
Predição de séries temporais por similaridade

Antonio Rafael Sabino Parmezan

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Antonio Rafael Sabino Parmezan

Predição de séries temporais por similaridade

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Mestre em Ciências – Ciências de Computação e Matemática Computacional. *VERSÃO REVISADA*

Área de Concentração: Ciências de Computação e Matemática Computacional

Orientador: Prof. Dr. Gustavo Enrique de Almeida Prado Alves Batista

USP – São Carlos
Junho de 2016

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados fornecidos pelo(a) autor(a)

P253p Parmezan, Antonio Rafael Sabino
Predição de séries temporais por similaridade /
Antonio Rafael Sabino Parmezan; orientador Gustavo
Enrique de Almeida Prado Alves Batista. - São Carlos
- SP, 2016.
219 p.

Dissertação (Mestrado - Programa de Pós-Graduação
em Ciências de Computação e Matemática Computacional)
- Instituto de Ciências Matemáticas e de Computação,
Universidade de São Paulo, 2016.

1. Predição de séries temporais. 2. Métodos
baseados em similaridade. 3. Aprendizado de máquina.
4. Mineração de dados. I. Batista, Gustavo Enrique de
Almeida Prado Alves, orient. II. Título.

Antonio Rafael Sabino Parmezan

Similarity-based time series prediction

Master dissertation submitted to the Instituto de Ciências Matemáticas e de Computação – ICMC-USP, in partial fulfillment of the requirements for the degree of the Master Program in Computer Science and Computational Mathematics. *FINAL VERSION*

Concentration Area: Computer Science and Computational Mathematics

Advisor: Prof. Dr. Gustavo Enrique de Almeida Prado
Alves Batista

USP – São Carlos
June 2016

*Aos meus pais,
Gilberto e Sandra.*

*Ao meu irmão,
Nicolas.*

*Aos laboratórios de pesquisa,
LABI/UNIOESTE e LABIC/ICMC-USP.*

AGRADECIMENTOS

A conclusão desta dissertação de mestrado demarca o término de uma importante e inesquecível etapa da minha vida acadêmica e pessoal. Sinto que, ao longo desses anos, fui agraciado por Deus com a amizade de pessoas extraordinárias, as quais me encorajaram a lutar pelos meus objetivos, sempre com muita força e alegria. A essas pessoas, cujo apoio e incentivo foram decisivos para a concretização de mais esse sonho, direciono algumas palavras de agradecimento.

Aos meus pais, Gilberto de Almeida Parmezan e Sandra Aparecida Sabino Parmezan, pela imensa dedicação, carinho, educação e amparo incondicional em cada passo da minha vida. Obrigado por sempre me apoiarem nas decisões que me levaram a trilhar este caminho.

Ao meu irmão Nicolas Sabino Parmezan pelo companheirismo, amizade e por sempre trazer a alegria nos dias difíceis. Você tem um grande coração e eu o admiro por isso.

Ao professor Gustavo Enrique de Almeida Prado Alves Batista pela amizade, respeito, confiança e ensinamentos transmitidos. É um privilégio trabalhar sob sua orientação. Agradeço também a sua esposa, professora Claudia Regina Milaré, pelo gentil apoio e incentivo.

A todos os meus amigos do Laboratório de Inteligência Computacional (LABIC) do ICMC-USP, em especial ao Alan Demetrios Baria Valejo, Bruno Magalhães Nogueira, Camila Vaccari Sundermann, Celso André Rodrigues de Sousa, Cristiano Inácio Lemes, Diego Furtado Silva, Igor Assis Braga, Ivone Penque Matsuno, Jorge Carlos Valverde Rebaza, Marcos Aurelio Domingues, Rafael Geraldeli Rossi, Rafael Giusti, Renan de Pádua, Roberta Akemi Sinoara, Thiago de Paulo Faleiros e Vinícius Mourão Alves de Souza. Agradeço pelos bons momentos compartilhados e pelas diversas oportunidades que tivemos para discutir questões de meu trabalho.

Não posso deixar de agradecer às professoras Maria Carolina Monard e Solange Oliveira Rezende, as quais preservam um legado de profissionalismo e ética.

Aos professores Wu Feng Chung e Huei Diana Lee que, mesmo de longe, têm estado tão perto. Obrigado pelo carinho, respeito, confiança, orientação e pelos valiosos ensinamentos, os quais transcendem a vida acadêmica.

Ao professor Renato Bobsin Machado pela amizade, preocupação e incentivo. Agradeço também pelos ensinamentos e pela notável dedicação para com o curso de Ciência da Computação da UNIOESTE.

A todos os meus amigos do Laboratório de Bioinformática (LABI) da UNIOESTE, em especial ao Moacir Fonteque Junior e Newton Spolaôr. Obrigado pela amizade e por compartilharem suas experiências de vida.

Aos amigos André Gustavo Maletzke, Adrieli Cristina da Silva, Barbara Lepretti de Nadai, Bianca Espíndola, Carlos Andrés Ferrero, Chris Mayara dos Santos Tibes, Dabna Hellen Tomim, Everton Alvares Cherman, Jorge Aikes Junior, Joylan Nunes Maciel, Juliano Koji Yugoshi, Vanize Meneghetti, Willian Zalewski e Wilson Jung. Sou muito grato por tê-los conhecido e por poder compartilhar momentos bons e outros turbulentos, mas todos de igual importância.

Aos meus estimados amigos Evelin Danielle Santos Matos, Jefferson Tales Oliva, Leandro Borges dos Santos, Ricardo Gil Belther Nabo e Simone Aparecida Pinto Romero. Obrigado pela amizade, companhia, apoio e incentivo durante todos esses anos.

Aos professores da Cultura Inglesa, Elielson Antonio Sgarbi, Tatiana Bauso Cardoso e Vera Mass. Gostaria também de agradecer ao André Henrique Guimarães Gabriel e Vitor Milanez, meus colegas de classe. Obrigado por tornarem minhas manhãs de sábado mais produtivas e agradáveis.

Ao ICMC-USP pela infraestrutura, professores de excelência e agentes universitários que o compõe. Particularmente, agradeço à Dóra Versetti e a Leiliane Ometto Ciamaricone que sempre me ajudaram no serviço de convênios, bolsas e auxílios.

A todos os professores justos e honestos que tive, no decorrer da minha trajetória acadêmica, a oportunidade de trabalhar. Agradeço por tentarem, diariamente e sem abrir mão de princípios, tornar esse mundo melhor. Sem vocês não haveria motivação para continuar nesta jornada.

Por fim, agradeço à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo suporte financeiro fornecido a mim para o desenvolvimento desta dissertação de mestrado.

*“Destino não é uma questão de sorte,
mas uma questão de escolha;
não é uma coisa que se espera,
mas que se busca”
(William Jennings Bryan)*

RESUMO

PARMEZAN, A. R. S.. **Predição de séries temporais por similaridade.** 2016. 219 f. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação (ICMC/USP), São Carlos – SP.

Um dos maiores desafios em Mineração de Dados é a integração da informação temporal ao seu processo. Esse fato tem desafiado profissionais de diferentes domínios de aplicação e recebido investimentos consideráveis da comunidade científica e empresarial. No contexto de predição de Séries Temporais, os investimentos se concentram no subsídio de pesquisas destinadas à adaptação dos métodos convencionais de Aprendizado de Máquina para a análise de dados na qual o tempo constitui um fator importante. À vista disso, neste trabalho é proposta uma nova extensão do algoritmo de Aprendizado de Máquina *k-Nearest Neighbors (kNN)* para predição de Séries Temporais, intitulado de *kNN - Time Series Prediction with Invariances (kNN-TSPI)*. O algoritmo concebido difere da versão convencional pela incorporação de três técnicas para obtenção de invariância à amplitude e deslocamento, invariância à complexidade e tratamento de casamentos triviais. Como demonstrado ao longo desta dissertação de mestrado, o uso simultâneo dessas técnicas proporciona ao *kNN-TSPI* uma melhor correspondência entre as subsequências de dados e a consulta de referência. Os resultados de uma das avaliações empíricas mais extensas, imparciais e compreensíveis já conduzidas no tema de predição de Séries Temporais evidenciaram, a partir do confronto de dez métodos de projeção, que o algoritmo *kNN-TSPI*, além de ser conveniente para a predição automática de dados a curto prazo, é competitivo com os métodos estatísticos estado-da-arte *ARIMA* e *SARIMA*. Por mais que o modelo *SARIMA* tenha atingido uma precisão relativamente superior a do método baseado em similaridade, o *kNN-TSPI* é consideravelmente mais simples de ajustar. A comparação objetiva e subjetiva entre algoritmos estatísticos e de Aprendizado de Máquina para a projeção de dados temporais vem a suprir uma importante lacuna na literatura, a qual foi identificada por meio de uma revisão sistemática seguida de uma meta-análise das publicações selecionadas. Os 95 conjuntos de dados empregados nos experimentos computacionais juntamente com todas as projeções analisadas em termos de Erro Quadrático Médio, coeficiente *U* de Theil e taxa de acerto *Prediction Of Change In Direction* encontram-se disponíveis no portal *Web ICMC-USP Time Series Prediction Repository*. A presente pesquisa abrange também contribuições e resultados significativos em relação às propriedades inerentes à predição baseada em similaridade, sobretudo do ponto de vista prático. Os protocolos experimentais delineados e as diversas conclusões obtidas poderão ser usados como referência para guiar o processo de escolha de modelos, configuração de parâmetros e aplicação dos algoritmos de Inteligência Artificial para predição de Séries Temporais.

Palavras-chave: Predição de séries temporais, Métodos baseados em similaridade, Aprendizado de máquina, Mineração de dados.

ABSTRACT

PARMEZAN, A. R. S.. **Predição de séries temporais por similaridade.** 2016. 219 f. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação (ICMC/USP), São Carlos – SP.

One of the major challenges in Data Mining is integrating temporal information into process. This difficulty has challenged professionals several application fields and has been object of considerable investment from scientific and business communities. In the context of Time Series prediction, these investments consist mainly of grants for designed research aimed at adapting conventional Machine Learning methods for data analysis problems in which time is an important factor. We propose a novel modification of the k-Nearest Neighbors (kNN) learning algorithm for Time Series prediction, namely the kNN - Time Series Prediction with Invariances (kNN-TSPI). Our proposal differs from the literature by incorporating techniques for amplitude and offset invariance, complexity invariance, and treatment of trivial matches. These three modifications allow more meaningful matching between the reference queries and Time Series subsequences, as we discuss with more details throughout this master's thesis. We have performed one of the most comprehensible empirical evaluations of Time Series prediction, in which we faced the proposed algorithm with ten methods commonly found in literature. The results show that the kNN-TSPI is appropriate for automated short-term projection and is competitive with the state-of-the-art statistical methods ARIMA and SARIMA. Although in our experiments the SARIMA model has reached a slightly higher precision than the similarity-based method, the kNN-TSPI is considerably simpler to adjust. The objective and subjective comparisons of statistical and Machine Learning algorithms for temporal data projection fills a major gap in the literature, which was identified through a systematic review followed by a meta-analysis of selected publications. The 95 data sets used in our computational experiments, as well all the projections with respect to Mean Squared Error, Theil's U coefficient and hit rate Prediction Of Change In Direction are available online at the *ICMC-USP* Time Series Prediction Repository. This work also includes contributions and significant results with respect to the properties inherent to similarity-based prediction, especially from the practical point of view. The outlined experimental protocols and our discussion on the usage of them, can be used as a guideline for models selection, parameters setting, and employment of Artificial Intelligence algorithms for Time Series prediction.

Key-words: Time series prediction, Similarity-based methods, Machine learning, Data mining.

LISTA DE ILUSTRAÇÕES

Figura 1 – Expressão de busca adotada no protocolo de revisão sistemática	45
Figura 2 – Distribuição do número de trabalhos por ano de publicação	46
Figura 3 – Percentagem do número de publicações que contemplaram uma determinada ferramenta computacional	48
Figura 4 – Produção mensal de chocolate na Austrália	53
Figura 5 – Preço anual de ingressos para o <i>Super Bowl</i>	55
Figura 6 – Tendência da série de preços de ingressos para o <i>Super Bowl</i>	57
Figura 7 – Exemplo de obtenção da tendência usando <i>MA</i> com $r = 3$	57
Figura 8 – Tendência da série de produção trimestral de cerveja nos EUA	58
Figura 9 – Tendência da série de produção mensal de chocolate na Austrália	60
Figura 10 – Técnica do <i>scatter plot</i> para a série de produção trimestral de cerveja nos EUA	62
Figura 11 – Técnica do <i>scatter plot</i> para a série de produção mensal chocolate na Austrália	63
Figura 12 – Sazonalidade da série de produção de cerveja nos EUA	65
Figura 13 – Sazonalidade e o efeito da desestacionalização da série de produção mensal de chocolate na Austrália	66
Figura 14 – Resíduo da série de produção trimestral de cerveja nos EUA	67
Figura 15 – Resíduo da série de produção mensal de chocolate na Austrália	68
Figura 16 – Processo de Mineração de Dados Temporais	70
Figura 17 – Processo de predição de valores em ST	74
Figura 18 – Hierarquia de abordagens para predição de ST	76
Figura 19 – Predições obtidas pelo modelo de <i>MA</i> com parâmetro $r = 1, 3$ e 5	78
Figura 20 – Predições computadas pelo modelo de <i>SES</i> com parâmetro $\alpha = 0, 1, 0, 3$ e $0, 9$	80
Figura 21 – Predições obtidas pelo modelo de <i>HES</i> parametrizado de dois modos: (i) $\alpha = 0, 3$ e $\beta = 0, 9$; (ii) $\alpha = 0, 9$ e $\beta = 0, 3$	82
Figura 22 – Representação dos tipos de variação sazonal	83
Figura 23 – Predições computadas pelos modelos aditivo e multiplicativo de HW, ambos com parâmetros $\alpha = 0, 3$, $\beta = 0, 5$, $\gamma = 0, 7$ e $s = 4$	85
Figura 24 – Diagrama de atividades para o fluxo de construção de um modelo <i>ARIMA</i> ou <i>SARIMA</i>	89
Figura 25 – Exemplificação do processo de predição de ST segundo a abordagem global	91
Figura 26 – Representação da estratégia de predição multi-etapa à frente para a abordagem global	92
Figura 27 – Estrutura do <i>Perceptron</i>	93

