MA</i>	25
6	Exemplo de obtenção do componente residual	28
7	Dados projetados usando o modelo de <i>MA</i> com parâmetro $r = 1, 3$ e 5	35
8	Dados projetados empregando o modelo de <i>SES</i> com parâmetro $\alpha = 0,1, 0,3$ e $0,9$	37
9	Dados projetados usando o modelo de <i>HES</i> com parâmetros $\alpha = 0,3$ e $\beta = 0,9$	39
10	Dados projetados empregando os modelos aditivo e multiplicativo de <i>HW</i> , ambos com parâmetros $\alpha = 0,3, \beta = 0,5, \gamma = 0,7$ e $s = 4$. .	42
11	Padrões temporais gerados a partir da composição de Fourier . .	69
12	Padrões temporais comumente observados em gráficos de controle	72
13	Sumário de características e de configurações dos conjuntos de dados sintéticos	80
14	Algoritmos utilizados e intervalos de variação numérica definidos para os seus parâmetros	81

Lista de Algoritmos

1	<i>kNN-TSPI</i>	62
2	<i>Holdout Validation</i>	64
3	<i>Cross-validation</i>	65
4	<i>Box-Jenkins Method</i>	67

Sumário

1	Introdução	1
1.1	Objetivos	3
1.2	Organização do Trabalho	4
2	Revisão Sistemática e Meta-análise da Literatura	4
3	Fundamentos de Séries Temporais	11
3.1	Componentes de uma Série Temporal	13
3.1.1	Tendência	14
3.1.2	Sazonalidade	20
3.1.3	Resíduo	26
3.2	Processamento e Análise de Séries Temporais	28
4	Construção de Modelos para Predição de Séries Temporais	31
4.1	Métodos Estatísticos	33
4.1.1	Médias Móveis	33
4.1.2	Suavização Exponencial Simples	35
4.1.3	Suavização Exponencial de Holt	37
4.1.4	Suavização Exponencial Sazonal de Holt-Winters	39
4.1.5	Modelos <i>ARIMA</i> e <i>SARIMA</i>	42
4.2	Métodos de Aprendizado de Máquina	48
4.2.1	<i>ANN</i>	50
4.2.2	<i>SVM</i>	53
4.2.3	<i>kNN-TSPI</i>	55
4.3	Técnicas para Estimação de Parâmetros	63
4.3.1	Validação <i>Holdout</i>	63
4.3.2	Validação Cruzada	64
4.3.3	Método Box-Jenkins	66
5	Avaliação Experimental	66
5.1	Descrição dos Conjuntos de Dados Sintéticos	68
5.1.1	Séries Determinísticas	68
5.1.2	Séries Estocásticas	71
5.1.3	Séries Caóticas	73
5.2	Configuração dos Experimentos	79
5.3	Resultados e Discussão	83
5.3.1	Análise Individual	83
5.3.2	Análise Geral	86
6	Considerações Finais	87

1 Introdução

A conversão de dados em informação e conhecimento úteis para o suporte à tomada de decisão só foi possível ser alavancada devido aos avanços tecnológicos na área da computação. Esses avanços graduais contribuíram para o desenvolvimento e a implantação de sistemas computacionais capazes de armazenar e gerenciar uma quantidade expansível de dados.

Atualmente, os dados podem assumir diferentes formatos, desde os mais usuais, como o numérico e o nominal, até os mais complexos, por exemplo áudio e vídeo. Contudo, a viabilização do armazenamento da informação temporal, que permite a organização cronológica dos dados coletados, compõe o formato de dados que mais tem atraído a atenção de pesquisadores e impulsionado a criação de grandes bases de dados para análises posteriores (Larose e Larose, 2014; Fu, 2011).

Embora ainda haja um esforço incipiente em determinados serviços no sentido de analisar dados adquiridos sequencialmente ao longo do tempo, grande parte das organizações já automatizaram essa tarefa e hoje utilizam o conhecimento embutido nesses dados para melhor compreender os fenômenos observados, bem como para embasar o planejamento de atividades e aprimorar processos decisórios (Montgomery et al., 2015; Chatfield, 2013; Cowpertwait e Metcalfe, 2009). Nesse contexto, a Mineração de Dados Temporais (MDT) consiste de um processo não trivial que tem como finalidade possibilitar, a partir de uma Série Temporal (ST), a extração de conhecimento que pode guiar decisões incumbidas à especialistas do domínio (Witten et al., 2011; Maimon e Rokach, 2010). Dentre as tarefas compreendidas pela MDT inclui-se a de predição cujas pesquisas são motivadas pelo desafio da redução da incerteza futura, sobretudo devido à volatilidade de alguns fenômenos.

Os métodos para predição de ST são baseados essencialmente na ideia de que dados históricos contemplam padrões intrínsecos, geralmente de difícil identificação e nem sempre interpretáveis, que se descobertos podem auxiliar na descrição futura do fenômeno investigado. Essa descrição constitui um dos principais objetivos do processamento de ST, pois visa responder em que circunstâncias os padrões encontrados irão se repetir e quais tipos de variações os mesmos poderão sofrer no decorrer do tempo (Chatfield, 2013).

A concepção de um modelo para predição de valores em ST incide na aplicação de algoritmos que realizam suposições acerca dos dados, a fim de capturar as variáveis envolvidas e modelar as relações dinâmicas existentes, resumindo-as em uma estrutura matemática robusta e potencialmente flexível. Tal estrutura, além de ajudar na compreensão do processo que originou os dados, pode ser usada para prever dados futuros. Essa predição é obtida a partir da extrapolação do modelo gerado para um momento futuro, de modo

que novos dados são projetados para o período subsequente à série de valores utilizada para o ajuste do modelo.

A aplicação de métodos estatísticos baseados em autorregressão e médias móveis têm sido considerados o estado-da-arte para a modelagem e a predição de ST por mais de meio século (Gooijer e Hyndman, 2006). Os algoritmos que implementam esses métodos assumem que os dados seguem alguma distribuição conhecida e, com base nessa informação, definem parâmetros de funções para ajustar um modelo aos dados. No entanto, o emprego desse tipo de abordagem, denominada de paramétrica, acaba se tornando um limitador por envolver alta complexidade matemática e exigir vasto conhecimento técnico para o estabelecimento dos parâmetros do modelo.

Em termos práticos, definir os valores dos parâmetros de um modelo estatístico reside em uma tarefa dispendiosa e normalmente constituída de quatro etapas (Box et al., 2015): (1) seleção da estrutura do modelo conforme as características dos dados; (2) identificação das ordens do modelo; (3) estimação dos coeficientes do modelo; e (4) diagnóstico do modelo ajustado. Todas essas etapas são guiadas por fundamentos da estatística descritiva e, na maioria dos casos, são realizadas de maneira semiautomática por meio do uso de funções baseadas em autocorrelação, cujos resultados podem ser interpretados via correlogramas, e da aplicação de técnicas para obtenção de argumentos de entrada a partir da minimização de critérios de informação, os quais penalizam o modelo pela quantidade de parâmetros suficientemente necessários para o seu ajustamento. Os procedimentos citados integram uma análise exaustiva que, além de demandar muita atenção e estar sujeita a subjetividade, requer profissionais especializados tanto no domínio de aplicação quanto na área computacional.

Como consequência da dificuldade em se estimar modelos estatísticos, diversos estudos vêm sendo empreendidos no intuito de criar uma modelagem não-paramétrica para a predição de ST (Parmezan e Batista, 2015; Claveria e Torra, 2014; Kandananond, 2012; Sapankevych e Sankar, 2009). Uma das principais vantagens da abordagem não-paramétrica é que esta não pressupõe sobre a natureza da distribuição dos dados, corroborando para que a descrição do fenômeno investigado possa ser expressa de modo intuitivo e simples, principalmente em relação ao número de parâmetros, o qual deve ser mínimo para o propósito almejado. Além disso, essa abordagem parte do preceito de que o modelo construído precisa, se possível, ser parametrizado de maneira que cada parâmetro possa ser interpretado facilmente e identificado conforme algum aspecto da realidade.

O Aprendizado de Máquina (AM) é uma subárea da Inteligência Artificial que apoia a MDT e oferece suporte à criação de modelos não-paramétricos

para a predição de ST. Sendo assim, uma das preocupações em AM está relacionada com a pesquisa e a busca constante pelo desenvolvimento de métodos que auxiliem no processamento de grandes bases de dados e, conseqüentemente, na construção de modelos que permitam a representação de novos conhecimentos adquiridos automaticamente, de modo mais eficiente e compreensível (Han et al., 2011).

Os métodos tradicionais de AM para a construção de modelos possuem restrições à dados que apresentam características temporais, haja vista que eles consideram os dados independentes e identicamente distribuídos. Esse fato tem desafiado profissionais de diferentes áreas do conhecimento e recebido investimentos consideráveis da comunidade científica e empresarial. Especificamente, os investimentos se concentram no subsídio de pesquisas destinadas à adaptação dos métodos convencionais de AM para a análise de dados na qual o tempo constitui um fator importante (Parmezan e Batista, 2014; Aikes Junior, 2012; Ferrero, 2009).

1.1 Objetivos

Neste trabalho, com base em uma meta-análise da literatura, são apresentados métodos para a tarefa de predição de Séries Temporais provenientes da Estatística e da Inteligência Artificial. Todos os procedimentos envolvidos na construção de um modelo preditivo, desde sua concepção até sua extrapolação para períodos subsequentes, são relatados sob uma perspectiva didática que pretende mostrar o quanto a subárea de Aprendizado de Máquina pode oferecer ao processo de predição de valores futuros. Particularmente, é descrito um método de Aprendizado de Máquina baseado em similaridade que utiliza-se de três técnicas para obtenção de invariância à deslocamento e amplitude, invariância à complexidade e tratamento de casamentos triviais. Proposto pelos autores deste trabalho, o referido método busca, a partir de uma subsequência de referência e com auxílio de uma medida de distância, as k subsequências mais similares dentro de uma determinada série e usa os valores seguintes dessas subsequências como entrada para uma função de predição, a qual realiza o cálculo do valor futuro.

O presente trabalho tem por objetivo caracterizar, em termos de vantagens e desvantagens de uso, os distintos métodos existentes para a construção de modelos preditivos. Complementarmente, considerando dados artificiais (sintéticos), uma avaliação experimental foi realizada com o propósito de evidenciar em que situações um algoritmo de predição supera o outro e quais aspectos dos dados tem maior influência no desempenho desses algoritmos.

1.2 Organização do Trabalho

O restante deste trabalho está organizado da seguinte maneira: na Seção 2 são exibidos os trabalhos relacionados e as estatísticas decorrentes de uma meta-análise da literatura, a qual foi conduzida pelos resultados de uma revisão sistemática. Na Seção 3 são relatadas as principais definições e notações acerca do tema de Séries Temporais. Na Seção 4 são descritas as abordagens para a predição de valores futuros com seus respectivos métodos usuais e algumas técnicas que auxiliam na estimação dos parâmetros adotados por esses métodos. Na Seção 5 é especificada a configuração dos experimentos e os resultados alcançados. Conclusões e trabalhos futuros são mostrados na Seção 6.

2 Revisão Sistemática e Meta-análise da Literatura

A área de processamento e análise de dados temporais tem sido influenciada, há mais de meio século, por métodos estatísticos baseados em autorregressão e médias móveis. Segundo [Gooijer e Hyndman \(2006\)](#), embora alguns estudos tenham afirmado, entre as décadas de setenta e oitenta, que os modelos paramétricos não poderiam ser prontamente adaptados para muitas aplicações reais, os mesmos resistiram ao longo dos anos. A resistência na preferência pelo uso desses métodos fez com que eles alcancem na literatura a condição de estado-da-arte para a modelagem e a predição de Séries Temporais (ST).

Nas últimas duas décadas, com a ascensão do processo Mineração de Dados (MD), houve um crescente aumento no interesse pela adaptação dos métodos usuais de Aprendizado de Máquina (AM), especialmente os de regressão, para dar suporte à análise de fenômenos com dependência temporal. Pela simplicidade e eventual compreensibilidade, os métodos não-paramétricos se estabeleceram como sérios candidatos aos modelos clássicos amplamente difundidos, de modo que concursos científicos vêm sendo realizados no intuito de incentivar tanto o aperfeiçoamento desses algoritmos quanto o desenvolvimento de novas soluções ([Ahmed et al., 2010](#)).

Os pesquisadores vinculados às comunidades de estatística e de AM têm dado atenção a diversos aspectos do processo de predição, tais como o auxílio na seleção do modelo mais promissor ([Lemke e Gabrys, 2010](#)), o estudo do efeito de dessazonalização na projeção de valores futuros ([Ben Taieb et al., 2012](#)) e a construção de modelos híbridos por meio da combinação de métodos estatísticos e de MD ([Andrawis et al., 2011](#)). Nesse sentido, a qualidade dos

modelos paramétricos e não-paramétricos tem sido explorada, principalmente, em competições anuais que visam avaliar o desempenho dos algoritmos de predição frente a uma quantidade considerável de ST (Ben Taieb et al., 2012; Crone et al., 2011).

No intuito de atualizar o conhecimento de fundo, identificar os recentes avanços da área e expor os benefícios que a subárea de AM pode oferecer ao processo de predição de ST, foi elaborada uma meta-análise da literatura relacionada. Essa meta-análise foi conduzida a partir de uma revisão sistemática guiada pelas questões sumarizadas na Tabela 1. Na última linha dessa tabela, na cor cinza, é indicada a questão central de pesquisa levantada neste trabalho.

Tabela 1: Questões de pesquisa que nortearam a revisão sistemática

ID	Questão de Pesquisa
1	Quais são os métodos mais utilizados para a tarefa de predição de ST?
2	Qual abordagem, paramétrica ou não-paramétrica, é frequentemente adotada em aplicações reais?
3	Quais medidas são usualmente empregadas para avaliar o desempenho dos algoritmos de predição?
4	Como são realizadas as configurações paramétricas dos algoritmos de predição?
5	Quais são os métodos de predição de ST usados como <i>baseline</i> em estudos empíricos?
6	Quais são os conjuntos de dados selecionados nestes estudos?
7	São construídos conjuntos de dados artificiais (sintéticos) nestes estudos?
8	Quais as desvantagens e/ou limitações dos métodos de predição de ST averiguados nestes estudos?
9	São ressaltadas vantagens na utilização de modelos de AM para predição de ST?
10	Foi realizada uma avaliação empírica robusta visando a comparação objetiva e subjetiva entre os métodos paramétricos e não-paramétricos?

As questões de pesquisa foram formuladas com base no *know-how* prévio dos autores e refinadas durante o estudo das publicações investigadas no decorrer da revisão sistemática, a qual foi executada em meados de novembro de 2014 e com o suporte dos motores de busca *ACM Digital Library*¹, *CiteSeerX*², *Google Scholar*³, *Scopus*⁴ e *Web of Science*⁵. Foram analisados somente trabalhos desenvolvidos nos últimos cinco anos, isto é, trabalhos publicados entre os anos de 2010 e 2014. Além disso, não foram admitidas publicações com mesmo título indexadas e detectadas simultaneamente por distintos motores de busca, nem trabalhos que são duplicados em relação aos resultados, com

¹<http://dl.acm.org>.

²<http://citeseerx.ist.psu.edu>.

³<http://scholar.google.com>.

⁴<http://www.scopus.com>.

⁵<http://webofknowledge.com>.

exceção da versão mais completa.

A expressão global de busca, ilustrada na Figura 1, foi construída por meio do uso de operadores booleanos sobre quatro listas de termos relacionados às questões de pesquisa. É importante ressaltar que adaptações na expressão original de consulta foram necessárias, visto que nem todos os motores de busca utilizados suportam expressões extensas, bem como apresentam limitações quanto à investigação separada de títulos, resumos e palavras-chave.

("time series" OR "time-series" OR "timeseries") AND ("prediction" OR "forecasting" OR "statistics" OR "statistical" OR "moving average" OR "exponential smoothing" OR "holt" OR "holt winters" OR "autoregressive model" OR "gaussian process" OR "ARIMA" OR "SARIMA" OR "artificial intelligence" OR "machine learning" OR "data mining" OR "knowledge discovery in databases" OR "ANN" OR "MLP" OR "kNN" OR "SVM") AND ("evaluation measure" OR "error measure" OR "MSE" OR "RMSE" OR "MAPE" OR "SMAPE" OR "Theil" OR "POCID" OR "correlation coefficient") AND ("data" OR "synthetic data" OR "artificial data" OR "real data" OR "dataset" OR "synthetic dataset" OR "artificial dataset" OR "real dataset" OR "datasets" OR "synthetic datasets" OR "artificial datasets" OR "real datasets" OR "data set" OR "synthetic data set" OR "artificial data set" OR "real data set" OR "data sets" OR "synthetic data sets" OR "artificial data sets" OR "real data sets")

Figura 1: Expressão de busca adotada no protocolo de revisão sistemática

Ao todo, foram identificadas 554 publicações, tendo esse número reduzido para 367 após a aplicação dos critérios de exclusão. Posteriormente, foi realizada a leitura do título e resumo de todos os trabalhos apurados. Quando a seleção não pôde ser concretizada indubitavelmente, considerou-se a leitura da introdução e da conclusão e, quando necessário, do restante da publicação, até que não restassem dúvidas sobre a relevância do trabalho. Ao final desse procedimento, obtiveram-se 42 publicações das quais foram extraídas informações para responder às questões de pesquisa. Essas informações foram organizadas em uma planilha eletrônica⁶ e, além de constituírem tópicos gerais dos trabalhos, como título, nome do evento e ano de publicação, compreendem observações pontuais atreladas a cada uma das questões listadas na Tabela 1.

Na Figura 2 é exibido um gráfico de barras contendo a distribuição, por ano de publicação, do número de trabalhos selecionados ao término da revisão sistemática. O gráfico confeccionado mostra que, no período de janeiro de 2011 a novembro de 2014, foram publicados anualmente cerca de sete trabalhos. Tal fato pode ser explicado devido à consolidação do interesse pelo tema pesquisado.

A análise dos 42 trabalhos descobertos na revisão sistemática permitiu elaborar a seguinte meta-análise:

⁶O formulário contendo as informações extraídas das publicações selecionadas pode ser solicitado aos autores deste relatório.

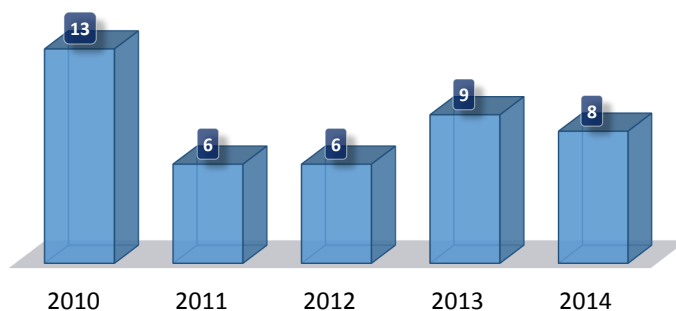


Figura 2: Distribuição do número de trabalhos por ano de publicação

Métodos de Predição de ST: Considerando a frequência com que os métodos apareceram nas publicações, 42,86% dos trabalhos empregaram Redes Neurais Artificiais (ANN), 42,86% construíram modelos Autorregressivos Integrados de Médias Móveis (ARIMA) ou Autorregressivos Integrados de Médias Móveis Sazonal (SARIMA), 26,19% usaram modelos híbridos, 23,81% aplicaram Máquinas de Suporte Vetorial (SVM), 14,29% executaram variações do algoritmo k -Vizinhos mais Próximos (kNN), 11,90% usufruíram dos métodos de Médias Móveis (MA) e de Suavização Exponencial Simples (SES), 7,14% abordaram técnicas bayesianas e 2,38% recorreram aos modelos de Holt-Winters (HW);

Abordagens para Predição de ST: Aplicações reais foram tema de 19 trabalhos, dos quais 57,89% utilizaram algoritmos decorrentes da abordagem não-paramétrica, 26,32% empregaram métodos da abordagem paramétrica e 15,79% construíram modelos híbridos baseados em ambas as abordagens;

Medidas de Avaliação: Observando a frequência com que as medidas foram usadas nas publicações, 33,33% dos trabalhos adotaram o Erro Absoluto Médio (MAE), 30,95% avaliaram a Raiz do Erro Quadrático Médio (RMSE), 30,95% verificaram o Erro Percentual Absoluto Médio (MAPE), 28,57% analisaram o Erro Percentual Absoluto Médio Simétrico (SMAPE), 26,19% utilizaram o Erro Quadrático Médio (MSE), 9,52% abordaram o Erro Médio Absoluto em Escala (MASE), 7,14% recorreram ao Coeficiente de Correlação (R) e 4,76% aplicaram o Erro Absoluto Relativo (RAE). As medidas U de Theil (U -Theil), *Prediction of Change in Direction* (POCID), Raiz do Erro Quadrático Médio Normalizado (NRMSE) e o Coeficiente de Determinação (R^2) apareceram em 2,38% das publicações;

Configuração dos Parâmetros: 76,19% dos trabalhos selecionados aplicaram alguma técnica de busca capaz de realizar a escolha dos parâmetros do modelo da melhor maneira possível. Entre essas técnicas, encontram-se a de treino e teste ou validação *holdout*, a de validação cruzada e a de

envolvida por [Box et al. \(2015\)](#), a qual é direcionada para modelos da categoria *ARIMA*;

Método *Baseline*: Avaliações empíricas que contemplam algoritmos estatísticos e de AM foram tema de 20 trabalhos, dos quais 70,00% adotaram como método *baseline* o modelo *ARIMA*, 20,00% empregaram o algoritmo de *MA*, 5,00% usufruíram do método de *SES* e 5,00% consideraram o modelo de *HW*;

Conjuntos de Dados Reais: O uso de conjuntos de dados reais esteve presente em 39 publicações, das quais 45,23% utilizaram dados produzidos em instituições, empresas ou indústrias, 23,81% analisaram dados concedidos em competições da área e 23,81% investigaram dados mantidos por repositórios, por exemplo os gerenciados pela *University of California at Irvine* ([Bache e Lichman, 2013](#)) e pela *Time Series Data Library*⁷;

Conjuntos de Dados Sintéticos: 7,14% dos trabalhos selecionados construíram conjuntos de dados sintéticos para avaliar o desempenho dos algoritmos de predição. Esses conjuntos de dados foram gerados computacionalmente para conter propriedades específicas, como a presença ou ausência de tendência e/ou sazonalidade;

Ferramentas: Entre os ambientes computacionais usados nas publicações estão *MATLAB*⁸, *R*⁹, *SPSS*¹⁰, *MINITAB*¹¹, *WEKA* ([Witten et al., 2011](#)), *STATISTICA*¹², *ForecastPRO*¹³ e *Phicast*¹⁴. Na Figura 3, o gráfico indica a frequência com que essas ferramentas apareceram nos trabalhos.

Durante a preparação da meta-análise verificou-se que seis trabalhos, do total de 42, envolviam avaliações experimentais focadas na interpretação do comportamento dos métodos de predição mais populares. Uma síntese das seis publicações identificadas é apresentada a seguir.

1. No trabalho de [Ahmed et al. \(2010\)](#) foi desenvolvido um estudo comparativo entre oito métodos de AM para regressão. Os métodos investigados foram *Multilayer Perceptron (MLP)*, *Gaussian Process (GP)*, *SVM*, *kNN*, *Redes Neurais Bayesianas (BNN)*, *Redes Neurais de Regressão Generalizada (GRNN)*, *Árvores de Regressão CART* e *Funções de Base Radial (RBF)*. Todos esses algoritmos foram aplicados sobre os conjuntos

⁷<http://robjhyndman.com/TSDL>.

⁸<http://www.mathworks.com/products/matlab>.

⁹<http://www.r-project.org>.

¹⁰<http://www.ibm.com/software/analytics/spss>.

¹¹<http://www.minitab.com>.

¹²<http://www.statsoft.com/Products/STATISTICA>.

¹³<http://www.forecastpro.com>.

¹⁴<http://phicast.software.informer.com>.

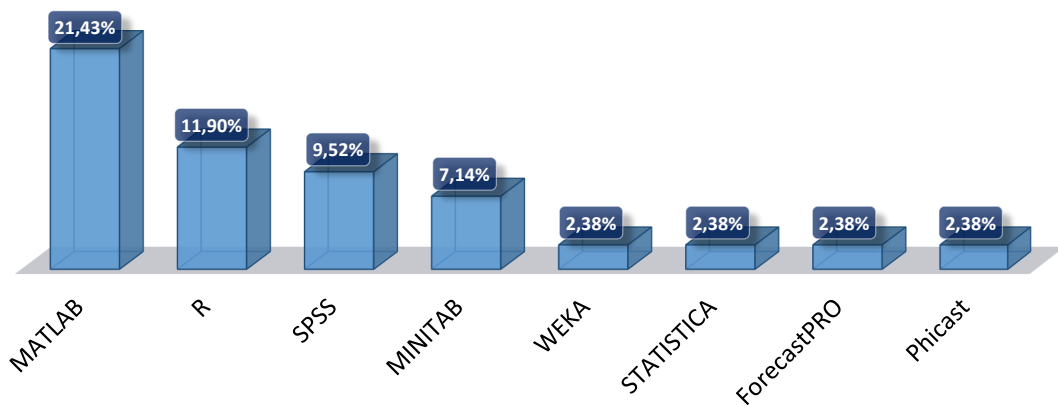


Figura 3: Percentagem do número de publicações que contemplaram uma determinada ferramenta computacional

de dados da *M3 Competition*¹⁵ e tiveram seus parâmetros estimados por meio de validação cruzada com dez partições. O panorama dos resultados mostrou que, além das técnicas de pré-processamento exercerem influência no desempenho dos algoritmos de predição, os modelos *MLP* e *GP* foram significativamente os mais promissores. Os autores explicaram que esse é um resultado interessante, haja vista que o potencial do método *GP* permaneceu inexplorado nos últimos anos;

2. Em Cortez (2010) foi proposto um algoritmo para a determinação do modelo não-paramétrico, *ANN* ou *SVM*, mais adequado para a predição de *ST* considerando múltiplas etapas à frente. Esse algoritmo procura, baseando-se na estratégia *grid search* e utilizando a direção de busca *backward selection* guiada por uma análise de sensibilidade, pelo valor subótimo de uma variável de defasagem (*time lag*). O protocolo experimental abrangeu oito *ST* sazonais e duas medidas para avaliação de desempenho, *MSE* e *SMAPE*. Os modelos de *AM* selecionados foram comparados com o método paramétrico de *HW*, o qual é recomendado para *ST* que apresentam tanto o componente de tendência quanto o de sazonalidade. Os resultados demonstraram a efetividade do modelo *SVM* ajustado por meio da estimativa do parâmetro *time lag*;
3. Realizou-se em Kandanond (2012) uma avaliação experimental que abrangeu a comparação de dois métodos de *AM* para regressão, *ANN* e *SVM*, e um modelo estatístico convencional, o *ARIMA*. Foram empregados seis conjuntos de dados reais referentes à demanda por produtos de consumo na Tailândia. Cada um desses conjuntos de dados foi analisado, previamente à construção dos modelos, segundo a estatística *Q* de Ljung-Box para verificar a existência de autocorrelação dos resíduos. O

¹⁵<http://forecasters.org/resources/time-series-data/m3-competition>.

método *SVM* proporcionou, conforme a medida *MAPE*, as melhores previsões sobre todas as categorias de produtos. Já o modelo *ARIMA*, que fundamenta-se na estrutura de autocorrelação, exibiu os piores resultados. Notou-se também que dados autocorrelacionados podem afetar o desempenho do algoritmo *SVM*, uma vez que o mais elevado grau de autocorrelação implicou em um menor número de vetores de suporte;

4. O trabalho de [Ristanoski et al. \(2013\)](#) destacou que integrar elementos do tempo no processo de aprendizagem constitui o maior desafio no uso de *SVM* para previsão de ST, pois elas são suscetíveis à erros quando mudanças de distribuição ocorrem com frequência ao longo da série. Para auxiliar nessa questão, foi investigada a distribuição de erros nas previsões obtidas pelo algoritmo *SVM*. Uma vez identificadas as amostras que produziam os maiores erros, observou-se a sua correlação com as mudanças que ocorriam na distribuição da série histórica. O entendimento desse comportamento motivou os autores a propor uma função de perda dependente do tempo, a qual viabiliza a inclusão de informações sobre mudanças de distribuição da série diretamente no processo de aprendizagem. Os experimentos foram conduzidos a partir de dados reais, 35 ST referentes à valores do mercado de ações e de medições de fenômenos físicos e químicos, e sintéticos, cinco versões de um conjunto de dados com diferentes níveis de distribuição. O método proposto foi comparado com sua versão alternativa, que adota média quadrática, e com outros seis algoritmos de previsão: *ANN*, *kNN*, *SVM*, *RBF*, *Robust Regression* e *SARIMA*. Os resultados, expressos de acordo com a medida *RMSE*, sugeriram que o uso de uma função de perda dependente do tempo pode reduzir a variância global dos erros e, portanto, acarretar em previsões mais precisas;
5. Em [Claveria e Torra \(2014\)](#) foi investigado o desempenho preditivo dos algoritmos *ANN*, *ARIMA* e *Self-Exciting Threshold Autoregressive (SETAR)*. Dados mensais pré-processados de dormidas¹⁶ e chegadas de turistas internacionais para Catalunha, entre os anos de 2001 a 2009, foram adotados no trabalho como indicadores oficiais da demanda turística. Ao comparar o desempenho dos métodos sobre distintos horizontes de previsão, o modelo *ARIMA* superou os algoritmos *ANN* e *SETAR*, principalmente para horizontes mais curtos. Os resultados obtidos via *ANN* apontaram um *trade-off* entre o grau de pré-processamento e a qualidade das previsões, as quais foram mais precisas na presença de não-linearidade

¹⁶Dormida refere-se à permanência de um indivíduo em um estabelecimento que fornece alojamento, por um período compreendido entre as 12 horas de um dia e as 12 horas do dia seguinte.

nos dados. Devido aos diferentes padrões de conduta do consumidor em turismo, verificou-se que as previsões de chegadas foram mais precisas do que as de dormidas de estrangeiros;

6. No trabalho de [Zhang et al. \(2014\)](#) foi concebido um estudo empírico que abrangeu quatro métodos de previsão de ST: Regressão Linear, *SES*, *SARIMA* e *SVM*. Empregaram-se nessa análise nove conjuntos de dados acerca de doenças infecciosas, os quais foram coletados por meio de um sistema nacional de vigilância em saúde pública na China. Os resultados obtidos, avaliados de acordo com as medidas *MAE*, *MAPE* e *MSE*, demonstraram que, embora o algoritmo *SVM* tenha superado, na maioria dos casos, o desempenho dos modelos estatísticos, nenhum dos métodos averiguados foi significativamente melhor que os demais.

Como pôde ser observado, os métodos de AM para previsão de valores em ST têm proporcionado resultados muito competitivos, frequentemente superando modelos estatísticos estado-da-arte. Ainda assim, não foram encontrados trabalhos contendo estudos empíricos robustos que propiciem o confronto entre os métodos paramétricos e não-paramétricos. Uma pesquisa minuciosa a respeito da aplicação desses algoritmos em dados seguramente conhecidos e na presença de propriedades que desafiam a modelagem de ST, como a tendência, a não-estacionariedade e as mudanças de distribuição, poderia consolidar a eficiência e a efetividade de cada método. O presente trabalho, além de preencher grande parte dessa lacuna, visa guiar o processo de escolha da estrutura matemática do modelo, de configuração de parâmetros e de execução dos algoritmos, em especial os de MD, para a tarefa de previsão de ST.

3 Fundamentos de Séries Temporais

Séries Temporais (ST) são comumente caracterizadas como um conjunto de observações obtidas de maneira sequencial ao longo do tempo. Desse modo, uma ST Z de tamanho m pode ser formulada como uma sequência ordenada de observações, ou seja, $Z = (z_1, z_2, \dots, z_m)$ onde $z_t \in \mathfrak{R}$ e representa uma observação z no instante de tempo t ([Chatfield, 2013](#)).

A formulação de dados no domínio do tempo é de grande importância, visto que a relação entre as observações cronologicamente adjacentes abrange a dependência que uma observação possui com outra. Demais exemplos de representação são o domínio das frequências intrínsecas na ST, e a junção de ambas, que contribui para a evidenciação de quais frequências estão inseridas em determinados intervalos de tempo.

De acordo com o intervalo em que os dados são adquiridos, as ST podem ser categorizadas em contínua e discreta. As ST contínuas são provenientes de sistemas analógicos, nos quais as observações dos dados são realizadas de maneira contínua sobre um intervalo de tempo específico. Já as ST discretas são decorrentes de sistemas digitais, onde as observações dos dados são convertidas em intervalos fixos de tempo, geralmente igualmente espaçado (Brockwell e Davis, 2002). É conveniente salientar a possibilidade de obtenção de uma ST discreta a partir de uma ST contínua. Para tanto, é necessário amostrar a série contínua em intervalos de tempos previamente definidos.

Quando os valores de uma série puderem ser sintetizados por meio de uma função matemática $y = f(\text{tempo})$, diz-se que a série é determinística. Diferentemente, quando a série comporta, além de uma função matemática do tempo, um termo aleatório ϵ , $y = f(\text{tempo}, \epsilon)$, chama-se a série de estocástica ou não-determinística.

Ainda como característica relevante de uma ST, pode-se citar a estacionariedade. Em conformidade com Morettin e Tolo (2006), uma série estacionária se desenvolve no tempo aleatoriamente ao redor de uma média constante, refletindo alguma forma de equilíbrio estável. Em alguns cenários é indispensável a atenção a essa propriedade, pois, sempre que um método de análise supor condição de estacionariedade, será vantajoso transformar os dados originais caso eles constituam uma série não-estacionária. A transformação mais comum consiste em tomar diferenças sucessivas da série original. Geralmente a primeira diferença, definida pela Equação 1, é suficiente para tornar a série estacionária.

$$\Delta z_t = z_t - z_{t-1} \quad (1)$$

Na Figura 4 é esquematizada uma ST real referente à produção mensal de chocolate em toneladas na Austrália, no período de janeiro de 1958 a dezembro de 1990. Essas observações foram cedidas pela agência australiana de estatísticas¹⁷ e, assim como todas as séries adotadas neste trabalho, encontram-se disponíveis no ICMC-USP *Time Series Prediction Repository* (Parmezan e Batista, 2014).

Observa-se na Figura 4 que os valores da ST não oscilam em torno de um nível fixo. Ao invés disso, eles apresentam um comportamento crescente cujo período de variação se mantém constante à medida que o nível aumenta. Essas e outras propriedades podem ser melhor exploradas com o uso de técnicas para decomposição de dados temporais. As técnicas de decomposição, além de permitirem a identificação dos componentes que atuam na série, possibilitam

¹⁷<http://www.abs.gov.au>.

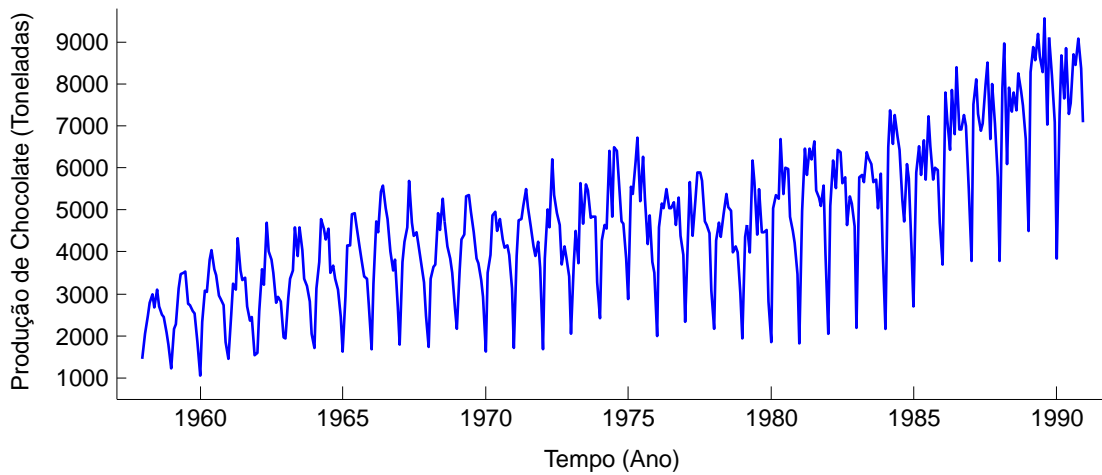


Figura 4: Produção mensal de chocolate na Austrália

a obtenção de índices e/ou equações que podem ser acoplados a um modelo computacional para predição de valores futuros.

3.1 Componentes de uma Série Temporal

Frequentemente na literatura, para uma melhor análise e compreensão de eventos representados por ST, é utilizado o conceito de decomposição da série em um conjunto finito de elementos independentes. Os principais componentes abordados são intitulados de tendência, sazonalidade e resíduo. Com base nesses elementos, a ST Z pode ser reformulada, conforme as Equações 2 e 3, por uma decomposição aditiva ou multiplicativa de seus componentes (Cov-pertwait e Metcalfe, 2009). Nessas equações, T , S e N correspondem à tendência, a sazonalidade e ao resíduo em um instante t , respectivamente.

$$Z_t = T_t + S_t + N_t \quad (2)$$

$$Z_t = T_t \times S_t \times N_t \quad (3)$$

No modelo aditivo (Equação 2), o valor da variável de interesse é constituído pelo resultado da soma dos valores dos componentes, os quais contemplam a mesma unidade da observação Z_t . Em contraste, no modelo multiplicativo (Equação 3), apenas a tendência possui a mesma unidade da variável investigada. Os demais componentes exibem valores que podem modificar a tendência, ou seja, assumem valores maiores, menores ou exatamente iguais a 1. É importante ressaltar que nem sempre uma ST, mesmo quando a decomposição clássica for considerada, irá abranger os três componentes mencionados.