Figura 28 – Estrutura de uma rede <i>MLP</i> com camada oculta única	95
Figura 29 – 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	97
Figura 30 – Mapeamento de dados para um espaço de características de mais alta dimensão utilizando como artifício a função <i>kernel</i>	97
Figura 31 – Hierarquia do aprendizado indutivo considerando o grau de supervisão dos dados	109
Figura 32 – Exemplo da aplicação do algoritmo <i>kNN</i> com parâmetro $k = 1, 3$ e 5	111
Figura 33 – Exemplo da aplicação do algoritmo <i>kNN-TSP</i> com parâmetros $k = 3$ e $l = 25$	113
Figura 34 – Exemplo de variância à amplitude e deslocamento	115
Figura 35 – Exemplo de variância à complexidade	116
Figura 36 – Exemplo de casamentos triviais	117
Figura 37 – Exemplificação da necessidade de invariância à amplitude e deslocamento .	124
Figura 38 – Exemplificação de invariância à escala local	125
Figura 39 – Exemplificação de invariância à escala uniforme	125
Figura 40 – Exemplificação de invariância à fase	126
Figura 41 – Exemplificação de invariância à oclusão	126
Figura 42 – Exemplificação da necessidade de invariância à complexidade	127
Figura 43 – Representação gráfica do alinhamento realizado, entre duas ST, pela distância euclidiana	128
Figura 44 – Efeito da variação de p na Norma L_p	129
Figura 45 – Representação da estimativa de complexidade adotada pela <i>CID</i>	133
Figura 46 – Esquematização da matriz de distâncias acumuladas, com rota de ajuste traçado, decorrente da aplicação da medida <i>DTW</i>	135
Figura 47 – Configuração experimental I	140
Figura 48 – Diagramas de distância crítica para os valores dos índices <i>MSE</i> , <i>TU</i> e <i>POCID</i> provenientes do <i>kNN-TSP</i> , usando distintas técnicas invariantes, sobre ST reais	145
Figura 49 – Desempenho em ST reais do <i>kNN-TSP</i> , empregando diferentes técnicas invariantes, para quatro faixas de valores do coeficiente <i>TU</i>	146
Figura 50 – Médias e desvios padrão das taxas de acerto <i>POCID</i> obtidas pelo <i>kNN-TSP</i> , utilizando distintas técnicas invariantes, em ST reais	147
Figura 51 – Diagramas de distância crítica para os valores dos índices <i>MSE</i> , <i>TU</i> e <i>POCID</i> provenientes dos métodos baseados em similaridade sobre ST reais	148
Figura 52 – Desempenho em ST reais dos métodos baseados em similaridade para quatro faixas de valores do coeficiente <i>TU</i>	148
Figura 53 – Médias e desvios padrão das taxas de acerto <i>POCID</i> obtidas pelos métodos baseados em similaridade em ST reais	149

Figura 54 – Diagramas de distância crítica para os valores dos índices <i>MSE</i> , <i>TU</i> e <i>POCID</i> provenientes do <i>kNN-TSPI</i> , usando diferentes medidas distâncias, sobre ST reais	150
Figura 55 – Desempenho em ST reais do <i>kNN-TSPI</i> , empregando distintas medidas de distância, para quatro faixas de valores do coeficiente <i>TU</i>	151
Figura 56 – Médias e desvios padrão das taxas de acerto <i>POCID</i> obtidas pelo <i>kNN-TSPI</i> , utilizando diferentes medidas de distâncias, em ST reais	152
Figura 57 – Diagramas de distância crítica para os valores dos índices <i>MSE</i> , <i>TU</i> e <i>POCID</i> provenientes do <i>kNN-TSPI</i> , usando distintas estimativas de complexidade, sobre ST reais	153
Figura 58 – Desempenho em ST reais do <i>kNN-TSPI</i> , empregando diferentes estimativas de complexidade, para quatro faixas de valores do coeficiente <i>TU</i>	154
Figura 59 – Médias e desvios padrão das taxas de acerto <i>POCID</i> obtidas pelo <i>kNN-TSPI</i> , utilizando distintas estimativas de complexidade, em ST reais	154
Figura 60 – Diagramas de distância crítica para os valores dos índices <i>MSE</i> , <i>TU</i> e <i>POCID</i> provenientes do <i>kNN-TSPI</i> usando, diferentes funções de predição, sobre ST reais	156
Figura 61 – Desempenho em ST reais do <i>kNN-TSPI</i> , empregando distintas funções de predição, para quatro faixas de valores do coeficiente <i>TU</i>	157
Figura 62 – Médias e desvios padrão das taxas de acerto <i>POCID</i> obtidas pelo <i>kNN-TSPI</i> , utilizando diferentes funções de predição, em ST reais	157
Figura 63 – Configuração experimental II	160
Figura 64 – 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	165
Figura 65 – Desempenho dos métodos de predição para quatro faixas de valores do coeficiente <i>TU</i> em ST determinísticas	166
Figura 66 – Médias e desvios padrão das taxas de acerto <i>POCID</i> obtidas pelos métodos de predição em ST determinísticas	166
Figura 67 – 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	167
Figura 68 – Desempenho dos métodos de predição para quatro faixas de valores do coeficiente <i>TU</i> em ST estocásticas	168
Figura 69 – Médias e desvios padrão das taxas de acerto <i>POCID</i> obtidas pelos métodos de predição em ST estocásticas	169
Figura 70 – 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	170
Figura 71 – Desempenho dos métodos de predição para quatro faixas de valores do coeficiente <i>TU</i> em ST caóticas	171

Figura 72 – Médias e desvios padrão das taxas de acerto <i>POCID</i> obtidas pelos métodos de predição em ST caóticas	171
Figura 73 – 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	172
Figura 74 – Desempenho dos métodos de predição para quatro faixas de valores do coeficiente <i>TU</i> em ST sintéticas	173
Figura 75 – Médias e desvios padrão das taxas de acerto <i>POCID</i> obtidas pelos métodos de predição em ST sintéticas	174
Figura 76 – 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 reais	175
Figura 77 – Desempenho dos métodos de predição para quatro faixas de valores do coeficiente <i>TU</i> em ST reais	176
Figura 78 – Médias e desvios padrão das taxas de acerto <i>POCID</i> obtidas pelos métodos de predição em ST reais	176
Figura 79 – 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 e reais	177
Figura 80 – Desempenho dos métodos de predição para quatro faixas de valores do coeficiente <i>TU</i> em ST sintéticas e reais	178
Figura 81 – Médias e desvios padrão das taxas de acerto <i>POCID</i> obtidas pelos métodos de predição em ST sintéticas e reais	179
Figura 82 – Conteúdo administrado pelo ICMC-USP <i>Time Series Prediction Repository</i>	203
Figura 83 – Séries com sazonalidade aditiva derivadas da composição de Fourier	206
Figura 84 – Séries com dependência sazonal	206
Figura 85 – Série com sazonalidade multiplicativa	207
Figura 86 – Série de alta frequência com sazonalidade multiplicativa	207
Figura 87 – Padrões temporais provenientes da categoria GCA	209
Figura 88 – Padrões temporais provenientes da categoria GCB	209
Figura 89 – Séries com dependência sazonal e ruído	210
Figura 90 – Mapa Logístico	211
Figura 91 – Mapa de Hénon	211
Figura 92 – Sistema de Mackey-Glass	212
Figura 93 – Sistema de Lorenz	213
Figura 94 – Sistema de Rössler	214
Figura 95 – Sinais Caóticos	215
Figura 96 – Série de valores ECGSYN	215

LISTA DE QUADROS

Quadro 1 – Questões de pesquisa que nortearam a revisão sistemática	45
Quadro 2 – Etapas do método da relação entre <i>MA</i>	63
Quadro 3 – Formato atributo-valor	108
Quadro 4 – Medidas de distância entre ST	130
Quadro 5 – Problemas técnicos encontrados na implementação de algumas medidas de distância	131
Quadro 6 – Sumário de características e de configurações dos conjuntos de dados reais .	141
Quadro 7 – Problemas em dados temporais e técnicas invariantes a essas adversidades .	143
Quadro 8 – Pesos considerados na aplicação da função de predição <i>DW</i>	155
Quadro 9 – Sumário de características e de configurações dos conjuntos de dados sintéticos	161
Quadro 10 – Algoritmos utilizados e intervalos de variação numérica definidos para os seus parâmetros	162
Quadro 11 – Sumário de informações dos trabalhos identificados	197
Quadro 12 – Padrões temporais gerados a partir da composição de Fourier	205
Quadro 13 – Padrões temporais comumente observados em gráficos de controle	208

LISTA DE ALGORITMOS

Algoritmo 1 – <i>Holdout Validation</i>	99
Algoritmo 2 – <i>Cross-validation</i>	101
Algoritmo 3 – Box-Jenkins <i>Method</i>	102
Algoritmo 4 – <i>kNN-TSP</i>	114
Algoritmo 5 – <i>kNN-TSPI</i>	118

LISTA DE TABELAS

Tabela 1 – Exemplo de obtenção da tendência por MA centrada com $r = 4$	59
Tabela 2 – Exemplo de obtenção dos índices sazonais pelo método da relação entre MA	64
Tabela 3 – Exemplo de obtenção das medidas de síntese dos índices para cada período da variação sazonal utilizando o método da relação entre MA	65
Tabela 4 – Exemplo de obtenção do componente residual	68
Tabela 5 – Dados projetados usando o modelo de MA com parâmetro $r = 1, 3$ e 5	78
Tabela 6 – Dados projetados empregando o modelo de SES com parâmetro $\alpha = 0, 1, 0, 3$ e $0, 9$	80
Tabela 7 – Dados projetados usando o modelo de HES configurado de duas maneiras: (i) $\alpha = 0, 3$ e $\beta = 0, 9$; (ii) $\alpha = 0, 9$ e $\beta = 0, 3$	82
Tabela 8 – Dados projetados empregando os modelos aditivo e multiplicativo de HW , ambos com parâmetros $\alpha = 0, 3$, $\beta = 0, 5$, $\gamma = 0, 7$ e $s = 4$	85

LISTA DE ABREVIATURAS E SIGLAS

<i>AD</i>	Average Distance
<i>AIC</i>	Critério de Informação de Akaike
<i>ANN</i>	Redes Neurais Artificiais
<i>ARIMA</i>	Autorregressivo Integrado de Médias Móveis
<i>ARMA</i>	Autorregressivo de Médias Móveis
<i>AHW</i>	Holt-Winters Aditivo
<i>BNN</i>	Redes Neurais Bayesianas
<i>CD</i>	Distância Crítica
<i>CID</i>	<i>Complexity-Invariant Distance</i>
<i>DTW-D</i>	<i>Dynamic Time Warping - Delta</i>
<i>DW</i>	<i>Distance Weighted</i>
<i>GP</i>	<i>Gaussian Process</i>
<i>GRNN</i>	Redes Neurais de Regressão Generalizada
<i>IW</i>	<i>Index Weighted</i>
<i>kNN-TSPI</i>	<i>k-Nearest Neighbors - Time Series Prediction with Invariances</i>
<i>kNN-TSP</i>	<i>k-Nearest Neighbors - Time Series Prediction</i>
<i>kNN</i>	<i>k-Nearest Neighbors</i>
<i>MAE</i>	Erro Absoluto Médio
<i>MAPE</i>	Erro Percentual Absoluto Médio
<i>MASE</i>	Erro Médio Absoluto em Escala
<i>MA</i>	Médias Móveis
<i>MLP</i>	<i>Multilayer Perceptron</i>
<i>MSE</i>	Erro Quadrático Médio
<i>MHW</i>	Holt-Winters Multiplicativo
<i>NFL</i>	Liga Nacional de Futebol Americano
<i>NRMSE</i>	Raiz do Erro Quadrático Médio Normalizado
<i>POCID</i>	<i>Prediction of Change in Direction</i>
<i>RAE</i>	Erro Absoluto Relativo
<i>RBF</i>	Funções de Base Radial
<i>RMSE</i>	Raiz do Erro Quadrático Médio
<i>SARIMA</i>	Autorregressivo Integrado de Médias Móveis Sazonal

SAX	<i>Symbolic Aggregate approXimation</i>
SES	Suavização Exponencial Simples
SETAR	<i>Self-Exciting Threshold Autoregressive</i>
SMAPE	Erro Percentual Absoluto Médio Simétrico
SVM	Máquinas de Suporte Vetorial
TU	<i>U</i> de Theil
AM	Aprendizado de Máquina
DE	Distância Euclidiana
EUA	Estados Unidos da América
FAPESP	Fundaçāo de Amparo à Pesquisa do Estado de São Paulo
HES	Suavização Exponencial de Holt
HW	Holt-Winters
LABI	Laboratório de Bioinformática
LABIC	Laboratório de Inteligência Computacional
MD	Mineração de Dados
MDT	Mineração de Dados Temporais
MVA	Média de Valores Absolutos
MVR	Média de Valores Relativos
QAS	Quadrática Aditiva Simétrica
ST	Série Temporal

LISTA DE SÍMBOLOS

R — Coeficiente de Correlação

R^2 — Coeficiente de Determinação

Z — Série Temporal Aleatória

m — Tamanho de uma Série Temporal Aleatória

\mathfrak{R} — Conjunto dos Números Reais

z — Valor Aleatório

t — Instante de Tempo

f — Função Matemática

ε — Valor Aleatório

y — Valor Aleatório

T — Componente de Tendência

S — Componente de Sazonalidade

N — Componente Residual

Y — Vetor Aleatório

X — Vetor Aleatório

b_0 — Coeficiente Linear da Reta

b_1 — Coeficiente Angular da Reta

i — Índice

x — Valor Aleatório

r — Parâmetro do Método de Médias Móveis

lag — Valor de Defasagem

h — Horizonte de Predição

α — Constante de Suavização Associada ao Nível

β — Constante de Suavização Associada à Tendência

γ — Constante de Suavização Associada à Sazonalidade

- s — Quantidade de Observações que Compõe um Período Sazonal
- p — Ordem do Procedimento de Autorregressão
- ϕ — Valor de Ponderação do Componente de Autorregressão
- δ — Nível Inicial do Modelo
- μ — Média do Processo Estacionário
- e — Ruído Branco
- q — Ordem do Procedimento de Médias móveis
- θ — Valor de Ponderação do Componente de Médias Móveis
- d — Grau do Operador de Diferenciação
- P — Ordem do Procedimento de Autorregressão Sazonal
- Φ — Valor de Ponderação do Componente de Autorregressão Sazonal
- D — Grau do Operador de Diferenciação Sazonal
- Q — Ordem do Procedimento de Médias Móveis Sazonal
- Θ — Valor de Ponderação do Componente de Médias Móveis Sazonal
- n — Número de Unidades na Camada Oculta
- K — Função *Kernel*
- \mathbb{C} — Parâmetro de Regularização
- σ — Largura da Gaussiana da Função *Kernel* de Base Radial
- h' — Horizonte de Predição Usado pelo Método de Estimação de Parâmetros
- \mathbb{P} — Lista de Parâmetros Subótimos
- \hat{z} — Valor Predito
- C — Classe de um Conjunto de Dados
- k — Quantidade de Vizinhos Próximos
- E — Exemplo de um Conjunto de Dados
- A — Atributo de um Conjunto de Dados
- l — Tamanho da Janela de Busca
- j — Índice
- $O(\cdot)$ — Custo Computacional de uma Determinada Operação
- d — Medida de Distância
- L_p — Norma L_p ou Métrica de Minkowski

p — Valor que Determina a Medida de Similaridade na Norma L_p

L_1 — Distância Manhattan

L_2 — Distância Euclidiana

L_3 — Métrica L_3

L_∞ — Distância de Chebychev

R — Rota de Alinhamento Traçada entre duas Séries Temporais

max_p — Número Máximo de Observações que Constituem um Período Sazonal na ST

SUMÁRIO

1	INTRODUÇÃO	35
1.1	<i>Justificativa e Motivação</i>	37
1.2	<i>Objetivos, Hipóteses e Suposições</i>	38
1.3	<i>Principais Contribuições</i>	40
1.4	<i>Organização do Trabalho</i>	41
2	REVISÃO SISTEMÁTICA E META-ANÁLISE DA LITERATURA	43
2.1	<i>Considerações Iniciais</i>	43
2.2	<i>Planejamento e Execução</i>	43
2.3	<i>Apresentação dos Resultados</i>	46
2.4	<i>Considerações Finais</i>	50
3	SÉRIES TEMPORAIS	51
3.1	<i>Considerações Iniciais</i>	51
3.2	<i>Definições e Notações</i>	51
3.3	<i>Componentes de Séries Temporais</i>	53
3.3.1	<i>Tendência</i>	54
3.3.2	<i>Sazonalidade</i>	60
3.3.3	<i>Resíduo</i>	66
3.4	<i>Análise de Séries Temporais</i>	69
3.5	<i>Mineração de Dados Temporais</i>	69
3.6	<i>Considerações Finais</i>	71
4	PREDIÇÃO DE SÉRIES TEMPORAIS	73
4.1	<i>Considerações Iniciais</i>	73
4.2	<i>O Problema da Predição de Dados</i>	73
4.3	<i>Métodos para Construção de Modelos Preditivos</i>	75
4.3.1	<i>Métodos Paramétricos</i>	75
4.3.1.1	<i>Médias Móveis</i>	76
4.3.1.2	<i>Suavização Exponencial Simples</i>	78
4.3.1.3	<i>Suavização Exponencial de Holt</i>	80
4.3.1.4	<i>Suavização Exponencial Sazonal de Holt-Winters</i>	81
4.3.1.5	<i>Modelos ARIMA e SARIMA</i>	86
4.3.2	<i>Métodos Não-paramétricos</i>	91