3.1.1 Tendência

A tendência pode ser definida como o movimento regular e lentamente desenvolvido ao longo da série. Em outras palavras, esse componente engloba um comportamento de extensa duração, podendo ser tanto crescente quanto decrescente e assumir uma grande variação de padrões, dentre os quais se sobressaem (Ehlers, 2009):

Crescimento Linear: Compreende uma taxa de crescimento constante para os dados, a qual obedece uma proporção linear;

Crescimento Exponencial: Caracterizado pelo progressivo aumento percentual dos dados por período de tempo. As particularidades da taxa de crescimento equivalem às propriedades de uma função exponencial;

Crescimento Amortecido: Ocorre quando a taxa de crescimento de dados futuros é menor que os dados atuais, como em situações nas quais para um determinado ano o crescimento esperado é 70% do ano anterior.

No mundo real, ST com tendência crescente podem ser oriundas de fenômenos interligados ao desenvolvimento demográfico, a mudança gradual de hábitos de consumo e a demanda por tecnologias nos setores da sociedade. Por outro lado, a tendência decrescente pode ser encontrada em séries relativas às taxas de mortalidade, epidemias e desemprego.

A identificação do componente de tendência inclui três objetivos (Moret-tin e Tolo, 2006): (1) avaliar o seu comportamento e integrá-lo a um modelo preditivo; (2) removê-lo da ST para promover a visibilidade dos demais componentes; ou (3) usá-lo para estimar o nível da série. Neste último caso, o nível refere-se ao valor ou a faixa típica de valores que a variável pode assumir caso não seja observada tendência crescente ou decrescente a longo prazo.

Na Figura 5 é mostrada uma ST anual, adquirida entre os anos de 1985 e 2006, cujos dados refletem os preços dos ingressos para o *Super Bowl*, o jogo do campeonato da Liga Nacional de Futebol Americano (NFL). Os valores dessa série configuram uma tendência levemente linear, a qual pode ser evidenciada empregando um modelo de regressão.

Regressão é o processo que relaciona o comportamento de uma variável resposta (ou dependente) com outras variáveis explicativas (ou independentes), a fim de se obter um modelo matemático que permite justificar as alternâncias da variável dependente com base na variação dos níveis das variáveis independentes. Esse modelo é designado de simples quando envolve uma relação entre duas variáveis, Y em consequência de X . Já quando o comportamento da variável resposta Y é descrito por mais de uma variável explicativa, tal

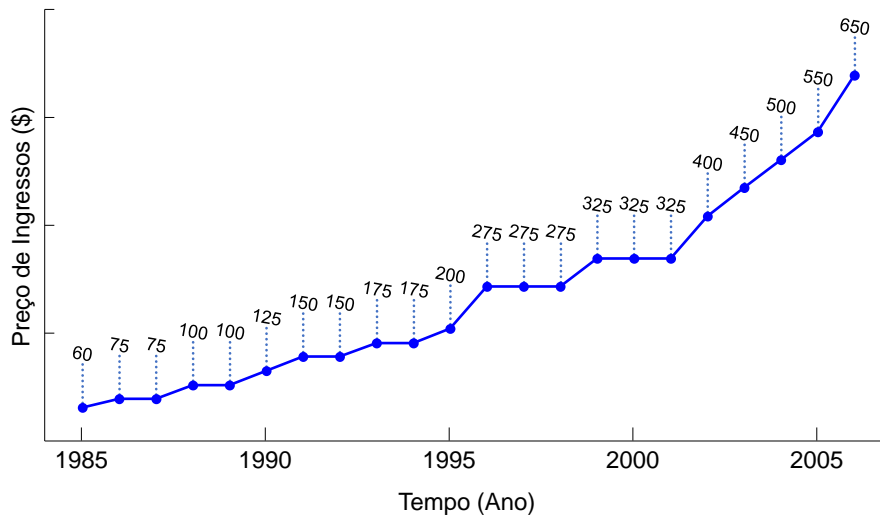


Figura 5: Preço anual de ingressos para o *Super Bowl*

como X_1, X_2, \dots, X_m , o modelo de regressão é denominado de múltiplo. Adicionalmente, o comportamento de Y em função de X pode assumir diferentes proporções, por exemplo linear, quadrática, cúbica, exponencial e logarítmica. Nessas condições, a escolha da estrutura de um modelo de regressão está associada ao tipo curva que as observações do fenômeno naturalmente tendem a se aproximar (Chatterjee e Hadi, 2012).

Uma das estruturas mais simples que relaciona duas variáveis X e Y pode ser visibilizada na Equação 4 (Barroso et al., 1987). Esse modelo, também conhecido como equação da reta, possui dois parâmetros: b_0 , que consiste no coeficiente linear da reta, e b_1 , que refere-se ao coeficiente angular da reta. O valor assumido por b_1 controla a direção pela qual os valores observados se movimentam.

$$Y = b_0 + b_1X \quad (4)$$

Em ST, o uso da Equação 4 considera a tendência como variável resposta e o tempo, em períodos, como variável explicativa. A determinação dos coeficientes b_0 e b_1 pode ser realizada por meio das Equações 5 e 6. Nessas equações, y_i compreende uma observação registrada na ST, x_i indica o período associado à observação y_i e m denota o número de períodos da série.

$$b_1 = \frac{m \sum x_i y_i - \sum x_i \sum y_i}{m \sum x_i^2 - (\sum x_i)^2} \quad (5)$$

$$b_0 = \frac{\sum y_i - (\sum x_i) b_1}{m} \quad (6)$$

As equações para determinação dos parâmetros b_0 e b_1 foram derivadas a partir do método dos mínimos quadrados, o qual minimiza a soma dos quadrados das distâncias entre os pontos da ST e os pontos da reta ajustada pelo

modelo de regressão (Chatterjee e Hadi, 2012). Em se tratando de um problema de otimização, os valores observados de X e Y nem sempre serão iguais aos valores de X' e Y' estimados. Essas diferenças podem significar que as variações de Y não são perfeitamente explicadas pelas variações de X , ou que os valores de X e Y são decorrentes de uma amostra específica que apresenta distorções em relação à realidade, ou ainda que existem outras variáveis das quais Y é dependente.

Embora tenha sido descrito os passos para a obtenção dos coeficientes do modelo de regressão linear, o qual visa ajustar uma reta aos dados, o mesmo procedimento empregando o método dos mínimos quadrados pode ser aplicado para equações matemáticas com maior número de parâmetros. Nesse sentido, quando $X_1 = X, X_2 = X^2, \dots, X_P = X^P$, pode-se explicitar a tendência de uma ST por meio do uso da Equação 7 (Barroso et al., 1987).

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \dots + \beta_P X^P \quad (7)$$

A Equação 7 compreende a estrutura do modelo de regressão polinomial que, além de incluir a equação da reta, é um caso especial de ajuste linear múltiplo. Os coeficientes desse modelo são determinados a partir resolução do seguinte sistema:

$$\begin{bmatrix} m & \sum x_i & \sum x_i^2 & \dots & \sum x_i^P \\ \sum x_i & \sum x_i^2 & \sum x_i^3 & \dots & \sum x_i^{P+1} \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 & \dots & \sum x_i^{P+2} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum x_i^P & \sum x_i^{P+1} & \sum x_i^{P+2} & & \sum x_i^{2P} \end{bmatrix} \times \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_P \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \\ \vdots \\ \sum x_i^P y_i \end{bmatrix}$$

Na Figura 6 é exibida as linhas de tendência resultantes da aplicação dos modelos de regressão linear e polinomial sobre a ST de preços de ingressos retratada na Figura 5. Em relação ao modelo polinomial, foi ajustada aos dados a equação $Y = b_0 + b_1 X + b_2 X^2$.

A aproximação polinomial, procedente do ajuste $Y = 76,01 - 0,06X + 1,07X^2$, foi mais adequada que a aproximação linear, proveniente da equação estimada $Y = -22,86 + 24,66X$, para determinar a tendência da ST de preços de ingressos. Esse fato, apesar de esperado, reforça a importância do critério adotado na escolha do modelo de regressão, o qual é dependente do tipo curva que as observações da ST mais se assemelham. No exemplo abordado, os valores observados seguem uma proporção quadrática.

Uma maneira alternativa para obtenção da tendência ou nível de uma ST é por meio da aplicação do método de Médias Móveis (MA). Esse método calcula a média das r primeiras observações da série e associa ao período central

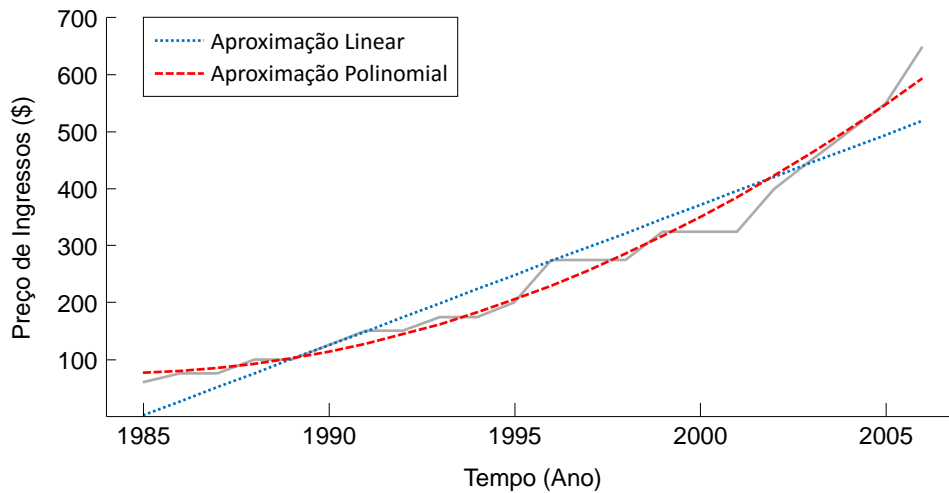


Figura 6: Linhas de tendência para a série de preços de ingressos para o *Super Bowl*

delas ($r/2$) o respectivo resultado. Iterativamente, é acrescentado o valor do período seguinte e desprezada a primeira observação abrangida no cálculo da média imediatamente anterior, de modo que novas médias são computadas à medida que esse procedimento move-se em direção ao último valor da série (Schiller et al., 2012).

Na Figura 7 é exemplificado, a partir de seis observações e supondo $r = 3$, o funcionamento do método de *MA* para a extração do componente de tendência. Nessa figura, ao passo que o procedimento percorre a *ST*, as observações são organizadas em grupos de três períodos e a média desses valores é colocada no período central do grupo corrente. Nota-se no término do procedimento que, devido aos resultados das médias serem posicionados no centro dos grupos, alguns períodos acabam sem tendência.

①		②		③		④		Tendência	
<i>t</i>	<i>y</i>	<i>t</i>	<i>y</i>	<i>t</i>	<i>y</i>	<i>t</i>	<i>y</i>	<i>t</i>	<i>MA(3)</i>
1	0,7	1	0,7	1	0,7	1	0,7	1	—
2	1,3	2	1,3	2	1,3	2	1,3	2	1,2
3	1,6	3	1,6	3	1,6	3	1,6	3	1,8
4	2,4	4	2,4	4	2,4	4	2,4	4	2,3
5	2,8	5	2,8	5	2,8	5	2,8	5	1,9
6	0,5	6	0,5	6	0,5	6	0,5	6	—

Figura 7: Exemplo de obtenção da tendência usando *MA* com $r = 3$

O parâmetro r é frequentemente escolhido como um valor ímpar para que o resultado da média possa ser vinculado a um período central com correspondente na série original. Porém, em algumas situações, para que a tendência seja obtida sem influência da sazonalidade, esse parâmetro precisa representar, em número de observações, uma variação sazonal. Por exemplo, se a *ST* for registrada trimestralmente (quatro trimestres por ano) ou mensalmente

(doze meses por ano) será necessário adotar *MA* com $r = 4$ e $r = 12$, respectivamente. A princípio, o uso desses valores pares impossibilitaria a remoção da tendência para observação de outros componentes, pois os períodos centrais que começariam, nessa ordem, em 2,5 e 6,5, não possuem correspondentes na série real. Um mecanismo matemático que permite contornar esse problema é o designado de centralização, por meio do qual são calculadas novas médias, a partir das obtidas com $r = 4$ e $r = 12$, empregando $r = 2$ e inserindo os resultados em períodos que têm correspondentes na série original.

Na Tabela 2 é apresentado um exemplo de aplicação do método de *MA* centrada, com $r = 4$, sobre uma ST de produção de cerveja, expressa em milhões de barris, nos Estados Unidos da América (EUA). Os dados foram coletados trimestralmente no período de janeiro de 1975 a dezembro de 1982. Na referida tabela, especificamente na quarta coluna, encontram-se as médias de valores obtidas a partir da aplicação de *MA* com quatro períodos. Observa-se que esses resultados não possuem correspondentes na série real e, portanto, como indicado na última coluna, *MA* de dois períodos, calculadas considerando as de quatro períodos, foram empregues para obter resultados centrados.

Na Figura 8 é mostrada a tendência, determinada conforme indicado na Tabela 2, da ST referente à produção trimestral de cerveja. Nessa figura, o componente de tendência assume um comportamento crescente entre os anos de 1975 a 1980. Posteriormente, o movimento ascendente se torna uniforme com nível em torno de 48 milhões de barris produzidos.

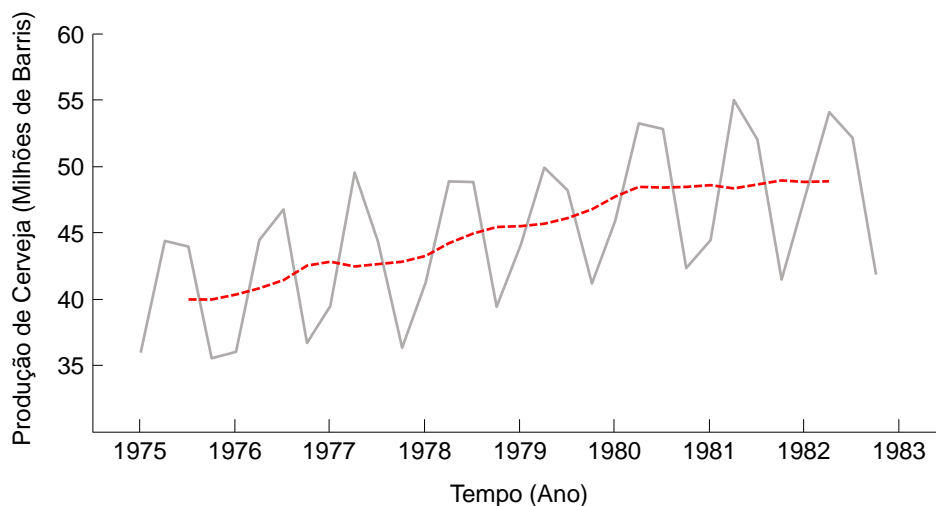


Figura 8: Tendência, indicada pela linha tracejada, da série de produção trimestral de cerveja nos EUA

Na Figura 9 são exibidas três linhas de tendência para a ST de produção mensal de chocolate na Austrália. Tais tendências foram obtidas usando *MA* com sete e nove períodos, além de doze períodos com centralização. Essa última configuração retrata adequadamente o comportamento de longo prazo da série supracitada.

Tabela 2: Exemplo de obtenção da tendência por *MA* centrada com $r = 4$

Ano	Trimestre	Observação	MA(4)	Centralização MA(2)
1975	I	36,14		—
	II	44,60		—
	III	44,15	40,15	40,16
	IV	35,72	40,17	40,17
1976	I	36,19	40,17	40,52
	II	44,63	40,87	41,02
	III	46,95	41,17	41,60
	IV	36,90	42,04	42,67
1977	I	39,66	43,31	43,00
	II	49,72	42,69	42,65
	III	44,49	42,60	42,83
	IV	36,54	43,05	42,97
1978	I	41,44	42,89	43,45
	II	49,07	44,01	44,39
	III	48,98	44,77	45,13
	IV	39,59	45,48	45,61
1979	I	44,29	45,74	45,67
	II	50,09	45,60	45,82
	III	48,42	46,05	46,28
	IV	41,39	46,50	46,92
1980	I	46,11	47,34	47,91
	II	53,44	48,49	48,63
	III	53,00	48,77	48,58
	IV	42,52	48,39	48,61
1981	I	44,61	48,83	48,73
	II	55,18	48,64	48,53
	III	52,24	48,42	48,83
	IV	41,66	49,23	49,12
1982	I	47,84	49,00	49,01
	II	54,27	49,02	49,07
	III	52,31	49,11	—
	IV	42,03		—

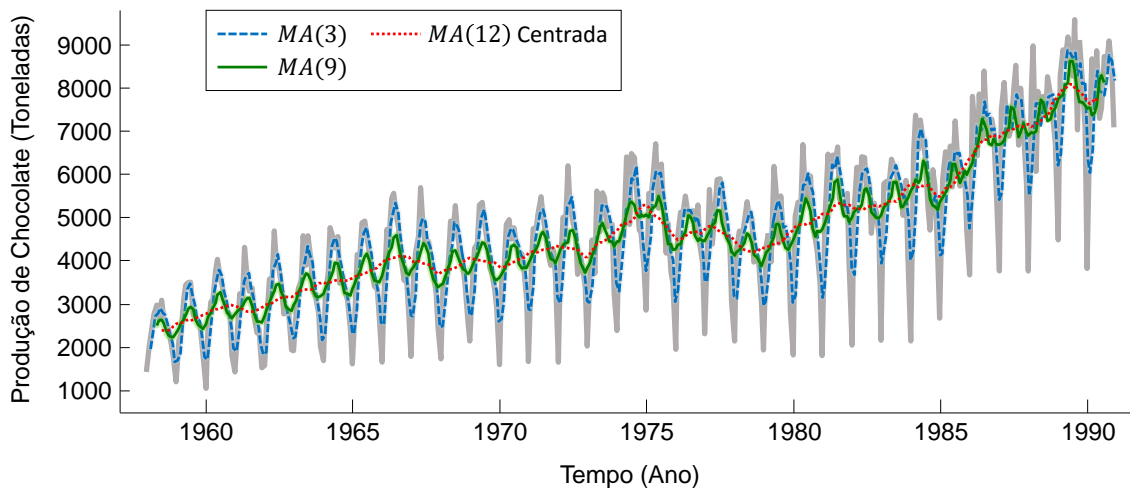


Figura 9: Linhas de tendência, obtidas utilizando MA , para a série de produção mensal de chocolate na Austrália

Como pode ser verificado na Figura 9, a linha de tendência acompanha gradativamente o comportamento dos dados à medida que o valor do parâmetro r é reduzido. Todavia, essa linha se torna mais suave quando ampliada a quantidade de observações agrupadas pela média. À vista disso, é comum a aplicação do método de MA para suavizar ST providas de irregularidades.

Uma vez identificada a tendência, seja via regressão ou por médias móveis, a mesma pode ser removida da série para realçar os demais componentes. A remoção pode ser realizada usando o modelo aditivo, subtraindo das observações originais a tendência ($Z_t - T_t = S_t + N_t$), ou empregando o modelo multiplicativo, dividindo as observações originais pela tendência ($Z_t \div T_t = S_t \times N_t$).

3.1.2 Sazonalidade

Um comportamento que tende a se repetir em diferentes períodos de tempo na ST é conhecido como sazonalidade. As variações sazonais são representadas pelas oscilações ao longo do componente de tendência segundo uma determinada particularidade. Essa particularidade, além de estar frequentemente relacionada às estações do ano, pode ser decorrente de causas naturais, econômicas e/ou sociais (Brocklebank e Dickey, 2003).

De maneira empírica, podem ser citados como exemplos de variações sazonais fenômenos que ocorrem de ano em ano (ou algum outro ciclo temporal), como aumento nas vendas de condicionadores de ar no verão e de agasalhos no inverno. Particularmente, essas relações podem ser visibilizadas em meses sucessivos de um ano exclusivo ou em um mesmo mês durante anos consecutivos (Morettin e Toloi, 2006).

A identificação de picos e depressões regularmente espaçados, que tem uma direção consistente e aproximadamente a mesma magnitude em cada ciclo, consiste em um procedimento importante no tema de análise de ST . Por

meio da explicitação do componente de sazonalidade é possível compreender aspectos relevantes do fenômeno observado. Contudo, dependendo do campo de aplicação, a existência de sazonalidade pode dificultar o reconhecimento e a interpretação de movimentos não-sazonais peculiares em uma série. Em resumo, a detecção do componente de sazonalidade pode revelar informações valiosas e sua remoção pode ressaltar padrões proveitosos da ST (Chatfield, 2013).

Outra característica do componente sazonal é que ele pode ser categorizado, segundo sua variação, em dois tipos (Ehlers, 2009):

Sazonalidade Aditiva: Neste tipo de sazonalidade a série apresenta uma flutuação sazonal estável, sem levar em consideração o nível global da série;

Sazonalidade Multiplicativa: Este tipo de sazonalidade ocorre quando o tamanho da flutuação sazonal varia de acordo com o nível global da série.

A determinação da variação sazonal na ST é a primeira análise que deve ser realizada em relação à sazonalidade. Uma maneira de detectar essa oscilação é por meio da própria inspeção visual da série no gráfico de variação da variável no tempo, tal como ilustrado na Figura 8 da página 18, na qual é notável que o padrão sazonal se repete a cada quatro períodos de tempo. Nos casos em que a ocorrência do padrão sazonal não seja evidente, deve-se aplicar técnicas mais sofisticadas para auxiliar na interpretação desse comportamento. Uma técnica amplamente difundida, que utiliza-se de um gráfico de dispersão (*scatter plot*), trabalha com o conceito de autocorrelação na ST. A ideia dessa técnica é encontrar o valor mínimo da norma do resíduo oriundo da regressão linear computada a partir das observações da série para vários valores de defasagem (*lag*) (Buffa e Sarin, 1987).

A norma do resíduo, definida pela Equação 8, baseia-se na diferença entre o valor real da variável investigada e o seu valor esperado.

$$RN = \sqrt{\sum |Z_{t+lag} - \hat{Z}_{t+lag}|^2} \quad (8)$$

No *scatter plot*, o eixo das abcissas contempla os valores da ST no instante Z_t , enquanto que o eixo das ordenadas abrange os valores da série no momento Z_{t+lag} . Por exemplo, para um valor de $lag = 4$, os pontos no *scatter plot* seriam representados pelas coordenadas (Z_t, Z_{t+4}) . Se os pontos indicados por essas coordenadas forem colineares, então a norma do resíduo da regressão linear resultará em zero, indicando que os dados da série são perfeitamente periódicos e que o número de períodos da variação sazonal é o próprio valor de *lag* (quatro períodos).

Como na prática dificilmente têm-se ST perfeitamente periódicas, a determinação do número de períodos do padrão sazonal é conduzida pelo conteúdo

de *lag* que fornecer o menor valor para a norma do resíduo da regressão linear.

Os gráficos de dispersão dispostos na Figura 10 ilustram, para valores de *lag* no intervalo de $[1, 4]$, o cálculo da variação sazonal da série de produção de cerveja apresentada na Figura 8 da página 18. A norma residual mínima foi computada usando *lag* de quatro períodos.

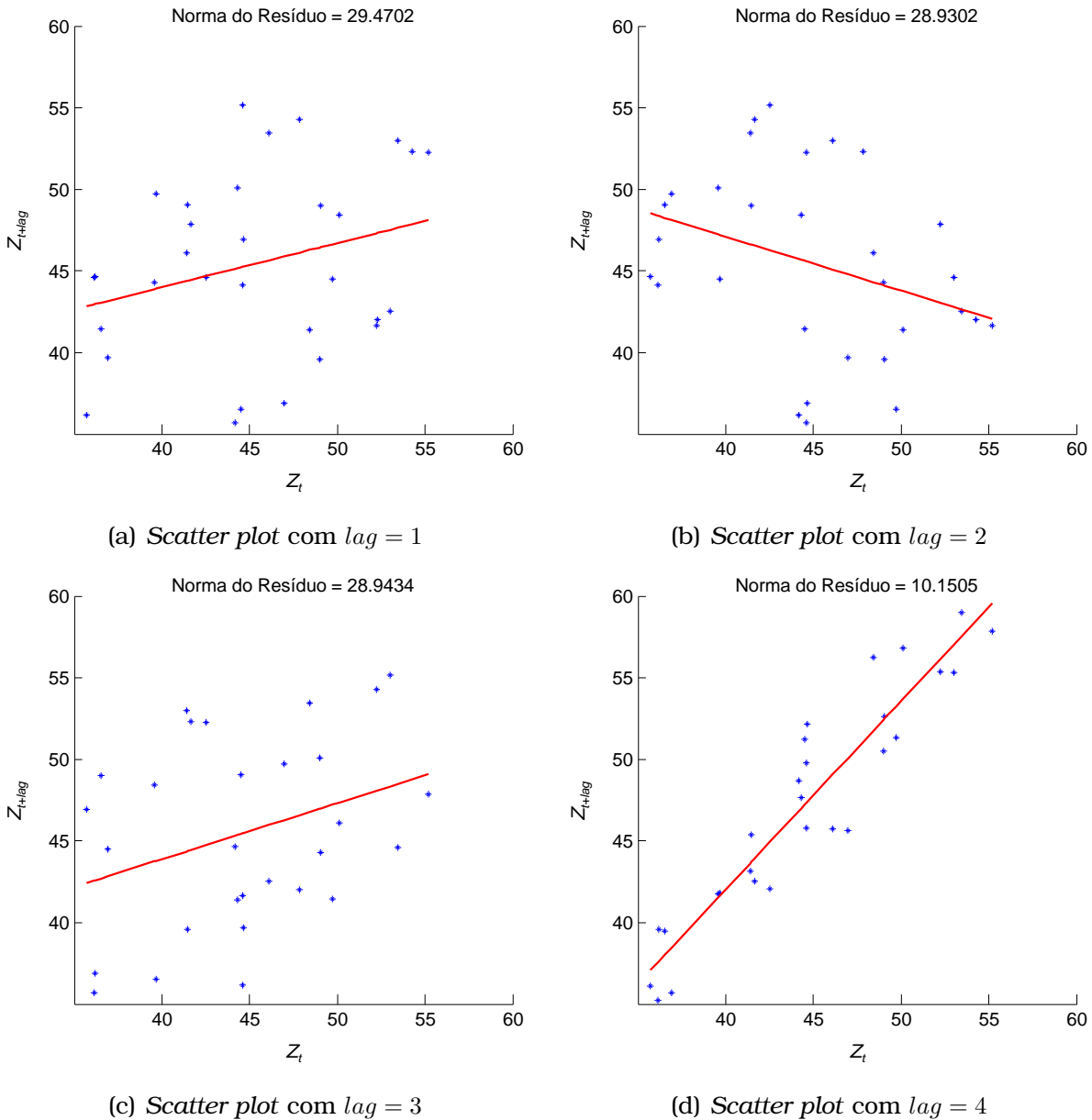
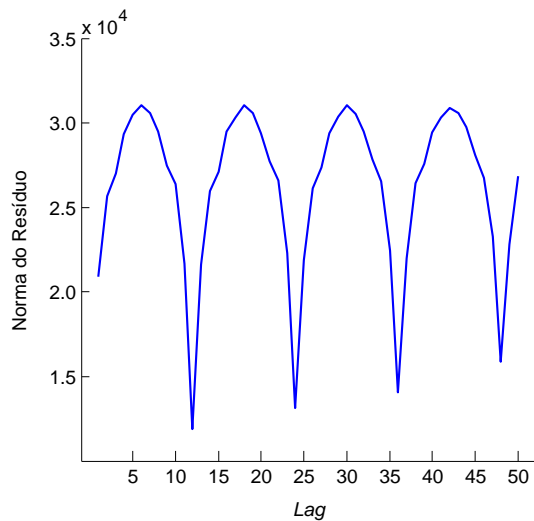
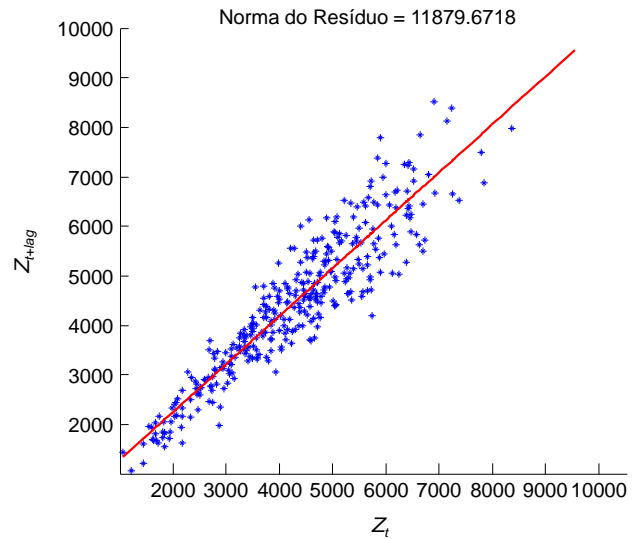


Figura 10: Técnica do *scatter plot* para a série de produção trimestral de cerveja nos EUA

A aplicação da técnica do *scatter plot* para a série de produção de chocolate, inicialmente mostrada na Figura 4 da página 13, é sintetizada na Figura 11. O gráfico da Figura 11(a) compreende a norma do resíduo para cada valor de *lag* no intervalo de $[1, 50]$. Nesse gráfico, o menor valor para a norma residual foi obtido com *lag* de 12 períodos. O *scatter plot* para *lag* = 12 é exibido na Figura 11(b).



(a) Norma do residuo em função do lag



(b) Scatter plot para lag = 12

Figura 11: Técnica do *scatter plot* para a série de produção mensal de chocolate na Austrália

Um índice sazonal equivale, dependendo do tipo de decomposição considerado, à diferença ou ao quociente entre o valor observado em um determinado período de tempo e a média das observações que compõem esse período. Na literatura há distintos métodos que possibilitam, a partir da extração dos índices sazonais, obter o componente de sazonalidade. À exemplo, têm-se o algoritmo da percentagem média, da relação percentual e dos elos relativos. Dentre todos esses métodos destaca-se o da relação entre *MA* ou da média móvel percentual, o qual pode ser executado em cinco etapas, como descrito na Tabela 3, considerando tanto a decomposição do modelo aditivo quanto a do multiplicativo (Schiller et al., 2012).

Tabela 3: Etapas do método da relação entre *MA*

Etapa	Modelo Aditivo	Modelo Multiplicativo
1	Obter <i>MA</i> com parâmetro r igual a ordem da sazonalidade na ST	Obter <i>MA</i> com parâmetro r igual a ordem da sazonalidade na ST
2	Calcular <i>MA</i> de dois períodos a partir dos resultados obtidos na Etapa 1	Calcular <i>MA</i> de dois períodos a partir dos resultados obtidos na Etapa 1
3	Determinar os índices de cada período por meio da subtração dos valores originais da série pelas <i>MA</i> centradas computadas na Etapa 2	Determinar os índices de cada período por meio da divisão dos valores originais da série pelas <i>MA</i> centradas computadas na Etapa 2
4	Calcular, para cada período da variação sazonal, a média aritmética simples dos índices obtidos na Etapa 3	Obter, para cada período da variação sazonal, a mediana dos índices estimados na Etapa 3
5	Somar todos os valores computados na Etapa 4 e dividir a soma pela ordem da sazonalidade. Após, subtrair esse fator de cada índice médio, garantindo que a soma deles seja igual a zero	Somar todos os valores computados na Etapa 4 e subtrair da soma a ordem da sazonalidade. Após, dividir o resultado da subtração pela ordem sazonal e subtrair o respectivo resultado de 1. Ao final, multiplicar esse fator por cada um dos índices medianos, garantindo que a soma deles seja igual a ordem da sazonalidade

Na Tabela 3, as primeiras duas etapas são idênticas ao procedimento para estimação da tendência por *MA* quando a ordem da *ST* é par. Na terceira etapa, para cada período cuja média centrada tem correspondente na série real, são calculados os índices considerando a decomposição aditiva ou multiplicativa. Esses índices são, na quarta etapa, agrupados em períodos de forma a compor uma variação sazonal. A partir de cada grupo de valores identificado, são obtidas as medidas de síntese dos índices que representam a variação sazonal do componente de sazonalidade. Por fim, na quinta e última etapa, são realizados os ajustes necessários para que a soma das medidas de síntese dos índices seja coerente.

Na Tabela 4 são apresentados os índices sazonais, computados de acordo com as Etapas 1 e 2 do método da relação entre *MA*, da série de produção de cerveja.

Tabela 4: Exemplo de obtenção dos índices sazonais pelo método da relação entre *MA*

Ano	Trimestre	Observação	<i>MA</i> (4) Centrada	Índice Sazonal (Modelo Aditivo)	Índice Sazonal (Modelo Multiplicativo)
1975	I	36,14	—	—	—
	II	44,60	—	—	—
	III	44,15	40,16	3,99	1,10
	IV	35,72	40,17	-4,45	0,89
1976	I	36,19	40,52	-4,33	0,89
	II	44,63	41,02	3,61	1,09
	III	46,95	41,60	5,35	1,13
	IV	36,90	42,67	-5,77	0,86
1977	I	39,66	43,00	-3,34	0,92
	II	49,72	42,65	7,07	1,17
	III	44,49	42,83	1,66	1,04
	IV	36,54	42,97	-6,43	0,85
1978	I	41,44	43,45	-2,01	0,95
	II	49,07	44,39	4,68	1,11
	III	48,98	45,13	3,85	1,09
	IV	39,59	45,61	-6,02	0,87
1979	I	44,29	45,67	-1,38	0,97
	II	50,09	45,82	4,27	1,09
	III	48,42	46,28	2,14	1,05
	IV	41,39	46,92	-5,53	0,88
1980	I	46,11	47,91	-1,80	0,96
	II	53,44	48,63	4,81	1,10
	III	53,00	48,58	4,42	1,09
	IV	42,52	48,61	-6,09	0,87
1981	I	44,61	48,73	-4,12	0,92
	II	55,18	48,53	6,65	1,14
	III	52,24	48,83	3,41	1,07
	IV	41,66	49,12	-7,46	0,85
1982	I	47,84	49,01	-1,17	0,98
	II	54,27	49,07	5,20	1,11
	III	52,31	—	—	—
	IV	42,03	—	—	—

No exemplo da Tabela 4, a tendência foi determinada por *MA* centrada de quatro períodos como descrito na Tabela 2. Para obter os índices sazonais por trimestre, os valores originais da série foram subtraídos, em conformidade com o modelo aditivo, e divididos, segundo o modelo multiplicativo, pelo componente de tendência. Em seguida, como sumarizado na Tabela 5, os índices computados foram agrupados por trimestre de modo a originar uma única

variação sazonal (quatro trimestres no ano).

Tabela 5: Exemplo de obtenção das medidas de síntese dos índices para cada período da variação sazonal utilizando o método da relação entre *MA*

Trimestre	Índices Sazonais								Modelo Aditivo	
									Média	Índice Sazonal Médio
I	-4,33	-3,34	-2,01	-1,38	-1,80	-4,12	-1,17	-2,59	$-2,59 - (0,17 \div 4) = -2,64$	
II	3,61	7,07	4,68	4,27	4,81	6,65	5,20	5,18	$5,18 - (0,17 \div 4) = 5,14$	
III	3,99	5,35	1,66	3,85	2,14	4,42	3,41	3,55	$3,55 - (0,17 \div 4) = 3,50$	
IV	-4,45	-5,77	-6,43	-6,02	-5,53	-6,09	-7,46	-5,96	$-5,96 - (0,17 \div 4) = -6,01$	
Total									0,17	-0,01

Trimestre	Índices Sazonais								Modelo Multiplicativo	
									Mediana	Índice Sazonal Mediano
I	0,89	0,92	0,95	0,97	0,96	0,92	0,98	0,95	$0,95 \times (1 - (4,01 - 4) \div 4) = 0,95$	
II	1,09	1,17	1,11	1,09	1,10	1,14	1,11	1,11	$1,11 \times (1 - (4,01 - 4) \div 4) = 1,10$	
III	1,10	1,13	1,04	1,09	1,05	1,09	1,07	1,09	$1,09 \times (1 - (4,01 - 4) \div 4) = 1,08$	
IV	0,89	0,86	0,85	0,87	0,88	0,87	0,85	0,87	$0,87 \times (1 - (4,01 - 4) \div 4) = 0,87$	
Total									4,01	4,00

Na nona coluna da Tabela 5 são exibidas as médias, para o modelo aditivo, e as medianas, para o modelo multiplicativo, obtidas a partir de cada grupo de índices. Essas medidas expressam, de acordo com o tipo de decomposição considerada, o padrão do componente de sazonalidade. Nota-se ainda que as somas totais dessas medidas foram 0,17, para o modelo aditivo, e 4,01, para o modelo multiplicativo. Sendo assim, na última coluna da Tabela 5 são mostrados os ajustes realizados para que a soma das medidas extraídas fosse aproximadamente igual a zero, no caso aditivo, ou equivalente a ordem da sazonalidade da ST, no caso multiplicativo.

As medidas extraídas para uma variação sazonal, em conformidade com cada tipo de decomposição, foram replicadas oito vezes, conforme o número de trimestres da ST original, para construir o componente de sazonalidade. O resultado dessa replicação é mostrado graficamente na Figura 12.