4.3.2.1	<i>Redes Neurais Artificiais</i>	93
4.3.2.2	<i>Máquinas de Suporte Vetorial</i>	96
4.3.2.3	<i>k-Vizinhos mais Próximos</i>	98
4.4	Técnicas para Estimação de Parâmetros	98
4.4.1	<i>Validação Holdout</i>	98
4.4.2	<i>Validação Cruzada</i>	100
4.4.3	<i>Método Box-Jenkins</i>	100
4.5	Avaliação da Qualidade de Predição	103
4.6	Considerações Finais	104
5	ALGORITMO <i>k</i>-NEAREST NEIGHBORS PARA PREDIÇÃO DE SÉRIES TEMPORAIS	107
5.1	Considerações Iniciais	107
5.2	Aprendizado de Máquina	107
5.3	O Algoritmo <i>kNN</i>	110
5.4	O Algoritmo <i>kNN-TSP</i>	112
5.5	O Algoritmo <i>kNN-TSPI</i>	114
5.6	Considerações Finais	120
6	SIMILARIDADE ENTRE SÉRIES TEMPORAIS	123
6.1	Considerações Iniciais	123
6.2	Invariâncias às Distorções Conhecidas em Dados Temporais	123
6.3	Medidas de Distância	128
6.4	Medidas de Distância Invariantes à Complexidade	132
6.4.1	<i>Complexity-Invariant Distance</i>	132
6.4.2	<i>Dynamic Time Warping - Delta</i>	134
6.5	Considerações Finais	136
7	AVALIAÇÃO EXPERIMENTAL I: EXPLORANDO AS PROPRIEDADES DA PREDIÇÃO POR SIMILARIDADE	139
7.1	Considerações Iniciais	139
7.2	Configuração Experimental	139
7.3	Resultados e Discussão	142
7.3.1	<i>Invariâncias às Distorções em Séries Temporais</i>	143
7.3.2	<i>Métodos Baseados em Similaridade</i>	146
7.3.3	<i>Medidas de Distância</i>	148
7.3.4	<i>Medidas de Complexidade Aplicadas à CID</i>	151
7.3.5	<i>Funções de Predição</i>	154
7.4	Considerações Finais	156

8	AVALIAÇÃO EXPERIMENTAL II: COMPARANDO O ALGORITMO <i>kNN-TSPI</i> COM MÉTODOS TRADICIONAIS DA LITERATURA .	159
8.1	Considerações Iniciais	159
8.2	Configuração Experimental	160
8.3	Resultados e Discussão	163
8.3.1	<i>Séries Sintéticas</i>	164
8.3.1.1	<i>Séries Determinísticas</i>	164
8.3.1.2	<i>Séries Estocásticas</i>	167
8.3.1.3	<i>Séries Caóticas</i>	169
8.3.1.4	Comparação Geral	172
8.3.2	<i>Séries Reais</i>	174
8.3.3	Comparação Geral	177
8.4	Considerações Finais	178
9	CONCLUSÃO	181
9.1	Limitações	184
9.2	Trabalhos Futuros	185
REFERÊNCIAS		187
APÊNDICE A	TRABALHOS SELECIONADOS NA REVISÃO SISTEMÁTICA	197
APÊNDICE B	ICMC-USP <i>TIME SERIES PREDICTION REPOSITORY</i>	203



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; 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; JENNINGS; KULAHCI, 2015; CHATFIELD, 2013; COWPERTWAIT; 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; FRANK; HALL, 2011; MAIMON; ROKACH, 2010). Dentro 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

interpretáveis¹, 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, summarizando-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 predizer dados futuros. Essa predição é obtida a partir da extração 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; 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) estimativa 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; BATISTA, 2015; CLAVERIA; TORRA, 2014; KANDANANOND, 2012; SAPANEVYCH; 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 des-

¹ O comportamento dos dados pode, por um considerável período de tempo, oscilar sistematicamente em decorrência de fatores externos fortuitos pouco ou não conhecidos.

crição 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, consequentemente, na construção de modelos que permitam a representação de novos conhecimentos adquiridos automaticamente, de modo mais eficiente e comprehensível (HAN; KAMBER; PEI, 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; BATISTA, 2014; JUNIOR, 2012; FERRERO, 2009).

Neste trabalho são investigadas adaptações de métodos provenientes da área de AM para o problema da predição de ST. Especificamente, tem-se interesse em métodos baseados em similaridade que buscam, 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 usam os valores seguintes dessas subsequências como entrada para uma função de predição, a qual realiza o cálculo do valor futuro.

1.1 Justificativa e Motivação

Na última década houve um crescente aumento no interesse por ST em Mineração de Dados (MD) e Descoberta de Conhecimento. Esse interesse culminou na proposta de literalmente centenas de algoritmos para tarefas como classificação, recuperação por conteúdo, agrupamento, identificação de padrões morfológicos (*motifs*), extração de regras de associação, detecção de anomalias e predição de valores futuros (FU, 2011).

Pesquisas empíricas na área de processamento e análise de ST têm demonstrado que métodos baseados em similaridade proporcionam resultados muito competitivos, frequentemente superando técnicas mais complexas. Por exemplo, em classificação, o algoritmo *k-Nearest Neighbors* (*kNN*) fornece resultados que são dificilmente superados (DING *et al.*, 2008); em agrupamento, alguns trabalhos sugerem que para agrupar ST, a escolha do método de *clustering*

é significativamente menos importante que a determinação da medida de distância utilizada entre as séries, com *Dynamic Time Warping* obtendo excelentes desempenhos (ZHU *et al.*, 2012); na detecção de anomalias, uma pesquisa mostrou, após uma extensa avaliação experimental, que algoritmos guiados por medidas de distância produzem os melhores resultados (CHAN-DOLA; CHEBOLI; KUMAR, 2009). Presume-se que a superioridade dos métodos baseados em similaridade pode ser explicada, em grande parte, devido ao trabalho incessante da comunidade em invariâncias de distância, tais como deformação (*warping*), linha de base, oclusão e rotação (BATISTA *et al.*, 2014; PRATI; BATISTA, 2012).

Em contraste, algoritmos estatísticos fundamentados 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; HYNDMAN, 2006). Nesta dissertação de mestrado é questionado se esse é realmente o caso, ou se a comunidade de MD tem mais a oferecer. Em particular, partindo da assertiva de que métodos por similaridade são simples e possuem poucos parâmetros para modelar o comportamento dos dados, é levantada a hipótese de que esses métodos podem proporcionar resultados competitivos quando comparados com os algoritmos estatísticos estado-da-arte na literatura.

Ainda que métodos baseados em similaridade para predição de ST tenham sido explorados no passado recente, acredita-se que os estudos anteriores não foram capazes de identificar as invariâncias necessárias para essa tarefa. É pressuposição central desta pesquisa que apenas com a combinação adequada de invariância à amplitude, invariância à deslocamento, e da recentemente proposta invariância de complexidade (BATISTA *et al.*, 2014), associada a uma política para evitar casamentos triviais, pode levar a previsões precisas e significativas.

1.2 Objetivos, Hipóteses e Suposições

Impulsionado pelos desafios e pelas necessidades da comunidade de predição de dados temporais, o escopo deste trabalho concentra-se nos seguintes objetivos:

Objetivo geral:

Investigar o uso de invariâncias na predição de Séries Temporais por similaridade e demonstrar empiricamente que esses métodos proporcionam resultados competitivos quando comparados com os métodos estatísticos considerados estado-da-arte na literatura.

Objetivos específicos:

- Realizar uma revisão sistemática para identificar os conceitos e os avanços científicos relacionados ao tema de predição de Séries Temporais;

- Averiguar como diferentes técnicas para obtenção de invariâncias às distorções indesejáveis em dados temporais podem influenciar o desempenho preditivo dos algoritmos baseados em similaridade;
- Estudar novas medidas de distância invariantes à complexidade projetadas para (pequenas) subsequências utilizando variantes das distâncias euclidiana e *Dynamic Time Warping*;
- Comparar experimentalmente os algoritmos estudados com abordagens tradicionais da literatura;
- Desenvolver um repositório e disponibilizá-lo na *Internet*, de maneira a tornar todos os resultados desta pesquisa amplamente reproduzíveis.

Dado o objetivo geral e os objetivos específicos, pode-se declarar a hipótese central deste trabalho:

Métodos para predição de Séries Temporais baseados em similaridade podem prover resultados competitivos em relação aos obtidos com a aplicação de métodos estatísticos estado-da-arte.

A hipótese supracitada é respaldada nas seguintes premissas:

1. Algoritmos guiados por medidas de distância são simples, de fácil codificação, e não dependem de muitos parâmetros para modelar o comportamento dos dados;
2. Algoritmos que trabalham com o grau de semelhança entre observações são conhecidos por sua eficácia em outras tarefas, de igual importância, para extração de padrões temporais.

As premissas expostas reforçam a ideia de que métodos por similaridade são de fácil compreensão para o ser humano, uma vez que os parâmetros requeridos por eles podem, na maioria dos casos, ser determinados sem a necessidade de conhecimento *a priori*. Além disso, a eficiência desses métodos está diretamente associada à medida de distância usada para detectar as subsequências de dados similares na série.

Segundo Batista *et al.* (2014), o desempenho de uma medida de distância reflete sua capacidade em capturar corretamente as invariâncias requeridas pelo domínio de aplicação. Neste trabalho, durante o processo de predição de ST baseado em similaridade, é defendida a ideia do emprego de técnicas que permitam desprezar informações² que não favoreçam o casamento entre subsequências visualmente similaridades. Desse modo, supõe-se que a combinação apropriada de invariância à amplitude e deslocamento, invariância à complexidade, e a desconsideração de casamentos triviais pode contribuir para a projeção de dados em um patamar mais preciso e significativo.

² Os termos dado e informação são utilizados indistintamente neste trabalho.

1.3 Principais Contribuições

As contribuições do presente trabalho podem ser elencadas da seguinte maneira:

- Planejamento e execução de uma revisão sistemática, seguida de uma meta-análise, para posicionar esta pesquisa no estado-da-arte correspondente. Foram selecionadas 42 publicações, entre os anos de 2010 e 2014, cujos conteúdos auxiliaram a responder dez questões de cunho científico. A interpretação dos resultados indicou que o estudo comparativo entre métodos estatísticos e de AM, concretizado nesta dissertação de mestrado, vem a suprir uma importante lacuna na literatura;
- Proposição de uma nova extensão do algoritmo *kNN* para predição de ST, intitulado de *k-Nearest Neighbors - Time Series Prediction with Invariances (kNN-TSPI)* ([PARMEZAN; BATISTA, 2015](#)). O algoritmo concebido difere da versão convencional pela incorporação de três técnicas para obtenção de invariância à amplitude e deslocamento, invariância à complexidade e tratamento de casamentos triviais. A aplicação do referido método evidenciou que a busca por similaridade com invariâncias possibilita uma correspondência mais significativa entre a subsequência de referência (consulta) e as subsequências de dados. Os resultados experimentais alcançados usando diferentes métodos de predição e o algoritmo aqui proposto, mostraram que o *kNN-TSPI* é competitivo e conveniente para a projeção automática de dados a curto prazo;
- Exploração das propriedades inerentes à predição de ST por similaridade, como invariâncias às distorções em dados temporais, medidas de distância, medidas de complexidade aplicadas à *Complexity-Invariant Distance* e funções de predição. Uma análise empírica dessas particularidades permitiu um melhor entendimento sobre o desempenho preditivo dos métodos baseados em similaridade e as conclusões extraídas poderão orientar futuras pesquisas com o algoritmo *kNN-TSPI*;
- Realização de uma das avaliações experimentais mais extensas, imparciais e compreensíveis já conduzidas no tema de predição de ST. Utilizando-se de 95 conjuntos de dados, foram confrontados dez algoritmos para a tarefa de predição de valores, dos quais sete são aplicados em conformidade com a abordagem paramétrica e três de acordo com a abordagem não-paramétrica. Essa análise comparativa, além de preencher uma lacuna na literatura, permitiu constatar em que circunstâncias um algoritmo de predição supera o outro e quais aspectos dos dados têm maior influência no desempenho desses algoritmos. O panorama dos resultados demonstrou que o Autorregressivo Integrado de Médias Móveis Sazonal (*SARIMA*) é o único método capaz de sobre-exceder, sem diferença estatisticamente significativa, o desempenho de predição do algoritmo *kNN-TSPI*. Por mais que o modelo *SARIMA* tenha atingido uma precisão superior a do método baseado em similaridade, o algoritmo com invariâncias proposto neste trabalho é consideravelmente

mais simples de ajustar. Enquanto SARIMA tem sete parâmetros a serem estimados, o método por similaridade tem apenas dois. O mais importante é que esses dois parâmetros são totalmente intuitivos e podem ser facilmente determinados apenas observando a sazonalidade dos dados;

- Construção de um portal *Web* ([PARMEZAN; BATISTA, 2014](#)), denominado de ICMC-USP *Time Series Prediction Repository*, que concede acesso aos materiais produzidos e também aos usados no decorrer das atividades contempladas por esta pesquisa. Atualmente, o repositório mantém 100 conjuntos de dados temporais (40 ST sintéticas e 60 ST reais) de acesso público e destinados às comunidades de Estatística e AM.