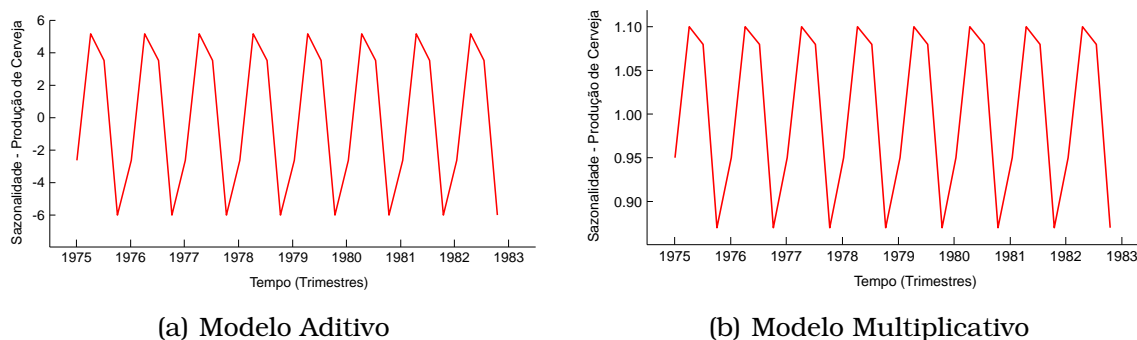


Figura 12: Sazonalidade, obtida pelo método da relação entre *MA*, da série de produção trimestral de cerveja nos EUA

Observa-se na Figura 12(a) que o componente sazonal modifica a tendên-

cia no modelo aditivo quando todos os índices médios são substancialmente maiores ou menores que zero. No modelo multiplicativo (Figura 12(b)), se os índices são diferentes de 1, pelo menos 5% acima ou abaixo em alguns dos trimestres, então os valores do componente de tendência sofrem perceptível influência da sazonalidade.

A remoção das variações sazonais (desestacionalização ou dessazonalização) pode ser realizada utilizando o modelo aditivo, subtraindo das observações reais os respectivos índices sazonais médios ($Z_t - S_t = T_t + N_t$), ou empregando o modelo multiplicativo, dividindo as observações originais pelo componente sazonal ($Z_t \div S_t = T_t \times N_t$).

Na Figura 13 é apresentada, considerando as decomposições aditiva e multiplicativa, a sazonalidade e o resultado da desestacionalização da ST de produção de chocolate.

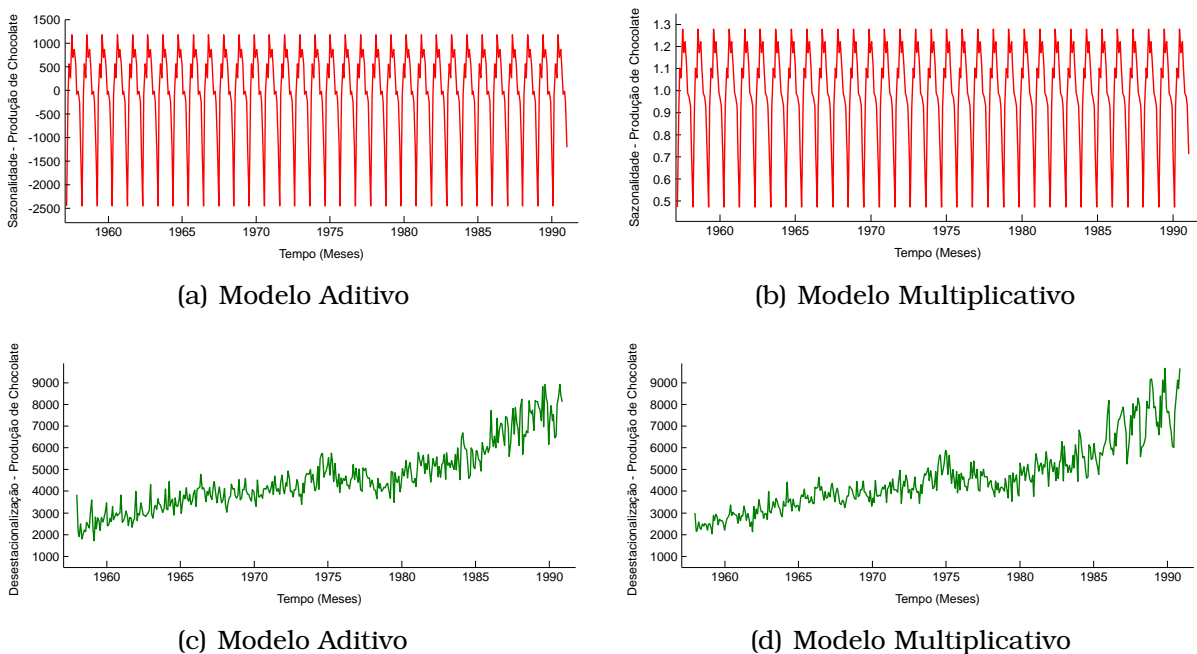


Figura 13: Sazonalidade, obtida pelo método da relação entre *MA*, e o efeito da desestacionalização da série de produção mensal de chocolate na Austrália

3.1.3 Resíduo

O resíduo é representado na ST pelos movimentos aleatórios causados por fatos fortuitos e inesperados, como catástrofes naturais, atentados terroristas, guerras, greves e decisões governamentais intempestivas. Esses fatos, que não são regulares e que também não se repetem em um padrão particular, podem comprometer o resultado de alguns estudos. Por exemplo, na estimação estatística de séries econômicas, a qualidade da predição é deteriorada pela existência de autocorrelação dos resíduos. Desse modo, a identificação

do componente residual é fundamental tanto para sua remoção quanto para a verificação de variações cíclicas que podem ocorrer no conjunto de dados categorizado como resíduo (Kirchgässner et al., 2013).

A análise de ST assume que os componentes sistemáticos, isto é tendência e sazonalidade, não são influenciados por distúrbios estocásticos e podem ser sumarizados por funções determinísticas de tempo. Sendo assim, o componente residual é o que resta depois que os componentes de tendência e sazonalidade foram estimados e removidos da série.

O ruído N de um instante de tempo t de uma ST pode ser estabelecido, em conformidade com os modelos clássicos de decomposição, pelas Equações 9 e 10. Nessas equações, Z refere-se à série e T e S indicam a tendência e a sazonalidade, respectivamente (Cowpertwait e Metcalfe, 2009).

$$N_t = Z_t - (T_t + S_t) \quad (9)$$

$$N_t = \frac{Z_t}{(T_t \times S_t)} \quad (10)$$

O impacto estocástico nas Equações 9 e 10 é restringido ao resíduo, o qual, por outro lado, é o resultado de flutuações de curto prazo que não são nem sistemáticas, nem previsíveis. Em uma série altamente irregular, essas flutuações podem dominar o comportamento intrínseco e ocultar tanto a tendência como a sazonalidade.

Na Tabela 6 é exibido o componente de tendência, estimado por MA com $r = 4$, o componente de sazonalidade, extraído conforme o método da relação entre MA, e o componente residual, resultado da aplicação das Equações 9 e 10.

O resíduo exposto na Tabela 6 é esquematizado graficamente na Figura 14. Por meio de gráficos como esse, é possível examinar eventuais quedas e altas nos valores do componente e relacionar tais comportamentos com acontecimentos ocorridos no mesmo período. Entretanto, em grande parte das situações, esses acontecimentos podem não causar efeito imediato, ou mesmo não causar efeito algum.

Na Figura 15 é novamente tratada da ST sobre chocolate apresentada na Figura 4 da página 13. Na Figura 15(a), de acordo com a decomposição aditiva, tem-se a referida série após a remoção dos componentes de tendência e de sazonalidade, resultando apenas nas variações cíclicas e irregulares. Na Figura 15(b) é ilustrado esse mesmo procedimento, mas empregando a decomposição multiplicativa.

É importante frisar que na estrutura do modelo de predição não é possível incluir o componente residual, pois ele é resultado de fatos imprevisíveis e, teoricamente, não modeláveis.

Tabela 6: Exemplo de obtenção do componente residual

Ano	Trimestre	Observação	Tendência	Modelo Aditivo		Modelo Multiplicativo	
				Sazonalidade	Resíduo	Sazonalidade	Resíduo
1975	I	36,14	—	-2,64	—	0,95	—
	II	44,60	—	5,14	—	1,10	—
	III	44,15	40,16	3,50	0,49	1,08	1,02
	IV	35,72	40,17	-6,01	1,56	0,87	1,03
1976	I	36,19	40,52	-2,64	-1,69	0,95	0,94
	II	44,63	41,02	5,14	-1,53	1,10	0,99
	III	46,95	41,60	3,50	1,85	1,08	1,04
	IV	36,90	42,67	-6,01	0,24	0,87	1,00
1977	I	39,66	43,00	-2,64	-0,70	0,95	0,97
	II	49,72	42,65	5,14	1,93	1,10	1,06
	III	44,49	42,83	3,50	-1,84	1,08	0,96
	IV	36,54	42,97	-6,01	-0,42	0,87	0,98
1978	I	41,44	43,45	-2,64	0,63	0,95	1,00
	II	49,07	44,39	5,14	-0,46	1,10	1,00
	III	48,98	45,13	3,50	0,35	1,08	1,00
	IV	39,59	45,61	-6,01	-0,01	0,87	1,00
1979	I	44,29	45,67	-2,64	1,26	0,95	1,02
	II	50,09	45,82	5,14	-0,87	1,10	0,99
	III	48,42	46,28	3,50	-1,36	1,08	0,97
	IV	41,39	46,92	-6,01	0,48	0,87	1,02
1980	I	46,11	47,91	-2,64	0,84	0,95	1,01
	II	53,44	48,63	5,14	-0,33	1,10	1,00
	III	53,00	48,58	3,50	0,92	1,08	1,01
	IV	42,52	48,61	-6,01	-0,08	0,87	1,01
1981	I	44,61	48,73	-2,64	-1,48	0,95	0,96
	II	55,18	48,53	5,14	1,51	1,10	1,03
	III	52,24	48,83	3,50	-0,09	1,08	0,99
	IV	41,66	49,12	-6,01	-1,45	0,87	0,98
1982	I	47,84	49,01	-2,64	1,47	0,95	1,03
	II	54,27	49,07	5,14	0,06	1,10	1,00
	III	52,31	—	3,50	—	1,08	—
	IV	42,03	—	-6,01	—	0,87	—

3.2 Processamento e Análise de Séries Temporais

A análise de ST consiste na aplicação de técnicas, provenientes de áreas relacionadas à Estatística e a Inteligência Artificial, que buscam descrever os componentes que se mostram como características de uma série em particular. Todavia, em grande parte dos problemas investigados, não é possível detectar pontualmente a atuação desses componentes na série, de modo que torna-se necessário isolar um componente de outro para uma melhor análise dos dados.

Em termos práticos, a análise de uma ST pode ser realizada com base em diversos objetivos, os quais podem ser divididos em quatro grupos ([Chatfield, 2013](#)):

Descrição: Esta análise visa descrever os comportamentos da ST, tais como a existência ou não de tendência, variação sazonal, observações discrepantes (*outliers*) e alterações estruturais da série, a qual inclui a existência de pontos de curva (mudança de padrão de uma tendência ou sazonalidade crescente para decrescente). Dentre as ferramentas disponíveis que fornecem apoio à análise descritiva estão os peridiogramas, os histogramas e os diagramas de dispersão;

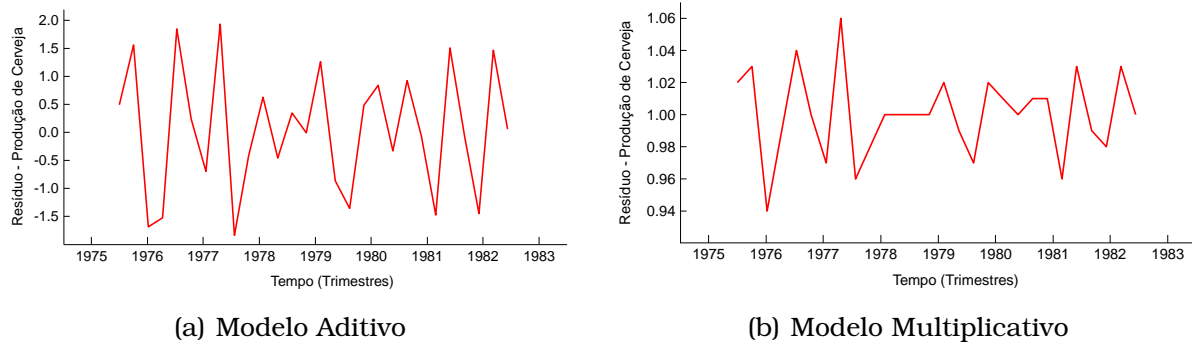


Figura 14: Resíduo da série de produção trimestral de cerveja nos EUA

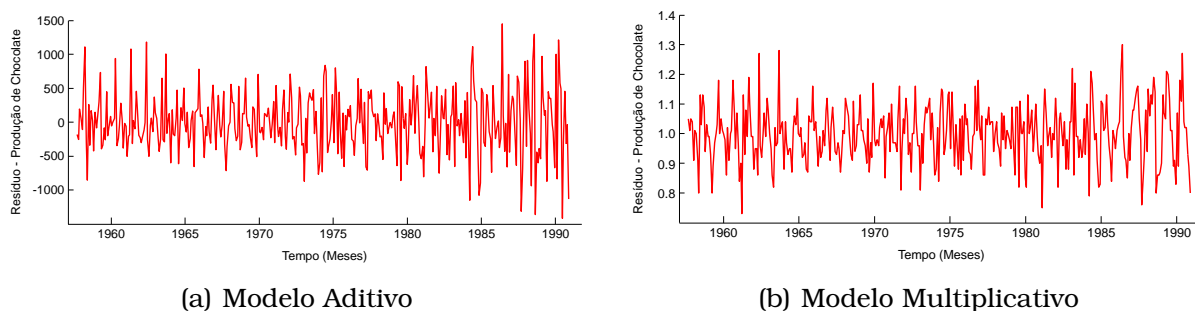


Figura 15: Resíduo da série de produção mensal de chocolate na Austrália

Explicação: Nesta análise é requerida a identificação de duas ou mais variáveis para auxiliar na determinação das paridades entre duas ST. O emprego dessa análise viabiliza a explicação da variação de uma série com base em outra;

Predição: Esta análise procura sumarizar as propriedades presentes na ST e caracterizar seu comportamento, identificando ou sugerindo um modelo que permita, a partir dos valores passados da série, a predição dos possíveis valores futuros da mesma;

Controle: Neste tipo de análise os valores de uma ST expressam dados de controle sobre um determinado processo e o objetivo incide em mensurar a qualidade desse processo.

Em diversos domínios de aplicação, tem-se observado uma ampla disponibilidade de métodos e processos que podem guiar essas análises. O processo de Mineração de Dados (MD), por exemplo, que surgiu no final da década de oitenta, focaliza a extração de conhecimento a partir de grandes volumes de dados usando computador. Entretanto, a maioria das técnicas propostas em MD considera os dados temporais como uma coleção de eventos desordenados, omitindo assim detalhes de informações atreladas à ordem e a data de aquisição dos dados (Han et al., 2011).

Impulsionada pelas limitações da MD tradicional, a área de Mineração de Dados Temporais (MDT) foi criada com a finalidade de possibilitar, a partir de eventos com dependência temporal, a extração de conhecimento que pode guiar especialistas do domínio no processo de tomada de decisão (Maimon e Rokach, 2010). Analogamente à MD, o processo de MDT é caracterizado como sendo interativo e iterativo, dividido, basicamente, em três fases, como ilustrado na Figura 16.

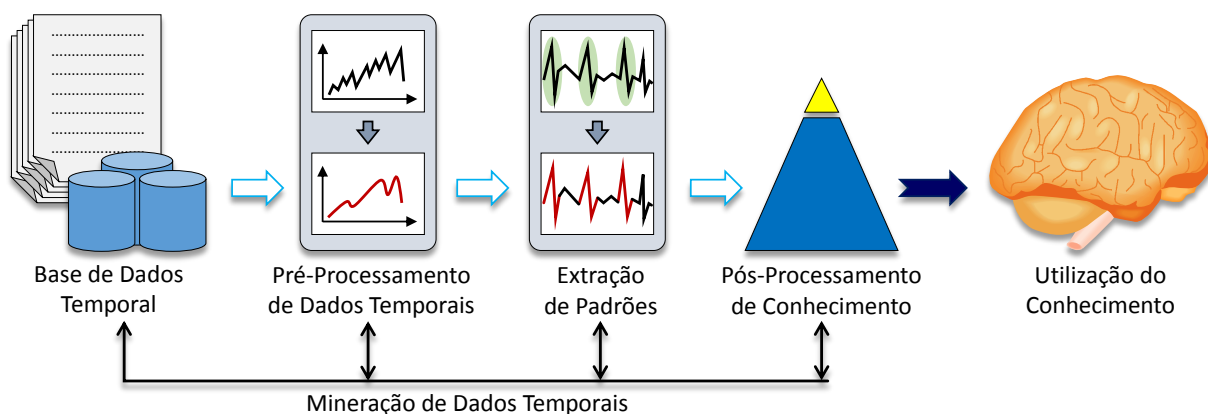


Figura 16: Processo de Mineração de Dados Temporais

A primeira fase, pré-processamento, possui essencialmente dois objetivos: conhecer o domínio da aplicação e a base de dados e prepará-los para a próxima fase. Dentre as diversas tarefas que podem ser executadas nessa fase destacam-se: a normalização de amplitude e/ou deslocamento; a normalização por escala; o tratamento de valores faltantes e a remoção de *outliers* utilizando, por exemplo, métodos de interpolação ou autorregressão; a transformação de ST com amostragem irregular para uma representação equidistante; a identificação e a remoção do componente tendência por meio de métodos de ajustes lineares; e a aplicação de técnicas para diminuição ou remoção do componente resíduo.

A segunda fase, extração de padrões, tem como objetivo a identificação, construção ou sugestão de modelos que possam extrair o conhecimento embutido nos dados. Essa fase comporta distintas tarefas como a classificação, a recuperação por conteúdo, o agrupamento, a identificação de padrões morfológicos (*motifs*), a extração de regras de associação, a detecção de anomalias e a predição de valores futuros. Comumente, essas tarefas são apoiadas por diversas áreas, entre as quais Aprendizado de Máquina, Base de Dados, Visualização e Estatística.

Na última fase, pós-processamento de conhecimento, os padrões extraídos e representados nos modelos construídos são avaliados, validados e consolidados. A avaliação é realizada, por exemplo, com a interpretação dos resultados por meio de visualização dos padrões extraídos, remoção de padrões irrelevantes.

tes ou redundantes e tradução de padrões úteis para formas compreensíveis para os usuários. Durante essa fase, os resultados devem ser avaliados para garantir que estes sejam estatisticamente significativos e confiáveis. O conhecimento extraído deve ser também validado com relação ao conhecimento prévio do domínio, com o auxílio de especialistas da área, para que possíveis conflitos sejam removidos e o conhecimento seja consolidado e disponibilizado ao usuário.

O planejamento e a execução de todas essas três fases são importantes para que o processo de MDT seja realizado com sucesso. No entanto, a fase de pré-processamento de dados, considerada como uma das mais custosas por consumir aproximadamente 80% de todo o processo (Pyle, 1999), é de fundamental importância para assegurar que os dados sejam de qualidade.

4 Construção de Modelos para Predição de Séries Temporais

A modelagem de Séries Temporais (ST) objetiva, entre outras tarefas, a construção de modelos explicativos para a projeção futura de dados históricos a curto, médio ou a longo prazo. Esses modelos podem ser catalogados em dois tipos: univariados, quando as observações do fenômeno são analisadas em função do tempo e sua predição é estabelecida a partir das propriedades inerentes às observações adquiridas no passado; e multivariados, em que outras variáveis independentes, além do tempo, são incluídas na estrutura do modelo (Reinsel, 2003). Embora os modelos multivariados não sejam tratados neste trabalho, é importante mencionar que eles possibilitam a compreensão dos fatores que influenciam o comportamento da variável de interesse.

O processo de predição de valores em ST abrange, em geral, seis etapas, as quais são esquematizadas na Figura 17.

Na primeira etapa ocorre a partição da ST em duas sequências, uma anterior ao horizonte de predição, que é destinada ao treinamento (construção e ajuste) do modelo, e outra posterior a esse período, a qual é usada para testar (avaliar) a qualidade do modelo ajustado.

Na segunda etapa, a estrutura do modelo preditivo é escolhida com base nas características dos dados e seus parâmetros são estimados empregando alguma técnica de busca. Usualmente, essa técnica é implementada por um algoritmo que recebe como entrada a sequência de treinamento, a qual é subdividida em subsequências (amostras) para treino e validação, e um conjunto de parâmetros pré-definido. A cada iteração, o algoritmo busca minimizar o erro atribuído ao desempenho do modelo ajustado pelo parâmetro (ou vetor de parâmetros) corrente.

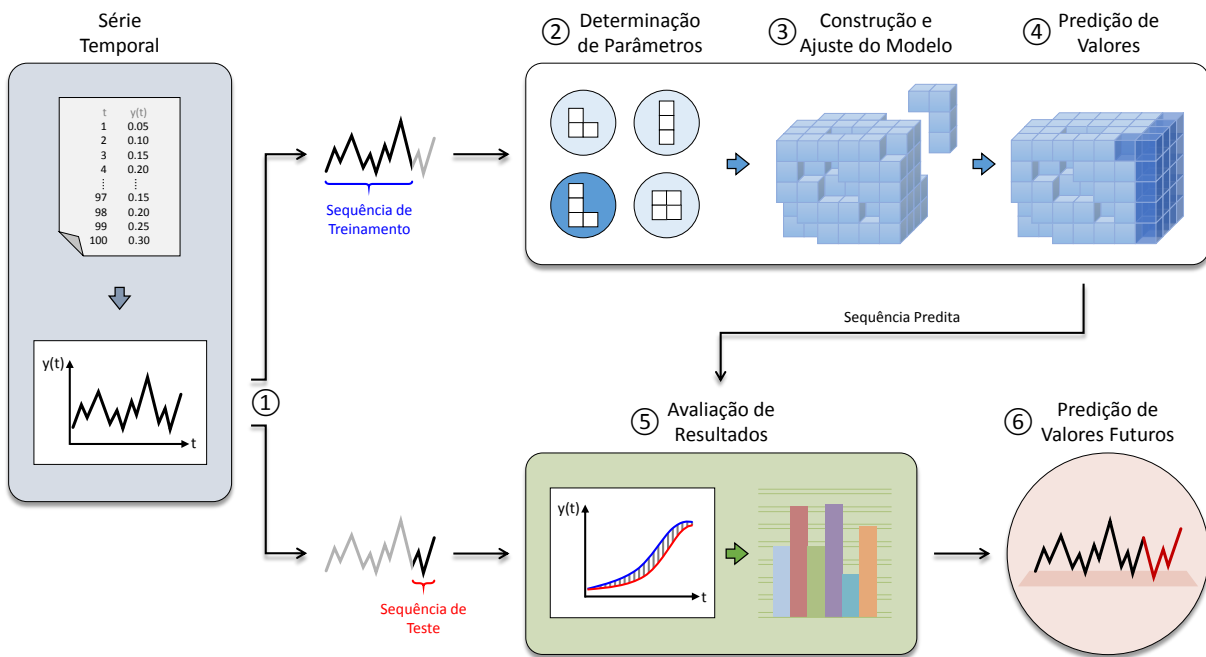


Figura 17: Processo de previsão de valores em ST

Na terceira etapa, considerando os parâmetros identificados na segunda etapa, o modelo de interesse é construído e ajustado aos dados da sequência de treinamento. Esse modelo é então extrapolado, na quarta etapa, para os períodos da sequência de teste. Evidentemente, as incertezas intrínsecas ao procedimento de estimativa de parâmetros são refletidas na previsão dos novos valores e, mais que isso, são ampliadas ao passo que a previsão se estende para períodos distantes no futuro.

Outra questão relevante atrelada à quarta etapa diz respeito à estratégia empregada para prever os valores de uma ST vários períodos à frente (horizonte de previsão $h > 1$). A estratégia mais intuitiva para essa finalidade é conhecida como recursiva (ou multi-etapa) e trata a previsão de $h > 1$ como uma aplicação, conduzida h vezes sucessivamente, do modelo preditivo considerando $h = 1$. Após a extrapolação do modelo, o valor predito ou o respectivo valor real é retroalimentado para o cálculo da previsão seguinte. Neste trabalho, quando há reutilização dos valores preditos, a referida estratégia é denominada de multi-etapa à frente com passo aproximado. Do contrário, quando ocorre retroalimentação dos valores reais, a estratégia é designada de multi-etapa à frente com passo atualizado.

Na quinta etapa, os valores preditos pelo modelo são analisados em termos de acurácia de previsão. Medir essa acurácia consiste em avaliar a extensão do erro de previsão, ou seja, mensurar o quanto os valores da sequência predita se distanciam dos valores da sequência de teste. A importância dessa análise está no fato de que distintos modelos podem ter ajustes semelhantes, mas resultarem em valores de previsão consideravelmente diferentes.

Na sexta etapa, após assegurada a qualidade do modelo, são realizadas previsões para períodos futuros da ST. Nessa etapa, o erro de previsão deve ser monitorado à medida que o banco de dados é completado com os valores reais observados. Por meio dessa monitoração é possível saber quando atualizar o modelo (agregar novos dados a sua estrutura e/ou reajustar seus parâmetros), uma vez que este pode vir a não representar a distribuição das observações recém-adquiridas.

4.1 Métodos Estatísticos

Os métodos estatísticos se distinguem por utilizar, na confecção do modelo de previsão, o conhecimento *a priori* sobre a natureza da distribuição dos dados. Esse controle probabilístico faz com que o modelo dependa explicitamente de um conjunto finito de parâmetros, os quais devem ser determinados de maneira a otimizar os resultados da previsão. Segundo [Morettin e Tolo \(2006\)](#), os modelos estatísticos podem ser divididos, conforme o seu grau de complexidade matemática, em dois grupos: modelos de suavização exponencial e modelos Autorregressivos Integrados de Médias Móveis (*ARIMA*).

Os modelos de suavização exponencial são caracterizados por decompor a ST em componentes (nível, tendência e sazonalidade) cujos valores são suavizados pela atribuição de pesos diferenciados que decaem exponencialmente com o tempo. Ao final, os componentes suavizados são recompostos, de acordo com uma estrutura aditiva ou multiplicativa, para predizer valores futuros ([Gardner, 1985](#)).

Em contraposição, os modelos da categoria *ARIMA* são conhecidos por envolver três procedimentos estatísticos ([Box et al., 2015](#)): (1) autorregressão, (2) integração e (3) Médias Móveis. No modelo *ARIMA*, a parte autorregressiva expressa a autocorrelação das observações, isto é, quanto o valor de uma observação influencia no valor da próxima. O procedimento de integração indica o número de diferenças necessário para garantir a estacionariedade da série. Por fim, a parte de Médias Móveis compreende fatores desconhecidos que não podem ser explicados pelos valores passados da ST. Uma generalização dos modelos *ARIMA*, intitulada de Autorregressivo Integrado de Médias Móveis Sazonal (*SARIMA*), foi desenvolvida para modelar séries com variações sazonais.

4.1.1 Médias Móveis

O modelo de Médias Móveis (*MA*) de ordem r , $MA(r)$, é resultado de uma técnica simples de previsão que realiza, a partir dos r últimos dados históricos da ST, uma média aritmética para predizer o próximo valor. A quantidade de

observações considerada em cada cálculo da média permanece constante, de modo a explorar a estrutura de autocorrelação dos resíduos de predição do período atual com aqueles ocorridos em períodos anteriores. O modelo de *MA* é definido pela Equação 11 (Montgomery et al., 2015), onde r representa o número de observações incluídas na média z_{t+1} .

$$z_{t+1} = \frac{z_t + z_{t-1} + z_{t-2} + \dots + z_{t-r+1}}{r} \quad (11)$$

Quanto maior for o valor do parâmetro r , mais uniforme (suavizado) será o comportamento dos dados preditos. Assim, quando a série exibir pequenas distorções em seus padrões ou flutuações aleatórias, recomenda-se usar um valor de r substancial para que a predição não fique vulnerável às mudanças de distribuição e ruídos. Do contrário, se a série for aproximadamente desprovida de aleatoriedade e apresentar alteração significativa nos pontos de inflexão das curvas, é indicado utilizar um valor de r menor para que a predição possa reagir de maneira rápida às mudanças de distribuição. Há dois aspectos importantes no que diz respeito ao valor assumido pelo parâmetro r (Morettin e Toloi, 2006):

1. Se $r = 1$, então o valor mais recente da ST será empregado como predição da próxima observação. Essa é a estratégia mais simples e ingênua para projeção de dados futuros;
2. Se $r = m$, então a predição do próximo valor será o resultado da média aritmética de todos os dados observados. A escolha de m como valor para r só é conveniente quando as mudanças de nível da ST são ocultadas pelo componente de aleatoriedade.

As desvantagens do modelo de *MA* estão atreladas a sua baixa acurácia ao lidar com ST que contenham os componentes de tendência e/ou de sazonalidade, haja vista que nesse método a predição do próximo valor envolve sempre a adição de novos dados e a desconsideração dos anteriores. Além disso, os pesos atribuídos às r observações são todos iguais e nenhum peso é dado às observações passadas. Uma alternativa que permite amenizar tal problema incide no emprego da média ponderada para construir um padrão que seja o mais próximo da realidade. Contudo, a adesão da média ponderada exige conhecimento de fundo para determinar os pesos de contribuição das observações históricas na projeção dos valores futuros (Gardner, 1985).

Na Tabela 7 é sintetizada uma aplicação do modelo de *MA*, considerando $r = 1, 3$ e 5 , sobre uma ST que retrata o número médio diário de defeitos de fabricação em caminhões nos Estados Unidos da América (EUA). As observações dessa série foram coletadas ao longo de 14 dias consecutivos no ano de 1994.

Tabela 7: Dados projetados usando o modelo de *MA* com parâmetro $r = 1, 3$ e 5

Dia (t)	Observação (z_t)	MA(1)	MA(3)	MA(5)
		Predição (z_{t+1})	Predição (z_{t+1})	Predição (z_{t+1})
1	3,44	—	—	—
2	2,83	3,44	—	—
3	1,76	2,83	—	—
4	2,00	1,76	2,68	—
5	2,09	2,00	2,20	—
6	1,89	2,09	1,95	2,42
7	1,80	1,89	1,99	2,11
8	1,25	1,80	1,93	1,91
9	1,58	1,25	1,65	1,81
10	2,25	1,58	1,54	1,72
11	2,50	2,25	1,69	1,75
12	2,05	2,50	2,11	1,88
13	1,46	2,05	2,27	1,93
14	1,54	1,46	2,00	1,97
15	—	1,54	1,68	1,96

Observa-se no exemplo da Tabela 7 que, de posse da ST composta por 14 observações (z_1, \dots, z_{14}), o valor predito de interesse é apenas o alusivo ao décimo quinto dia (z_{14+1} ou z_{15}). No entanto, valores referentes aos meses anteriores foram projetados para viabilizar a comparação com os valores históricos. Essas comparações são mostradas graficamente na Figura 18.

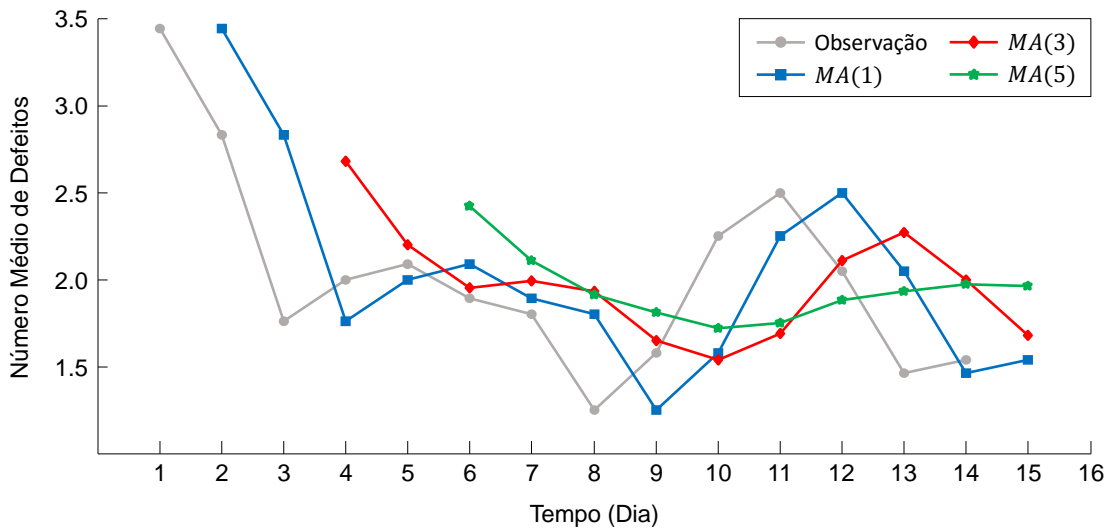


Figura 18: Predições obtidas pelo modelo de *MA* com parâmetro $r = 1, 3$ e 5

4.1.2 Suavização Exponencial Simples

O método de Suavização Exponencial Simples (*SES*) é equivalente ao de *MA* com $r = m$, exceto pelo fato de que cada valor da série recebe um peso diferente. Os pesos estipulados para os valores observados crescem exponencialmente no decorrer do tempo, de maneira que as observações mais recentes

exercem maior influência no cálculo das futuras previsões (Montgomery et al., 2015; Hyndman et al., 2002).

A estrutura do modelo de *SES* pode ser expressa conforme a Equação 12 (Gardner, 1985), em que L denota o nível no instante t , α ($0 < \alpha < 1$) simboliza o peso (ou constante de suavização) atribuído aos valores históricos da ST e z_t corresponde ao último valor observado.

$$L_t = \alpha z_t + \alpha(1 - \alpha)z_{t-1} + \alpha(1 - \alpha)^2 z_{t-2} + \dots + \alpha(1 - \alpha)^{m-1} z_1 \quad (12)$$

No intuito de evitar a utilização de todas as observações a cada nova estimativa L , a Equação 12 é reduzida em função do valor da série no instante atual e do nível computado no momento antecedente. O resultado dessa simplificação reside na relação de recorrência do algoritmo, a qual é formalizada pela Equação 13 (Gardner, 1985).

$$L_t = \alpha z_t + (1 - \alpha)L_{t-1} \quad (13)$$

Frequentemente, no início do processo de previsão via *SES*, supõe-se que o primeiro valor ajustado será igual ao primeiro valor da ST, ou seja, $L_1 = z_1$ (Morettin e Tolo, 2006). Nesse caso, o procedimento de ajuste só começa a partir da segunda observação da série e a previsão do instante $m + 1$ é dada pelo alisamento exponencial do último valor observado ($z_{m+1} = L_m$). Esse modo de previsão é designado de um passo à frente e a sua extensão para horizontes maiores que um não é suportada pelo método. Quando múltiplos horizontes são considerados, a previsão de todos os valores futuros é dada pelo valor ajustado L_m .

O desempenho preditivo do modelo de *SES* depende do valor escolhido para a constante de suavização α . Essa seleção pode ser conduzida de maneira objetiva, aplicando alguma técnica para estimação de parâmetros, ou subjetiva, estabelecendo uma regra de escolha (Morettin e Tolo, 2006):

- Quando o interesse consistir na obtenção do componente de tendência, sugere-se que o valor de α seja próximo de zero para amortizar o efeito da sazonalidade e do resíduo. É importante mencionar que valores pequenos de α produzem projeções com maior influência das observações antigas da ST;
- Quando houver interesse na tarefa de previsão, recomenda-se que α seja próximo do valor 1. Quanto mais próximo a constante de suavização estiver desse valor, maior será o impacto das observações recentes na projeção do valor futuro.

Conforme relatado em Montgomery et al. (2015), a execução do algoritmo

de *SES* adotando $\alpha = 2 \div (r - 1)$ fornece resultados semelhantes aos obtidos com o modelo $MA(r)$.

A popularidade conferida ao método de *SES* pode ser explicada pela sua flexibilidade, simplicidade matemática e razoável precisão. Para efetuar uma nova predição, o algoritmo necessita apenas da observação mais recente, do último valor predito e do parâmetro α . Dentre as desvantagens do método, sobressai-se a dificuldade em encontrar o valor mais adequado para a constante de suavização (Hyndman et al., 2002).

Na Tabela 8 é exibida uma execução do modelo de *SES*, supondo $\alpha = 0,1, 0,3$ e $0,9$, sobre as 14 observações da ST relativa ao número médio diário de defeitos de fabricação em caminhões nos EUA.

Tabela 8: Dados projetados empregando o modelo de *SES* com parâmetro $\alpha = 0,1, 0,3$ e $0,9$

Dia (t)	Observação (z_t)	SES(0,1)		SES(0,3)		SES(0,9)	
		L_t	Predição (z_{t+1})	L_t	Predição (z_{t+1})	L_t	Predição (z_{t+1})
1	3,44	3,44	—	3,44	—	3,44	—
2	2,83	3,38	—	3,26	—	2,89	—
3	1,76	3,22	3,38	2,81	3,26	1,87	2,89
4	2,00	3,10	3,22	2,57	2,81	1,99	1,87
5	2,09	2,99	3,1	2,42	2,57	2,08	1,99
6	1,89	2,88	2,99	2,26	2,42	1,91	2,08
7	1,80	2,78	2,88	2,12	2,26	1,81	1,91
8	1,25	2,62	2,78	1,86	2,12	1,31	1,81
9	1,58	2,52	2,62	1,78	1,86	1,55	1,31
10	2,25	2,49	2,52	1,92	1,78	2,18	1,55
11	2,50	2,49	2,49	2,09	1,92	2,47	2,18
12	2,05	2,45	2,49	2,08	2,09	2,09	2,47
13	1,46	2,35	2,45	1,89	2,08	1,52	2,09
14	1,54	2,27	2,35	1,79	1,89	1,54	1,52
15	—	—	2,27	—	1,79	—	1,54

As predições do modelo de *SES*, dispostas na Tabela 8 em conformidade com os valores atribuídos à constante de suavização α , são esquematizadas na Figura 19.

4.1.3 Suavização Exponencial de Holt

O modelo de *SES* quando aplicado sobre dados temporais que apresentem comportamento linear crescente (ou decrescente), fornece predições que subestimam (ou superestimam) os valores reais. A fim de esquivar-se desse erro sistemático, outros métodos, por exemplo o de Suavização Exponencial de Holt (HES) (Holt, 2004; Gardner, 1985), podem ser usados.

A estrutura do modelo de HES, estabelecida pelas Equações 14 e 15 (Holt, 2004), é similar, em princípio, a do modelo de *SES*. No entanto, além de usufruir do parâmetro α para suavizar o componente de nível, o algoritmo utiliza uma segunda constante de suavização (β) para modelar a tendência da ST.

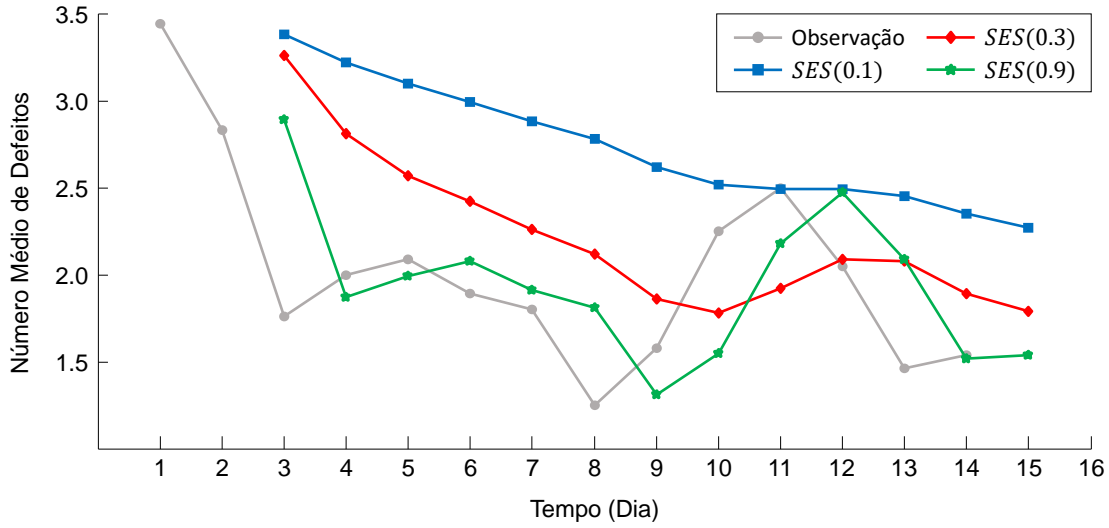


Figura 19: Predições computadas pelo modelo de SES com parâmetro $\alpha = 0,1, 0,3$ e $0,9$

$$L_t = \alpha z_t + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (14)$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \quad (15)$$

$$z_{t+h} = L_t + hT_t \quad (16)$$

Os valores das constantes de suavização α e β encontram-se no intervalo de $[0, 1]$ e os componentes de nível e tendência são estimados pelas Equações 14 e 15, respectivamente. Essas equações, assim como em qualquer método de suavização exponencial, modificam estimativas prévias quando uma nova observação é calculada. Complementarmente, na Equação 16, z_{t+h} indica a predição do valor z para o instante $t + h$, onde h representa o horizonte de predição.

Para que a relação de recorrência do algoritmo de HES possa ser implementada, necessita-se pressupor seus valores iniciais. Uma regra amplamente aceita na literatura é assumir $L_1 = z_1$ e $T_1 = z_2 - z_1$ (Morettin e Toloi, 2006). Como o método baseia-se no conceito de autoaprendizagem, isto é, ele aprende com seus próprios erros, os valores iniciais não prejudicam as projeções. Porém, esse fato não se aplica às constantes de suavização, que são de difícil especificação e, dependendo dos valores elegidos, podem degradar o desempenho preditivo do algoritmo.

Na Tabela 9 é apresentada uma aplicação do modelo de HES, considerando $\alpha = 0,3$ e $\beta = 0,9$, sobre a série que reflete o número médio diário de defeitos de fabricação em caminhões nos EUA. Os valores preditos por esse modelo são ilustrados na Figura 20.

Tabela 9: Dados projetados usando o modelo de HES com parâmetros $\alpha = 0,3$ e $\beta = 0,9$

Dia (t)	Observação (z_t)	L_t	T_t	Predição (z_{t+1})
1	3,44	3,44	-0,61	—
2	2,83	2,83	-0,61	—
3	1,76	2,08	-0,73	2,22
4	2,00	1,54	-0,56	1,35
5	2,09	1,32	-0,26	0,99
6	1,89	1,31	-0,03	1,06
7	1,80	1,43	0,11	1,27
8	1,25	1,45	0,03	1,54
9	1,58	1,51	0,06	1,48
10	2,25	1,77	0,24	1,57
11	2,50	2,16	0,37	2,01
12	2,05	2,39	0,24	2,53
13	1,46	2,28	-0,07	2,63
14	1,54	2,01	-0,25	2,21
15	—	—	—	1,75

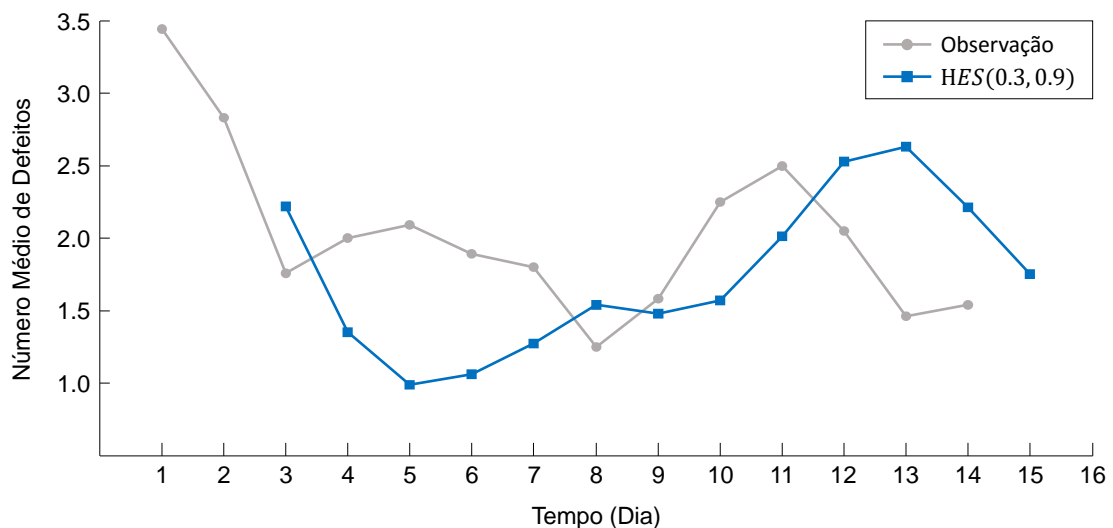


Figura 20: Predições obtidas pelo modelo de HES com parâmetros $\alpha = 0,3$ e $\beta = 0,9$

4.1.4 Suavização Exponencial Sazonal de Holt-Winters

Métodos Holt-Winters (HW) são recomendados para ST que contemplam tanto o componente de tendência quanto o de sazonalidade. Uma série com essas características é descrita pela ocorrência de padrões cíclicos de variação que se repetem em intervalos relativamente constantes de tempo (Montgomery et al., 2015).

A estrutura dos modelos de HW abrange três equações com constantes de suavização distintas, as quais são vinculadas aos componentes básicos da ST. À vista dessa estrutura, os modelos são divididos em dois grupos: multiplicativo e aditivo (Winters, 1960). A preferência pela escolha de uma estrutura em detrimento da outra está relacionada com o tipo de padrão sazonal da série investigada (Figura 21).

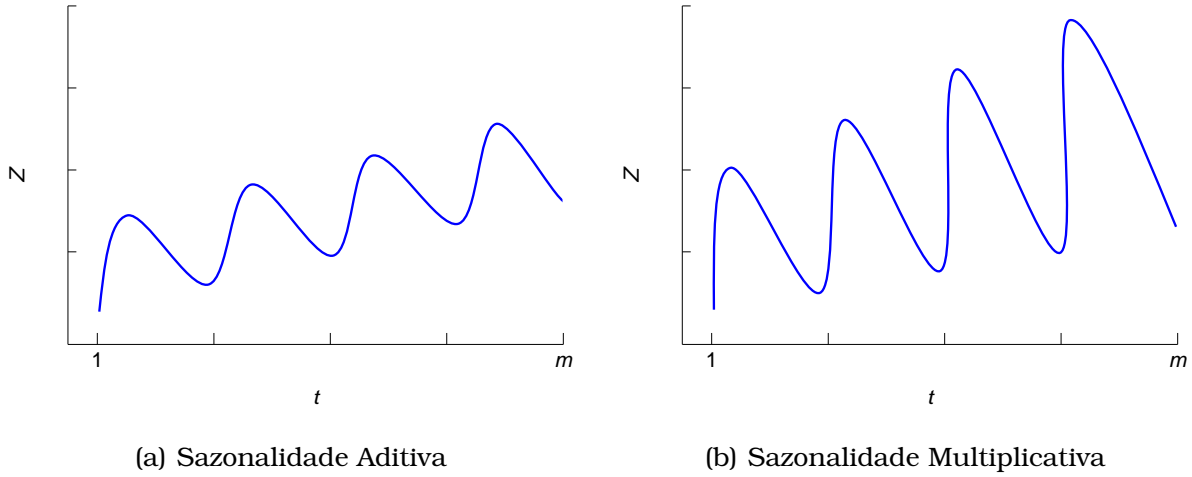


Figura 21: Representação dos tipos de variação sazonal

O modelo de HW Multiplicativo (MHW) é usado para ajustar ST que possuem tendência e sazonalidade multiplicativa, ou seja, aquelas em que a amplitude da variação sazonal aumenta com o acréscimo do nível médio da série. O algoritmo de predição que implementa esse modelo emprega as seguintes equações (Winters, 1960):

$$L_t = \alpha \frac{z_t}{S_{t-s}} + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (17)$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \quad (18)$$

$$S_t = \gamma \frac{z_t}{L_t} + (1 - \gamma)S_{t-s} \quad (19)$$

$$z_{t+h} = (L_t + hT_t)S_{t-s+h} \quad (20)$$

Nessas equações, α , β e γ são constantes de suavização, cujos valores situam-se no intervalo $[0, 1]$, s indica a quantidade de observações que compõe um período sazonal e z_{t+h} representa a predição do valor z para o período $t + h$. É importante ressaltar que a escolha dos valores para as constantes de suavização é condicionada a algum critério, por exemplo, a minimização do erro quadrático médio atribuído ao desempenho do modelo (Morettin e Tolo, 2006).

Analogamente aos métodos de SES e de HES, o funcionamento do algoritmo MHW ocorre pela aplicação recursiva das três equações sobre os dados da ST. Tal aplicação deve ser iniciada em algum período no passado, onde os valores de L , T e S precisam ser previamente estimados. Uma maneira simples de realizar essa estimativa é por meio da inicialização do nível e da tendência no mesmo período s . Desse modo, o nível pode ser determinado a partir da média da primeira estação (Equação 21).

$$L_s = \frac{1}{s}(z_1 + z_2 + \dots + z_s) \quad (21)$$

A tendência pode ser inicializada utilizando duas estações completas, como definido pela Equação 22.

$$T_s = \frac{1}{s} \left(\frac{z_{s+1} - z_1}{s} + \frac{z_{s+2} - z_2}{s} + \dots + \frac{z_{s+s} - z_s}{s} \right) \quad (22)$$

Os índices sazonais iniciais podem ser determinados pela razão entre as primeiras observações e a média do primeiro período, como mostrado pela Equação 23.

$$S_1 = \frac{z_1}{L_s}, S_2 = \frac{z_2}{L_s}, \dots, S_s = \frac{z_s}{L_s} \quad (23)$$

O modelo de HW Aditivo (AHW) tem maior capacidade de explicação em séries com tendência e sazonalidade aditiva, onde a diferença entre o maior e o menor valor de demanda dentro das estações permanece constante no tempo. O algoritmo de predição que implementa esse modelo usufrui das seguintes equações:

$$L_t = \alpha(z_t - S_{t-s}) + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (24)$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \quad (25)$$

$$S_t = \gamma(z_t - L_t) + (1 - \gamma)S_{t-s} \quad (26)$$

$$z_{t+h} = L_t + hT_t + S_{t-s+h} \quad (27)$$

Observa-se que a Equação 18 do modelo multiplicativo de HW é idêntica à Equação 25 do modelo aditivo. A diferença está no uso das outras equações, nas quais os índices de sazonalidade são somados e subtraídos, ao invés de multiplicados e divididos como no modelo multiplicativo.

As inicializações de L e T podem ser realizadas aplicando as mesmas equações do modelo multiplicativo. Entretanto, aconselha-se que os valores iniciais dos índices sazonais sejam calculados conforme a Equação 28.

$$S_1 = z_1 - L_s, S_2 = z_2 - L_s, \dots, S_s = z_s - L_s \quad (28)$$

A correta aplicação dos modelos de HW está associada à morfologia das variações sazonais na ST, independentemente da existência do componente de tendência. Nesses modelos, quando $\gamma = 0$ não significa que a série é desprovida de sazonalidade, mas sim que os índices sazonais foram inicializados com valores que não precisam ser corrigidos ao longo da predição.

Na Tabela 10, utilizando a ST que expressa o número médio diário de defeitos de fabricação em caminhões nos EUA, é resumida uma execução dos modelos aditivo e multiplicativo de HW. Foram adotados, para ambos os modelos, $\alpha = 0,3$, $\beta = 0,5$, $\gamma = 0,7$ e $s = 4$.

A fim de oportunizar comparações, as predições elencadas na Tabela 10

Tabela 10: Dados projetados empregando os modelos aditivo e multiplicativo de HW, ambos com parâmetros $\alpha = 0,3$, $\beta = 0,5$, $\gamma = 0,7$ e $s = 4$

Dia (t)	Observação (z_t)	Modelo Aditivo				Modelo Multiplicativo			
		L_t	T_t	S_t	Predição (z_{t+1})	L_t	T_t	S_t	Predição (z_{t+1})
1	3,44	—	—	0,93	—	—	—	1,37	—
2	2,83	—	—	0,32	—	—	—	1,13	—
3	1,76	—	—	-0,75	—	—	—	0,70	—
4	2,00	2,51	-0,19	-0,51	—	2,51	-0,19	0,80	—
5	2,09	1,97	-0,36	0,36	—	2,08	-0,31	1,11	—
6	1,89	1,60	-0,37	0,30	—	1,74	-0,32	1,10	—
7	1,80	1,62	-0,17	-0,10	—	1,76	-0,15	0,92	—
8	1,25	1,55	-0,12	-0,36	—	1,60	-0,16	0,79	—
9	1,58	1,36	-0,16	0,26	1,78	1,44	-0,16	1,11	1,61
10	2,25	1,43	-0,04	0,67	1,51	1,51	-0,04	1,37	1,40
11	2,50	1,75	0,14	0,50	1,28	1,83	0,14	1,23	1,35
12	2,05	2,04	0,22	-0,10	1,53	2,17	0,24	0,90	1,55
13	1,46	1,94	0,06	-0,26	2,52	2,08	0,07	0,82	2,65
14	1,54	1,66	-0,11	0,11	2,67	1,84	-0,08	1,00	2,96
15	—	—	—	—	2,05	—	—	—	2,17

são mostrados graficamente na Figura 22.

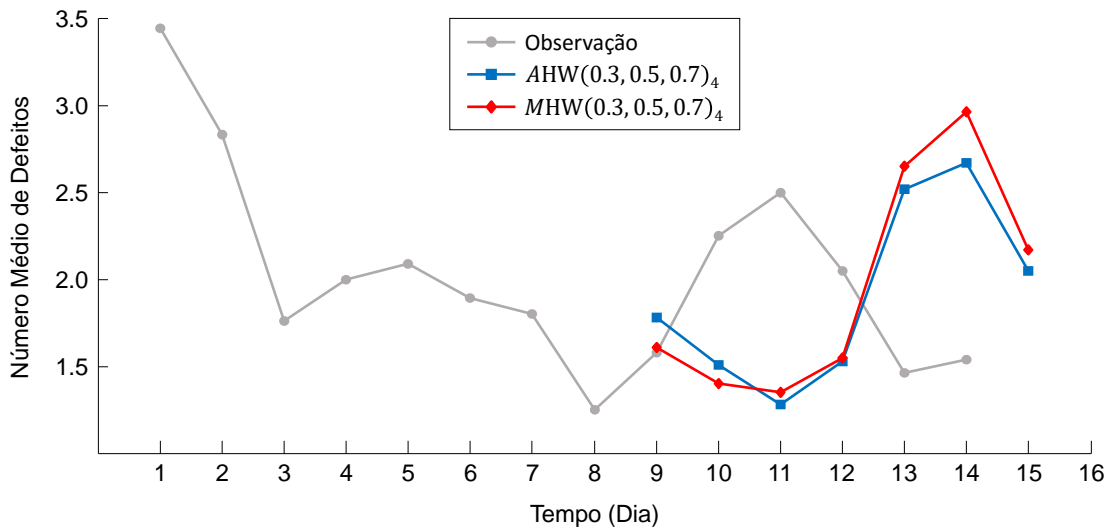


Figura 22: Predições computadas pelos modelos aditivo e multiplicativo de HW, ambos com parâmetros $\alpha = 0,3$, $\beta = 0,5$, $\gamma = 0,7$ e $s = 4$

4.1.5 Modelos ARIMA e SARIMA

Os modelos ARIMA de ordem (p, d, q) , isto é, $ARIMA(p, d, q)$, resultam da combinação de três procedimentos estatísticos (Box et al., 2015): (1) autorregressão ($AR(p)$), (2) integração¹⁸ e (3) Médias Móveis ($MA(q)$). O uso simultâneo desses três componentes não constitui uma regra para modelar ST com ausência de padrões sazonais, uma vez que eles podem ser executados de maneira conjugada, ou seja, um complementando o outro. Por essa perspectiva

¹⁸Integração é o nome dado à operação de diferenciação, a qual consiste em tomar diferenças sucessivas da série original $Z = (z_1, z_2, \dots, z_m)$. A primeira diferença é denotada por $\Delta z_t = z_t - z_{t-1}$; a segunda diferença é definida como $\Delta^2 z_t = \Delta(\Delta z_t) = \Delta(z_t - z_{t-1})$; por fim, a d -ésima diferença equivale à $\Delta^d z_t = \Delta(\Delta^{d-1} z_t)$.

e como o procedimento de integração pode ser realizado em um passo de pré-processamento, a nomenclatura *ARIMA* também é utilizada para referir-se às seguintes estruturas (Montgomery et al., 2015):

- $ARIMA(p, 0, 0) = AR(p)$;
- $ARIMA(0, 0, q) = MA(q)$;
- $ARIMA(p, 0, q) = ARMA(p, q)$ ¹⁹.

No modelo Autorregressivo de ordem p , $AR(p)$ (Equação 29), o valor atual da série é expresso como um agregado linear de valores históricos e um fator inovação (sequência de valores aleatórios). Esse fator, chamado por alguns autores de ruído branco, representa qualquer particularidade desconhecida que não pode ser explicada pelos valores passados da ST.

$$z_t = \delta + \sum_{i=1}^p \phi_i z_{t-i} + e_t \quad (29)$$

Na Equação 29, z_t corresponde à observação da série no momento t ; p denota o número de observações consideradas; ϕ_i indica o i -ésimo coeficiente autorregressivo; δ reflete o nível inicial do modelo (exerce a mesma função que o intercepto em regressão linear) e é calculado conforme a Equação 30, em que μ representa a média do processo estacionário; e e_t compreende o ruído branco em uma distribuição com média zero e variância constante σ_e^2 . Ainda, para cada instante t do tempo, assume-se que e_t é independente dos valores passados da ST ($z_{t-1}, z_{t-2}, \dots, z_{t-m+1}$).

$$\delta = \mu(1 - \phi_1 - \phi_2 - \dots - \phi_p) \quad (30)$$

Similarmente ao AR , o modelo de Médias Móveis com ordem q , $MA(q)$, estabelece o valor atual da série por meio de uma combinação linear dos fatores de inovação do período corrente com aqueles ocorridos em períodos anteriores. A estrutura desse modelo é descrita conforme a Equação 31, onde θ_i é o i -ésimo valor de ponderação usado e e_t corresponde ao ruído branco no instante t .

$$z_t = \mu + \sum_{i=1}^q \theta_i e_{t-i} + e_t \quad (31)$$

Observa-se na Equação 31 que os pesos $\theta_1, \theta_2, \dots, \theta_q$, além de serem aplicados sobre os fatores de inovação $e_{t-1}, e_{t-2}, \dots, e_{t-q}$, podem assumir valores positivos ou negativos. Quando se deseja prever o valor z_{t+1} , move-se os pesos utilizados, aplicando-os nos ruídos $e_{t+1}, e_t, e_{t-1}, \dots, e_{t-q+1}$.

¹⁹Autorregressivo de Médias Móveis (*ARMA*).

A fim de aumentar a flexibilidade na modelagem de ST, as Equações 29 e 31 são combinadas para obter o modelo Autorregressivo de Médias Móveis com ordem (p, q) , $ARMA(p, q)$ (Equação 32), que incorpora modelos mistos $AR(p)$ e $MA(q)$ (Shumway e Stoffer, 2011). Em termos práticos, o modelo $ARMA$ procura estimar um valor considerando que a série é descrita em parte por um processo autorregressivo, e em parte por um processo envolvendo Médias Móveis.

$$z_t = \delta + \sum_{i=1}^p \phi_i z_{t-i} + \sum_{i=1}^q \theta_i e_{t-i} + e_t \quad (32)$$

O principal benefício do $ARMA$ é que, para ajustar sua estrutura à ST estacionárias complexas, ele frequentemente faz uso de uma quantidade de termos menor que a exigida pelos modelos puramente AR ou puramente MA (Chatfield, 2013).

A seleção de modelos AR , MA e $ARMA$ é apropriada quando a série em estudo é estacionária, ou seja, suas propriedades estatísticas básicas, como média, variância e covariância, permanecem constantes ao longo do tempo. Entretanto, quando a ST não é estacionária, ela pode ser transformada empregando um procedimento de diferenciação dos dados que assegura a propriedade de estacionariedade. O acréscimo desse procedimento de integração na estrutura do $ARMA$, resulta no modelo Autorregressivo Integrado de Médias Móveis com ordem (p, d, q) , $ARIMA(p, d, q)$, determinado pela Equação 33 (Montgomery et al., 2015).

$$I'_t = \delta + \sum_{i=1}^p \phi_i I'_{t-i} + \sum_{i=1}^q \theta_i e_{t-i} + e_t \quad (33)$$

Na Equação 33, $I'_t = \Delta^d z_t = \Delta(\Delta^{d-1} z_t)$ e d indica o grau do operador de diferença; ϕ_p e θ_q são, nessa ordem, os parâmetros dos procedimentos autorregressivo, com comprimento de defasagem p , e de Médias Móveis, com comprimento de defasagem q ; e e_t corresponde ao fator de inovação que não pode ser explicado pelo modelo. Em resumo, para a utilização do $ARIMA$, supõe-se que a d -ésima diferença entre as observações da série pode ser representada por um processo estacionário capaz de ser estimado por um modelo $ARMA$. Sendo assim, ST que apresentam tendência não-explosiva, isto é, não-estacionariedade homogênea, bem como séries estacionárias podem ser modeladas pelo $ARIMA$.

A constante δ pode ser omitida no modelo $ARIMA$ quando a ST em estudo é não-estacionária por natureza e, conseqüentemente, foi preciso diferenciá-la para obter estacionariedade ($d > 1$). Se a série for estacionária em sua forma original ($d = 0$), mas não com média zero e desvio padrão unitário, a constante é necessária. Adicionalmente, quando o modelo é desprovido da

parte autorregressiva ($AR(p)$), assume-se que a constante é igual a média dos valores da ST ($\delta = \mu$) (Cowpertwait e Metcalfe, 2009).

O *ARIMA* e suas variações exploram a autocorrelação entre os valores da série em instantes sucessivos, porém quando os dados são observados em períodos inferiores a um ano, a ST também pode apresentar autocorrelação para uma estação de sazonalidade s . Nesse contexto, o modelo Autorregressivo Integrado de Médias Móveis Sazonal (*SARIMA*), por ser uma generalização do *ARIMA*, contém em sua estrutura uma parte não-sazonal (Equação 33), com parâmetros (p, d, q) , e uma sazonal (Equação 34), com parâmetros $(P, D, Q)_s$.

$$I_t'' = \delta + \sum_{i=1}^P \Phi_{is} I_{t-is}'' + \sum_{i=1}^Q \Theta_{is} e_{t-is} + e_t \quad (34)$$

Na Equação 34, $I_t'' = \Delta^D z_t = \Delta(\Delta^{D-1} z_t)$ e D indica o grau do operador de diferença sazonal; a constante δ é computada segundo a Equação 35 e seu uso segue as mesmas regras que as impostas para a estrutura *ARIMA*, mas agora considerando D ; Φ_P e Θ_Q são, nessa ordem, os parâmetros dos procedimentos autorregressivo sazonal, com comprimento de defasagem P , e de Médias Móveis sazonal, com comprimento de defasagem Q ; e e_t corresponde ao fator de inovação que não pode ser explicado pelo modelo.

$$\delta = \mu(1 - \Phi_1 - \Phi_2 - \dots - \Phi_p) \quad (35)$$

O *SARIMA* $(p, d, q) \times (P, D, Q)_s$ é formalizado pela Equação 36, onde somam-se as partes não-sazonal e sazonal.

$$I_t = \delta + \sum_{i=1}^p \phi_i I_{t-i}' + \sum_{i=1}^q \theta_i e_{t-i} + \sum_{i=1}^P \Phi_{is} I_{t-is}'' + \sum_{i=1}^Q \Theta_{is} e_{t-is} + e_t \quad (36)$$

No modelo *SARIMA*, a constante δ é calculada por meio da Equação 37 e ela pode ser omitida quando $d + D > 1$. Caso contrário, isto é, se $d + D \leq 1$, a utilização da constante é exigida. Obviamente, quando o modelo é desprovido das partes autorregressiva e autorregressiva sazonal, assume-se que a constante é igual a média dos valores da série ($\delta = \mu$) (Cowpertwait e Metcalfe, 2009).

$$\delta = \mu(1 - \phi_1 - \phi_2 - \dots - \phi_p)(1 - \Phi_1 - \Phi_2 - \dots - \Phi_p) \quad (37)$$

A aplicação do *SARIMA* é apropriada, sobretudo, em cenários nos quais os dados possuem variações sazonais que não são adequadamente tratadas pela primeira diferença ($\Delta z_t = z_t - z_{t-1}$). Um exemplo comum dessa característica compreende ST reais que retratam dados mensais. Nessas séries, uma dependência entre as observações z_t e z_{t-12} é passível de ser encontrada.

Cada valor predito pelos modelos estatísticos supracitados precisará ser

mapeado para o espaço de valores reais quando a ST original sofrer integração ($d > 0$ e/ou $D > 0$). Tal transformação é alcançada a partir do emprego da operação inversa das diferenciações tomadas (Cowpertwait e Metcalfe, 2009).

Segundo Morettin e Tolo (2006), a concepção de um modelo ARIMA ou SARIMA é respaldada no ciclo iterativo de Box-Jenkins (Box et al., 2015). Esse método, cujas quatro etapas são detalhadas a seguir e esquematizadas na Figura 23, viabiliza a identificação do processo estocástico gerador dos dados e dos seus parâmetros de ajuste.

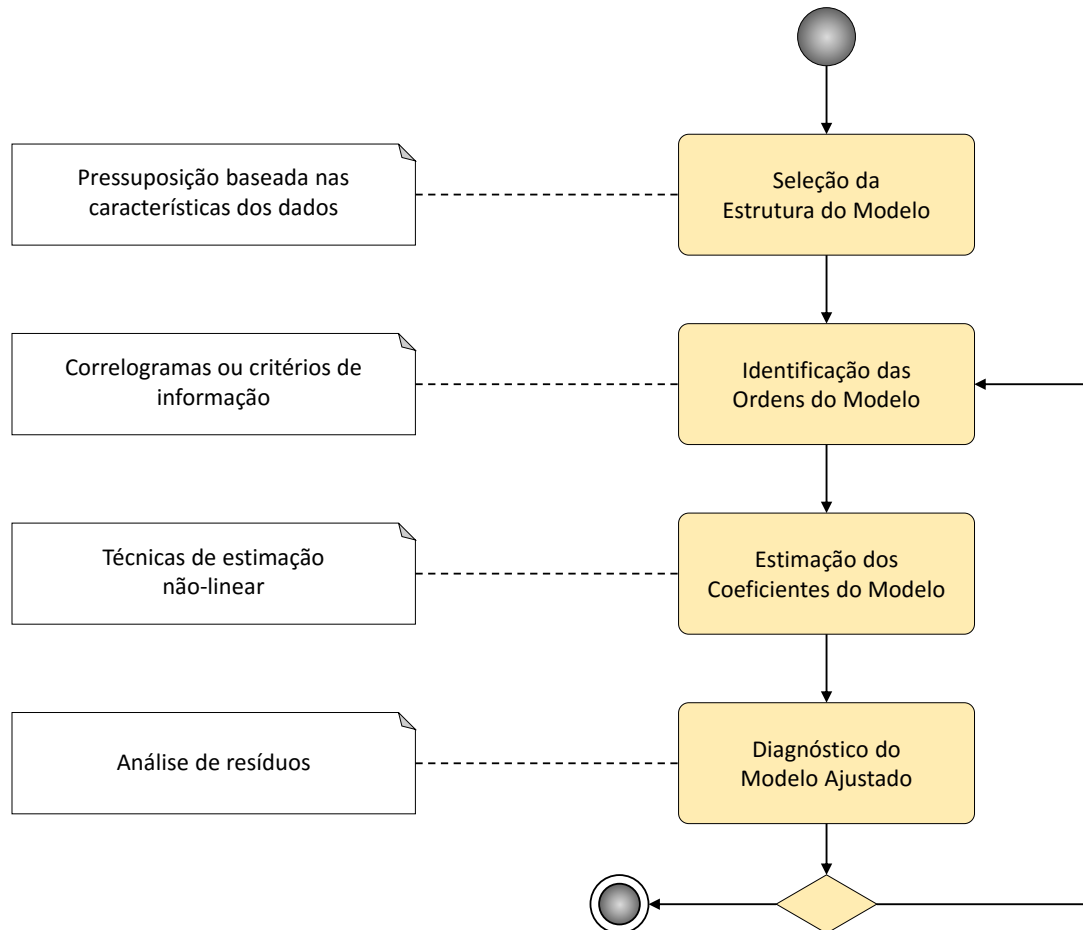


Figura 23: Diagrama de atividades para o fluxo de construção de um modelo ARIMA ou SARIMA (modificado de Box et al. (2015))

- 1. Seleção da Estrutura do Modelo:** A estrutura do modelo é escolhida com base nas características dos dados. Desse modo, quando a série contemplar o componente de tendência, recomenda-se o uso da estrutura ARIMA. Por outro lado, se a série apresentar tanto o componente de tendência quanto o de sazonalidade, sugere-se a adoção da estrutura SARIMA;
- 2. Identificação das Ordens do Modelo:** Os valores de p , d e q , do $ARIMA(p, d, q)$, ou os valores de p , d , q , P , D e Q , do $SARIMA(p, d, q) \times$

$(P, D, Q)_s$, são estabelecidos utilizando-se de correlogramas ou de critérios de informação. Inicialmente, o número de integrações (d ou D) pode ser contabilizado por meio de um mecanismo iterativo, no qual a sequência de dados é diferenciada tantas vezes quantas necessárias até que sua variância seja menor que a variância computada para a sua versão original (sem diferenciação). Posteriormente, inspecionam-se as funções de autocorrelação e autocorrelação parcial amostrais da ST adequadamente diferenciada nas defasagens $1, 2, 3, \dots$ para obter o valor de p e q e nas defasagens $s, s \times 2, s \times 3, \dots$ para conseguir o valor de P e Q . Uma maneira alternativa de encontrar os valores desses parâmetros consiste na aplicação de critérios de informação. Nesse sentido, o Critério de Informação de Akaike (AIC), expresso pela Equação 38 (Cowpertwait e Metcalfe, 2009), penaliza a qualidade do ajuste de modelos com muitos parâmetros.

$$AIC = -2 \times LL + (\log(n) + 1) \times NP \quad (38)$$

Na Equação 38, LL refere-se ao logaritmo da função de verossimilhança, n compreende o número de observações da série de treinamento e NP indica a quantidade de parâmetros tida como necessária. A ideia é selecionar o modelo cujo AIC calculado seja mínimo. Obviamente, um modelo com mais parâmetros pode ter um ajuste melhor, mas não necessariamente será preferível em termos de AIC ;

- 3. Estimação dos Coeficientes do Modelo:** As estimativas preliminares dos coeficientes ϕ_p e Φ_P , do componente autorregressivo, e dos coeficientes θ_q e Θ_Q , do componente de Médias Móveis, podem ser estabelecidas usando as autocorrelações da ST integrada na etapa de identificação do modelo. Após a atribuição desses valores iniciais, os coeficientes são estimados maximizando-se a função de verossimilhança (Morettin e Tolloi, 2006). Como os estimadores de máxima verossimilhança podem ser aproximados por estimadores de mínimos quadrados, a referida função é normalmente maximizada por meio de mínimos quadrados não-lineares utilizando o algoritmo de Levenberg-Marquardt (Moré, 1978);
- 4. Diagnóstico do Modelo Ajustado:** O modelo apontado como o mais promissor é examinado para assegurar que a dinâmica dos dados foi representada satisfatoriamente. Na prática, as estimativas dos erros (resíduos) são analisadas pela aplicação de testes de autocorrelação residual. Esses testes têm por objetivo verificar se os resíduos apresentam comportamento de ruído branco, isto é, se suas autocorrelações comportam-se de forma aleatória e não são significativas. Em caso afirmativo, o modelo pode ser extrapolado para momentos futuros. Em caso negativo, será

necessário selecionar outro modelo e repetir as etapas de identificação, estimação e diagnóstico.

Nota-se que o emprego de modelos da categoria *ARIMA* exige conhecimento especializado tanto no domínio de aplicação quanto em matemática computacional. Além disso, a percepção e a experiência do analista são fundamentais para que o processo de modelagem se torne mais prático e menos dispendioso.

4.2 Métodos de Aprendizado de Máquina

Os métodos de Aprendizado de Máquina (AM) para predição, em oposição aos modelos estatísticos, buscam descrever as propriedades dos dados sem o conhecimento prévio da distribuição dos mesmos. Por não dependerem explicitamente de parâmetros para modelar o comportamento do fenômeno, esses métodos são mais simples de serem ajustados e demonstram considerável desempenho mesmo quando aplicados à séries complexas e altamente não-lineares. De acordo com [Islam e Sivakumar \(2002\)](#), pelo modo como as observações da ST são aproveitadas no modelo preditivo, os métodos de AM podem ser divididos em duas abordagens: global e local.

A aproximação global é uma abordagem na qual os métodos de AM constroem modelos a partir de um procedimento de treinamento que recebe como entrada todas as observações da série. Normalmente, o uso dessa abordagem envolve a transposição da sequência de dados para uma tabela atributo-valor, de maneira que seja possível fornecê-la como entrada para os algoritmos convencionais de AM destinados à tarefa de regressão. Na Figura 24 é ilustrado o processo de predição de ST com o emprego dessa abordagem.

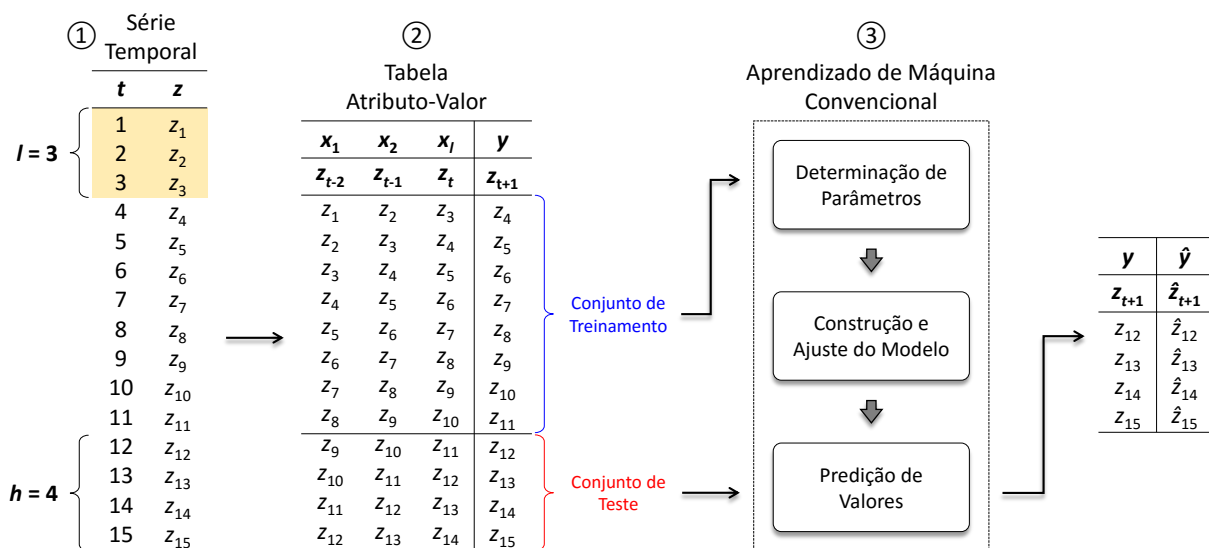


Figura 24: Exemplificação do processo de predição de ST segundo a abordagem global

Na Figura 24, o procedimento de transposição da sequência Z de tamanho $m = 15$ para o formato atributo-valor pode ser visibilizado como uma janela deslizante de comprimento $l = 3$. Essa janela é iterativamente deslocada sobre a ST com o propósito de coletar todas as subsequências originadas pelas l observações consecutivas. Cada subsequência extraída remete a um par (X_i, y_i) , onde: $X_i = (x_{i1}, x_{i2}, x_{il})$ e corresponde ao padrão temporal de comprimento l ; e y_i indica o valor subsequente à X_i , observado no instante $l + 1$. O conjunto de pares (X_{ij}, y_{ij}) , em que $j \in [1, m - l]$, configura a tabela atributo-valor. A ideia por trás dessa conversão é utilizar as observações do passado para prever uma observação no futuro.