1.4 Organização do Trabalho

O restante trabalho está organizado do seguinte modo:

Capítulo 2 — Revisão Sistemática e Meta-análise da Literatura: Neste capítulo 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;

Capítulo 3 — Séries Temporais: Neste capítulo são apresentadas as principais definições e notações acerca do tema de Séries Temporais. Além disso, são descritos os componentes básicos que constituem essas séries, bem como os objetivos da análise de Séries Temporais e do processo de descoberta de conhecimento designado de Mineração de Dados Temporais;

Capítulo 4 — Predição de Séries Temporais: Neste capítulo, o problema da predição de Séries Temporais é introduzido e, em seguida, formulado como um processo de busca que visa, a partir de um conjunto de informações conhecidas, estimar dados desconhecidos. Complementarmente, são especificadas as abordagens para predição de valores futuros com seus respectivos métodos usuais e algumas técnicas que auxiliam na determinação dos parâmetros requeridos por esses métodos;

Capítulo 5 — Algoritmo *kNN* para Predição de Séries Temporais: Neste capítulo são abordados conceitos referentes à subárea da Inteligência Artificial conhecida como Aprendizado de Máquina. Esses conceitos são explanados considerando a hierarquia do aprendizado indutivo e os paradigmas de aprendizado. Posteriormente, são definidos os algoritmos *kNN* e sua respectiva adaptação, *k-Nearest Neighbors - Time Series Prediction (kNN-TSP)*, para a predição de Séries Temporais. Ao final deste capítulo é proposta uma variação do algoritmo *kNN-TSP*, nominada de *kNN-TSP with Invariances*, que tem como finalidade contornar os problemas encontrados no método original por meio da combinação apropriada de invariância à amplitude e deslocamento, invariância à complexidade e eliminação de casamento triviais;

Capítulo 6 — Similaridade entre Séries Temporais: Neste capítulo, a partir da especificação de semelhança entre subsequências de dados, são sumarizadas algumas das distorções conhecidas em Séries Temporais, assim como as técnicas levantadas neste trabalho para se obter invariância a esses efeitos indesejáveis. Aliada a essa questão, são descritas diversas medidas de similaridade, incluindo distâncias invariantes à complexidade, que podem absorver de um mesmo domínio características incomuns. Em resumo, uma parte das medidas examinadas obedecem ao alinhamento linear, enquanto as outras proporcionam um casamento não-linear mais robusto às distorções no eixo temporal;

Capítulo 7 — Avaliação Experimental I (Explorando as Propriedades da Predição por Similaridade): Neste capítulo, utilizando-se de 55 conjuntos de dados reais, é apresentada uma sequência de experimentos com intuito de verificar os distintos aspectos intrínsecos à predição de Séries Temporais por similaridade. Entre os fatores inspecionados estão invariâncias às distorções em dados temporais, medidas de distância, medidas de complexidade aplicadas à *Complexity-Invariant Distance* e funções de predição. Ainda nesse capítulo, o algoritmo proposto neste trabalho foi comparado com outros dois métodos baseados em similaridade e os resultados foram analisados em termos de desempenho preditivo;

Capítulo 8 — Avaliação Experimental II (Comparando o Algoritmo *kNN-TSPI* com M odos Tradicionais da Literatura): Neste capítulo, considerando 95 conjuntos de dados (55 Séries Temporais reais e 40 Séries Temporais sintéticas), o algoritmo de predição por similaridade com invariâncias foi confrontado com nove métodos amplamente difundidos na literatura. Essa avaliação empírica foi planejada para: (1) viabilizar a consolidação da eficiência e da efetividade de cada algoritmo apurado; e (2) caracterizar, em termos de vantagens e desvantagens de uso, os distintos métodos existentes para a construção de modelos preditivos;

Capítulo 9 — Conclusão: Neste capítulo são enunciadas as considerações finais desta dissertação de mestrado, bem como as principais limitações e trabalhos futuros.



REVISÃO SISTEMÁTICA E META-ANÁLISE DA LITERATURA

2.1 Considerações Iniciais

A revisão sistemática consiste em um método para exploração bibliográfica que permite a resolução de questões de pesquisa por meio de procedimentos explícitos para a identificação, seleção e avaliação de publicações. Esse processo é realizado com a intenção de descobrir trabalhos relevantes e avaliar o tema pesquisado de modo justo, rigoroso e replicável ([KITCHENHAM, 2007](#)).

Em estudos científicos arrojados, onde são exigidos ineditismo e originalidade na contribuição, a revisão da literatura desempenha um papel de grande importância. Por isso, operá-la de maneira sistemática é uma decisão inevitável quando deseja-se, além da construção de uma base sólida de conhecimento, aprimorar a teoria em setores ativos e/ou descobrir domínios com oportunidade de pesquisa. Cabe ressaltar que uma revisão sistemática pode ainda incluir uma meta-análise, a qual corresponde a uma síntese dos resultados obtidos usando técnicas estatísticas.

Neste capítulo são descritos, a partir da formulação de dez questões de pesquisa, o planejamento e a execução de uma revisão sistemática de trabalhos desenvolvidos na área de predição de Séries Temporais (ST). Uma meta-análise das publicações selecionadas foi preparada para facilitar a identificação da abrangência e do estado atual do tema abordado.

2.2 Planejamento e Execução

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 [Go-](#)

oijer 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 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 comprehensibilidade, 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; GABRYS, 2010), o estudo do efeito de dessazonalização na projeção de valores futuros (TAIEB *et al.*, 2012) e a construção de modelos híbridos por meio da combinação de métodos estatísticos e de MD (ANDRAWIS; ATIYA; EL-SHISHINY, 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 (TAIEB *et al.*, 2012; CRONE; HIBON; NIKOLOPOULOS, 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 summarizadas no Quadro 1. Na última linha desse quadro, na cor cinza, é indicada a questão central de pesquisa levantada neste trabalho.

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 Library¹, CiteSeerX², Google Scholar³, Scopus⁴ e Web of Science⁵. 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

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>>.

Quadro 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?

em relação aos resultados, 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.

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

```
(“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”)
```

Fonte: Elaborada pelo autor.

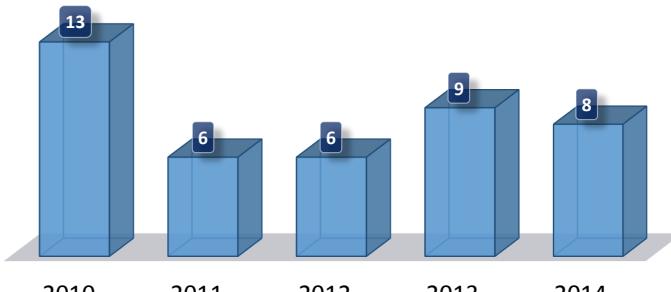
Ao todo, foram identificadas 554 publicações, tendo esse número reduzido para 367 após a aplicação dos critérios de exclusão. Posteriormente, foi realizada a leitura do título e resumo de todos os trabalhos apurados. Quando a seleção não pôde ser concretizada indubitavelmente, considerou-se a leitura 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 ([Apêndice A](#)) 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 no Quadro 1.

2.3 Apresentação dos Resultados

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.

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



Fonte: Elaborada pelo autor.

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

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;

⁶ O formulário contendo as informações extraídas das publicações selecionadas pode ser solicitado por intermédio do e-mail: antoniop@icmc.usp.br.

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 (*R²*) 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 desenvolvida 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; 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 MATLAB⁸, R⁹, SPSS¹⁰, MINITAB¹¹, WEKA ([WITTEN; FRANK; HALL, 2011](#)), STATISTICA¹², ForecastPRO¹³ e Phicast¹⁴. Na [Figura 3](#), o gráfico indica a frequência com que essas ferramentas apareceram nos trabalhos.

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>>.

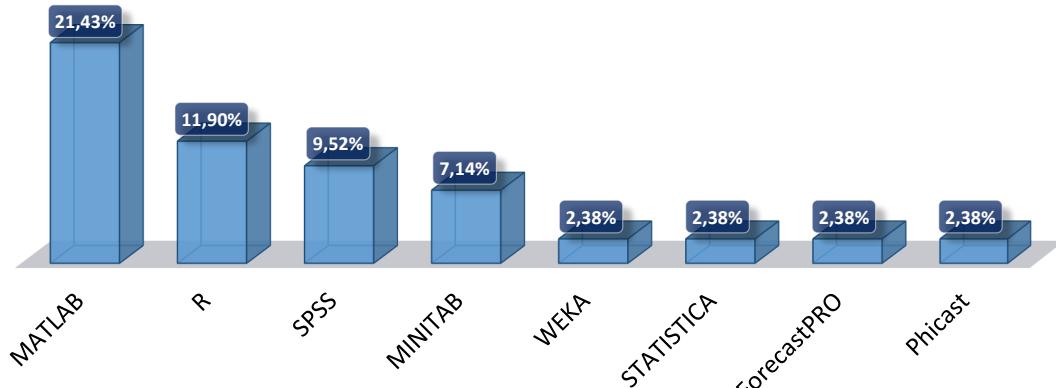
¹¹ <<http://www.minitab.com>>.

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



Fonte: Elaborada pelo autor.

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 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*;

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

3. Realizou-se em [Kandaninanond \(2012\)](#) 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 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, Liu e Bailey \(2013\)](#) 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

¹⁶ Dormida 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 significantemente 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 esta 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.

2.4 Considerações Finais

O estudo bibliográfico é incontestavelmente o primeiro passo para o desdobramento de qualquer pesquisa, seja ela de caráter científico ou filosófico. Todavia, sua qualidade pode ser contestada em virtude da carência de planejamento e rigor técnico. Uma maneira de prevenir esse questionamento é aplicar o método de revisão sistemática, o qual possibilita uma exploração ampla, justa e reproduzível da literatura.

A revisão sistemática realizada neste trabalho permitiu identificar 42 publicações pertinentes ao tema de previsão de ST. Uma meta-análise do conteúdo desses artigos, além de auxiliar na resolução de dez questões de pesquisa, proporcionou uma visão contemporânea da área de interesse. No que se refere ao objetivo, o levantamento da literatura apontou a necessidade de avaliações empíricas massivas que priorizem o confronto entre métodos estatísticos e de AM para a projeção de valores.

No próximo capítulo são apresentados os conceitos e definições sobre ST e seus componentes. Adicionalmente, são caracterizadas as etapas da análise de dados sequenciais e introduzido o processo de Mineração de Dados Temporais.



SÉRIES TEMPORAIS

3.1 Considerações Iniciais

Avanços tecnológicos na área da computação, como o aumento da capacidade de processamento e de armazenamento de dados, têm incentivado a utilização de sistemas computacionais para a aquisição e o gerenciamento de dados nas mais diversas áreas do conhecimento. Tais sistemas possibilitam, dependendo do objetivo da aplicação, o armazenamento de grandes quantidades de dados expressos em diferentes formatos. No entanto, em determinados domínios, além do armazenamento de dados em formatos usuais, como o numérico e o nominal, pode ainda ser de interesse o armazenamento da informação temporal que permite a organização cronológica dos dados coletados.

Distintas tarefas vêm sendo propostas para auxiliar no descobrimento e na compreensão de conhecimento embutido em dados temporais. Entre essas tarefas, destacam-se a classificação, o agrupamento, a detecção de anomalias e a predição de valores futuros (FU, 2011). Normalmente, a execução dessas tarefas implica no uso de algoritmos fundamentados em conceitos de áreas correlatas à Matemática e a Ciências de Computação, como a Estatística e a Inteligência Artificial.

Neste capítulo são abordadas definições associadas à representação de dados sequenciais por meio de Séries Temporais. Essas definições são apresentadas considerando as características e os componentes básicos que constituem as séries, assim como os objetivos dos métodos de análise de dados temporais presentes na literatura.

3.2 Definições e Notações

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

$z_t \in \Re$ 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; 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 , $y = f(\text{tempo}, \varepsilon)$, 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 Toloi (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 3.1, é suficiente para tornar a série estacionária.

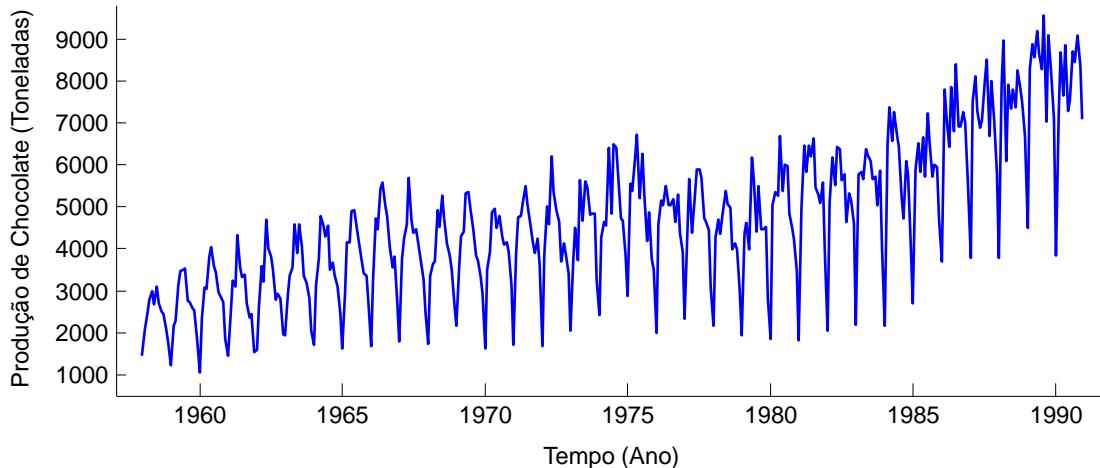
$$\Delta z_t = z_t - z_{t-1} \quad (3.1)$$

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 estatísticas¹ e, assim como todas as séries adotadas neste trabalho, encontram-se disponíveis no ICMC-USP Time Series Prediction Repository (PARMEZAN; 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

¹ <http://www.abs.gov.au>.

Figura 4 – Produção mensal de chocolate na Austrália



Fonte: Elaborada pelo autor.

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

3.3 Componentes de Séries Temporais

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 3.2 e 3.3, por uma decomposição aditiva ou multiplicativa de seus componentes (COWPERTWAIT; METCALFE, 2009). Nessas equações, T , S e N correspondem à tendência, a sazonalidade e ao resíduo em um instante t , respectivamente.

$$Z_t = T_t + S_t + N_t \quad (3.2)$$

$$Z_t = T_t \times S_t \times N_t \quad (3.3)$$

No modelo aditivo (Equação 3.2), 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 3.3), 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.

3.3.1 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;

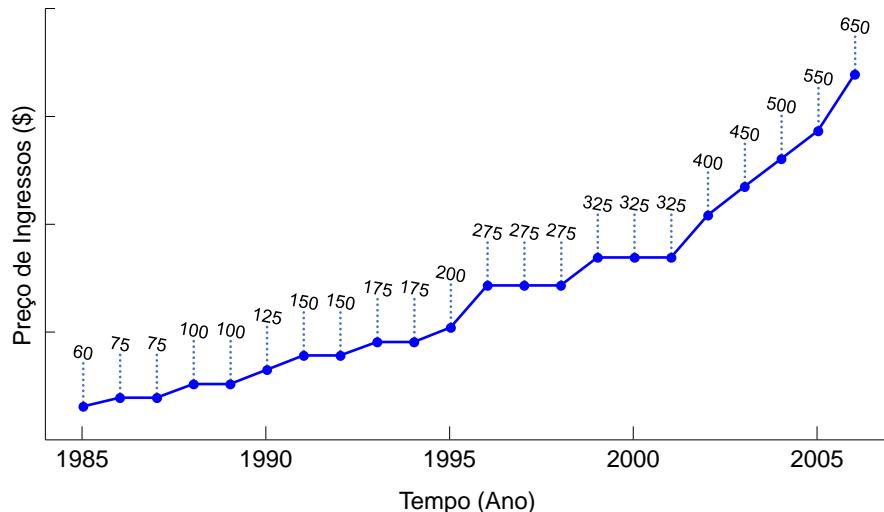
Crescimento Amortecido: Ocorre quando a taxa de crescimento de dados futuros é menor que os dados atuais, como em situações nas quais para um determinado ano o crescimento esperado é 70% do ano anterior.

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 (MORETTIN; TOLOI, 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.

Régressão 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 como X_1, X_2, \dots, X_m , o modelo

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

Fonte: Elaborada pelo autor.

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; HADI, 2012).

Uma das estruturas mais simples que relaciona duas varáveis X e Y pode ser visibilizada na Equação 3.4 (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.

$$Y = b_0 + b_1 X \quad (3.4)$$

Em ST, o uso da Equação 3.4 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 3.5 e 3.6. 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.

$$b_1 = \frac{m \sum x_i y_i - \sum x_i \sum y_i}{m \sum x_i^2 - (\sum x_i)^2} \quad (3.5)$$

$$b_0 = \frac{\sum y_i - (\sum x_i) b_1}{m} \quad (3.6)$$

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; 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 3.7 (BARROSO *et al.*, 1987).

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \dots + \beta_P X^P \quad (3.7)$$

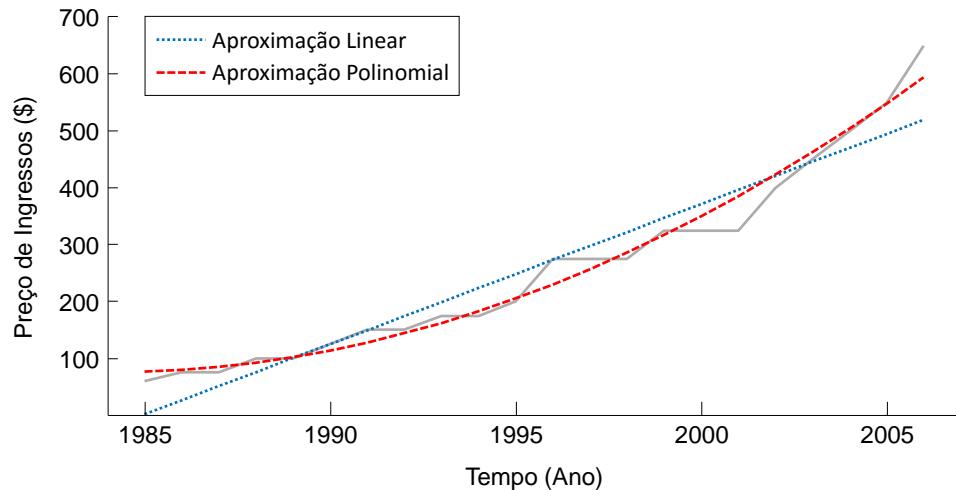
A Equação 3.7 comprehende a estrutura do modelo de regressão polinomial que, além de incluir a equação da reta, é um caso especial de ajuste linear múltiplo. Os coeficientes desse modelo são determinados a partir resolução do seguinte sistema:

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

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 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

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

Fonte: Elaborada pelo autor.

que esse procedimento move-se em direção ao último valor da série ([SCHILLER; SPIEGEL; SRINIVASAN, 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.

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

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	—

Fonte: Elaborada pelo autor.

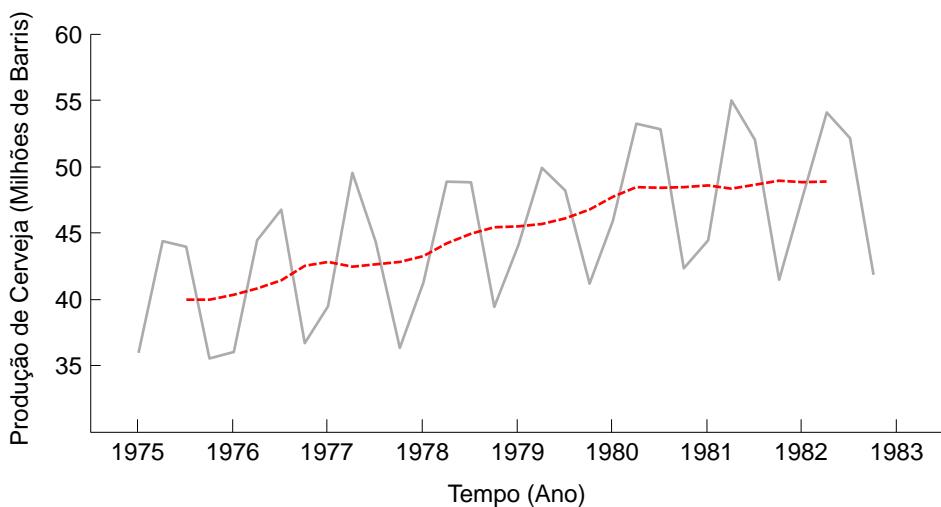
O parâmetro r é frequentemente escolhido como um valor ímpar para que o resultado da média possa ser vinculado a um período central com correspondente na série original. Porém, em algumas situações, para que a tendência seja obtida sem influência da sazonalidade, esse parâmetro precisa representar, em 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 1](#) 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 1](#), 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



Fonte: Elaborada pelo autor.

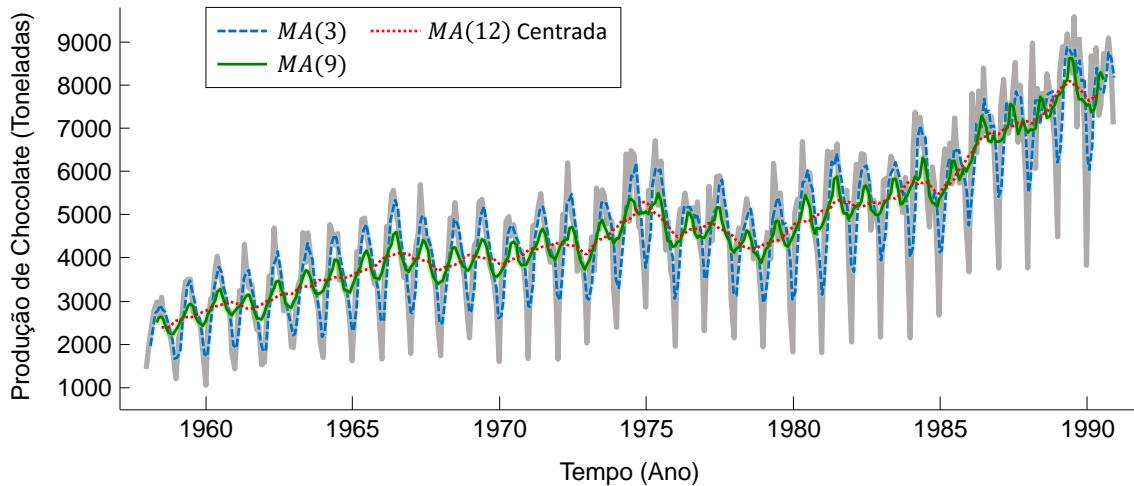
Na [Figura 9](#) são exibidas três linhas de tendência para a ST de produção mensal de chocolate na Austrália. Tais tendências foram obtidas usando *MA* com sete e nove períodos, além de doze períodos com centralização. Essa última configuração retrata adequadamente o comportamento de longo prazo da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

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

Tabela 1 – 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	40,15	—
	III	44,15	40,17	40,16
	IV	35,72	40,17	40,17
1976	I	36,19	40,87	40,52
	II	44,63	41,17	41,02
	III	46,95	42,04	41,60
	IV	36,90	43,31	42,67
1977	I	39,66	42,69	43,00
	II	49,72	42,60	42,65
	III	44,49	43,05	42,83
	IV	36,54	42,89	42,97
1978	I	41,44	44,01	43,45
	II	49,07	44,77	44,39
	III	48,98	45,48	45,13
	IV	39,59	45,74	45,61
1979	I	44,29	45,60	45,67
	II	50,09	46,05	45,82
	III	48,42	46,50	46,28
	IV	41,39	47,34	46,92
1980	I	46,11	48,49	47,91
	II	53,44	48,77	48,63
	III	53,00	48,39	48,58
	IV	42,52	48,83	48,61
1981	I	44,61	48,64	48,73
	II	55,18	48,42	48,53
	III	52,24	49,23	48,83
	IV	41,66	49,00	49,12
1982	I	47,84	49,02	49,01
	II	54,27	49,11	49,07
	III	52,31		—
	IV	42,03		—

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



Fonte: Elaborada pelo autor.

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 ($Z_t \div T_t = S_t \times N_t$).

3.3.2 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; 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; 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 relevar 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 58](#), 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; SARIN, 1987).

A norma do resíduo, definida pela [Equação 3.8](#), baseia-se na diferença entre o valor real da variável investigada e o seu valor esperado.

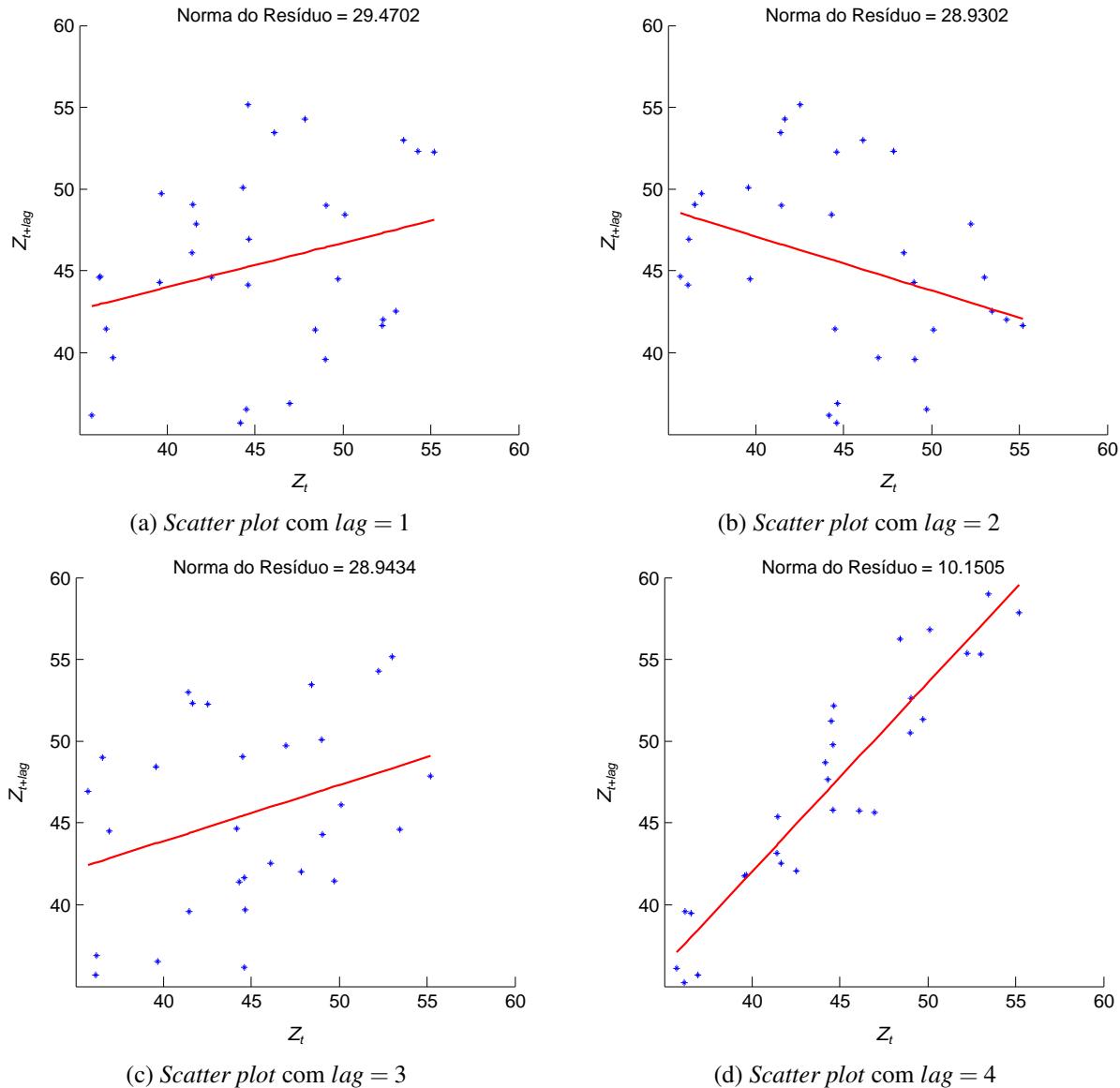
$$RN = \sqrt{\sum |Z_{t+lag} - \hat{Z}_{t+lag}|^2} \quad (3.8)$$

No *scatter plot*, o eixo das abscissas 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 58](#). A norma residual mínima foi computada usando *lag* de quatro períodos.

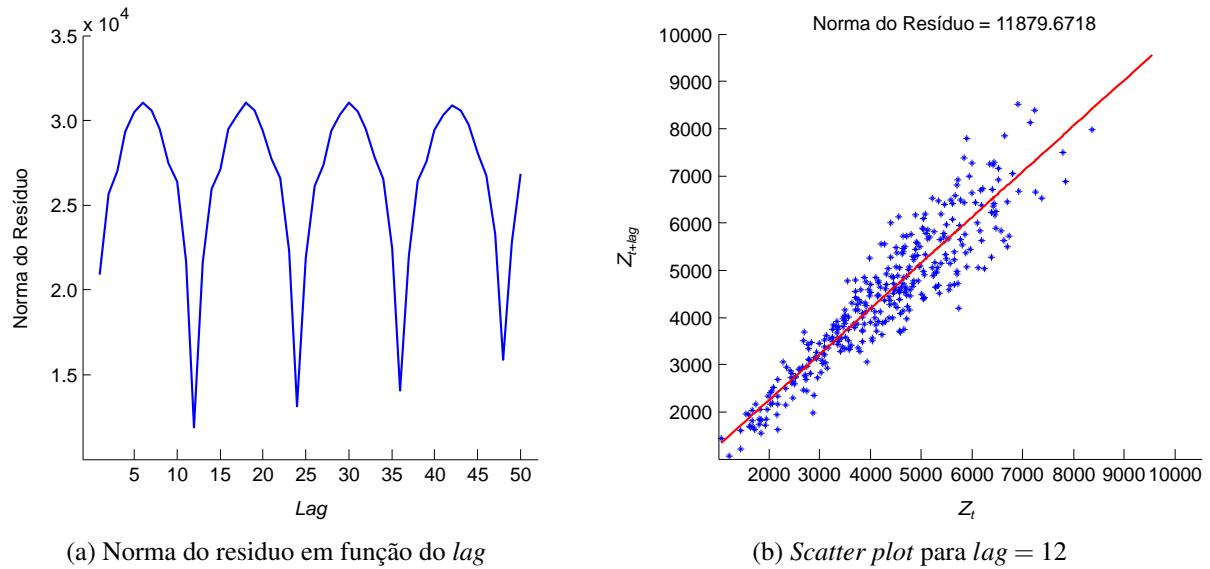
Figura 10 – Técnica do *scatter plot* para a série de produção trimestral de cerveja nos EUA



Fonte: Elaborada pelo autor.

A aplicação da técnica do *scatter plot* para a série de produção de chocolate, inicialmente mostrada na Figura 4 da página 53, é sintetizada na Figura 11. O gráfico da Figura 11a comprehende 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 11b.

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. A exemplo, têm-se o algoritmo da percentagem média, da relação percentual e dos elos relativos. Dentre

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

Fonte: Elaborada pelo autor.

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 no [Quadro 2](#), considerando tanto a decomposição do modelo aditivo quanto a do multiplicativo ([SCHILLER; SPIEGEL; SRINIVASAN, 2012](#)).

Quadro 2 – 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

No [Quadro 2](#), 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 2](#) 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 2 – Exemplo de obtenção dos índices sazonais pelo método da relação entre *MA*

Ano	Trimestre	Observação	<i>MA(4)</i> 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 2](#), a tendência foi determinada por *MA* centrada de quatro períodos como descrito na [Tabela 1](#). 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 3](#), os índices computados foram agrupados por trimestre de modo a originar uma única variação

sazonal (quatro trimestres no ano).