Supondo um horizonte de predição $h = 4$, os dados da referida tabela são particionados em dois conjuntos: o conjunto de treinamento, que é atribuído à confecção do modelo, e o conjunto de teste, que é conferido à validação do desempenho preditivo do modelo ajustado. A acurácia de predição é estimada por meio da comparação entre os valores \hat{y}_k preditos com os respectivos valores reais y_k observados, onde $k \in [1, h]$. Na Figura 25 é exemplificada a aplicação da estratégia de predição multi-etapa à frente, com passos aproximado e atualizado, para calcular as estimativas \hat{y}_k .

x_1	x_2	x_l	y		\hat{y}	x_1	x_2	x_l	y		\hat{y}
z_{t-2}	z_{t-1}	z_t	z_{t+1}		\hat{z}_{t+1}	z_{t-2}	z_{t-1}	z_t	z_{t+1}		\hat{z}_{t+1}
z_9	z_{10}	z_{11}	z_{12}	\rightarrow	\hat{z}_{12}	z_9	z_{10}	z_{11}	z_{12}	\rightarrow	\hat{z}_{12}
z_{10}	z_{11}	\hat{z}_{12}	z_{13}	\rightarrow	\hat{z}_{13}	z_{10}	z_{11}	z_{12}	z_{13}	\rightarrow	\hat{z}_{13}
z_{11}	\hat{z}_{12}	\hat{z}_{13}	z_{14}	\rightarrow	\hat{z}_{14}	z_{11}	z_{12}	z_{13}	z_{14}	\rightarrow	\hat{z}_{14}
\hat{z}_{12}	\hat{z}_{13}	\hat{z}_{14}	z_{15}	\rightarrow	\hat{z}_{15}	z_{12}	z_{13}	z_{14}	z_{15}	\rightarrow	\hat{z}_{15}

(a) Passo aproximado
(b) Passo atualizado

Figura 25: Representação da estratégia de predição multi-etapa à frente para a abordagem global

É sabido que a predição multi-etapa à frente com passo aproximado enfrenta algumas dificuldades, como acúmulo de erro, desempenho reduzido e aumento da incerteza. Esses problemas tornam-se mais visíveis à medida que o horizonte de predição cresce e os valores preditos acabam por não serem substituídos pelos valores reais observados. Apesar dessas desvantagens, a estratégia multi-etapa à frente com passo aproximado viabiliza o estudo dos algoritmos quanto à propagação do erro de predição.

Obviamente, a aproximação global não está isenta de limitações. A mais evidente delas é o fato de que os pares (X_{ij}, y_{ij}) são considerados independentes e identicamente distribuídos pelos algoritmos de AM tradicionais. Essa suposição acarreta em perda de informação temporal, o que implica na degra-

dação do desempenho preditivo do modelo de regressão construído. Entre os métodos aplicados em conformidade com essa abordagem, podem ser citados os que empregam funções polinomiais e racionais, além dos embasados nas Redes Neurais Artificiais (ANN) e Máquinas de Suporte Vetorial (SVM).

Diferentemente da aproximação global, a abordagem local contempla algoritmos de AM que foram adaptados para incluir, no processo de predição, a informação temporal associada aos dados. Tais métodos particionam a ST original em subsequências cujos valores mais próximos ou mais importantes em relação ao valor atual são combinados para produzir o valor futuro. Essas combinações são empreendidas por funções de aproximação, como a média local simples e a ponderada. Dentre os métodos aplicados conforme essa abordagem encontram-se variações do algoritmo *k-Nearest Neighbors*.

4.2.1 ANN

As ANN são modelos computacionais que buscam simular o processamento de informação realizado pelo cérebro humano. Elas são compostas por unidades simples de processamento, os neurônios, que se unem por meio de conexões sinápticas (Zhang et al., 1998). Cada conexão, além de ser altamente especializada, é responsável pelo envio de sinais de um neurônio para outro. Segundo Haykin (2009), os neurônios e suas conexões podem ser implementados utilizando-se de componentes eletrônicos ou via simulação programada em computador.

O *Perceptron*, exibido na Figura 26, constitui a forma mais simples de uma ANN usada para, além de outras tarefas, a classificação de padrões denominados de linearmente separáveis. Um conjunto de dados é dito linearmente separável quando ele pode ser particionado, por intermédio de um hiperplano, em dois subconjuntos disjuntos (Rosenblatt, 1958; McCulloch e Pitts, 1943).

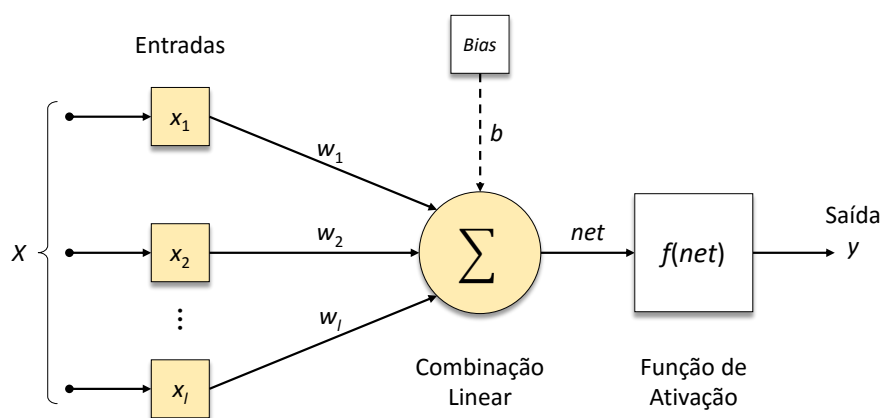


Figura 26: Estrutura do *Perceptron* (modificado de Haykin (2009))

Na Figura 26, o único neurônio abrange l entradas de dados $x_i \in X$. O i -ésimo elemento de X , fornecido eventualmente pelos neurônios adjacentes,

encontra-se associado a um peso sináptico w_i . Esse peso pode assumir, dependendo do tipo de conexão estabelecida (inibitória ou excitatória), um valor negativo ou positivo que reflete a importância da respectiva entrada para o processamento. A combinação linear das entradas com os pesos, adicionada de um limiar (*bias*) $b \in \mathfrak{R}$, resulta no valor *net* (Equação 39). Tal valor é enviado para uma função de ativação f que define a saída y do neurônio.

$$net = \sum_{i=1}^l w_i x_i + b \quad (39)$$

O *bias* tem como finalidade corrigir, aumentar ou diminuir, o valor de *net*. Essa correção contribui para que o resultado de $f(net)$ seja o mais próximo do esperado. No modelo de McCulloch e Pitts (1943), f corresponde a uma função degrau do tipo:

$$f(net) = \begin{cases} 1 & \text{se } net > 0 \\ 0 & \text{se } net \leq 0 \end{cases}$$

Em relação ao domínio do valor y , este pode ser tanto binário, com $y \in \{0, 1\}$ ou $y \in \{-1, 1\}$, quanto contínuo, em que $y \in \mathfrak{R}$. Além disso, outros tipos de função de ativação podem ser aplicados (Haykin, 2009). A exemplo, tem-se a função linear, que trabalha com a multiplicação do valor *net* por um coeficiente linear contínuo; a função rampa, que limita o valor de y no intervalo de $[-1, 1]$; e a função sigmoideal, a qual baseia-se na aplicação de uma função exponencial para produzir valores contínuos com variações suaves.

Os pesos sinápticos do *Perceptron* podem ser adaptados empregando um processo de aprendizado com um número finito de iterações. A aprendizagem é conduzida pela regra de correção de erro conhecida como algoritmo de convergência do *Perceptron*. Esse algoritmo visa encontrar um vetor de pesos w tal que as duas igualdades da função degrau sejam satisfeitas (Lippmann, 1987).

Apesar de ter causado grande impacto nas pesquisas de Inteligência Artificial, o estudo do *Perceptron* foi subitamente interrompido devido às críticas relatadas em Minsky e Papert (1969). No referido trabalho, a partir de uma análise matemática rigorosa, foi demonstrado que o modelo de um neurônio não era capaz de resolver questões não-linearmente separáveis, como simular o comportamento de uma função *XOR* ou “ou-exclusivo”. O problema do *Perceptron* de duas camadas (entrada e saída) poderia ser solucionado acrescentando camadas intermediárias a sua estrutura. Porém, conforme exposto por Minsky e Papert (1969), haviam dúvidas sobre a possibilidade de se construir um algoritmo de aprendizado que garantisse, com baixo custo computacional, a convergência para modelos com múltiplas camadas.

O desenvolvimento das redes neurais com mais de duas camadas só foi re-

almente retomado após a proposição do algoritmo de aprendizado *Backpropagation* (Rumelhart et al., 1986). Esse algoritmo permite que uma rede neural com camadas intermediárias possa ser treinada em uma sequência de dois passos. Inicialmente, um padrão é apresentado à camada de entrada de dados. O processamento propaga-se ao longo da rede, camada por camada, até que a resposta seja produzida pela camada de saída. Em seguida, um erro é calculado por meio da comparação entre a saída obtida com a saída esperada para o padrão investigado. O erro é retropropagado da camada de saída para a camada de entrada, e os pesos das conexões das unidades das camadas intermediárias são ajustados utilizando a regra delta generalizada (Haykin, 2009). Esse processo é repetido até a convergência da rede para um estado que propicia a codificação de todos os padrões do conjunto de treinamento.

Um tipo de rede neural com múltiplas camadas, usualmente treinado pelo algoritmo *Backpropagation*, é denominado de *Multilayer Perceptron (MLP)*. Na Figura 27 é mostrada a estrutura de uma rede *MLP* com três camadas.

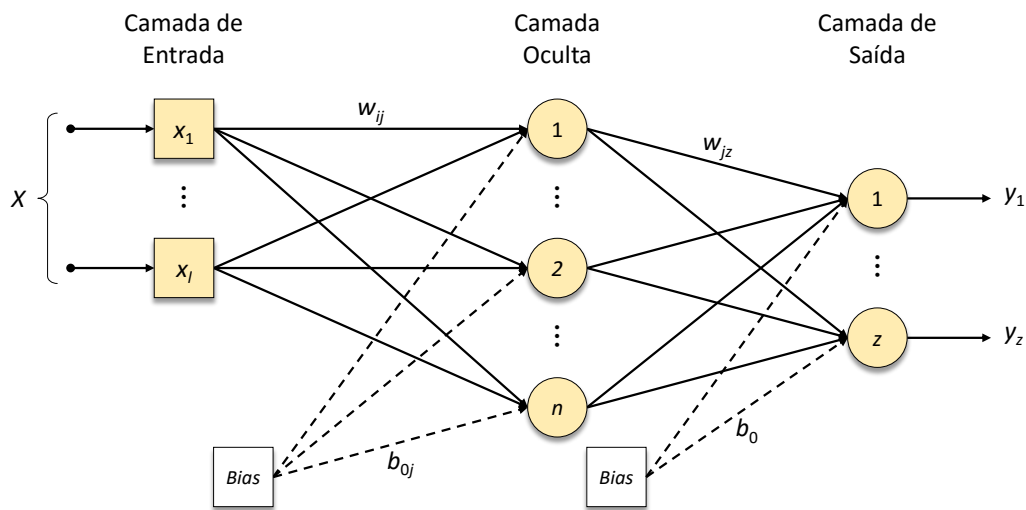


Figura 27: Estrutura de uma rede *MLP* com camada oculta única

A saída do modelo esquematizado na Figura 27, porém considerando um único neurônio na camada de saída, pode ser representada equacionalmente do seguinte modo:

$$y = f \left(\sum_{j=1}^n w_j f \left(\sum_{i=1}^l w_{ij} x_i + b_{0j} \right) + b_0 \right)$$

No *MLP*, podem existir uma ou mais camadas de neurônios entre as camadas de entrada de dados e de saída dos resultados. Essas camadas intermediárias são unidades que não interagem diretamente com o ambiente e funcionam como extratoras de características. Se existirem conexões apropriadas entre as unidades de entrada e um conjunto considerável de unidades intermediárias, pode-se sempre encontrar a representação que irá produzir o

mapeamento correto entre a entrada de dados e a saída dos resultados. Embora resulte em modelos pouco compreensíveis, a aplicação do *MLP* possibilita a manipulação eficiente de grandes volumes de dados, sendo a capacidade de generalização uma das suas principais características.

A precisão do *MLP* está associada à três aspectos topológicos: (1) determinação do número de camadas ocultas; (2) definição do número de neurônios em cada uma das camadas; e (3) especificação dos pesos sinápticos que interconectam os neurônios nas diferentes camadas da rede. Conforme documentado em [Cybenko \(1989\)](#), são necessárias no máximo duas camadas intermediárias, com um número suficiente de unidades por camada, para se produzir quaisquer mapeamentos. Por outro lado, com apenas uma única camada intermediária é possível aproximar qualquer função contínua.

4.2.2 SVM

Uma *SVM* constitui uma técnica de AM que, baseada na Teoria do Aprendizado Estatístico ([Vapnik, 1999](#)), tem a capacidade de resolver diferentes problemas, incluindo os de classificação e regressão ([Gunn et al., 1998](#)).

Apesar dos modelos *SVM* apresentarem uma estrutura semelhante a das redes neurais, eles divergem no modo como o aprendizado é conduzido. Enquanto as *ANN* trabalham com a minimização do risco empírico, ou seja, a minimização do erro do modelo induzido sobre os dados de treinamento, as *SVM* são fundamentadas no princípio da minimização do risco estrutural, o qual busca pelo menor erro de treino ao passo que minimiza um limite superior para o erro de generalização do modelo (erro do modelo quando aplicado aos dados de teste) ([Cristianini e Shawe-Taylor, 2000](#); [Vapnik, 1999](#); [Shawe-Taylor et al., 1998](#)).

O conceito de generalização é melhor compreendido para o caso de classificação binária. Sendo assim, dadas duas classes e um conjunto de pontos que pertencem a estas, as *SVM* determinam o hiperplano que os separam, de maneira a colocar a maior quantidade possível de pontos da mesma classe do mesmo lado, ao mesmo tempo que a distância de cada classe a essa superfície de decisão é maximizada. Para fins didáticos, na [Figura 28](#) é ilustrado um conjunto de retas que discriminam os dados em duas classes. Entre essas retas, apenas uma maximiza a margem de separação (distância entre o hiperplano e a amostra mais próxima de cada classe). A reta com margem máxima, designada de hiperplano de separação ótima, é nada mais que o objeto a ser procurado durante o treinamento do modelo.

A técnica indicada na [Figura 28](#) é restrita aos problemas linearmente separáveis. No entanto, em situações onde as amostras não são linearmente separáveis, a solução incide em mapear os dados de entrada para um espaço

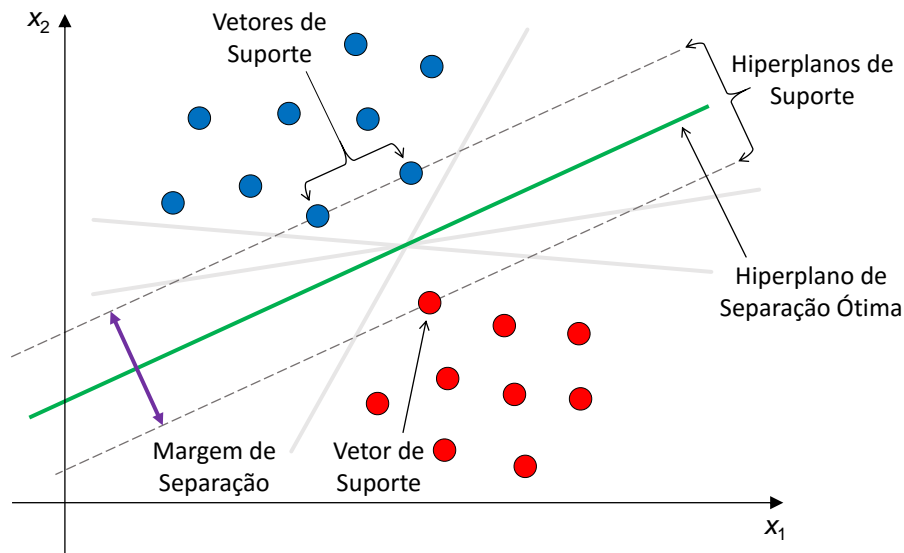


Figura 28: Hiperplano de separação ótima e seus hiperplanos de suporte. Os eixos ordenados x_1 e x_2 representam as dimensões das amostras no espaço 2D

de dimensão maior (espaço de características). Tal mapeamento é alcançado mediante ao uso de uma função *kernel* (Cristianini e Shawe-Taylor, 2000). O processo de transformação, por aumento da dimensão, de um domínio não-linearmente separável, em um problema linearmente separável, é retratado na Figura 29. Nessa figura, o mapeamento foi realizado pela função *kernel* $K(x_i, x_j)$.

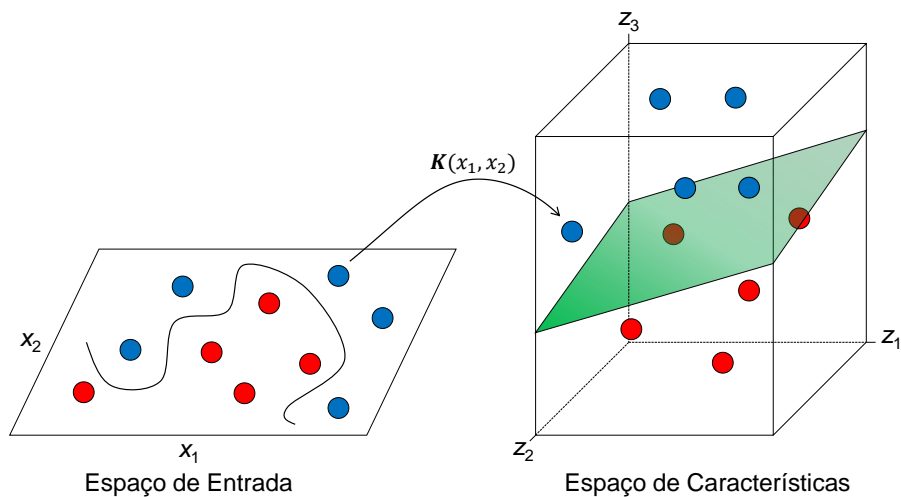


Figura 29: Mapeamento de dados para um espaço de características de mais alta dimensão utilizando como artifício a função *kernel* (modificado de Kaundal et al. (2006))

Os *kernels* mais utilizados na prática são os polinomiais, os gaussianos ou Funções de Base Radial (*RBF*) e os sigmoidais. Cada um deles abrange parâmetros que precisam ser estabelecidos pelo usuário. Por exemplo, o modelo *SVM* com *kernel RBF* exige dois parâmetros, C e σ . A constante C é um termo de regularização que impõe um peso à minimização dos erros no conjunto

de treinamento em relação à minimização da complexidade do modelo. Já o valor de σ reflete a largura da gaussiana da função *kernel* (Lorena e de Carvalho, 2007). O número de funções radiais e os seus respectivos centros são determinados pelos vetores de suporte encontrados.

A construção de uma *SVM* implica na resolução de um problema quadrático, com restrições lineares, que depende do conjunto de dados de entrada, de alguns parâmetros e da margem de separação. Durante o treinamento desse modelo são obtidos os multiplicadores de Lagrange que caracterizam os vetores de suporte, os quais, por sua vez, definem as margens do hiperplano de separação ótima.

De modo análogo às redes neurais, as *SVM* são capazes de conduzir à solução de problemas de elevada complexidade sem, todavia, permitir interpretar adequadamente os resultados ou entender a evolução do processo adaptativo que acarretou na resposta. Na realidade, a carência de legibilidade da solução está atrelada a sua própria estrutura, a qual reside em uma combinação linear de funções *kernel*.

Os modelos *SVM* podem ser estendidos para resolver tarefas de regressão e, conseqüentemente, de predição de ST (Ristanoski et al., 2013). Contudo, o problema de otimização que viabiliza o seu treinamento deve ser reformulado para lidar com as particularidades e propósitos dessas tarefas. Dos algoritmos que codificam *SVM*, o *Sequential Minimal Optimization (SMO)* (Platt, 1999), desenvolvido para decompor o problema de minimização quadrático em vários subproblemas de programação quadrática, dispõe dos recursos suficientemente necessários para a projeção de dados temporais.

4.2.3 *kNN-TSPI*

Uma das estratégias mais populares para projeção local de dados reside em uma modificação do método *k-Nearest Neighbors (kNN)*. O *kNN* é um algoritmo de aprendizado supervisionado do paradigma baseado em exemplos que visa encontrar, segundo alguma medida de similaridade, os k exemplos mais próximos de um exemplo ainda não rotulado e, baseado nos rótulos desses k exemplos próximos, rotular o novo exemplo (Fix e Hodges, 1951).

A adequação do *kNN* para predição de ST é bastante simples. Dada uma série $Z = (z_1, z_2, \dots, z_m)$, o problema é predizer o valor z_{m+h} , onde h indica o horizonte de predição. Em termos práticos, a predição do momento z_{m+h} é tipicamente denotada por $\hat{z}(m, h)$ ou $\hat{z}(h)$. Por simplicidade, mas sem perda de generalidade, a ideia da referida adaptação será discutida com um horizonte unitário ($h = 1$), ou seja, considerando a predição do próximo valor da ST.

Tendo por base a ST Z , o objetivo consiste em predizer o próximo (não observado) ponto de dados $\hat{z}(m, 1)$. O método considera as últimas l observações

como consulta Q , e procura as k subsequências mais parecidas a Q , usando uma janela deslizante de tamanho l . Seja $S_{1..l}^{(1)}, \dots, S_{1..l}^{(k)}$ as k subsequências mais similares, utiliza-se as observações seguintes de cada subsequência $S_{l+1}^{(j)}$ com $1 \leq j \leq k$ para prever $\hat{z}(m, 1)$. Desse modo, os valores de $S_{l+1}^{(j)}$ são fornecidos como entrada para uma função de predição f , por exemplo a Média de Valores Absolutos (Equação 40), que tem como finalidade aproximar o valor de $\hat{z}(m, 1)$.

$$f(S) = \frac{1}{k} \sum_{j=1}^k S_{l+1}^{(j)} \quad (40)$$

Na Equação 40, f corresponde à função de predição, S indica o conjunto de subsequências mais similares e $S^{(j)}$ refere-se ao j -ésimo vizinho mais próximo. Essa é a maneira mais condensada de projetar dados, pois consiste em calcular a média das predições considerando que todos os valores a serem combinados são igualmente prováveis de ocorrerem no futuro.

Um exemplo de aplicação do método de projeção baseado em similaridade com $k = 3$ e $l = 25$ é exibido na Figura 30. No gráfico dessa figura, a linha pontilhada na cor cinza representa as observações que compõem a ST; a linha em verde indica a subsequência de tamanho 25 tida como consulta de referência; as linhas tracejadas azuis expressam as subsequências mais similares encontradas pelo algoritmo de acordo com uma medida de similaridade, nesse caso a distância euclidiana; os círculos nas cores azuis correspondem às observações seguintes de cada subsequência encontrada; e o quadrado na cor vermelha reflete o valor a ser predito.

Há três problemas com o procedimento de busca por similaridade esquematizado na Figura 30 e a combinação desses problemas faz com que as k subsequências mais parecidas sejam usualmente não significativas ou completamente diferentes da consulta Q . À vista disso, os autores deste trabalho formularam técnicas para solucionar esses problemas. As técnicas desenvolvidas, discutidas a seguir, foram conciliadas para originar o algoritmo *kNN - Time Series Prediction with Invariances (kNN-TSPI)* (Parmezan e Batista, 2015).

A primeira questão é a falta de invariância à amplitude e deslocamento. Ao comparar uma consulta Q a uma subsequência S , ambas devem ser invariantes à amplitude e deslocamento. Esse tipo de invariância pode ser obtido por vários métodos, porém uma maneira simples é utilizando a normalização em z -scores (ou z -normalização), a qual é definida pela Equação 41 (Rakthanmanon et al., 2012).

$$z'_t = \frac{z_t - \mu}{\sigma} \quad (41)$$

Na Equação 41, z'_t e z_t referem-se, nessa ordem, ao valor normalizado e a observação da ST Z , ambos no instante de tempo t . Do mesmo modo, μ indica

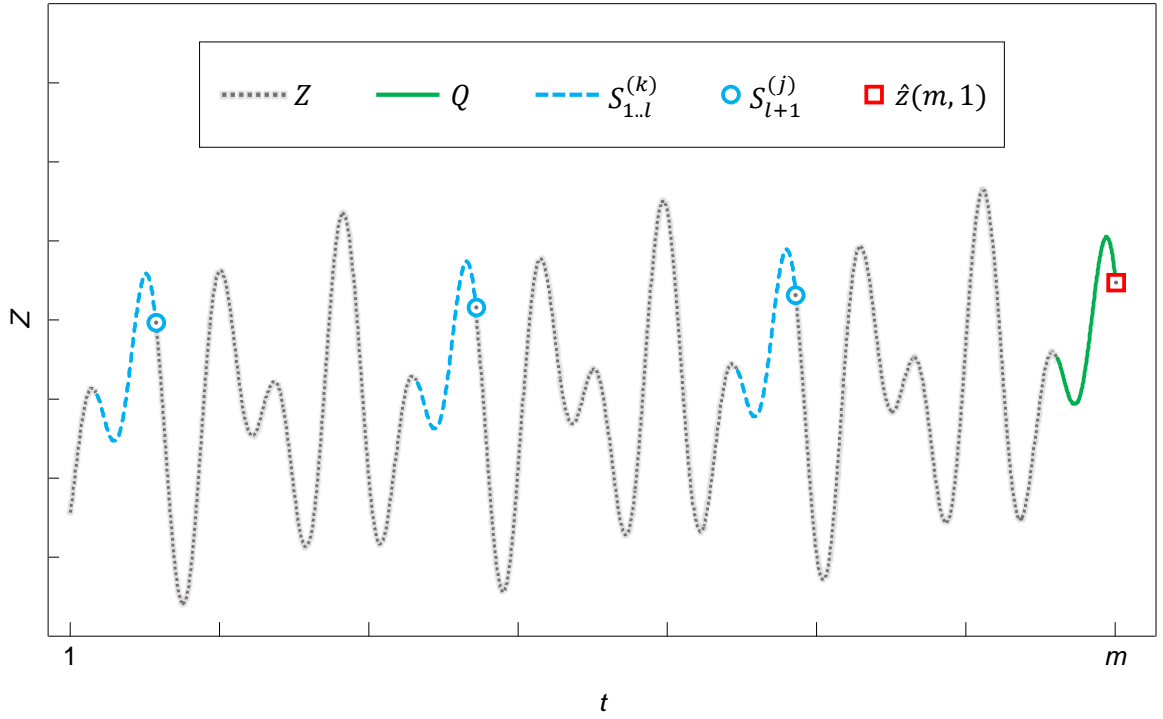


Figura 30: Exemplo da aplicação do método baseado em similaridade para predição de ST com parâmetros $k = 3$ e $l = 25$

a média e σ o desvio padrão dos valores de uma dada subsequência que inclui a observação z_t . A z -normalização, que transforma os dados para garantir média zero e desvio padrão unitário, tem sido fortemente defendida na busca por subsequências em ST (Rakthanmanon et al., 2012).

Na Figura 31 é ilustrada uma subsequência $S^{(1)}$ que é um casamento exato com a consulta Q . A subsequência $S^{(1)}$ apresenta os mesmos valores de Q , com apenas uma diferença, um pequeno deslocamento. Adicionalmente, foi incluído uma segunda subsequência $S^{(2)}$ que é totalmente diferente de Q . De acordo com a Distância Euclidiana (DE), a subsequência mais parecida com Q é a linha reta $S^{(2)}$ ($DE(Q, S^{(2)}) = 84,26$ e $DE(Q, S^{(1)}) = 114,54$), mesmo que a diferença de deslocamento entre Q e $S^{(1)}$ seja só de 10 unidades. A razão para esse acontecimento é simples e passível de explicação: pequenas variações de deslocamento são drasticamente acumuladas fazendo com que a DE final cresça muito rápido.

Embora para algumas aplicações a amplitude e o deslocamento sejam atributos importantes para caracterizar a subsequência, na maior parte dos domínios de aplicação, a busca sem invariância à amplitude e deslocamento não leva a resultados significativos. O motivo é que, mesmo em tais domínios, as subsequências de dados semelhantes ocorrem raramente com o mesmo deslocamento.

Na Figura 31 foi retratado um cenário ideal, pois $S^{(1)}$ é uma cópia perfeita de Q . No entanto, as pequenas diferenças de deslocamento já são suficientes

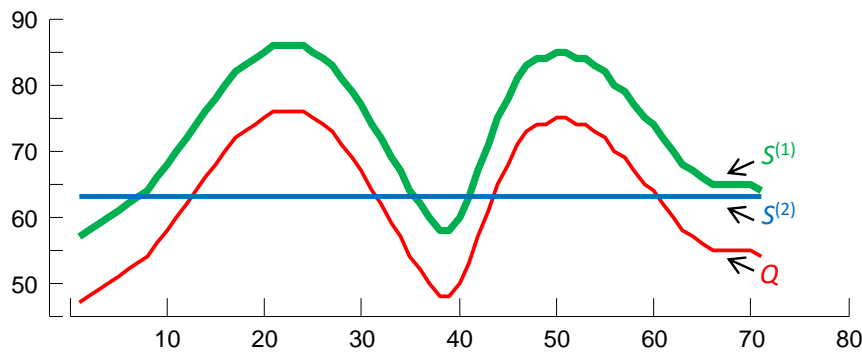


Figura 31: Exemplo de variância à amplitude e deslocamento: Pequeno deslocamento faz considerar Q mais parecida a $S^{(2)}$ do que a $S^{(1)}$

para produzir um casamento incorreto. Se fossem inseridos ruído, diferenças de amplitude e até mesmo deformação, o casamento entre essas subsequências seria mais improvável. No exemplo esquematizado foi usada a DE para esboçar esse problema, porém outras distâncias populares em ST, como *Dynamic Time Warping (DTW)* (Petitjean et al., 2011), também resultariam no impasse supracitado.

A segunda questão aborda a necessidade de invariância à complexidade ao comparar subsequências. Em resumo, o problema reside no fato de que pares de objetos complexos, mesmo aqueles que subjetivamente sejam similares, tendem a estar mais separados em termos de distância do que pares de objetos simples (Batista et al., 2014). Esse fato introduz erros na busca de similaridade, devido à circunstância de que subsequências complexas são consideradas mais similares a outras mais simples. Uma ideia geral desse problema é exemplificada na Figura 32.

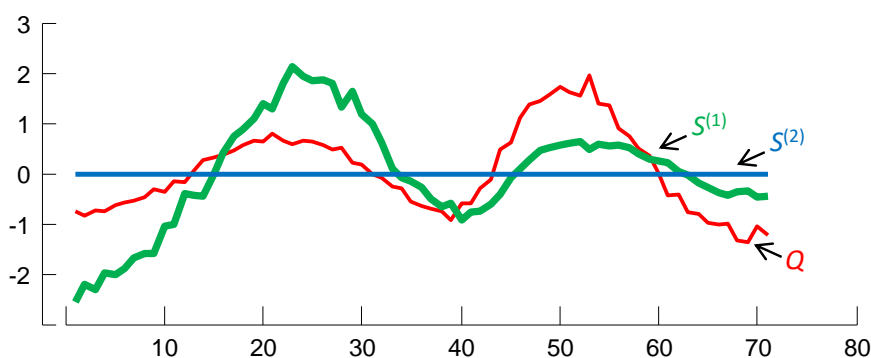


Figura 32: Exemplo de variância à complexidade: A subsequência simples $S^{(2)}$ é considerada o melhor casamento para Q , embora $S^{(1)}$ tenha um comportamento parecido

Na Figura 32 foi empregada z -normalização sobre versões das consultas da Figura 31 e introduzidas corrupções nos dados (ruído, diferenças de fase e distorção de amplitude), em pequenas quantidades, de modo que Q e $S^{(1)}$ ainda sejam parecidas. É importante ressaltar que essa configuração é mais realista

do que a exibida na Figura 31 pois, na maior parte dos domínios de aplicação, espera-se que o fenômeno de interesse gere subsequências semelhantes, mas nunca cópias exatas. Novamente, a distância entre Q e $S^{(1)}$ ($DE(Q, S^{(1)})$) é maior do que a distância entre Q e a linha reta $S^{(2)}$. Isso acontece porque as formas simples como $S^{(2)}$ geralmente apresentam um bom comportamento médio que casa bem com formas complexas. Em contraste, as formas complexas geralmente possuem várias características, tais como os picos e as depressões, as quais dificultam o casamento entre si, mesmo quando as formas parecem semelhantes ao olho humano.

Como uma janela deslizante é utilizada ao longo da ST para encontrar as subsequências mais parecidas com Q , haverá subsequências com uma enorme diversidade de formas e as medidas de distância atuais tendem a escolher formas simples como melhores casamentos. Nessas condições, as formas extremamente simples dificilmente serão úteis para propósitos de predição. A solução para esse problema reside no uso de distâncias com invariância de complexidade, como a *Complexity-Invariant Distance (CID)*.

Definida a partir da distância euclidiana e em conformidade com a Equação 42 (Batista et al., 2014), a medida *CID* considera a morfologia das sequências sendo comparadas e atribui distâncias maiores a sequências com complexidades diferentes.

$$CID(Q,C) = ED(Q,C) \times CF(Q,C) \quad (42)$$

Na Equação 42, Q e C são duas ST, ED corresponde à distância euclidiana e CF é um fator de correção de complexidade calculado pela Equação 43, na qual $CE(Q)$ e $CE(C)$ refletem estimativas da complexidade das sequências Q e C , respectivamente. Caso Q e C tenham a mesma complexidade, a *CID* simplesmente degenera para a distância euclidiana.

$$CF(Q,C) = \frac{\max(CE(Q),CE(C))}{\min(CE(Q),CE(C))} \quad (43)$$

A estimativa de complexidade usada pela medida *CID* é bastante simples. Ela é apoiada na intuição de que uma ST pode ser “esticada” até que se torne um segmento de reta. Consequentemente, uma sequência complexa resultaria em um segmento de reta mais longo que uma sequência simples. Na Figura 33, essa ideia é sintetizada com um exemplo.

O emprego da Equação 44 fornece a estimativa de complexidade ilustrada na Figura 33.

$$CE(S) = \sqrt{\sum_{i=1}^{l-1} (s_i - s_{i+1})^2} \quad (44)$$

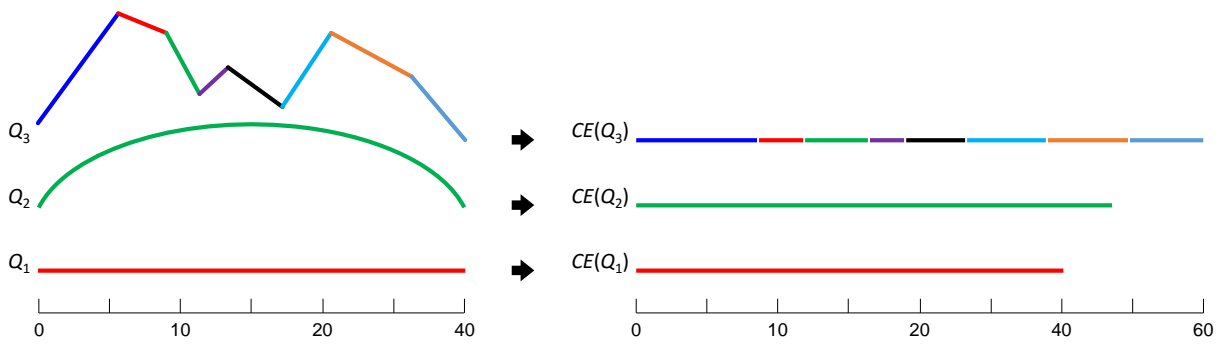


Figura 33: Representação da estimativa de complexidade adotada pela CID: Três sequências podem ter suas complexidades estimadas “esticando” cada uma delas e medindo o comprimento dos segmentos de reta resultantes

A terceira questão que deve ser levada em consideração contempla a eliminação de casamentos triviais (Mueen et al., 2009). A subsequência gerada por uma janela deslizante que começa na observação m é muito semelhante à subsequência que começa na observação $m + 1$ (ou $m - 1$). Isso ocorre porque essas subsequências estão deslocadas por uma unidade em relação ao tempo, o que faz com que elas compartilhem uma grande quantidade de observações. Essa situação é retratada na Figura 34, na qual a consulta Q é deslocada para a direita e após à esquerda por uma observação.

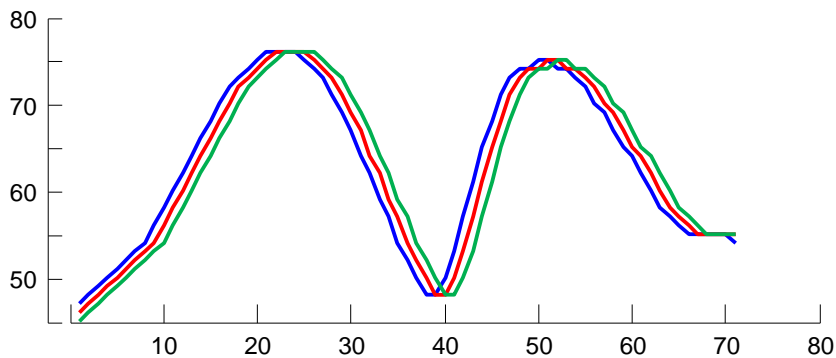


Figura 34: Exemplo de casamentos triviais: A distância entre a consulta Q e Q deslocada para a direita ou para a esquerda por uma observação é muito pequena, visto que essas três subsequências compartilham a maior parte das observações

O exemplo exibido na Figura 34 mostra que se Q fosse incluído como parte dos dados nos quais a busca será realizada, o algoritmo tende a retornar sempre como subsequência mais parecida, o casamento trivial constituído pela consulta Q deslocada por algumas poucas observações. Comumente, é um desperdício de tempo computacional procurar em toda a sequência de dados quando os resultados são altamente prováveis de serem pré-definidos. Contudo, essa não é a única desvantagem da consideração de casamentos triviais. Um problema parecido acontece quando a consulta casa com qualquer subsequência S da ST. Nesse caso, é provável que a distância entre Q e S seja

muito parecida com a distância entre Q e S deslocada à direita ou esquerda por poucas observações. Se S é uma das subsequências mais similares à Q , então os casamentos triviais de S provavelmente irão aparecer entre as subsequências mais similares. Essa condição torna-se uma contradição, pois a ideia de utilizar k no lugar de uma única subsequência mais similar é para incluir alguma diversidade e, portanto, tornar o critério de busca mais robusto contra escolhas errôneas das subsequências mais similares. No entanto, casamentos triviais fazem exatamente o contrário, ou seja, introduzem pouca diversidade, acarretando na geração de várias cópias da mesma subsequência com alguma pequena variação. Uma maneira de assegurar essa diversidade é por meio da exclusão de casamentos triviais por checagem iterativa.

Dada uma subsequência $S_{i..(i+l-1)}^{(1)}$ e uma subsequência $S_{j..(j+l-1)}^{(2)}$, tal que $j > i$, diz-se que $S^{(2)}$ é um casamento trivial de $S^{(1)}$ se $|j - i| \leq l$. Obviamente essa condição implica na identificação de casamentos triviais para sua posterior exclusão.

Todas as técnicas descritas foram empregadas para aprimorar o desempenho da predição de ST por similaridade e criar o Algoritmo 1: *kNN-TSPI*.

Na segunda linha do Algoritmo 1 é atribuída à variável S todas as subsequências de tamanho l extraídas da ST Z , a qual terá seu próximo valor predito. Em seguida, na terceira linha, todas as subsequências de dados geradas a partir da série original são z -normalizadas. No decorrer dessas normalizações, as médias e os desvios padrão extraídos são também alocados para uso posterior. Na quarta linha é armazenada a consulta de referência Q , ou seja, a subsequência que será utilizada como parâmetro para a busca por subsequências similares no passado da série. Na quinta linha são computadas as distâncias com invariância à complexidade entre a consulta Q z -normalizada e todas as subsequências de tamanho l previamente normalizadas. É relevante salientar que a z -normalização proposta não é uma normalização da série inteira, mas sim da consulta de referência Q e da janela deslizante, dessa maneira a busca por similaridade é feita com invariância à amplitude e deslocamento. Na sexta linha, as k subsequências mais parecidas com a consulta Q z -normalizada são buscadas. No conjunto de subsequências mais semelhantes não há casamentos triviais, isto é, não ocorre sobreposição dos dados nem com a consulta de referência Q , nem entre os próprios dados das k subsequências. Os valores seguintes a cada uma das k subsequências encontradas são, na sétima linha, obtidos e, na oitava linha, z -normalizados usando a média e o desvio padrão da sua respectiva subsequência de origem. Essa normalização tem por objetivo fazer com que o valor $S_{l+1}^{(k)}$ tenha a mesma distribuição que a subsequência $S_{1..l}^{(k)}$. Posteriormente, na nona linha, os valores subsequentes z -normalizados são mapeados segundo a Equação 45 para

Algoritmo 1: *k*NN-TSPI

```
// Z representa uma s3rie temporal com m observa33es
// l refere-se ao tamanho da janela de busca
// k indica a quantidade de vizinhos mais pr3ximos
Input:  $Z, l, k$ 
Output:  $\hat{z}(m, 1)$ 
1 begin
  /*  $S_{1..l}^{(i)}$  cont3m a subsequ3ncia de comprimento l que come3a na
  observa33o i da s3rie temporal Z */
2  $S \leftarrow \text{generate\_subsequences}(Z_{1..(m-l)}, l)$ ;
  /*  $S^{(i)'}$  3 a z-normaliza33o da subsequ3ncia  $S^{(i)}$ . A m3dia e o
  desvio padr3o usado na normaliza33o s3o tamb3m alocados
  para re3uso */
3  $[S', \sigma, \mu] \leftarrow z\_scores(S)$ ;
  // Obten33o da consulta Q
4  $Q \leftarrow Z_{(m-l+1)..m}$ ;
  /*  $D^{(i)}$  armazena a dist3ncia com invari3ncia 3 complexidade
  entre a consulta Q e  $S^{(i)}$ , ambas z-normalizadas */
5  $D \leftarrow CID(z\_scores(Q), S')$ ;
  /* Escolha das k subsequ3ncias mais semelhantes
  (desconsidera33o de casamentos triviais) */
6  $P \leftarrow \text{search\_nearest\_neighbors}(S, D, k)$ ;
  /* Obten33o do valor seguinte  $S_{l+1}^{(k)}$  de cada uma das k
  subsequ3ncias mais similares  $\in P$  */
7  $R \leftarrow \{S_{l+1}^{(1)}, \dots, S_{l+1}^{(k)}\}$ ;
  /* Normaliza33o em z-scores de cada valor  $S_{l+1}^{(k)} \in R$  utilizando a
  m3dia e o desvio padr3o da sua respectiva subsequ3ncia  $S_{1..l}^{(k)}$ 
  */
8  $R' \leftarrow \left\{ \frac{S_{l+1}^{(1)} - \mu(S_{1..l}^{(1)})}{\sigma(S_{1..l}^{(1)})}, \dots, \frac{S_{l+1}^{(k)} - \mu(S_{1..l}^{(k)})}{\sigma(S_{1..l}^{(k)})} \right\}$ ;
  /* Mapeamento, segundo a Equa33o 45, das z-normaliza33es para
  os valores da consulta Q */
9  $R \leftarrow \text{map\_query\_values}(Q, R')$ ;
  // C3lculo da predi33o
10  $\hat{z}(m, 1) \leftarrow f(R)$ ;
11 return  $\hat{z}(m, 1)$ ;
12 end
```

o espa3o de valores da consulta Q e utilizados, na d3cima linha, pela fun33o de predi33o f (Equa33o 40) para o c3lculo do valor futuro. Ao final, na d3cima primeira linha, o valor estimado 3 retornado pelo algoritmo.

$$R^{(k)} = \sigma(Q) \times R^{(k)'} + \mu(Q) \quad (45)$$

Na Equa33o 45, σ e μ correspondem, nessa ordem, ao desvio padr3o e a m3dia da consulta de refer3ncia Q . Essa equa33o compreende 3 fun33o inversa da z-normaliza33o e permite mapear, para o espa3o de valores da

subsequência Q , o conteúdo da variável $R^{(k)'}$ que foi previamente normalizado, isto é, teve seu valor $S_{l+1}^{(k)}$ subtraído da média e dividido pelo desvio padrão da subsequência de dados $S_{1..l}^{(k)}$.

O *kNN-TSPI* é um método univariado e apropriado para predição automática a curto prazo. O algoritmo é de fácil implementação e possui apenas dois parâmetros: a cardinalidade do conjunto de subsequências similares (k) e a quantidade de observações utilizadas como referência (l) para a busca por subsequências semelhantes. A complexidade de tempo do *kNN-TSPI* é $O(m \cdot l)$, onde m indica o tamanho da ST.

4.3 Técnicas para Estimação de Parâmetros

Uma das principais dificuldades enfrentadas por pesquisadores no tema de predição de dados temporais incide na determinação da melhor configuração de parâmetros, entre as muitas possíveis, para ajustar o modelo de interesse ao conjunto dados investigado.

Em termos teóricos, o estabelecimento de todos os parâmetros de um modelo poderia necessitar a exploração completa do espaço de estados. Como tal procedimento é impraticável no mundo real, sugere-se usar algum tipo de algoritmo de busca, que se apoia em um método de amostragem de dados, para encontrar uma solução subótima (ou até mesmo ótima), com desempenho satisfatório, a um custo computacional aceitável.

4.3.1 Validação *Holdout*

O procedimento de estimação de parâmetros por validação *holdout* busca, por numeração intervalar, tornar mínimo o erro de ajuste do modelo preditivo sobre os dados de treinamento. Para tanto, uma parcela desses dados é utilizada para induzir o modelo, enquanto a outra é usada para verificar o potencial do modelo ajustado. O Algoritmo 2 transcreve a lógica dessa técnica de amostragem para estabelecer os parâmetros α , β , γ e s do método multiplicativo de Holt-Winters.

No Algoritmo 2 avalia-se iterativamente um conjunto de parâmetros previamente delimitado. A cada iteração, um novo modelo é construído e ajustado, segundo a combinação de parâmetros corrente, sobre as $m - h$ observações da subsequência S , ou seja, z_1, \dots, z_{m-h} ($S \in Z$). Posteriormente, o referido modelo é extrapolado para um horizonte de predição h' cujo comprimento equivale ao da subsequência de validação ($s_{n-h'+1}, \dots, s_n$). Ao término da busca, os parâmetros mais promissores (\mathbb{P}) são aqueles que minimizam o erro entre a subsequência predita e a subsequência de validação. Existem muitas medidas que podem ser aplicadas para mensurar esse erro, porém a mais usual é

Algoritmo 2: Holdout Validation

```
/* S representa uma subsequência de treinamento de comprimento n
   extraída da ST Z de tamanho m */
/* max_p é limite superior para o número de observações que
   constituem uma variação sazonal na série histórica */
/* h indica a quantidade de valores a serem preditos pelo melhor
   modelo identificado */
/* P compreende a lista de parâmetros que acarretou no menor erro
   de predição */
Input:  $S, max\_p, h$ 
Output:  $\mathbb{P}$ 
1 begin
   /* realmax corresponde ao maior número do sistema de ponto
      flutuante */
2  $min\_error = realmax;$ 
3  $h' = (max\_p + h) \div 2;$ 
4 for  $s \leftarrow 3 : 2 : max\_p$  do
5     for  $\alpha \leftarrow 0 : 0.25 : 1$  do
6         for  $\beta \leftarrow 0 : 0.25 : 1$  do
7             for  $\gamma \leftarrow 0 : 0.25 : 1$  do
8                  $error = holt\_winters(S, s, \alpha, \beta, \gamma, h', type = "MULTI");$ 
9                 if  $error < min\_error$  then
10                     $min\_error = error;$ 
11                     $s_{best} = s;$ 
12                     $\alpha_{best} = \alpha;$ 
13                     $\beta_{best} = \beta;$ 
14                     $\gamma_{best} = \gamma;$ 
15                end
16            end
17        end
18    end
19 end
20  $\mathbb{P} \leftarrow \{s_{best}, \alpha_{best}, \beta_{best}, \gamma_{best}\};$ 
21 return  $\mathbb{P};$ 
22 end
```

designada de Erro Quadrático Médio (MSE).

O número de observações que compõe o horizonte de predição h' pode ser escolhido de distintas maneiras, neste trabalho foi adotado $h' = (max_p + h) \div 2$, onde h indica a quantidade de valores a serem projetados na etapa de teste pelo método de predição utilizando o melhor conjunto de parâmetros encontrado.

4.3.2 Validação Cruzada

A validação cruzada é uma das técnicas de amostragem mais usadas para a avaliação de modelos em Mineração de Dados. Na área de predição de ST, ela é normalmente empregada para auxiliar na estimação de parâmetros dos

métodos aplicados conforme a abordagem global, onde um conjunto de dados no formato atributo-valor precisa ser gerado. O Algoritmo 3 exemplifica essa técnica para identificar os parâmetros l , \mathbb{C} e σ do algoritmo de regressão SVM.

Algoritmo 3: *Cross-validation*

```

// S representa uma subsequência de treinamento
/* max_p é limite superior para o número de observações que
   constituem uma variação sazonal na série histórica */
/* kFolds especifica o número de partições na qual a amostra de
   dados de treinamento será dividida */
/* P compreende a lista de parâmetros que acarretou no menor erro
   de predição */
Input:  $S, max\_p, h$ 
Output:  $\mathbb{P}$ 
1 begin
   /* realmax corresponde ao maior número do sistema de ponto
      flutuante */
2    $min\_error = realmax;$ 
3   for  $l \leftarrow 3 : 2 : max\_p$  do
     // T corresponde a uma tabela atributo-valor
4      $T \leftarrow generate\_data\_table(S, l);$ 
5     for  $\mathbb{C} \leftarrow 0 : 0.25 : 1$  do
6       for  $\sigma \leftarrow 0.005 : 0.05 : 0.25$  do
7          $error = cross\_validation(T, kFolds, \mathbb{C}, \sigma, model = "SVM");$ 
8         if  $error < min\_error$  then
9            $min\_error = error;$ 
10           $l_{best} = l;$ 
11           $\mathbb{C}_{best} = \mathbb{C};$ 
12           $\sigma_{best} = \sigma;$ 
13        end
14      end
15    end
16  end
17   $\mathbb{P} \leftarrow \{l_{best}, \mathbb{C}_{best}, \sigma_{best}\};$ 
18  return  $\mathbb{P};$ 
19 end

```

No Algoritmo 3, a ideia por trás da busca pelo conjunto de parâmetros mais adequado é similar à visibilizada no Algoritmo 2, exceto pelo conteúdo da quarta e sétima linhas. Na quarta linha ocorre a transposição da subsequência de treino S para o formato atributo-valor utilizando uma janela deslizante de tamanho l , tal como ilustrado pelo conjunto de treinamento na Figura 24 da página 48. Na sétima linha, a tabela atributo-valor T é dividida aleatoriamente em k amostras ($kFolds$) mutuamente exclusivas, sendo todas aproximadamente do mesmo tamanho. A k -ésima amostra é usada como conjunto de validação e as $k - 1$ amostras restantes formam o conjunto de treino. Para cada combinação das $k - 1$ amostras, um modelo é construído e ajustado segundo a combinação de parâmetros corrente. O erro de projeção é verificado

sobre o conjunto de validação k por intermédio, por exemplo, da medida MSE . Notoriamente, nesse cenário, o erro de parametrização (*error*) é estimado como a média dos MSE dos k modelos gerados e é considerado como uma estimativa do erro verdadeiro.

4.3.3 Método Box-Jenkins

Os parâmetros de um modelo da categoria $ARIMA$ podem ser determinados computacionalmente por meio de um mecanismo de busca guiado por um critério de informação que penaliza o ajuste de modelos com muitos parâmetros. O Algoritmo 4 exemplifica o uso desse método, intitulado de Box-Jenkins e introduzido na página 46, para determinar os sete parâmetros do modelo $SARIMA$ de ordem $(p, d, q) \times (P, D, Q)_s$.

No Algoritmo 4, p, d, q, P, D e Q são específicos retardos temporais relevantes da ST dentro de um espaço de busca pré-determinado pelo usuário. O valor de max_p corresponde, em número de observações, a um período sazonal na série e pode ser adquirido, em situações onde essa informação não é evidente, aplicando a técnica do *scatter plot* (Seção 3). Na décima linha, uma condição para inclusão ou omissão da constante δ , que representa o nível inicial do modelo, foi inserida respeitando as regras de diferenciação para o $SARIMA$. Na décima sexta linha, os parâmetros mais promissores são escolhidos de maneira a minimizar o AIC (Equação 38 da página 47).

Como mencionado, os valores de d e D podem ser estabelecidos, em um passo de pré-processamento, integrando a ST até que sua variância se torne menor que a variância calculada sobre sua versão original (não diferenciada). O estabelecimento prévio desses valores é de grande importância, pois contribui para redução do tempo de processamento do Algoritmo 4.

5 Avaliação Experimental

Pesquisas empíricas são necessárias para ancorar e comprovar, por meio de experimentos ou observação, aquilo formalizado conceitualmente. Nesse tipo de pesquisa, as hipóteses são construídas a partir das premissas e as variáveis que possivelmente exercem influência no comportamento do fenômeno investigado são manipuladas. A manipulação na quantidade e qualidade dessas variáveis proporciona o estudo da relação entre causas e efeitos no fenômeno.

Na área de predição de Séries Temporais (ST), avaliações experimentais constituem um importante instrumento da estimativa de que ou quais algoritmos são mais adequados diante de um problema específico. Com base nisso e utilizando-se de 40 conjuntos de dados sintéticos, nesta seção é realizada uma avaliação empírica visando à comparação objetiva e subjetiva entre os

Algoritmo 4: Box-Jenkins Method

```
// S representa uma subsequência de treinamento
/* max_ord1 e max_ord2 são vetores de três posições cujos valores
   indicam os comprimentos de defasagem máximos das partes
   não-sazonal e sazonal, respectivamente */
/* max_p é o número de observações que constituem uma variação
   sazonal na série histórica */
/* P compreende a lista de parâmetros que acarretou no menor erro
   de predição */
Input:  $S, max\_ord1, max\_ord2, max\_p$ 
Output:  $\mathbb{P}$ 
1 begin
   /* realmax corresponde ao maior número do sistema de ponto
      flutuante */
2  $best\_aic \leftarrow realmax;$ 
   // Obtenção do comprimento da subsequência de treino
3  $n \leftarrow length(S);$ 
4 for  $p \leftarrow 0 : max\_ord1[1]$  do
5   for  $d \leftarrow 0 : max\_ord1[2]$  do
6     for  $q \leftarrow 0 : max\_ord1[3]$  do
7       for  $P \leftarrow 0 : max\_ord2[1]$  do
8         for  $D \leftarrow 0 : max\_ord2[2]$  do
9           for  $Q \leftarrow 0 : max\_ord2[3]$  do
10            if  $d + D \leq 1$  then
11               $fit = sarima(S, [p, d, q], [P, D, Q], max\_p, \delta = TRUE);$ 
12            else
13               $fit = sarima(S, [p, d, q], [P, D, Q], max\_p, \delta = FALSE);$ 
14            end
              /* Computação do Critério de Informação de
                 Akaike */
15               $fit\_aic = -2 * fit.loglik + (log(n) + 1) * length(fit.coef);$ 
16              if  $fit\_aic < best\_aic$  then
17                 $best\_aic \leftarrow fit\_aic;$ 
18                 $best\_fit \leftarrow fit;$ 
19                 $best\_model \leftarrow [p, d, q, P, D, Q, max\_p];$ 
20              end
            end
          end
        end
      end
    end
  end
26 end
27  $\mathbb{P} \leftarrow \{best\_aic, best\_fit, best\_model\};$ 
28 return  $\mathbb{P};$ 
29 end
```

métodos descritos na Seção 4. O uso de dados artificiais foi escolhido para possibilitar a execução de experimentos em um ambiente relativamente controlado, o que permite analisar e compreender o desempenho dos algoritmos de predição em ST com características distintas.

5.1 Descrição dos Conjuntos de Dados Sintéticos

Os conjuntos de dados artificiais construídos neste trabalho são comumente referenciados na literatura correlata e podem ser agrupados, conforme seu processo originário, em três categorias: (1) determinística, (2) estocástica e (3) caótica.

5.1.1 Séries Determinísticas

Uma série determinística possui comportamento recorrente, ou seja, ela repete um conjunto de observações no tempo, em escala similar ou considerando diferentes proporções (Box et al., 2015). Como os padrões ocorrem entre intervalos de tempo fixo, esse tipo de ST é caracterizado como previsível e sua avaliação empírica pode contribuir para o entendimento do desempenho dos algoritmos de predição.

Composição de Fourier: As séries dessa categoria possibilitam a representação de funções periódicas, contínuas e discretas, a partir da composição aditiva de sinais cossenoidais e senoidais com distintas amplitudes e frequências. Uma série de Fourier que exibe período T pode ser expressa pela Equação 46 (Bloomfield, 2000), onde a_i e b_j referem-se aos pesos atribuídos à contribuição de cada componente trigonométrico (cosseno e seno) e c indica um valor constante.

$$S_t = \sum_{i=0}^{\infty} a_i \cos\left(\frac{2\pi i t}{T}\right) + \sum_{j=0}^{\infty} b_j \sin\left(\frac{2\pi j t}{T}\right) + c \quad (46)$$

Algumas séries de Fourier podem ser explicadas pela física moderna, como os sinais musicais e elétricos. Nesse sentido, elas podem ser empregadas na resolução de equações diferenciais parciais, por exemplo nas equações que modelam o movimento de uma onda e a condução de calor em um sólido homogêneo. Neste trabalho, utilizando-se da Equação 46 com $t \in [1, 790]$, foram confeccionadas 12 ST cujas estruturas são listadas na Tabela 11.

Todas as equações dispostas na Tabela 11 abrangem o componente de sazonalidade e foram organizadas, de acordo com seus comportamentos, em três grupos: séries sazonais com nível constante, séries sazonais com tendência crescente e séries sazonais com tendência decrescente. Os padrões temporais dessas séries são mostrados graficamente na Figura 35.

Tabela 11: Padrões temporais gerados a partir da composição de Fourier

Composição de Fourier	Sazonalidade Aditiva		
	Nível Constante	Tendência Crescente	Tendência Decrescente
A	$S_t^{(A_1)} = -15\text{sen}\left(\frac{2\pi t}{25}\right) + 25$	$S_t^{(A_2)} = S_t^{(A_1)} + \frac{t}{50}$	$S_t^{(A_3)} = S_t^{(A_1)} - \frac{t}{50}$
B	$S_t^{(B_1)} = 2,5\text{cos}\left(\frac{2\pi t}{25}\right) + 5\text{sen}\left(\frac{2\pi t}{75}\right) + 25$	$S_t^{(B_2)} = S_t^{(B_1)} + \frac{t}{50}$	$S_t^{(B_3)} = S_t^{(B_1)} - \frac{t}{50}$
C	$S_t^{(C_1)} = 3,5\text{cos}\left(\frac{2\pi t}{65}\right) + \text{sen}\left(\frac{2\pi t}{25}\right) + 35$	$S_t^{(C_2)} = S_t^{(C_1)} + \frac{t}{50}$	$S_t^{(C_3)} = S_t^{(C_1)} - \frac{t}{50}$
D	$S_t^{(D_1)} = 12\left[\text{sen}\left(\frac{2\pi t}{75}\right) + \frac{1}{3}\text{sen}\left(\frac{6\pi t}{75}\right)\right] + 35$	$S_t^{(D_2)} = S_t^{(D_1)} + \frac{t}{50}$	$S_t^{(D_3)} = S_t^{(D_1)} - \frac{t}{50}$

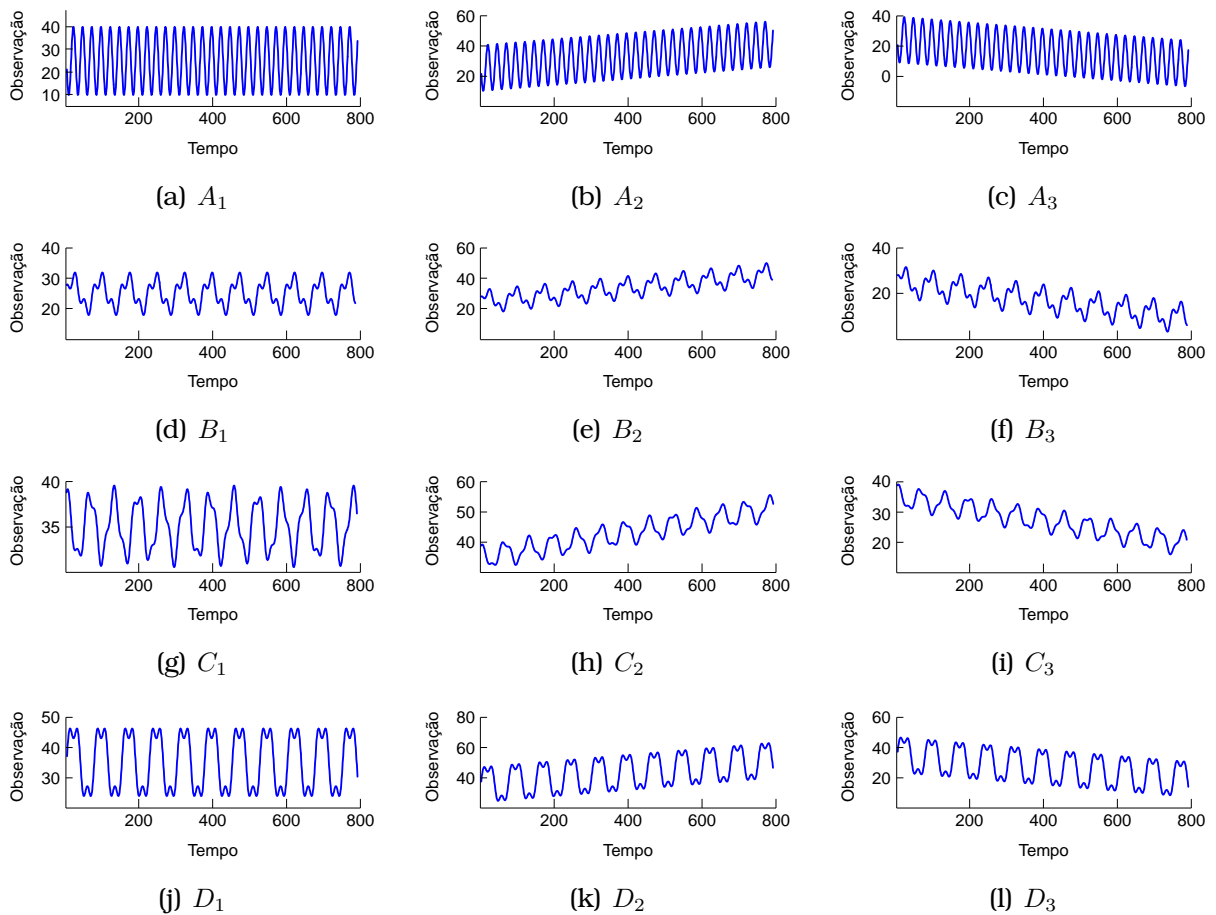


Figura 35: Séries com sazonalidade aditiva derivadas da composição de Fourier

Dependência Sazonal: Esse tipo de padrão temporal, estabelecido pela Equação 47, possui tendência linear e sazonalidade constante (Kulesh et al., 2008).

$$S_t = \cos\left(\frac{t}{25}\right)\text{sen}\left(\frac{t}{100}\right) \quad (47)$$

Na Figura 36 é ilustrada três ST geradas usando a Equação 47 com $t \in$

[1, 2200]. Os comportamentos crescente e decrescente foram obtidos com a referida equação acrescida dos termos $(t \div 1000) + 1$ e $((t \times -1) \div 1000) - 1$, respectivamente. A ausência de variações na amplitude do componente sazonal ao longo do tempo reside em outra característica relevante dessas séries.

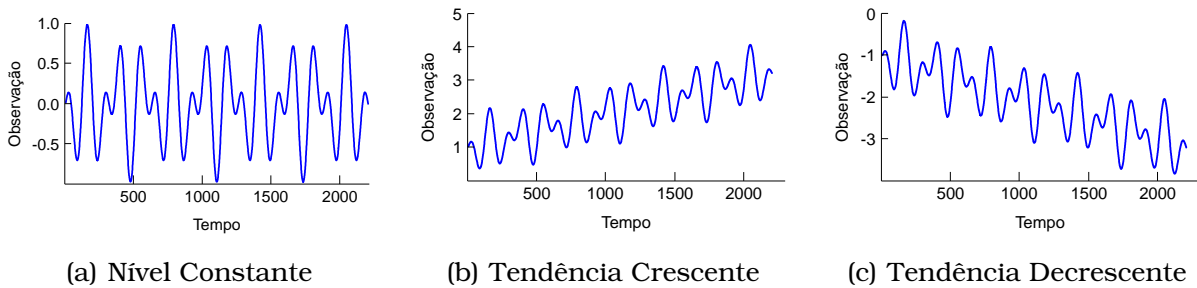


Figura 36: Séries com dependência sazonal

Sazonalidade Multiplicativa: Modelado por meio da Equação 48, esse tipo de padrão temporal abrange uma tendência não-linear e a amplitude das oscilações sazonais aumenta com o tempo (Kulesh et al., 2008).

$$S_t = \begin{cases} R_t, & t \in [1, 79] \\ \frac{(S_{t-t_0})^2}{S_{t-2t_0}}, & t \in [80, 590], \quad t_0 = 14 \end{cases} \quad (48)$$

Complementarmente à Equação 48, tem-se que:

$$R_t = \frac{t}{70000} \left[\cos\left(\frac{9t}{7}\right) \text{sen}\left(\frac{t}{350}\right) + 10 \right]$$

O resultado da aplicação das equações supracitadas é esquematizado na Figura 37. Como pode ser observado nessa figura, a série sintética expõe um crescimento levemente curvo e de ordem quadrática.

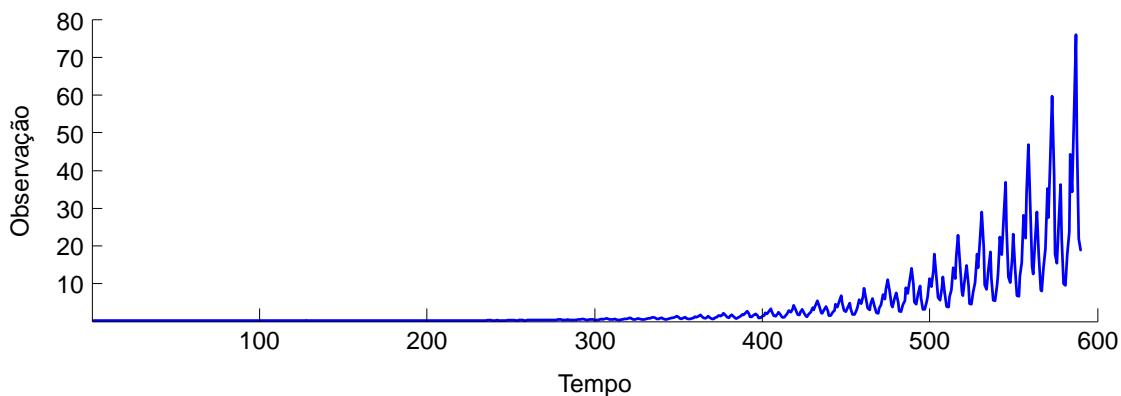


Figura 37: Série com sazonalidade multiplicativa

Alta Frequência: Descritos pela Equação 49, os dados desse padrão temporal assumem sazonalidade multiplicativa e constante aumento de amplitude (Kulesh et al., 2008).

$$S_t = \frac{t}{100} |\text{sen}(t/2)| + |\text{cos}(t/20)|, \quad t \in [0, 550] \quad (49)$$

Na Figura 38 é exibida uma série sintética constituída de 550 amostras que retrata o padrão de alta frequência.

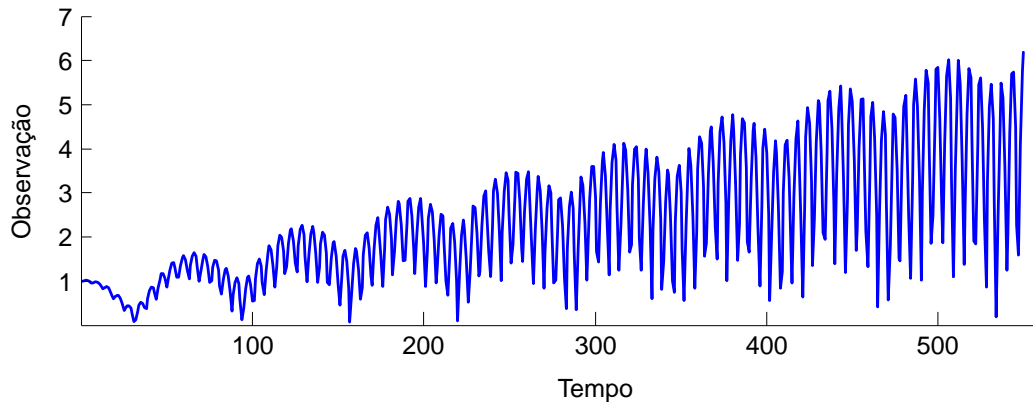


Figura 38: Série de alta frequência com sazonalidade multiplicativa

5.1.2 Séries Estocásticas

Séries estocásticas são originárias de eventos aleatórios e, conseqüentemente, imprevisíveis. A imprevisibilidade é uma propriedade que dificulta a modelagem de dados temporais, haja vista que conhecer o comportamento passado da série já não é suficiente para prever sua evolução no futuro.

Gráficos de Controle: Esse tipo de gráfico permite monitorar o comportamento de um processo por meio da construção de uma faixa de valores estatisticamente limitada por duas linhas, uma superior e outra inferior. Uma terceira linha pode ainda ser incluída no gráfico a fim de possibilitar a verificação do comportamento médio do processo. As equações que reproduzem os padrões temporais usualmente observados em gráficos de controle são listadas na Tabela 12. Nessa tabela, as equações foram agrupadas em duas categorias: Gráficos de Controle A (GCA), que contém as equações originais de modelagem decorrentes da literatura (Montgomery, 2009); e Gráficos de Controle B (GCB), que abrange modificações das equações originais.

Nas equações organizadas na Tabela 12, y_t indica o valor amostrado no instante de tempo t , μ e σ representam, nessa ordem, a média e o desvio padrão do processo simulado, r compreende uma função geradora

Tabela 12: Padrões temporais comumente observados em gráficos de controle

Padrão Temporal	GCA	GCB
Nível Constante	$y_t = \mu + r_t\sigma$	$y_t = \mu + r_t + \sigma \text{sen}\left(\frac{2\pi t}{T}\right)$
Padrões Sazonais	$y_t = \mu + r_t\sigma + a \text{sen}\left(\frac{2\pi t}{T}\right)$	$y_t = \mu + r_t + \sigma \text{sen}\left(\frac{2\pi t}{200}\right)$
Tendência Crescente	$y_t = \mu + r_t\sigma + gt$	$y_t = \mu + r_t + \sigma \text{sen}\left(\frac{2\pi t}{T}\right) + gt$
Tendência Decrescente	$y_t = \mu + r_t\sigma - gt$	$y_t = \mu + r_t + \sigma \text{sen}\left(\frac{2\pi t}{T}\right) - gt$
Deslocamento para Cima	$y_t = \mu + r_t\sigma + bs$	$y_t = \mu + r_t + \sigma \text{sen}\left(\frac{2\pi t}{T}\right) + bs$
Deslocamento para Baixo	$y_t = \mu + r_t\sigma - bs$	$y_t = \mu + r_t + \sigma \text{sen}\left(\frac{2\pi t}{T}\right) - bs$

de números aleatórios normalmente distribuídos no intervalo de $[-3, 3]$, α reflete a amplitude das variações sazonais, T expressa o período do componente sazonal, g corresponde ao gradiente do componente de tendência, b refere-se ao deslocamento dos dados para cima ou para baixo ($b = 0$ antes da mudança e $b = 1$ durante e após a ocorrência desse padrão) e s simboliza a magnitude do deslocamento. Neste trabalho, utilizando-se das equações dispostas na Tabela 12, foram geradas 12 ST sintéticas compostas por 1000 amostras cada. Em relação aos parâmetros dessas equações, foram adotados os seguintes valores: $\mu = 80$, $\sigma = 5$, $\alpha = 15$, $T = 30$, $g = 0,5$, $s = 20$ e b assumiu valor unitário no momento $t = 500$.

As séries provenientes da aplicação das equações catalogadas como GCA são retratadas na Figura 39, enquanto que as séries geradas usando as equações da categoria GCB são esquematizadas na Figura 40.