Tabela 3 – Exemplo de obtenção das medidas de síntese dos índices para cada período da variação sazonal utilizando o método da relação entre MA

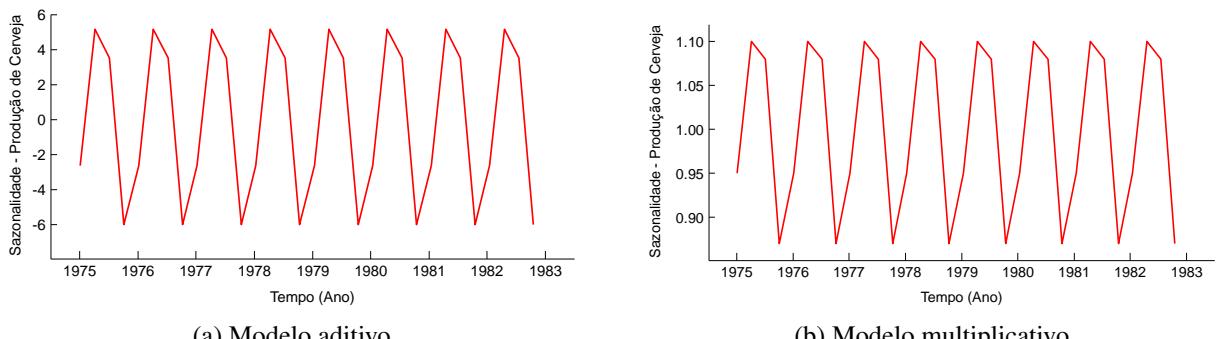
Trimestre	Índices Sazonais								Média	Índice Sazonal Médio	Modelo Aditivo	
	I	-4,33	-3,34	-2,01	-1,38	-1,80	-4,12	-1,17			-2,59 - (0,17 ÷ 4) = -2,64	5,18 - (0,17 ÷ 4) = 5,14
II	3,61	7,07	4,68	4,27	4,81	6,65	5,20	5,18			3,55 - (0,17 ÷ 4) = 3,50	-5,96 - (0,17 ÷ 4) = -6,01
III	3,99	5,35	1,66	3,85	2,14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

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

Na nona coluna da [Tabela 3](#) 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 3](#) são mostrados os ajustes realizados para que a soma das medidas extraídas fosse aproximadamente igual a zero, no caso aditivo, ou equivalente à 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



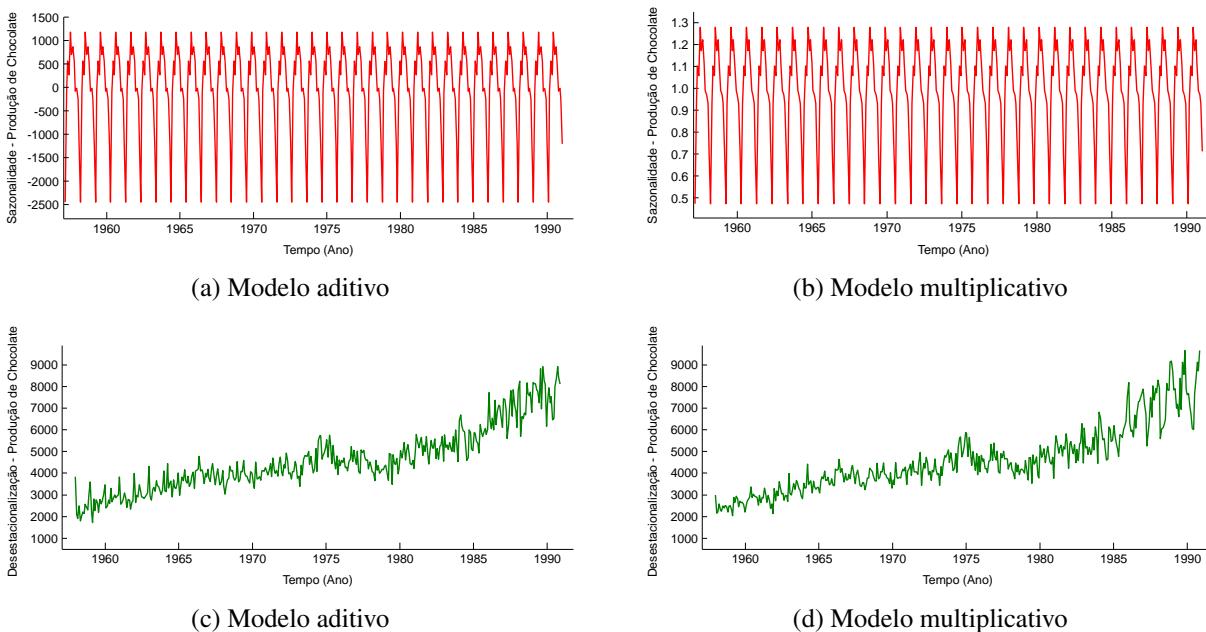
Fonte: Elaborada pelo autor.

Observa-se na [Figura 12a](#) que o componente sazonal modifica a tendência no modelo aditivo quando todos os índices médios são substancialmente maiores ou menores que zero. No modelo multiplicativo ([Figura 12b](#)), 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



Fonte: Elaborada pelo autor.

3.3.3 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 estimativa 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; WOLTERS; HASSSLER, 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 3.9 e 3.10. Nessas equações, Z refere-se à série e T e S indicam a tendência e a sazonalidade, respectivamente (COWPERTWAIT; METCALFE, 2009).

$$N_t = Z_t - (T_t + S_t) \quad (3.9)$$

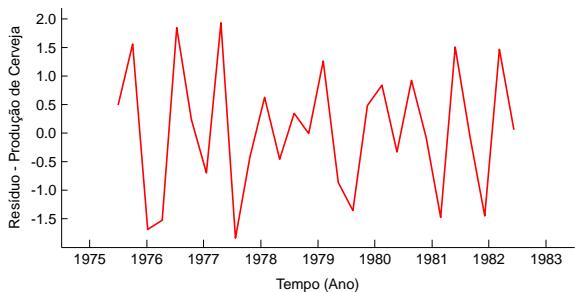
$$N_t = \frac{Z_t}{(T_t \times S_t)} \quad (3.10)$$

O impacto estocástico nas Equações 3.9 e 3.10 é restrinido 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.

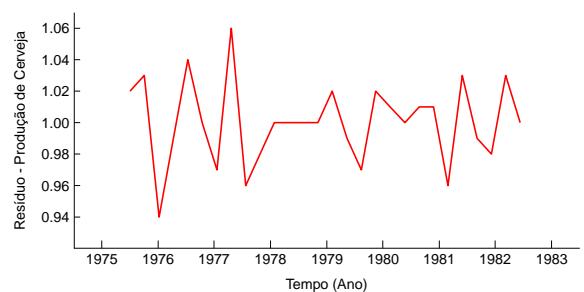
Na Tabela 4 é exibido o componente de tendência, estimado por MA com $r = 4$, o componente de sazonalidade, extraído conforme o método da relação entre MA , e o componente residual, resultado da aplicação das Equações 3.9 e 3.10.

O resíduo exposto na Tabela 4 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.

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



(a) Modelo aditivo



(b) Modelo multiplicativo

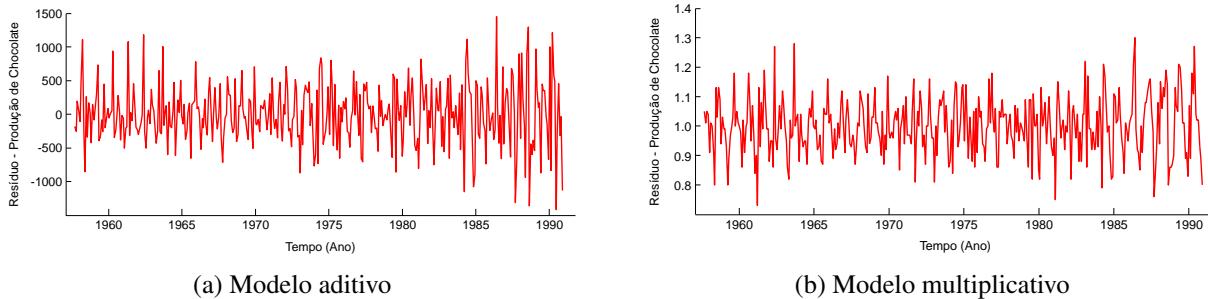
Fonte: Elaborada pelo autor.

Tabela 4 – 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
	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	—

Na [Figura 15](#) é novamente tratada da ST sobre chocolate apresentada na [Figura 4](#) da página 53. Na [Figura 15a](#), 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 15b](#) é ilustrado esse mesmo procedimento, mas empregando a decomposição multiplicativa.

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



Fonte: Elaborada pelo autor.

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.

3.4 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;

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 summarizar 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.

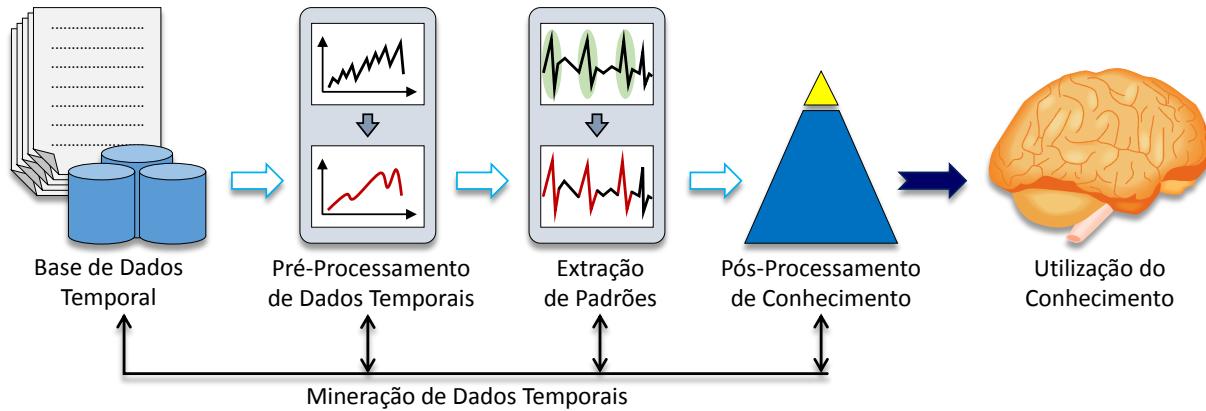
Em diversos domínios de aplicação, tem-se observado uma ampla disponibilidade de métodos e processos que podem guiar essas análises. O processo de Mineração de Dados (MD), por exemplo, que surgiu no final da década de oitenta, focaliza a extração de conhecimento a partir de grandes volumes de dados usando 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; KAMBER; PEI, 2011](#)).

3.5 Mineração de Dados Temporais

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; 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



Fonte: Elaborada pelo autor.

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 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.

3.6 Considerações Finais

No decorrer deste capítulo foram abordadas algumas das questões que norteiam a representação de dados coletados em ordem cronológica. Com base nessas questões, foi possível observar que a análise manual de dados que variam ao longo do tempo pode ser uma tarefa dispendiosa e sujeita a subjetividade, tornando-se, em alguns casos, impraticável devido à alta complexidade da relação entre os dados. Por outro lado, a possibilidade de representação desses dados por meio de ST, aliada à eminente necessidade de transformar tais dados em conhecimento útil para o suporte à decisão, tem despertado interesse e recebido investimentos significativos da comunidade científica e empresarial.

Grande parte dos investimentos concedidos concentra-se no planejamento e na execução de processos automatizados que colaboram para extração de conhecimento e análise inteligente de dados. Entretanto, os métodos tradicionais para descoberta de conhecimento possuem restrições à dados que apresentam características temporais. Nesse sentido, pesquisas vêm sendo desenvolvidas com a finalidade de permitir a adaptação desses métodos para a análise de dados cujo tempo constitui um fator importante.

No próximo capítulo o problema da predição de dados temporais é introduzido e, em seguida, formulado como um processo de busca que visa, a partir de um conjunto de informações conhecidas, estimar dados desconhecidos.



PREDIÇÃO DE SÉRIES TEMPORAIS

4.1 Considerações Iniciais

A predição de Séries Temporais (ST) permite que valores futuros de uma variável quantitativa sejam estimados tomando-se por base apenas seus dados históricos e respectivos inter-relacionamentos ([SORJAMAA et al., 2007](#)). Devido as suas distintas aplicações, essa tarefa é investigada por múltiplas áreas do conhecimento, por exemplo: em aviação, para predizer a demanda por viagens aéreas nacionais e internacionais ([CHANG; LIAO, 2010](#)); em medicina, para obter a quantidade diária aproximada de pacientes que visitam um determinado departamento ([VERPLANCKE et al., 2010](#)); em turismo, na predição do número de estrangeiros que realizam solicitações de reserva ou estadia ([CLAVERIA; TORRA, 2014](#)); e em economia, para estimar as taxas de câmbio ([AMAT; MICHALSKI; STOLTZ, 2014](#)).

Neste capítulo, em concordância com o arcabouço do processo de predição de ST, são descritos os principais métodos para projeção de valores futuros. Todos os procedimentos envolvidos na construção de um modelo preditivo, desde sua concepção até sua extração para períodos subsequentes, são relatados sob uma perspectiva didática que pretende mostrar o quanto a subárea de Aprendizado de Máquina (AM) pode contribuir para com a análise inteligente de dados sequenciais.

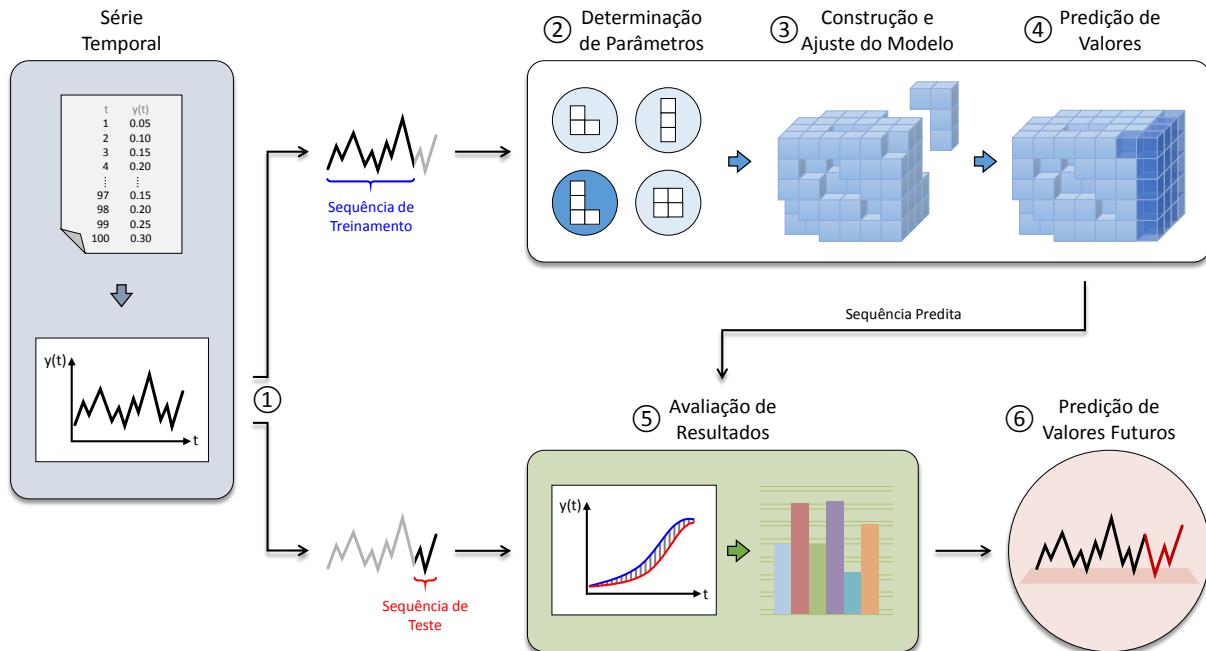
4.2 O Problema da Predição de Dados

A modelagem de 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](#).

Figura 17 – Processo de predição de valores em ST



Fonte: Elaborada pelo autor.

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.

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 predição dos novos valores e, mais que isso, são ampliadas ao passo que a predição se estende para períodos distantes no futuro.

Outra questão relevante atrelada à quarta etapa diz respeito à estratégia empregada para predizer os valores de uma ST vários períodos à frente (horizonte de predição $h > 1$). A estratégia mais intuitiva para essa finalidade é conhecida como recursiva (ou multi-etapa) e trata a predição de $h > 1$ como uma aplicação, conduzida h vezes sucessivamente, do modelo preditivo considerando $h = 1$. Após a extração do modelo, o valor predito ou o respectivo valor real é retroalimentado para o cálculo da predição 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 predição. Medir essa acurácia consiste em avaliar a extensão do erro de predição, 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 predição consideravelmente diferentes.

Na sexta etapa, após assegurada a qualidade do modelo, são realizadas predições para períodos futuros da ST. Nessa etapa, o erro de predição 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.

4.3 Métodos para Construção de Modelos Preditivos

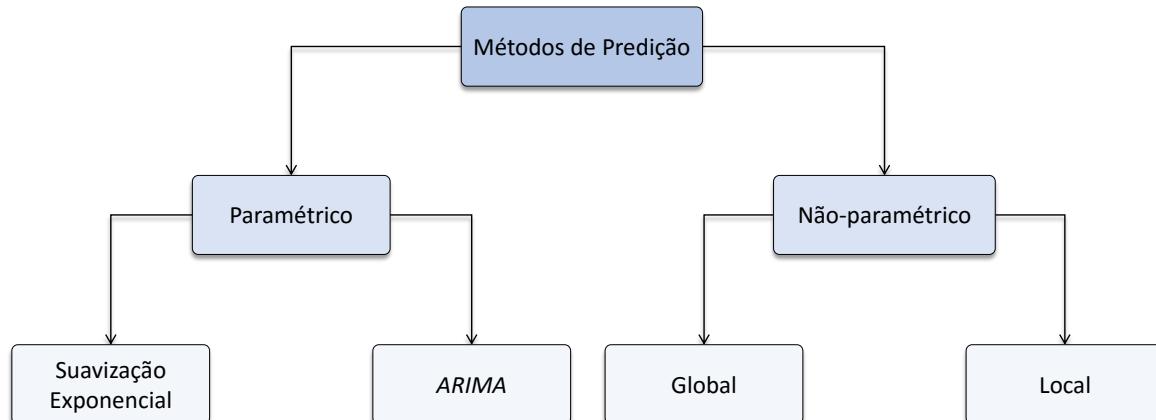
Os métodos de predição de ST evoluíram no decorrer dos anos, passando de simples técnicas de regressão para algoritmos robustos provenientes da Estatística e da Inteligência Artificial. Dependendo do conhecimento prévio acerca da distribuição dos dados, esses métodos podem ser agrupados (Figura 18) nas abordagens paramétrica (suavização exponencial ou baseada em autorregressão e médias móveis) e não-paramétrica (global ou local) (CHATFIELD, 2013; ISLAM; SIVAKUMAR, 2002).

Nas Subseções 4.3.1 e 4.3.2, as abordagens supracitadas são explicadas juntamente com os algoritmos mais populares para a projeção de dados temporais.

4.3.1 Métodos Paramétricos

Os métodos estatísticos se distinguem por utilizar, na confecção do modelo de predição, 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 predição. Segundo Morettin e Toloi (2006), os modelos estatísticos podem ser divididos, conforme o seu grau de complexidade

Figura 18 – Hierarquia de abordagens para predição de ST



Fonte: Elaborada pelo autor.

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 comprehende 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.

4.3.1.1 Médias Móveis

O modelo de Médias Móveis (*MA*) de ordem r , $MA(r)$, é resultado de uma técnica simples de predição 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 4.1](#) ([MONTGOMERY; JENNINGS; KULAHCI, 2015](#)), onde r

representa o número de observações incluídas na média z_{t+1} .

$$z_{t+1} = \frac{z_t + z_{t-1} + z_{t-2} + \dots + z_{t-r+1}}{r} \quad (4.1)$$

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; 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.

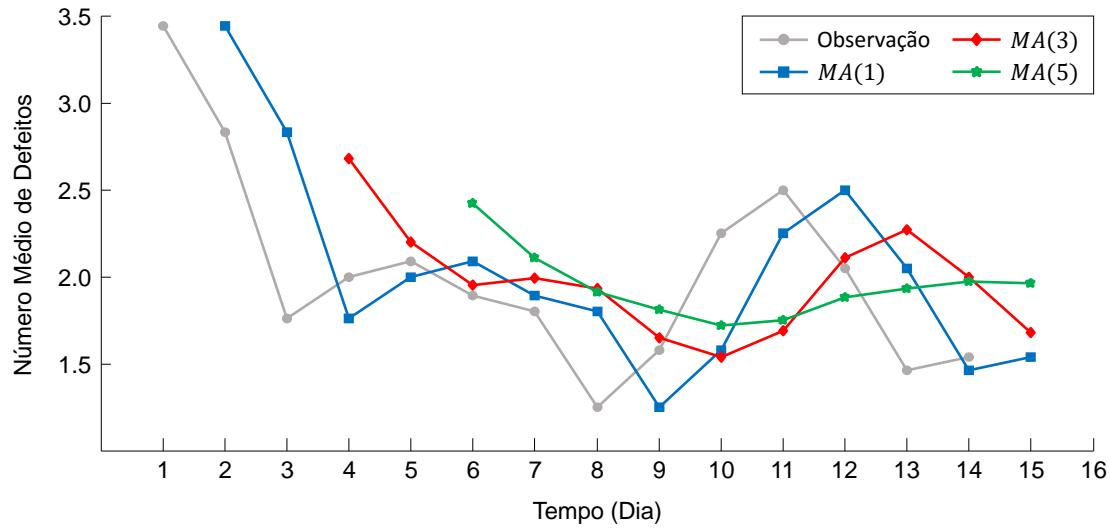
As desvantagens do modelo de *MA* estão atreladas a sua baixa acurácia ao lidar com ST que contenham os componentes de tendência e/ou de sazonalidade, haja vista que nesse método a predição do próximo valor envolve sempre a adição de novos dados e a desconsideração dos anteriores. Além disso, os pesos 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 5](#) 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.

Observa-se no exemplo da [Tabela 5](#) 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 19](#).

Tabela 5 – 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

Figura 19 – Predições obtidas pelo modelo de MA com parâmetro $r = 1, 3$ e 5 

Fonte: Elaborada pelo autor.

4.3.1.2 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 exerçam maior influência no cálculo das futuras previsões (MONTGOMERY; JENNINGS; KULAHCI, 2015; HYNDMAN *et al.*, 2002).

A estrutura do modelo de SES pode ser expressa conforme a Equação 4.2 (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.

$$L_t = \alpha z_t + \alpha(1 - \alpha)z_{t-1} + \alpha(1 - \alpha)^2z_{t-2} + \dots + \alpha(1 - \alpha)^{m-1}z_1 \quad (4.2)$$

No intuito de evitar a utilização de todas as observações a cada nova estimativa L , a Equação 4.2 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 4.3 (GARDNER, 1985).

$$L_t = \alpha z_t + (1 - \alpha)L_{t-1} \quad (4.3)$$

Frequentemente, no início do processo de predição via SES, supõe-se que o primeiro valor ajustado será igual ao primeiro valor da ST, ou seja, $L_1 = z_1$ (MORETTIN; TOLOI, 2006). Nesse caso, o procedimento de ajuste só começa a partir da segunda observação da série e a predição do instante $m + 1$ é dada pelo alisamento exponencial do último valor observado ($z_{m+1} = L_m$). Esse modo de predição 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 predição 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; TOLOI, 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 predição, 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, Jennings e Kulahci (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).

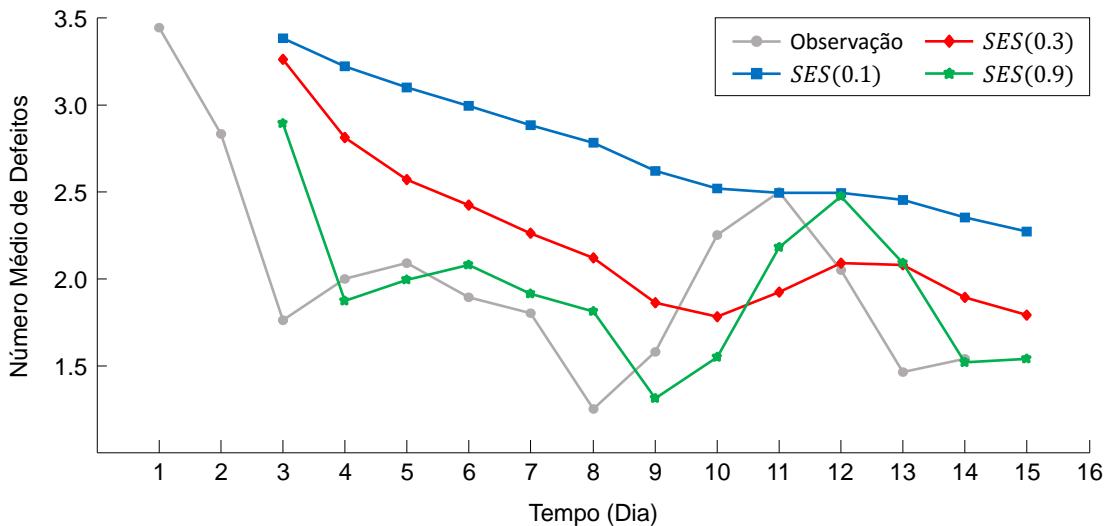
Na [Tabela 6](#) é exibida uma execução do modelo de *SES*, supondo $\alpha = 0, 1, 0,3$ e $0,9$, sobre as 14 observações da ST relativa ao número médio diário de defeitos de fabricação em caminhões nos EUA.

Tabela 6 – Dados projetados empregando o modelo de *SES* com parâmetro $\alpha = 0, 1, 0,3$ e $0,9$

Dia (t)	Observação (z_t)	<i>SES(0,1)</i>		<i>SES(0,3)</i>		<i>SES(0,9)</i>	
		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 6](#) em conformidade com os valores atribuídos à constante de suavização α , são esquematizadas na [Figura 20](#).

Figura 20 – Predições computadas pelo modelo de *SES* com parâmetro $\alpha = 0, 1, 0,3$ e $0,9$



Fonte: Elaborada pelo autor.

4.3.1.3 Suavização Exponencial de Holt

O modelo de *SES* quando aplicado sobre dados temporais que apresentam 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 4.4 e 4.5 (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.

$$L_t = \alpha z_t + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (4.4)$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \quad (4.5)$$

$$z_{t+h} = L_t + hT_t \quad (4.6)$$

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 4.4 e 4.5, respectivamente. Essas equações, assim como em qualquer método de suavização exponencial, modificam estimativas prévias quando uma nova observação é calculada. Complementarmente, na Equação 4.6, 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; 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 previsor do algoritmo.

Na Tabela 7 é apresentada uma aplicação do modelo de HES, considerando duas parametrizações ((i) $\alpha = 0,3$ e $\beta = 0,9$; (ii) $\alpha = 0,9$ e $\beta = 0,3$), 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 21.

4.3.1.4 Suavização Exponencial Sazonal de Holt-Winters

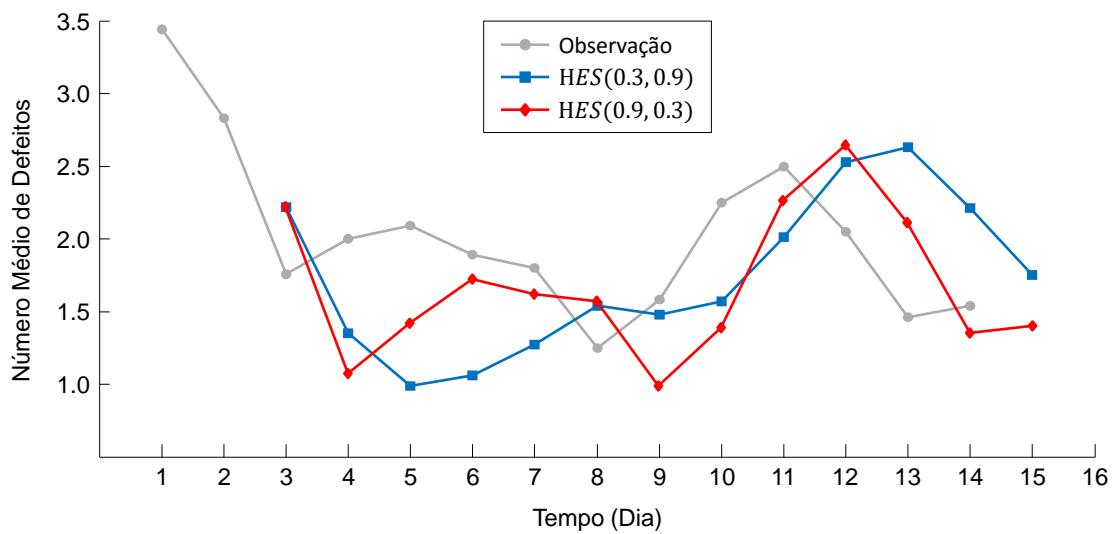
Métodos Holt-Winters (HW) são recomendados para ST que contemplam tanto o componente de tendência quanto o de sazonalidade. Uma série com essas características é descrita pela ocorrência de padrões cíclicos de variação que se repetem em intervalos relativamente constantes de tempo (MONTGOMERY; JENNINGS;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

Tabela 7 – Dados projetados usando o modelo de HES configurado de duas maneiras: (i) $\alpha = 0,3$ e $\beta = 0,9$; (ii) $\alpha = 0,9$ e $\beta = 0,3$

Dia (t)	Observação (z_t)	$\alpha = 0,3$ e $\beta = 0,9$			$\alpha = 0,9$ e $\beta = 0,3$		
		L_t	T_t	Predição (z_{t+1})	L_t	T_t	Predição (z_{t+1})
1	3,44	3,44	-0,61	—	3,44	-0,61	—
2	2,83	2,83	-0,61	—	2,83	-0,61	—
3	1,76	2,08	-0,73	2,22	1,81	-0,73	2,22
4	2,00	1,54	-0,56	1,35	1,91	-0,48	1,07
5	2,09	1,32	-0,26	0,99	2,02	-0,30	1,42
6	1,89	1,31	-0,03	1,06	1,87	-0,26	1,72
7	1,80	1,43	0,11	1,27	1,78	-0,21	1,62
8	1,25	1,45	0,03	1,54	1,28	-0,30	1,57
9	1,58	1,51	0,06	1,48	1,52	-0,14	0,99
10	2,25	1,77	0,24	1,57	2,16	0,10	1,39
11	2,50	2,16	0,37	2,01	2,48	0,16	2,26
12	2,05	2,39	0,24	2,53	2,11	0,00	2,64
13	1,46	2,28	-0,07	2,63	1,53	-0,17	2,11
14	1,54	2,01	-0,25	2,21	1,52	-0,12	1,35
15	—	—	—	1,75	—	—	1,40

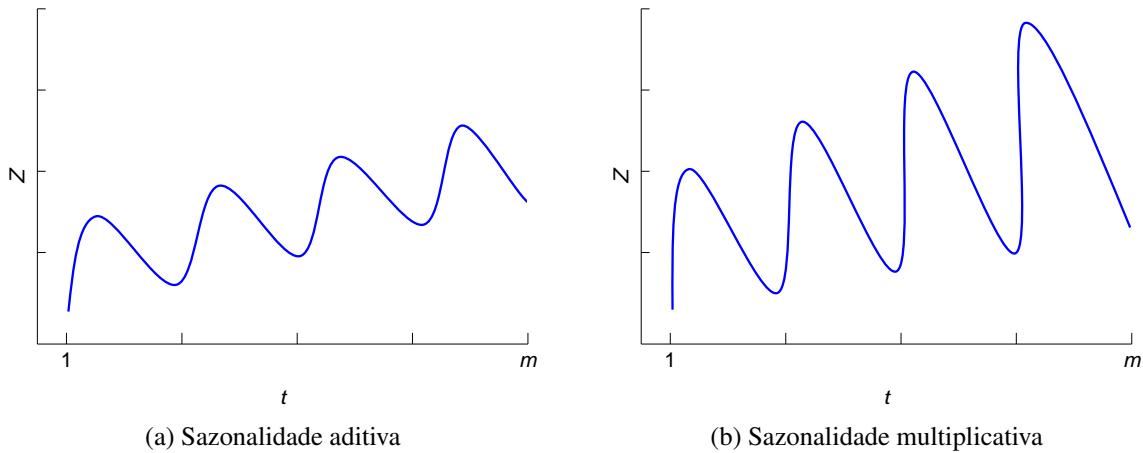
Figura 21 – Predições obtidas pelo modelo de HES parametrizado de dois modos: (i) $\alpha = 0,3$ e $\beta = 0,9$; (ii) $\alpha = 0,9$ e $\beta = 0,3$



Fonte: Elaborada pelo autor.

pela escolha de uma estrutura em detrimento da outra está relacionada com o tipo de padrão sazonal da série investigada ([Figura 22](#)).