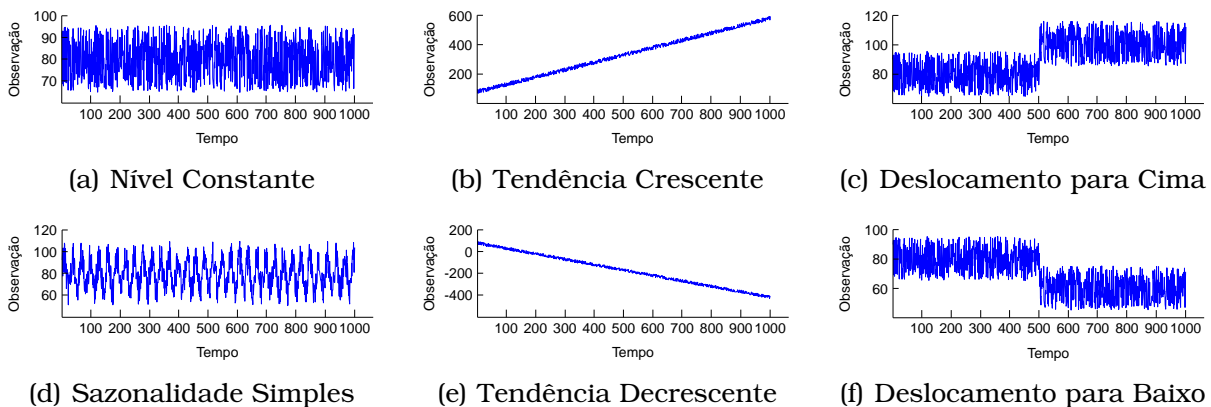


Figura 39: Padrões temporais provenientes da categoria GCA

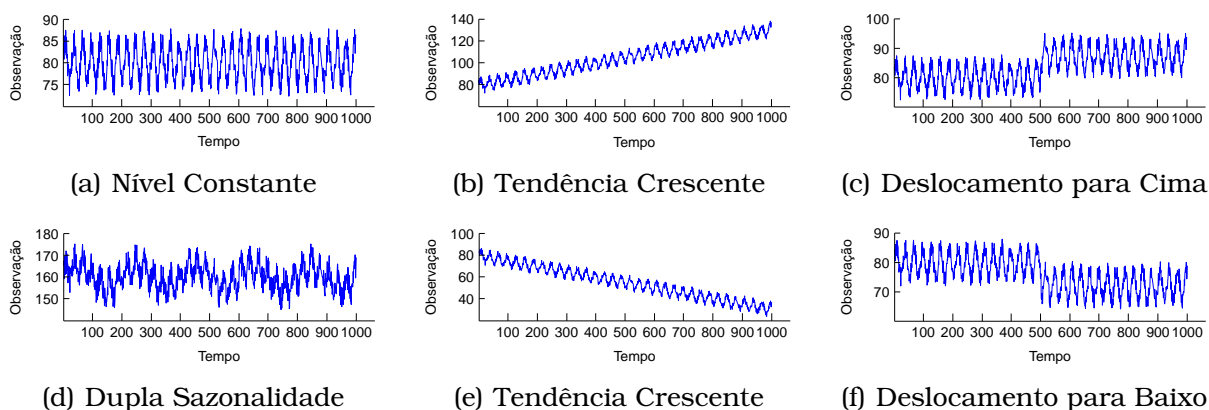


Figura 40: Padrões temporais provenientes da categoria GCB

Dependência Sazonal e Ruído: Esse padrão temporal é resultado da utilização da Equação 47 acrescida de uma função r que produz, a cada instante t , um número aleatório (ruído). Na Figura 41, para $t \in [1, 2200]$ e r no intervalo de $[-1/7, 1/7]$, são mostradas as séries com dependência sazonal e ruído confeccionadas. Os comportamentos crescente e decrescente foram obtidos do mesmo modo que o padrão de dependência sazonal exposto na página 69.

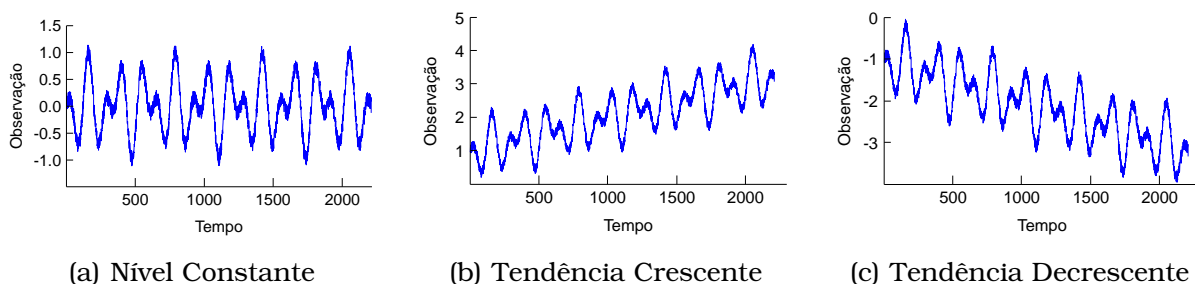


Figura 41: Séries com dependência sazonal e ruído

5.1.3 Séries Caóticas

Séries caóticas são habitualmente confundidas com séries estocásticas (não-determinísticas), uma vez que ambas carecem de regularidade. A diferença crucial entre essas duas categorias está no fato de que as séries estocásticas são provenientes de um processo aleatório, como o ruído branco, enquanto que as séries caóticas são geradas de modo determinístico a partir de sistemas dinâmicos. Um sistema que possui dinâmica caótica tem como principal característica sua sensibilidade à variação das condições iniciais e ao parâmetro de controle, ou seja, uma pequena variação nesses fatores faz com que o sistema evolua de forma imprevisível (Strogatz, 2014).

Mapa Logístico: Introduzido no tema de Teoria do Caos, o mapa logístico foi originalmente proposto para mapear o crescimento populacional de

espécies de insetos. O mapeamento, representado de maneira discreta pela Equação 50 (Verhulst, 1845), determina a quantidade de indivíduos em um dado instante (x_{n+1}) a partir de informações do momento anterior (x_n). Essas informações são associadas à taxa de crescimento dos indivíduos (r), a qual é comumente chamada de potencial biótico da população.

$$x_{t+1} = rx_t(1 - x_t) \quad (50)$$

Geometricamente, o mapa logístico pode ser visibilizado como uma parábola cuja concavidade é definida segundo o valor de r . Dependendo do valor assumido por esse parâmetro, o comportamento futuro do mapa perde a regularidade e passa a ser altamente sensível às condições iniciais. Por isso, na proposta original, assume-se que um limite máximo de população pode ser suportado pelo meio ambiente. A ultrapassagem desse limite pelos indivíduos poderá implicar em um desastre, como a rápida diminuição de alimentos, o que corrobora para a extinção das espécies.

Na Figura 42 é mostrada uma série sintética, constituída por 551 amostras, resultante da aplicação da Equação 50 com $r = 4$ e $x_1 = 0,01$.

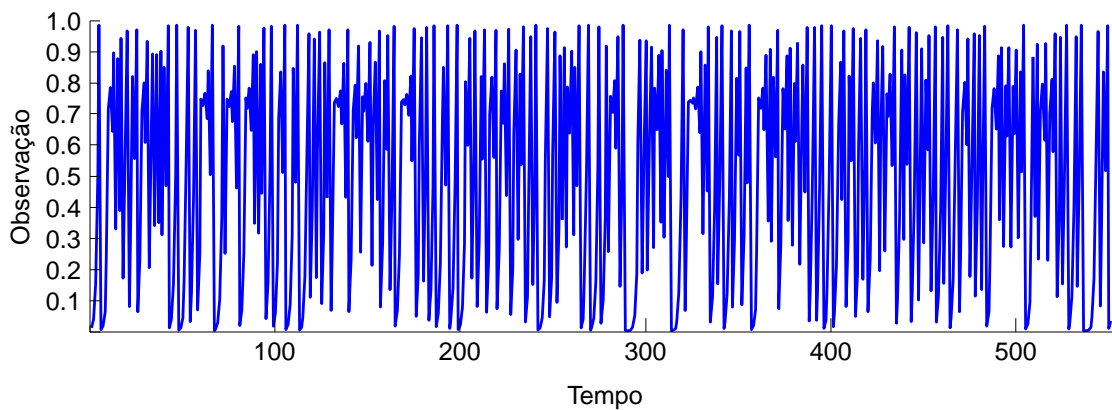


Figura 42: Mapa Logístico

Mapa de Hénon: Consiste de um sistema discreto bidimensional, expresso por meio da Equação 51 (Strogatz, 2014), que não foi derivado de nenhum fenômeno natural em particular, mas sim proposto como um modelo para descrever a conjectura de Poincaré do sistema em tempo contínuo de Lorenz. Esse mapa (quadrático e bidimensional) é representado na área de sistemas dinâmico pelas seguintes equações:

$$\begin{cases} x_{t+1} = 1 - ax_t^2 + by_t \\ y_{t+1} = x_t \end{cases} \quad (51)$$

Na Equação 51, x_t e y_t denotam os valores das variáveis dinâmicas no instante de tempo t , a indica o parâmetro de não-linearidade e b denota o parâmetro de dissipação do sistema. Para determinados conjuntos de valores de parâmetros o mapa de Hénon pode apresentar comportamento caótico, intermitente ou convergir para uma órbita periódica.

Na Figura 43 é ilustrada uma série sintética, constituída por 3000 amostras, decorrente do mapa de Hénon considerando $a = 1,4$, $b = 0,3$, e $x_1 = y_1 = 0,1$. Um exemplo de atrator caótico do mapa de Hénon também é retratado nessa figura.

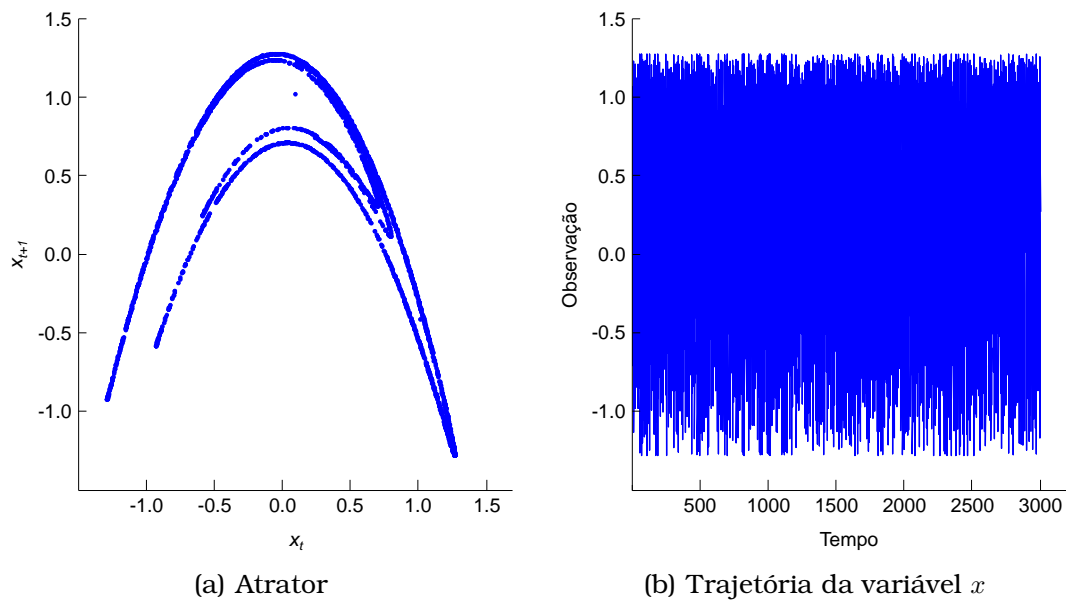


Figura 43: Mapa de Hénon

Sistema de Mackey-Glass: Permite a criação de séries caóticas por meio da integração da Equação 52 (Glass e Mackey, 1997), a qual foi originalmente desenvolvida para modelar a formação de linfócitos. Na referida equação diferencial, os termos β , θ , n , e γ são argumentos de ajuste do sistema.

$$\frac{d(x)}{dt} = \frac{\beta\theta^n x(t - \tau)}{x(t - \tau)^m + \theta^m} - \gamma x \quad (52)$$

O parâmetro de bifurcação τ permite a navegação do valor do sistema entre estados estáveis, de ciclo limite e instáveis. Neste trabalho, assumindo $\beta = 0,2$, $\theta = 1$, $n = 10$, $\tau = 17$ e $\gamma = 0,1$, foi construída uma série sintética composta por 3000 amostras. Essa série, visibilizada na Figura 44, foi obtida com auxílio do método iterativo de Runge-Kutta de quarta ordem (Christodoulou, 2009) para a aproximação numérica da Equação 52.

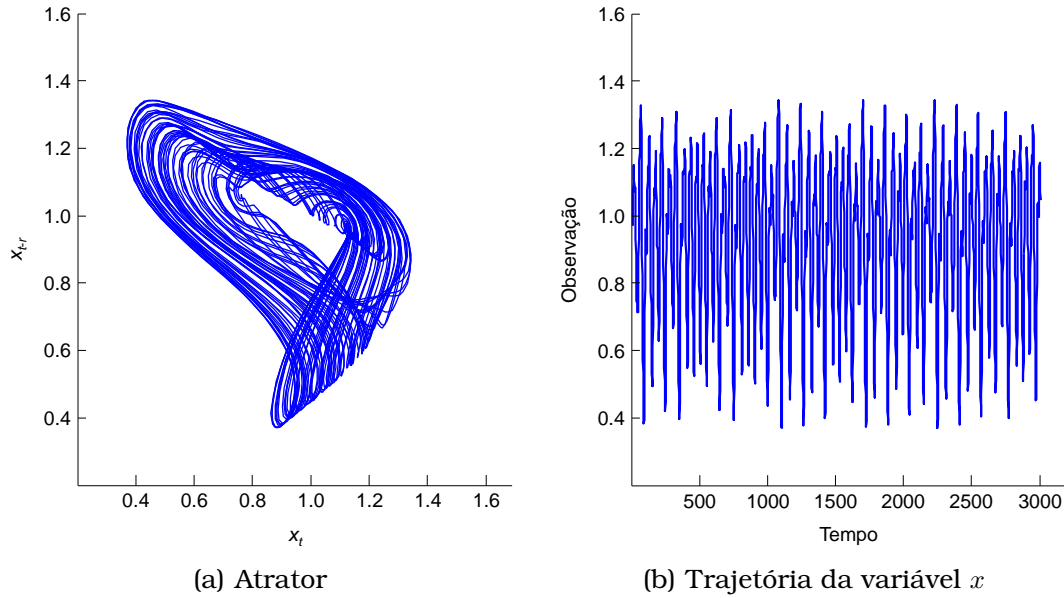


Figura 44: Sistema de Mackey-Glass

Sistema de Lorenz: Esse sistema foi derivado de um modelo matemático tridimensional relacionado ao estudo do movimento convectivo de fluidos na atmosfera terrestre. O sistema de Lorenz é composto por três equações que podem ser expressas do seguinte modo (Tahersima et al., 2010):

$$\begin{cases} x(t+1) = x(t) + d\alpha(y(t) - x(t)) \\ y(t+1) = y(t) + d(x(t)(r - z(t)) - y(t)) \\ z(t+1) = z(t) + d(x(t)y(t) - \beta z(t)) \end{cases} \quad (53)$$

Nesse sistema, x corresponde à intensidade dos movimentos convectivos. Esses movimentos acontecem em sentido horário quando $x > 0$, caso contrário eles ocorrem em sentido anti-horário. Enquanto y reflete a diferença de temperatura entre as correntes ascendentes e descendentes, z é proporcional à distorção no perfil de temperatura vertical. Os parâmetros α e β representam os números de Prandtl e Rayleigh, respectivamente. Por fim, r é um fator geométrico e d indica o tempo de amostragem (Lorenz, 1963).

Na Figura 45 é esquematizado o atrator e a série sintética de 3000 pontos com comportamento não-periódico e imprevisível obtidos por meio da aplicação do sistema de Lorenz usando $\alpha = 10$, $\beta = 8/3$, $r = 28$, $d = 0,01$ e $x_1 = y_1 = z_1 = 1$.

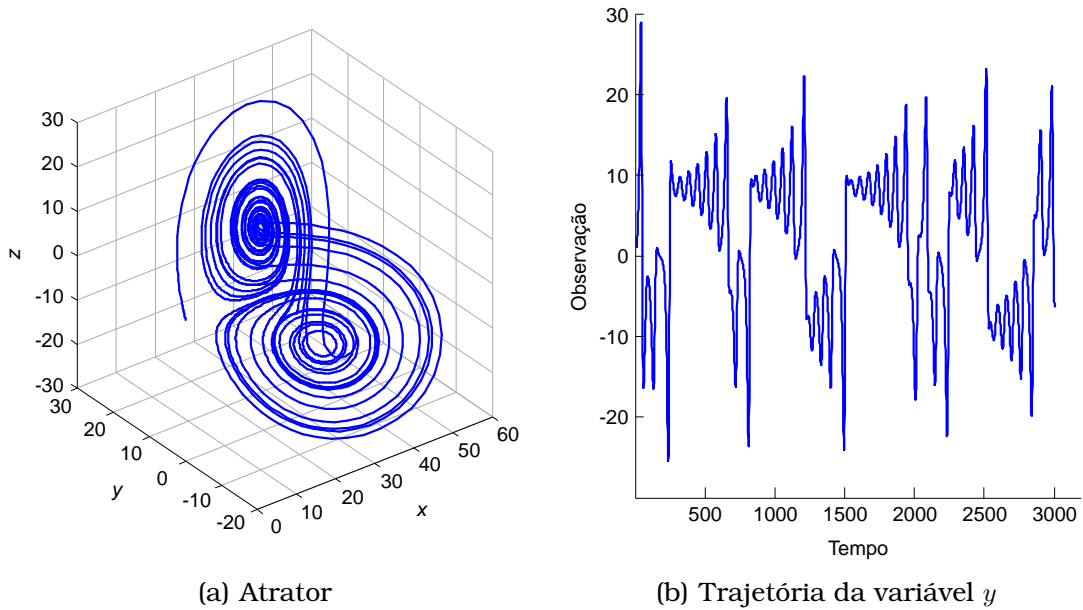


Figura 45: Sistema de Lorenz

Sistema de Rössler: Formulado com base no modelo de Lorenz, o sistema de Rössler não objetiva a simulação de nenhum fenômeno físico real conhecido e pode ser determinado em conformidade com as seguintes equações (Frunzete et al., 2015):

$$\begin{cases} x(t+1) = x(t) - d(y(t) + z(t)) \\ y(t+1) = y(t) + d(x(t) + ay(t)) \\ z(t+1) = z(t) + d(b + z(t)(x(t) - c)) \end{cases} \quad (54)$$

Nessas equações, a varável d representa o tempo de amostragem e os termos a , b e c correspondem aos parâmetros de controle do sistema, os quais devem assumir valores obrigatoriamente positivos. A fim de gerar uma série composta por 3000 valores, as equações de Rössler foram empregadas adotando $a = b = 0,2$, $c = 8,7$, $d = 0,02$, $x_1 = -2$, $y_1 = -10$ e $z_1 = 0,2$. O resultado dessa aplicação é mostrado graficamente na Figura 46.

Sinais Caóticos: Esses sinais, além de apresentarem dependência sensível às condições iniciais e parametrizações, são determinísticos e aperiódicos. O estudo dos sinais caóticos é útil, sobretudo, para algumas subáreas da Engenharia de Telecomunicações por abranger espectro de Fourier plano, dificuldade de predição e serem facilmente confundíveis com ruído (Eisencraft, 2001). Neste trabalho, o Sinal Caótico A (SCA) foi construído utilizando a Equação 55, enquanto que o Sinal Caótico B (SCB) foi construído usando a Equação 56.

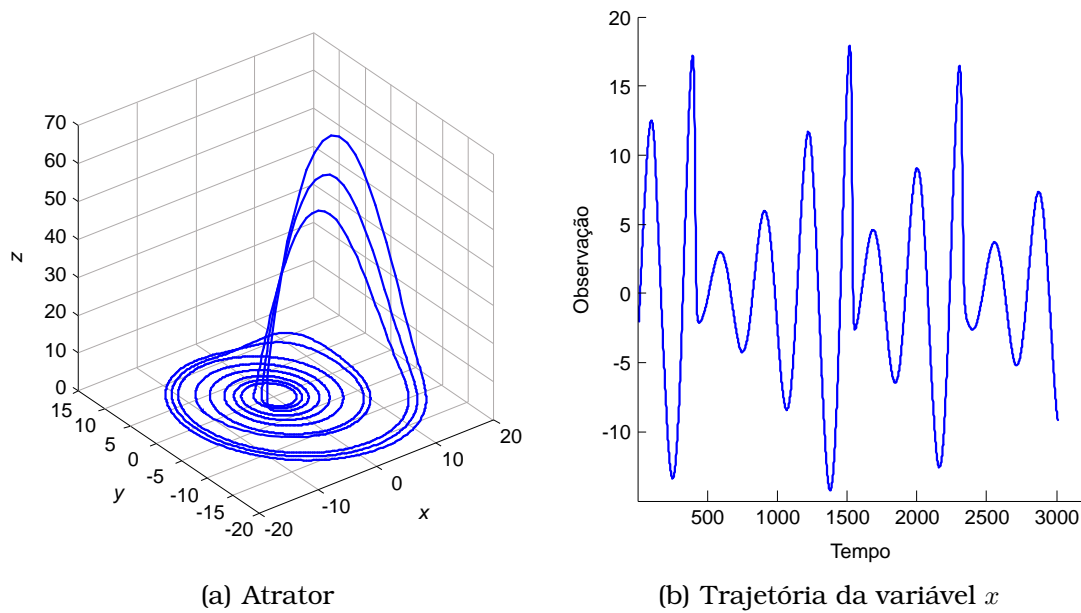


Figura 46: Sistema de Rössler

$$S = \frac{5}{2} \text{sen}\left(\frac{2\pi t}{100}\right) + \frac{3}{2} \text{sen}\left(\frac{2\pi t}{41}\right) + \text{sen}\left(\frac{2\pi t}{21}\right) \quad (55)$$

$$x_t = \frac{x_{t-1}}{2} + 25 \frac{x_{t-1}}{1 + x_{t-1}^2} + 7 \cos(1,2t) \quad (56)$$

Na Figura 47 são exibidos graficamente os sinais caóticos sintéticos projetados para 550 observações. É importante destacar que a aplicação da Equação 56 requer a definição de um valor para a variável x no instante $t = 1$. Nesse contexto, adotou-se $x_1 = 15$ para a confecção do SCB. É interessante notar que as séries geradas permitem avaliar os algoritmos de predição diante de comportamentos pouco previsíveis e que contemplam ciclos não repetitivos.

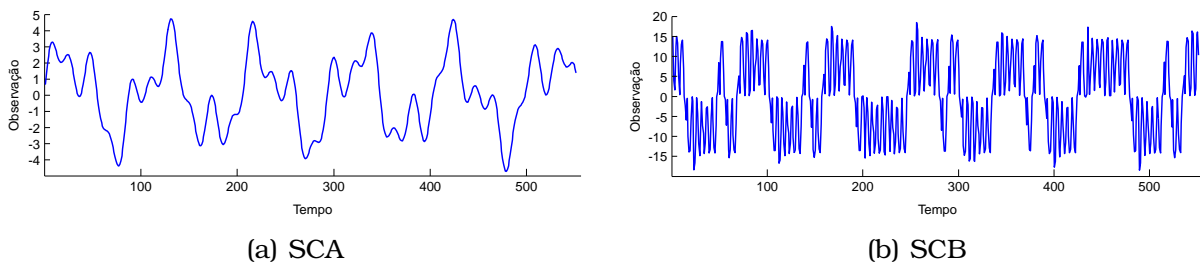


Figura 47: Sinais Caóticos

ECGSYN: Série constituída por 3000 observações e cujos valores foram obtidos a partir de um simulador de eletrocardiograma sustentado por três equações diferenciais ordinárias (McSharry et al., 2003). O comportamento dessa série é ilustrado na Figura 48.

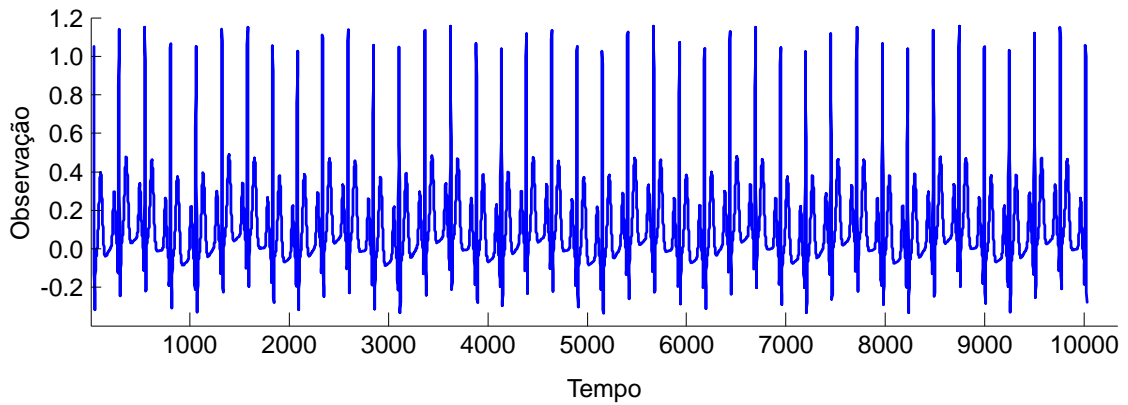


Figura 48: Série de valores ECGSYN

5.2 Configuração dos Experimentos

O protocolo para avaliação do desempenho preditivo dos algoritmos investigados foi organizado em três etapas, as quais são ilustradas na Figura 49.

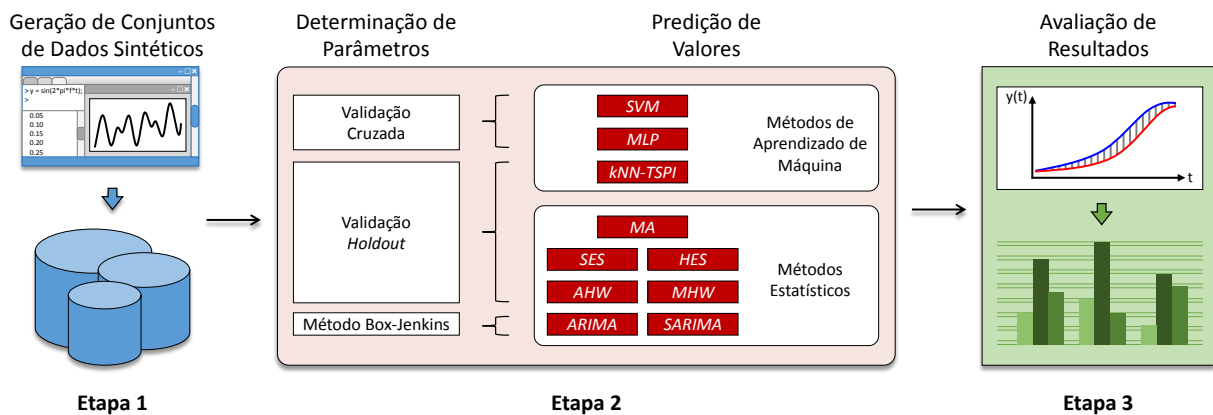


Figura 49: Configuração experimental

Na Etapa 1, tal como introduzido na página 68, houve a construção de 40 conjuntos de dados sintéticos. Um resumo das particularidades e configurações dessas sequências de dados artificiais é mostrado na Tabela 13. Nessa tabela, para cada conjunto de dados gerado, são exibidos o tamanho da ST sintética (m), o número máximo de observações que constituem um período sazonal (max_p) na série e o horizonte de predição (h), o qual corresponde à quantidade de valores a serem preditos. É importante notar que h foi fixado para equivaler à 5% do tamanho da sequência de dados ($h = m \times 5\%$).

Na Etapa 2, os parâmetros de ajuste foram estimados em conformidade com as características individuais de cada método. Nesse contexto, os algoritmos não-paramétricos *SVM* e *MLP*, os quais são aplicados de acordo com a abordagem global, tiveram seus parâmetros determinados por meio de validação cruzada em dez partições com minimização do Erro Quadrático Médio (*MSE*) (Algoritmo 3 da página 65). Já os parâmetros dos modelos paramétricos *ARIMA* e *SARIMA* foram definidos usando o método Box-Jenkins

Tabela 13: Sumário de características e de configurações dos conjuntos de dados sintéticos

ID	Conjunto de Dados	m	max_p	$h (m \times 5\%)$
	Fourier A:	790		
01.D	• Nível Constante		25	40
02.D	• Tendência Crescente		25	40
03.D	• Tendência Decrescente		25	40
	Fourier B:	790		
04.D	• Nível Constante		38	40
05.D	• Tendência Crescente		38	40
06.D	• Tendência Decrescente		38	40
	Fourier C:	790		
07.D	• Nível Constante		34	40
08.D	• Tendência Crescente		34	40
09.D	• Tendência Decrescente		34	40
	Fourier D:	790		
10.D	• Nível Constante		38	40
11.D	• Tendência Crescente		38	40
12.D	• Tendência Decrescente		38	40
	Dependência Sazonal:	2200		
13.D	• Nível Constante		25	110
14.D	• Tendência Crescente		25	110
15.D	• Tendência Decrescente		25	110
16.D	Sazonalidade Multiplicativa	590	14	30
17.D	Alta Frequência	550	63	28
	GCA:	1000		
18.E	• Nível Constante		12	50
19.E	• Padrões Sazonais		30	50
20.E	• Tendência Crescente		12	50
21.E	• Tendência Decrescente		12	50
22.E	• Deslocamento para Cima		12	50
23.E	• Deslocamento para Baixo		12	50
	GCB:	1000		
24.E	• Nível Constante		30	50
25.E	• Dupla Sazonalidade		30	50
26.E	• Tendência Crescente		30	50
27.E	• Tendência Decrescente		30	50
28.E	• Deslocamento para Cima		30	50
29.E	• Deslocamento para Baixo		30	50
	Dependência Sazonal e Ruído:	2200		
30.E	• Nível Constante		25	110
31.E	• Tendência Crescente		25	110
32.E	• Tendência Decrescente		25	110
33.C	Mapa Logístico	550	4	28
34.C	Mapa de Hénon	3000	3	150
35.C	Sistema de Mackey-Glass	3000	7	150
36.C	Sistema de Lorenz	3000	25	150
37.C	Sistema de Rössler	3000	14	150
	Sinais Caóticos:	550		
38.C	• A		22	28
39.C	• B		7	28
40.C	ECGSYN	3000	60	150

com minimização do Critério de Informação de Akaike (*AIC*) e máxima verossimilhança (Algoritmo 4 da página 67). Os métodos restantes tiveram seus parâmetros estabelecidos empregando validação *holdout* com minimização do *MSE* (Algoritmo 2 da página 64).

Os algoritmos de predição adotados na Etapa 2, em conjunto com seus

parâmetros, são listados na Tabela 14. Nessa tabela, na terceira coluna, são indicados os intervalos de variação numérica, estipulados para cada parâmetro, que foram testados pelas técnicas de estimação.

Tabela 14: Algoritmos utilizados e intervalos de variação numérica definidos para os seus parâmetros

Algoritmo	Parâmetro	Varição (início : passo : fim)
SVM	Tamanho da janela de busca (l)	$l = 3 : 2 : max_p$
	Largura da gaussiana da função <i>kernel</i> de base radial (σ)	$\sigma = 0,005 : 0,05 : 0,25$
MLP	Tamanho da janela de busca (l)	$l = 3 : 2 : max_p$
	Número de unidades na camada oculta (n)	$n = 3 : 2 : max_p$
kNN-TSPI	Tamanho da janela de busca (l) Quantidade de vizinhos mais próximos (k)	$l = 3 : 2 : max_p$ $k = 1 : 2 : 9$
MA	Quantidade de observações utilizadas na média (r)	$r = 3 : 2 : max_p$
SES	Constante de suavização associada ao nível (α)	$\alpha = 0 : 0,25 : 1$
HES	Constante de suavização associada ao nível (α)	$\alpha = 0 : 0,25 : 1$
	Constante de suavização associada à tendência (β)	$\beta = 0 : 0,25 : 1$
AHW	Constante de suavização associada ao nível (α)	$\alpha = 0 : 0,25 : 1$
MHW	Constante de suavização associada à tendência (β)	$\beta = 0 : 0,25 : 1$
	Constante de suavização associada à sazonalidade (γ)	$\gamma = 0 : 0,25 : 1$
	Quantidade de observações que compõe um período sazonal (s)	$s = 3 : 2 : max_p$
ARIMA	Ordem do procedimento de autorregressão (p)	$p = 0 : 1 : \sqrt{\log(m-h)}$
	Grau do operador de diferenciação (d)	$d = 0 : 1 : 2$
	Ordem do procedimento de médias móveis (q)	$q = 0 : 1 : \sqrt{\log(m-h)}$
SARIMA	Ordem do procedimento de autorregressão (p)	$p = 0 : 1 : \sqrt{\log(m-h)}$
	Grau do operador de diferenciação (d)	$d = 0 : 1 : 2$
	Ordem do procedimento de médias móveis (q)	$q = 0 : 1 : \sqrt{\log(m-h)}$
	Ordem do procedimento de autorregressão sazonal (P)	$P = 0 : 1 : \sqrt{\log(m-h)}$
	Grau do operador de diferenciação sazonal (D)	$D = 0 : 1 : 2$
	Ordem do procedimento de médias móveis sazonal (Q)	$Q = 0 : 1 : \sqrt{\log(m-h)}$
	Quantidade de observações que compõe um período sazonal (s)	$s = max_p$

Após a seleção dos melhores parâmetros, os modelos preditivos foram construídos e ajustados aos dados de treinamento. Cada modelo confeccionado foi, posteriormente, extrapolado h períodos à frente conforme as duas estratégias de projeção relatadas na Seção 4 e designadas de:

- Multi-etapa à frente com passo aproximado;
- Multi-etapa à frente com passo atualizado.

Na Etapa 3, os dados projetados foram comparados com os dados de teste por meio da utilização da medida MSE , do coeficiente U de Theil (TU) e do índice de desempenho *Prediction Of Change In Direction* ($POCID$).

A medida MSE , denotada pela Equação 57 (Hyndman e Koehler, 2006), trabalha com a diferença (erro) entre o valor real observado (z_t) e o valor predito (\hat{z}_t).

$$MSE = \frac{1}{h} \sum_{t=1}^h (z_t - \hat{z}_t)^2 \quad (57)$$

Na Equação 57, a soma quadrática dos erros de predição é dividida pela quantidade de observações analisadas. Logo, espera-se que a soma de todos os erros computados seja próxima do valor zero. Nessa mesma perspectiva, o coeficiente TU , expresso pela Equação 58 (Theil, 1971), baseia-se no MSE do preditor, normalizado pelo erro de predição de um modelo trivial ou ingênuo. O modelo trivial assume que o melhor valor para o tempo $t + 1$ é o valor obtido no tempo t .

$$TU = \frac{\sum_{t=1}^h (z_t - \hat{z}_t)^2}{\sum_{t=1}^h (z_t - z_{t-1})^2} \quad (58)$$

De acordo com o resultado da Equação 58, se $TU > 1$, o desempenho do algoritmo de interesse é inferior ao do modelo ingênuo; se $TU = 1$, o desempenho do algoritmo investigado é igual ao do modelo ingênuo; se $TU < 1$, o desempenho do algoritmo de interesse é superior ao do modelo ingênuo; e se $TU \leq 0,55$, o algoritmo investigado é confiável para efetuar predições futuras.

Outro índice de desempenho considerado foi o $POCID$, que é formalizado pela Equação 59. Nessa equação, o termo D_t armazena o valor 1 quando $(\hat{z}_t - \hat{z}_{t-1})(z_t - z_{t-1}) > 0$, e o valor 0 caso contrário. A ideia por trás desse índice é estimar a precisão das alterações da direção dos dados projetados, ou seja, se o valor futuro irá aumentar ou diminuir em relação ao valor atual.

$$POCID = \frac{\sum_{t=1}^h D_t}{h} \times 100 \quad (59)$$

A partir dos valores das medidas de avaliação adotadas foi possível confrontar os algoritmos estudados. Tais confrontos foram analisados empregando o teste estatístico não-paramétrico de Friedman para dados emparelhados e comparações múltiplas, com nível de significância de 5% (p -valor $< 0,05$), seguido do pós-teste de Nemenyi²⁰.

A execução do protocolo experimental exposto abrangeu o uso dos seguintes ambientes computacionais e linguagens de alto nível: (1) MATLAB²¹ e GNU Octave²², assim como seus respectivos pacotes de funções relacionados ao tema de predição de ST; (2) linguagem R²³ aliada ao pacote *Forecast*²⁴, ambos integrados ao ambiente de desenvolvimento RStudio²⁵; e (3) linguagem Java (Deitel e Deitel, 2012), no ambiente *Eclipse*²⁶, com a biblioteca Weka (Witten et al., 2011).

²⁰Testes estatísticos realizados utilizando *KEEL Software Tool* para Windows, <http://www.keel.es>.

²¹<http://www.mathworks.com/products/matlab>.

²²<https://www.gnu.org/software/octave>.

²³<https://www.r-project.org>.

²⁴<https://cran.r-project.org/web/packages/forecast>.

²⁵<https://www.rstudio.com/>.

²⁶<http://www.eclipse.org>.

5.3 Resultados e Discussão

Os testes computacionais foram executados utilizando 40 sequências de dados sintéticas, das quais 17 são determinísticas, 15 são estocásticas e oito são caóticas. Os resultados, além de examinados individualmente para cada um dos grupos de ST, foram percorridos considerando todos os conjuntos de dados.

É importante frisar que, para cada grupo de experimentos, as configurações investigadas foram comparadas, a partir das medidas *MSE*, *TU* e *POCID*, usando o teste estatístico de Friedman, com nível de significância de 5%, seguido do pós-teste de Nemenyi. Os resultados dessa validação estatística foram esquematizados em diagramas de Distância Crítica (*CD*) (Demšar, 2006) cuja escala indica o *ranking* do desempenho médio de cada configuração. Ainda nesses diagramas, as configurações unidas por uma linha ondulada não demonstram diferenças estatisticamente significativas de qualidade.

5.3.1 Análise Individual

Os experimentos conduzidos sobre as sequências de dados determinísticas envolveram 340 configurações (10 modelos preditivos \times 2 estratégias de predição \times 17 conjuntos de dados). Na Figura 50 são mostrados os diagramas de distância crítica com relação aos valores das medidas *MSE*, *TU* e *POCID* procedentes das configurações pesquisadas.

Em conformidade com as escalas representadas na Figura 50, o *kNN-TSPI* com passo aproximado ficou em terceiro lugar segundo as medidas *MSE* (Figura 50(a)) e *TU* (Figura 50(c)), perdendo sem diferenças estatisticamente significativas para os modelos *SARIMA* e *SVM*. Já para o índice *POCID* (Figura 50(e)), o *kNN-TSPI* com passo aproximado não conseguiu, por uma pequena margem de diferença, superar as taxas de acerto obtidas pelo método *SARIMA*.

O algoritmo *MLP* com passo atualizado apresentou, em média, os melhores resultados utilizando as medidas *MSE* (Figura 50(b)) e *TU* (Figura 50(d)). No entanto, quando analisadas as taxas de acerto quanto à tendência dos horizontes projetados (Figura 50(f)), essa configuração ocupou a segunda posição no *ranking* médio, não exibindo diferenças estatisticamente significativas em comparação com o modelo *SVM*.

Os experimentos realizados usando as sequências de dados estocásticas englobaram 300 configurações (10 modelos preditivos \times 2 estratégias de predição \times 15 conjuntos de dados). Os resultados desses testes computacionais, expressos conforme os índices *MSE*, *TU* e *POCID*, são indicados nos diagramas de distância crítica da Figura 51.

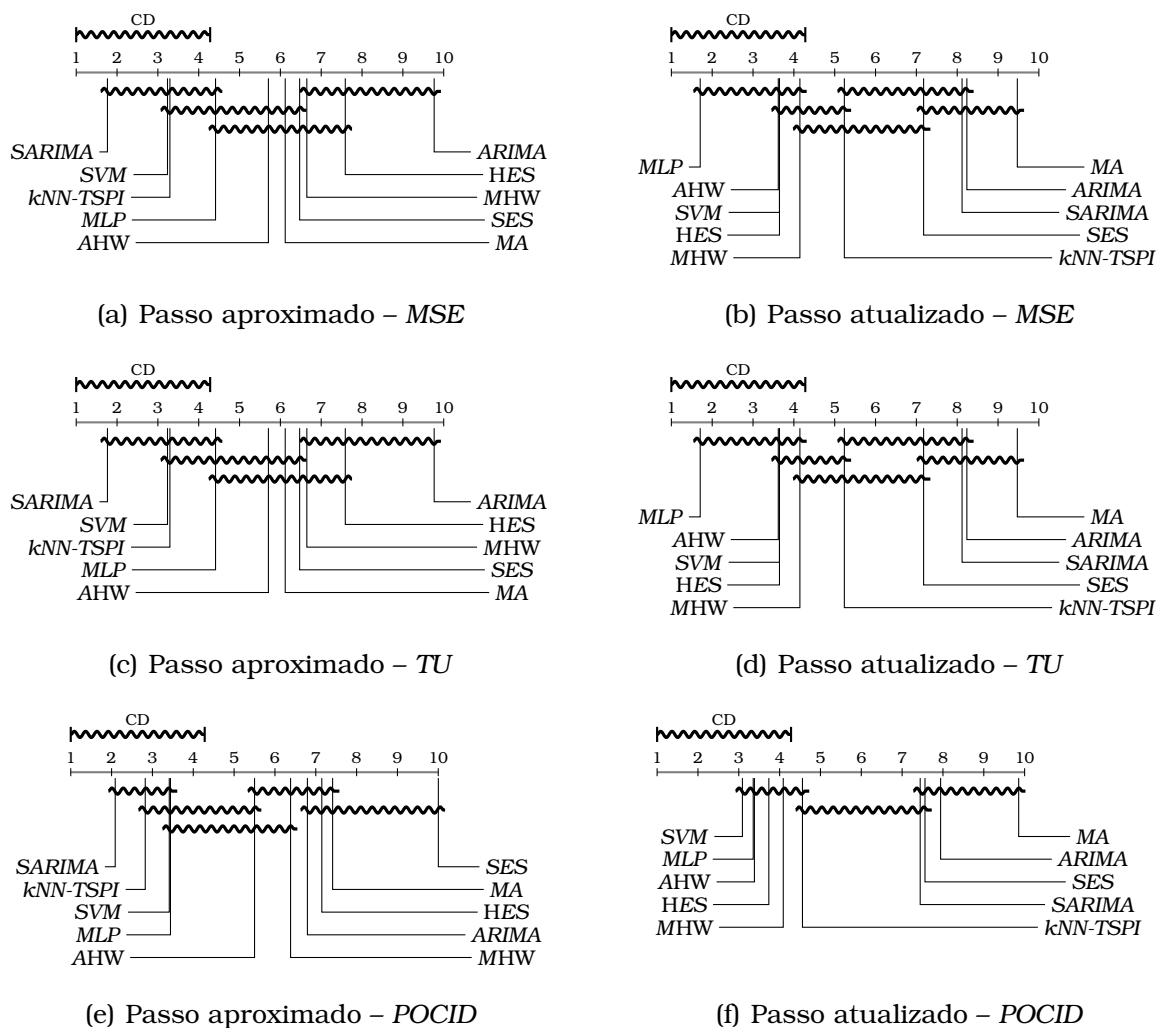


Figura 50: Diagramas de distância crítica para os valores dos índices *MSE*, *TU* e *POCID* provenientes dos métodos de predição sobre ST determinísticas

Na Figura 51, o algoritmo *kNN-TSPI* com passo aproximado ocupou a terceira posição no *ranking* dos valores de *MSE* (Figura 51(a)) e *TU* (Figura 51(c)). Tal configuração perdeu, nessa mesma perspectiva e por uma pequena margem de diferença, para os modelos *SARIMA* e *SVM*. Particularmente, o *SVM* com passo aproximado apresentou, em média, as melhores taxas de acerto *POCID* (Figura 51(e)), seguido do método *SARIMA*.

Para o cenário no qual os modelos são atualizados a cada nova predição com os valores reais observados, o algoritmo *SVM* expôs o terceiro melhor resultado tanto pela medida *MSE* (Figura 51(b)) quanto pelo coeficiente *TU* (Figura 51(d)). Ainda considerando essas duas medidas de erro, os modelos *SARIMA* e *ARIMA* assumiram, respectivamente, a primeira e segunda posições nos *rankings* de desempenho. No que diz respeito ao índice *POCID*, o *kNN-TSPI* superou, sem diferenças estatisticamente significativas, o desempenho dos métodos *SARIMA* e *SVM*.

Os experimentos conduzidos sobre as sequências de dados caóticas con-

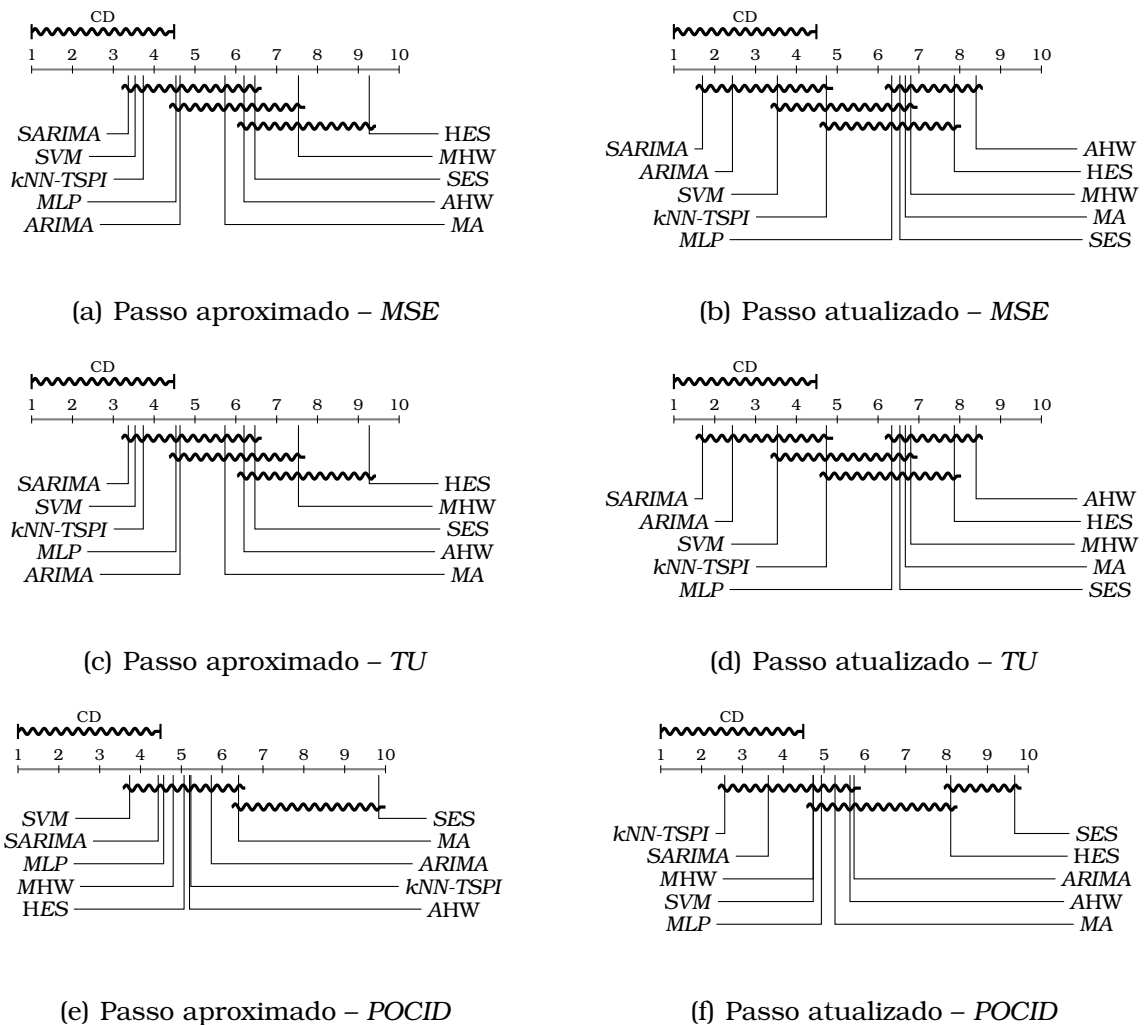
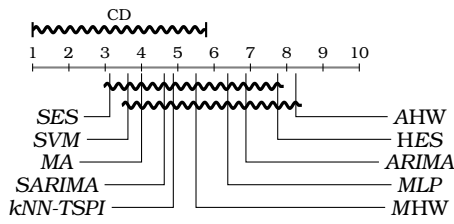


Figura 51: Diagramas de distância crítica para os valores dos índices *MSE*, *TU* e *POCID* provenientes dos métodos de predição sobre ST estocásticas

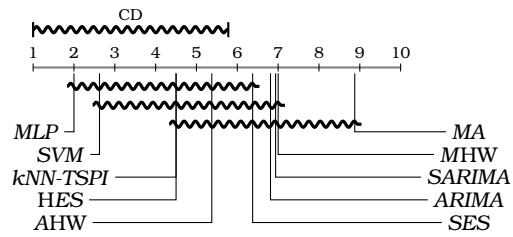
templaram 160 configurações (10 modelos preditivos \times 2 estratégias de predição \times 8 conjuntos de dados). Conforme retratado nos diagramas de distância crítica da Figura 52, os modelos preditivos com passo aproximado *MA*, *SVM* e *SES* exibiram, em ordem crescente e sem diferenças estatisticamente significativas, os melhores resultados de *MSE* (Figura 52(a)) e *TU* (Figura 52(c)). Os preditores com passo aproximado *SVM*, *MLP* e *kNN-TSPI* ocuparam, respectivamente, a primeira, segunda e terceira posições no *ranking* médio oriundo das taxas de acerto *POCID* (Figura 52(e)).

No que se refere à estratégia de predição multi-etapa à frente com passo atualizado, o algoritmo *MLP* acarretou, em conformidade com os três índices de desempenho (Figuras 52(b), 52(d) e 52(f)), nos melhores resultados. Nessa mesma linha de raciocínio, o modelo *MA* demonstrou os mais altos erros de predição e as mais baixas taxas de acerto quanto à tendência dos horizontes projetados.

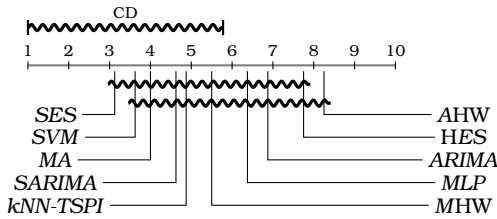
Verificando simultaneamente as duas estratégias de predição da Figura 52,



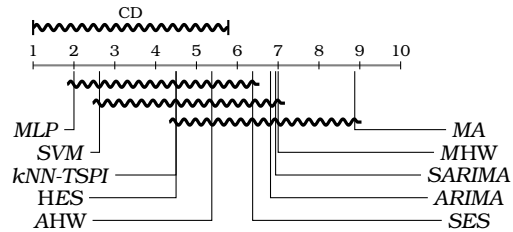
(a) Passo aproximado - MSE



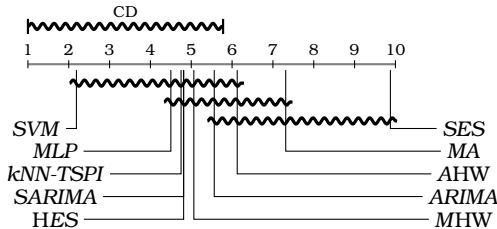
(b) Passo atualizado - MSE



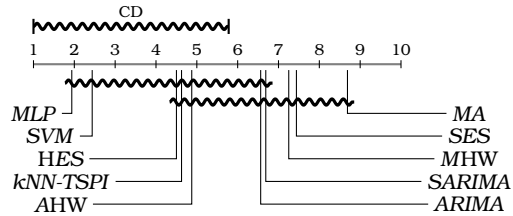
(c) Passo aproximado - TU



(d) Passo atualizado - TU



(e) Passo aproximado - POCID



(f) Passo atualizado - POCID

Figura 52: Diagramas de distância crítica para os valores dos índices *MSE*, *TU* e *POCID* provenientes dos métodos de predição sobre *ST* caóticas

o desempenho do algoritmo não-paramétrico *SVM* foi, em geral, mais estável que os outros métodos para predizer *ST* caóticas.

5.3.2 Análise Geral

Os resultados das 800 configurações (10 modelos preditivos \times 2 estratégias de predição \times 40 conjuntos de dados) são resumidos nos diagramas de distância crítica da Figura 53.

Na Figura 53, o modelo *SARIMA* com passo aproximado apresentou os melhores resultados utilizando as medidas *MSE* (Figura 53(a)) e *TU* (Figura 53(c)). Por outro lado, quando analisadas as taxas de acerto sobre as tendências dos horizontes projetados (Figura 53(e)), a referida configuração ocupou a segunda posição no *ranking* médio, não exibindo diferenças estatisticamente significativas em comparação com o modelo *SVM*. Considerando a estratégia de predição multi-etapa à frente com passo atualizado, os algoritmos *SVM*, *MLP* e *kNN-TSPI* ocuparam, nessa ordem e por uma pequena margem de diferença,

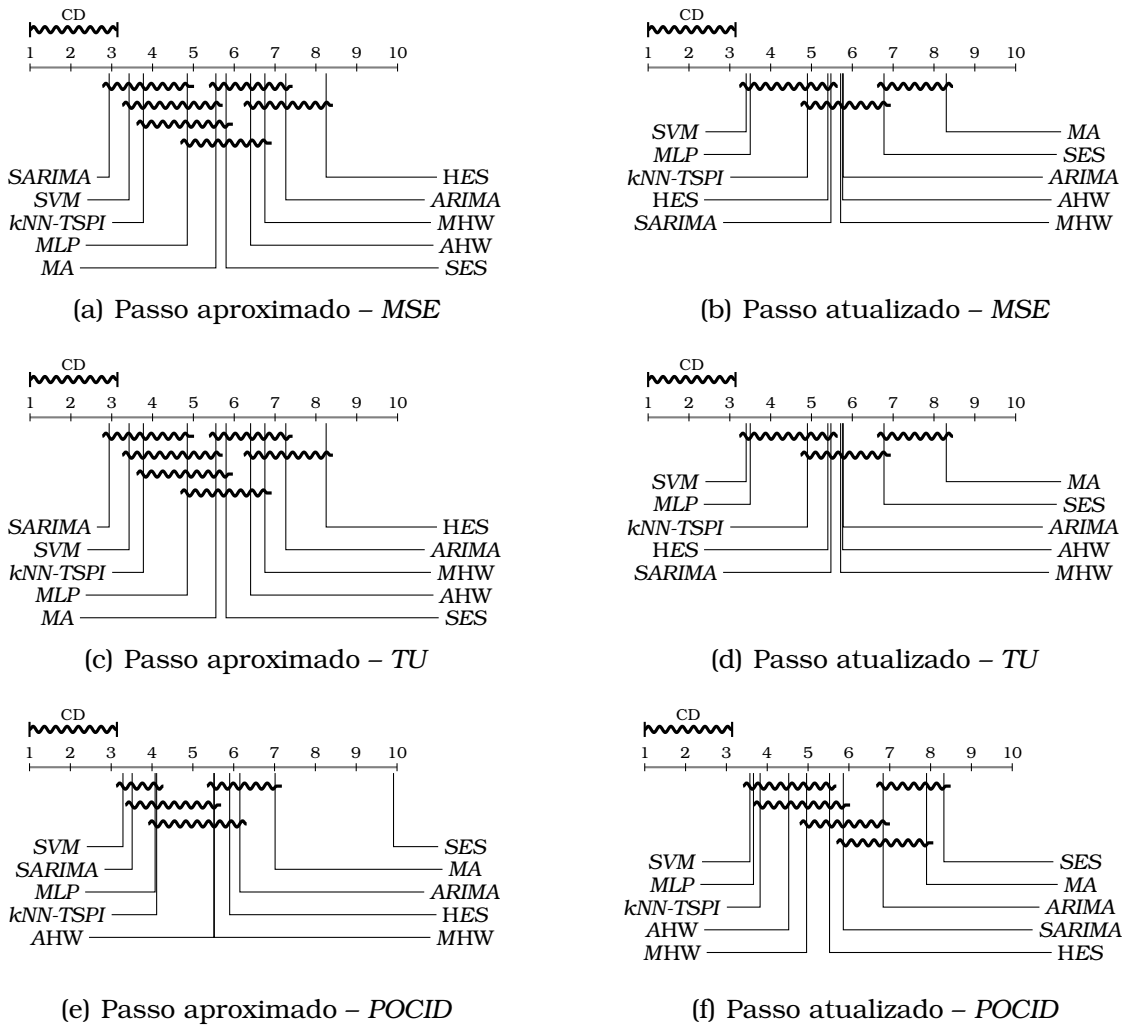


Figura 53: Diagramas de distância crítica para os valores dos índices *MSE*, *TU* e *POCID* provenientes dos métodos de predição sobre ST sintéticas

a primeira, segunda e terceira posições nos *rankings* derivados dos índices *MSE* (Figura 53(b)), *TU* (Figura 53(d)) e *POCID* (Figura 53(f)).

Examinando simultaneamente as duas estratégias de projeção da Figura 53, o *kNN-TSPI* exibiu o terceiro melhor desempenho para todas as medidas de avaliação, exceto quando tratado do índice *POCID* para a estratégia de predição multi-etapa à frente com passo atualizado. Adicionalmente, o *SVM* mostrou-se bastante competitivo em relação aos resultados do modelo *SARIMA*.

6 Considerações Finais

No presente trabalho foram investigados diversos métodos estatísticos e de Aprendizado de Máquina (AM) para a tarefa de predição de Séries Temporais (ST). A partir do planejamento e da execução de uma revisão sistemática, seguida de uma meta-análise, foi possível detectar uma importante lacuna na

literatura relacionada. Essa lacuna de pesquisa, parcialmente suprida neste relatório técnico, faz referência ao desprovimento de uma comparação objetiva e subjetiva entre modelos estatísticos e de AM para projeção de valores futuros.

Impulsionado pelas necessidades da comunidade de predição de dados temporais, o escopo deste trabalho se limitou à exposição sucinta dos modelos preditivos mais populares e ao confronto destes empregando 40 ST sintéticas, das quais 17 são determinísticas, 15 são estocásticas e oito são caóticas. Todos os conjuntos de dados considerados encontram-se disponíveis no portal *Web ICMC-USP Time Series Prediction Repository* (Parmezan e Batista, 2014). Esse repositório foi desenvolvido para viabilizar o acesso aos materiais produzidos e também aos usados no decorrer das atividades contempladas por esta pesquisa.

Para as ST determinísticas, o modelo *SARIMA* forneceu os melhores resultados quando utilizada a estratégia de predição multi-etapa à frente com passo aproximado. Por outro lado, considerando a estratégia de predição multi-etapa à frente de com passo atualizado, o *MLP* providenciou os menores erros de projeção. Ainda assim, foi o algoritmo *kNN-TSPI* que obteve, para ambas as estratégias de predição, o desempenho mais estável em termos de erro de predição e taxa de acerto quanto à tendência dos horizontes projetados.

Considerando as ST estocásticas, o *SARIMA* com passos aproximado e atualizado demonstrou ser o modelo de predição mais promissor. No entanto, ele não apresentou diferenças estatisticamente significativas em relação aos métodos *SVM* e *kNN-TSPI*. O fato do *SARIMA* ter proporcionado, em média, menores taxas de erro pode ser esclarecido tendo em vista a sua estrutura, a qual é constituída por um procedimento de Médias Móveis que contempla estimativas do fator de inovação que não pode ser explicado pelo modelo.

No que diz respeito às ST caóticas, os algoritmos não-paramétricos *MLP*, *SVM* e *kNN-TSPI* obtiveram, sem diferenças estatisticamente significativas, os melhores resultados. Como esses modelos não fazem suposições acerca da distribuição dos dados, era esperado que eles apresentassem um desempenho superior ao dos métodos estatísticos.

Analisando os 40 conjuntos de dados sintéticos (17 ST determinísticas, 15 ST estocásticas e 8 ST caóticas), o *kNN-TSPI* com passos aproximado e atualizado registrou o terceiro melhor desempenho em relação a todas as medidas de avaliação consideradas, exceto quando tratado o índice *POCID* para a estratégia de predição multi-etapa à frente com passo atualizado. Os menores erros de projeção foram fornecidos pelo modelo *SARIMA*, seguido do algoritmo *SVM*. Por mais que *SARIMA* e *SVM* tenham atingido uma precisão superior a do método baseado em similaridade, o algoritmo com invariâncias é considera-

velmente mais simples de compreender, codificar e ajustar. Enquanto *SARIMA* tem sete parâmetros e *SVM* possui três, o *kNN-TSPI* tem apenas dois. O mais importante é que esses dois parâmetros são totalmente intuitivos e podem ser facilmente estimados apenas observando a sazonalidade dos dados.

Como trabalhos futuros, pretende-se ampliar o protocolo experimental proposto empregando ST de domínio real. Além disso, é de interesse explorar as demais particularidades dos índices de avaliação de desempenho adotados. Dentre essas propriedades, destacam-se as quatro faixas de valores do coeficiente *U* de Theil e a taxa de acerto quanto à tendência dos horizontes projetados (*Prediction Of Change In Direction*).

Referências

- Ahmed, N. K., Atiya, A. F., Gayar, N. E., e El-Shishiny, H. (2010). An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, 29(5-6):594–621. Citado nas páginas 4 e 8.
- Aikes Junior, J. (2012). Estudo da influência de diversas medidas de similaridade na previsão de séries temporais utilizando o algoritmo kNN-TSP. Dissertação de Mestrado em Engenharia de Sistemas Dinâmicos e Energéticos, Universidade Estadual do Oeste do Paraná, Foz do Iguaçu, Brasil. Citado na página 3.
- Andrawis, R. R., Atiya, A. F., e El-Shishiny, H. (2011). Forecast combinations of computational intelligence and linear models for the NN5 time series forecasting competition. *International Journal of Forecasting*, 27(3):672–688. Citado na página 4.
- Bache, K. e Lichman, M. (2013). UCI machine learning repository. School of Information and Computer Sciences, University of California, Irvine, United States of America. Disponível em: <http://archive.ics.uci.edu/ml>. Citado na página 8.
- Barroso, L. C., Barroso, M. M. A., Campos Filho, F. F., Carvalho, M. L. B., e Maia, M. L. (1987). *Cálculo numérico com aplicações*. Harbra, São Paulo, Brasil, 2 edição. Citado nas páginas 15 e 16.
- Batista, G. E. A. P. A., Keogh, E. J., Tataw, O. M., e Souza, V. M. A. (2014). CID: An efficient complexity-invariant distance for time series. *Data Mining and Knowledge Discovery*, 28(3):634–669. Citado nas páginas 58 e 59.
- Ben Taieb, S., Bontempi, G., Atiya, A. F., e Sorjamaa, A. (2012). A review and comparison of strategies for multi-step ahead time series forecasting

- based on the NN5 forecasting competition. *Expert Systems with Applications*, 39(8):7067–7083. Citado nas páginas 4 e 5.
- Bloomfield, P. (2000). *Fourier analysis of time series: An introduction*. Wiley Series in Probability and Statistics. Wiley, New York, United States of America, 2 edição. Citado na página 68.
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., e Ljung, G. M. (2015). *Time series analysis: Forecasting and control*. Wiley Series in Probability and Statistics. Wiley, New Jersey, United States of America, 5 edição. Citado nas páginas 2, 8, 33, 42, 46, e 68.
- Brocklebank, J. e Dickey, D. (2003). *SAS for forecasting time series*. SAS Institute, Cary, United States of America, 2 edição. Citado na página 20.
- Brockwell, P. J. e Davis, R. A. (2002). *Introduction to time series forecasting*. Springer, New York, United States of America, 2 edição. Citado na página 12.
- Buffa, E. S. e Sarin, R. K. (1987). *Modern production/operations management*. Wiley, New York, United States of America. Citado na página 21.
- Chatfield, C. (2013). *The analysis of time series: An introduction*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, Boca Raton, United States of America. Citado nas páginas 1, 11, 21, 28, e 44.
- Chatterjee, S. e Hadi, A. S. (2012). *Regression analysis by example*. Wiley, New York, United States of America, 5 edição. Citado nas páginas 15 e 16.
- Christodoulou, N. S. (2009). An algorithm using Runge-Kutta methods of orders 4 and 5 for systems of ODEs. *International Journal of Numerical Methods and Applications*, 2(1):47–57. Citado na página 75.
- Claveria, O. e Torra, S. (2014). Forecasting tourism demand to Catalonia: Neural networks vs. time series models. *Economic Modelling*, 36(C):220–228. Citado nas páginas 2 e 10.
- Cortez, P. (2010). Sensitivity analysis for time lag selection to forecast seasonal time series using neural networks and support vector machines. Em *International Joint Conference on Neural Networks*, páginas 3694–3701, Barcelona, Spain. IEEE. Citado na página 9.
- Cowpertwait, P. S. P. e Metcalfe, A. V. (2009). *Introductory time series with R*. Springer, New York, United States of America. Citado nas páginas 1, 13, 27, 45, 46, e 47.

- Cristianini, N. e Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge University Press, Cambridge, United Kingdom. Citado nas páginas 53 e 54.
- Crone, S. F., Hibon, M., e Nikolopoulos, K. (2011). Advances in forecasting with neural networks? Empirical evidence from the NN3 competition on time series prediction. *International Journal of Forecasting*, 27(3):635–660. Citado na página 5.
- Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2(4):303–314. Citado na página 53.
- Deitel, P. J. e Deitel, H. M. (2012). *Java: How to Program*. Prentice Hall, Massachusetts, United States of America, 9 edição. Citado na página 82.
- Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30. Citado na página 83.
- Ehlers, R. S. (2009). Análise de séries temporais. Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, Brasil. Disponível em: <http://www.icmc.usp.br/ehlers/stemp/stemp.pdf>. Citado nas páginas 14 e 21.
- Eisencraft, M. (2001). Sistemas de comunicação utilizando sinais caóticos. Dissertação de Mestrado, Escola Politécnica, Universidade de São Paulo, São Paulo, Brasil. Citado na página 77.
- Ferrero, C. A. (2009). Algoritmo kNN para previsão de dados temporais: Funções de previsão e critérios de seleção de vizinhos próximos aplicados à variáveis ambientais em limnologia. Dissertação de Mestrado, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, Brasil. Citado na página 3.
- Fix, E. e Hodges, J. L. (1951). Discriminatory analysis, nonparametric discrimination, consistency properties. Relatório Técnico 4, Project 21-49-004, US Air Force School of Aerospace Medicine, Randolph Field, United States of America. Citado na página 55.
- Frunzete, M., Popescu, A., e Barbot, J. P. (2015). Dynamical discrete-time Rössler map with variable delay. Em *Computational Science and Its Applications*, volume 9155 of *Lecture Notes in Computer Science*, páginas 431–446. Springer International Publishing, Banff, Canada. Citado na página 77.
- Fu, T. (2011). A review on time series data mining. *Engineering Applications of Artificial Intelligence*, 24(1):164–181. Citado na página 1.

- Gardner, E. S. (1985). Exponential smoothing: The state of the art. *Journal of forecasting*, 4(1):1–28. Citado nas páginas 33, 34, 36, e 37.
- Glass, L. e Mackey, M. C. (1997). *Dos relógios ao caos: Os ritmos da vida*. Editora da Universidade de São Paulo, São Paulo, Brasil. Citado na página 75.
- Gooijer, J. G. D. e Hyndman, R. J. (2006). 25 years of time series forecasting. *International Journal of Forecasting*, 22(3):443–473. Citado nas páginas 2 e 4.
- Gunn, S. R. et al. (1998). Support vector machines for classification and regression. Relatório técnico, Southampton, United Kingdom. Faculty of Engineering, Science and Mathematics, University of Southampton. Citado na página 53.
- Han, J., Kamber, M., e Pei, J. (2011). *Data mining: Concepts and techniques*. Morgan Kaufmann, San Francisco, United States of America, 3 edição. Citado nas páginas 3 e 29.
- Haykin, S. S. (2009). *Neural networks and learning machines*. Prentice Hall, Upper Saddle River, United States of America, 3 edição. Citado nas páginas 50, 51, e 52.
- Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International journal of forecasting*, 20(1):5–10. Citado na página 37.
- Hyndman, R. J. e Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688. Citado na página 81.
- Hyndman, R. J., Koehler, A. B., Snyder, R. D., e Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 18(3):439–454. Citado nas páginas 36 e 37.
- Islam, M. e Sivakumar, B. (2002). Characterization and prediction of runoff dynamics: A nonlinear dynamical view. *Advances in Water Resources*, 25(2):179–190. Citado na página 48.
- Kandananond, K. (2012). A comparison of various forecasting methods for autocorrelated time series. *International Journal of Engineering Business Management*, 4:1–6. Citado nas páginas 2 e 9.

- Kaundal, R., Kapoor, A. S., e Raghava, G. P. S. (2006). Machine learning techniques in disease forecasting: A case study on rice blast prediction. *BMC Bioinformatics*, 7(1):1–16. Citado na página 54.
- Kirchgässner, G., Wolters, J., e Hassler, U. (2013). *Introduction to modern time series analysis*. Springer, New York, United States of America, 2 edição. Citado na página 27.
- Kulesh, M., Holschneider, M., e Kurennaya, K. (2008). Adaptive metrics in the nearest neighbours method. *Physica D: Nonlinear Phenomena*, 237(3):283–291. Citado nas páginas 69, 70, e 71.
- Larose, D. T. e Larose, C. D. (2014). *Discovering knowledge in data: An introduction to data mining*. Wiley Series on Methods and Applications in Data Mining. Wiley, New Jersey, United States of America. Citado na página 1.
- Lemke, C. e Gabrys, B. (2010). Meta-learning for time series forecasting and forecast combination. *Neurocomputing*, 73(10-12):2006–2016. Citado na página 4.
- Lippmann, R. P. (1987). An introduction to computing with neural nets. *IEEE ASSP Magazine*, 4(2):4–22. Citado na página 51.
- Lorena, A. C. e de Carvalho, A. C. P. L. F. (2007). Uma introdução às support vector machines. *Revista de Informática Teórica e Aplicada*, 14(2):43–67. Citado na página 55.
- Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20(2):130–141. Citado na página 76.
- Maimon, O. e Rokach, L., editores (2010). *Data mining and knowledge discovery handbook*. Springer, Secaucus, United States of America, 2 edição. Citado nas páginas 1 e 30.
- McCulloch, W. S. e Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133. Citado nas páginas 50 e 51.
- McSharry, P. E., Clifford, G. D., Tarassenko, L., e Smith, L. A. (2003). A dynamical model for generating synthetic electrocardiogram signals. *Transactions on Biomedical Engineering*, 50(3):289–294. Citado na página 78.
- Minsky, M. e Papert, S. (1969). Perceptrons: An introduction to computational geometry. Citado na página 51.
- Montgomery, D. (2009). *Introduction to statistical quality control*. Wiley, New York, United States of America, 6 edição. Citado na página 71.

- Montgomery, D. C., Jennings, C. L., e Kulahci, M. (2015). *Introduction to time series analysis and forecasting*. Wiley Series in Probability and Statistics. Wiley, New Jersey, United States of America, 2 edição. Citado nas páginas 1, 34, 36, 39, 43, e 44.
- Moré, J. J. (1978). Numerical analysis. Em Watson, G. A., editor, *Proceedings of the Biennial Conference Held at Dundee*, volume 630 of *Lecture Notes in Mathematics*, capítulo The Levenberg-Marquardt algorithm: Implementation and theory, páginas 105–116. Springer, Heidelberg, Germany. Citado na página 47.
- Morettin, P. A. e Toloi, C. M. C. (2006). *Análise de séries temporais*. Blucher, São Paulo, Brasil, 2 edição. Citado nas páginas 12, 14, 20, 33, 34, 36, 38, 40, 46, e 47.
- Mueen, A., Keogh, E., Zhu, Q., e Cash, S. (2009). Exact discovery of time series motifs. Em *SIAM International Conference on Data Mining*, Sparks, United States of America. Society Industrial Applied Mathematics. Citado na página 60.
- Parmezan, A. R. S. e Batista, G. E. A. P. A. (2014). ICMC-USP time series prediction repository. Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, Brasil. Disponível em: http://sites.labcic.icmc.usp.br/icmc_tspr. Citado nas páginas 3, 12, e 88.
- Parmezan, A. R. S. e Batista, G. E. A. P. A. (2015). A study of the use of complexity measures in the similarity search process adopted by kNN algorithm for time series prediction. Em *International Conference on Machine Learning and Applications*, páginas 45–51, Miami, United States of America. IEEE. Citado nas páginas 2 e 56.
- Petitjean, F., Ketterlin, A., e GANÇARSKI, P. (2011). A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognition*, 44(3):678–693. Citado na página 58.
- Platt, J. C. (1999). Using analytic QP and sparseness to speed training of support vector machines. Em *Advances in Neural Information Processing Systems*, páginas 557–563. MIT Press, Cambridge, United States of America. Citado na página 55.
- Pyle, D. (1999). *Data preparation for data mining*. Morgan Kaufmann, California, United States of America. Citado na página 31.

- Rakthanmanon, T., Campana, B., Mueen, A., Batista, G., Westover, B., Zhu, Q., Zakaria, J., e Keogh, E. (2012). Searching and mining trillions of time series subsequences under dynamic time warping. Em *International Conference on Knowledge Discovery and Data Mining*, páginas 262–270, Beijing, China. ACM. Citado nas páginas 56 e 57.
- Reinsel, G. C. (2003). *Elements of multivariate time series analysis*. Springer Series in Statistics. Springer, New York, United States of America, 2 edição. Citado na página 31.
- Ristanoski, G., Liu, W., e Bailey, J. (2013). A time-dependent enhanced support vector machine for time series regression. Em *International Conference on Knowledge Discovery and Data Mining*, páginas 946–954, Chicago, United States of America. ACM. Citado nas páginas 10 e 55.
- Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408. Citado na página 50.
- Rumelhart, D. E., Hinton, G. E., e Williams, R. J. (1986). Parallel distributed processing: explorations in the microstructure of cognition. volume 1, capítulo Learning internal representations by error propagation, páginas 318–362. MIT Press, Cambridge, United States of America. Citado na página 52.
- Sapankevych, N. I. e Sankar, R. (2009). Time series prediction using support vector machines: A survey. *Computational Intelligence Magazine*, 4(2):24–38. Citado na página 2.
- Schiller, J., Spiegel, M., e Srinivasan, R. (2012). *Schaum's outline of probability and statistics*. McGraw-Hill, New York, United States of America. Citado nas páginas 17 e 23.
- Shawe-Taylor, J., Bartlett, P. L., Williamson, R. C., e Anthony, M. (1998). Structural risk minimization over data-dependent hierarchies. 44(5):1926–1940. Citado na página 53.
- Shumway, R. H. e Stoffer, D. S. (2011). *Time series analysis and its applications: With R examples*. Springer Texts in Statistics. Springer Science & Business Media, New York, United States of America, 3 edição. Citado na página 44.
- Strogatz, S. H. (2014). *Nonlinear dynamics and chaos: With applications to physics, biology, chemistry, and engineering*. Studies in Nonlinearity. Westview Press, Boulder, United States of America, 2 edição. Citado nas páginas 73 e 74.

- Tahersima, H., Tahersima, F., Sohani, A., Jafar, M., e Saleh, K. (2010). Prediction of Lorenz chaotic time series via genetic algorithm. Em *International Conference on Computational Intelligence for Measurement Systems and Applications*, páginas 13–17, Taranto, Italia. IEEE. Citado na página 76.
- Theil, H. (1971). *Principles of econometrics*. Wiley, New York, United States of America. Citado na página 82.
- Vapnik, V. N. (1999). *The nature of statistical learning theory*. Information Science and Statistics. Springer Science & Business Media, New York, United States of America, 2 edição. Citado na página 53.
- Verhulst, P. F. (1845). Recherches mathématiques sur la loi d'accroissement de la population. *Nouveaux mémoires de l'Académie Royale des Sciences et Belles-Lettres de Bruxelles*, 18:14–54. Citado na página 74.
- Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, 6(3):324–342. Citado nas páginas 39 e 40.
- Witten, I. H., Frank, E., e Hall, M. A. (2011). *Data mining: Practical machine learning tools and techniques*. The Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann, San Francisco, United States of America, 3 edição. Citado nas páginas 1, 8, e 82.
- Zhang, G., Patuwo, B. E., e Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14(1):35–62. Citado na página 50.
- Zhang, X., Zhang, T., Young, A. A., e Li, X. (2014). Applications and comparisons of four time series models in epidemiological surveillance data. *PLOS ONE*, 9(2). Citado na página 11.