[Figura 22 – Representação dos tipos de variação sazonal](#)



Fonte: Elaborada pelo autor.

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](#)):

$$L_t = \alpha \frac{z_t}{S_{t-s}} + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (4.7)$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \quad (4.8)$$

$$S_t = \gamma \frac{z_t}{L_t} + (1 - \gamma)S_{t-s} \quad (4.9)$$

$$z_{t+h} = (L_t + hT_t)S_{t-s+h} \quad (4.10)$$

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; TOLOI, 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 4.11](#)).

$$L_s = \frac{1}{s}(z_1 + z_2 + \dots + z_s) \quad (4.11)$$

A tendência pode ser inicializada utilizando duas estações completas, como definido pela [Equação 4.12](#).

$$T_s = \frac{1}{s} \left(\frac{z_{s+1} - z_1}{s} + \frac{z_{s+2} - z_2}{s} + \dots + \frac{z_{s+s} - z_s}{s} \right) \quad (4.12)$$

Os índices sazonais iniciais podem ser determinados pela razão entre as primeiras observações e a média do primeiro período, como mostrado pela [Equação 4.13](#).

$$S_1 = \frac{z_1}{L_s}, S_2 = \frac{z_2}{L_s}, \dots, S_s = \frac{z_s}{L_s} \quad (4.13)$$

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:

$$L_t = \alpha(z_t - S_{t-s}) + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (4.14)$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \quad (4.15)$$

$$S_t = \gamma(z_t - L_t) + (1 - \gamma)S_{t-s} \quad (4.16)$$

$$z_{t+h} = L_t + hT_t + S_{t-s+h} \quad (4.17)$$

Observa-se que a [Equação 4.8](#) do modelo multiplicativo de HW é idêntica à [Equação 4.15](#) 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 4.18](#).

$$S_1 = z_1 - L_s, S_2 = z_2 - L_s, \dots, S_s = z_s - L_s \quad (4.18)$$

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.

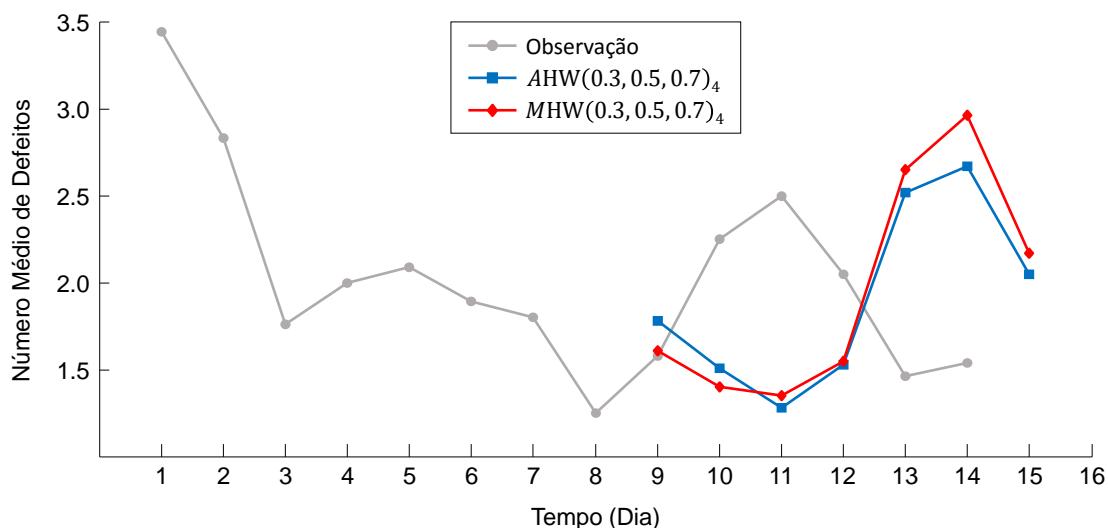
Na [Tabela 8](#), utilizando a ST que expressa o número médio diário de defeitos de fabricação em caminhões nos EUA, é resumida uma execução dos modelos aditivo e multiplicativo de HW. Foram adotados, para ambos os modelos, $\alpha = 0,3$, $\beta = 0,5$, $\gamma = 0,7$ e $s = 4$.

Tabela 8 – 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

A fim de oportunizar comparações, as predições elencadas na [Tabela 8](#) são mostrados graficamente na [Figura 23](#).

Figura 23 – Predições computadas pelos modelos aditivo e multiplicativo de HW, ambos com parâmetros $\alpha = 0,3$, $\beta = 0,5$, $\gamma = 0,7$ e $s = 4$



Fonte: Elaborada pelo autor.

4.3.1.5 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) integração¹ 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 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; JENNINGS; KULAHCI, 2015):

- $ARIMA(p, 0, 0) = AR(p)$;
- $ARIMA(0, 0, q) = MA(q)$;
- $ARIMA(p, 0, q) = ARMA(p, q)$ ².

No modelo Autorregressivo de ordem p , $AR(p)$ (Equação 4.19), 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.

$$z_t = \delta + \sum_{i=1}^p \phi_i z_{t-i} + e_t \quad (4.19)$$

Na Equação 4.19, 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 à Equação 4.20, em que μ representa a média do processo estacionário; e e_t comprehende 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 ($z_{t-1}, z_{t-2}, \dots, z_{t-m+1}$).

$$\delta = \mu(1 - \phi_1 - \phi_2 - \dots - \phi_p) \quad (4.20)$$

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

¹ Integração 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 $\Delta^d z_t = \Delta(\Delta^{d-1} z_t)$.

² Autorregressivo de Médias Móveis (*ARMA*).

conforme a Equação 4.21, onde θ_i é o i -ésimo valor de ponderação usado e e_t corresponde ao ruído branco no instante t .

$$z_t = \mu + \sum_{i=1}^q \theta_i e_{t-i} + e_t \quad (4.21)$$

Observa-se na Equação 4.21 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 predizer o valor z_{t+1} , move-se os pesos utilizados, aplicando-os nos ruídos $e_{t+1}, e_t, e_{t-1}, \dots, e_{t-q+1}$.

A fim de aumentar a flexibilidade na modelagem de ST, as Equações 4.19 e 4.21 são combinadas para obter o modelo Autorregressivo de Médias Móveis com ordem (p, q) , $ARMA(p, q)$ (Equação 4.22), que incorpora modelos mistos $AR(p)$ e $MA(q)$ (SHUMWAY; 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.

$$z_t = \delta + \sum_{i=1}^p \phi_i z_{t-i} + \sum_{i=1}^q \theta_i e_{t-i} + e_t \quad (4.22)$$

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 4.23 (MONTGOMERY; JENNINGS; KULAHCI, 2015).

$$I'_t = \delta + \sum_{i=1}^p \phi_i I'_{t-i} + \sum_{i=1}^q \theta_i e_{t-i} + e_t \quad (4.23)$$

Na Equação 4.23, $I'_t = \Delta^d z_t = \Delta(\Delta^{d-1} z_t)$ 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, consequentemente, 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; 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 4.23), com parâmetros (p, d, q) , e uma sazonal (Equação 4.24), com parâmetros $(P, D, Q)_s$.

$$I_t'' = \delta + \sum_{i=1}^P \Phi_{is} I_{t-is}'' + \sum_{i=1}^Q \Theta_{is} e_{t-is} + e_t \quad (4.24)$$

Na Equação 4.24, $I_t'' = \Delta^D z_t = \Delta(\Delta^{D-1} z_t)$ e D indica o grau do operador de diferença sazonal; a constante δ é computada segundo a Equação 4.25 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.

$$\delta = \mu(1 - \Phi_1 - \Phi_2 - \cdots - \Phi_p) \quad (4.25)$$

O *SARIMA* $(p, d, q) \times (P, D, Q)_s$ é formalizado pela Equação 4.26, onde somam-se as partes não-sazonal e sazonal.

$$I_t = \delta + \sum_{i=1}^p \phi_i I_{t-i}' + \sum_{i=1}^q \theta_i e_{t-i} + \sum_{i=1}^P \Phi_{is} I_{t-is}'' + \sum_{i=1}^Q \Theta_{is} e_{t-is} + e_t \quad (4.26)$$

No modelo *SARIMA*, a constante δ é calculada por meio da Equação 4.27 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; METCALFE, 2009).

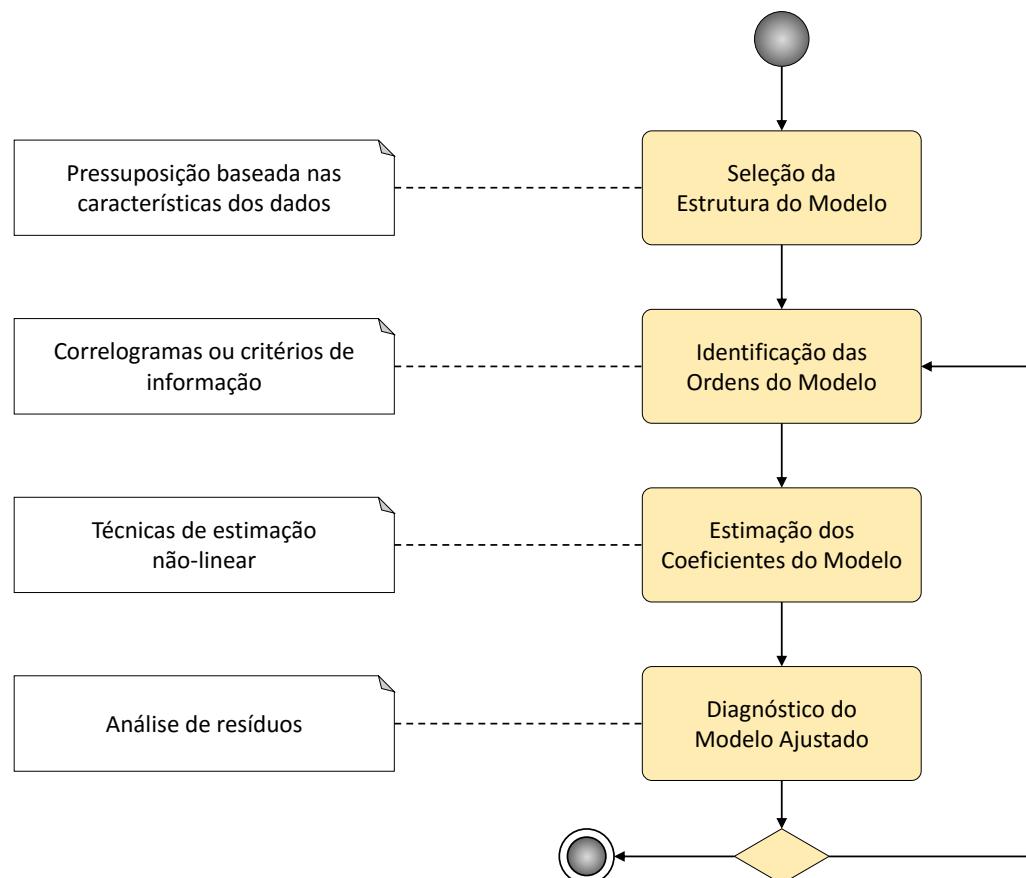
$$\delta = \mu(1 - \phi_1 - \phi_2 - \cdots - \phi_p)(1 - \Phi_1 - \Phi_2 - \cdots - \Phi_p) \quad (4.27)$$

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; METCALFE, 2009).

Segundo Morettin e Toloi (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 24, viabiliza a identificação do processo estocástico gerador dos dados e dos seus parâmetros de ajuste.

Figura 24 – Diagrama de atividades para o fluxo de construção de um modelo *ARIMA* ou *SARIMA*



Fonte: Adaptada 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 4.28](#) (COWPERTWAIT; METCALFE, 2009), penaliza a qualidade do ajuste de modelos com muitos parâmetros.

$$AIC = -2 \times LL + (\log(n) + 1) \times NP \quad (4.28)$$

Na [Equação 4.28](#), LL refere-se ao logaritmo da função de verossimilhança, n comprehende 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; TOLOI, 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.

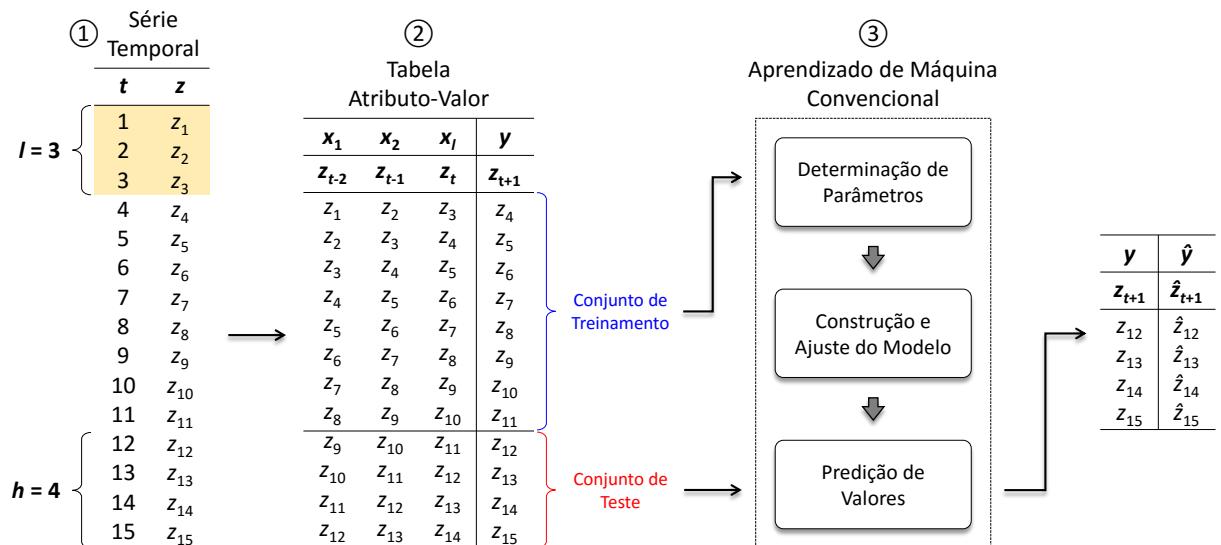
Nota-se que o emprego de modelos da categoria *ARIMA* exige conhecimento especializado tanto no domínio de aplicação quanto em matemática computacional. Além disso, a percepção e a experiência do analista são fundamentais para que o processo de modelagem se torne mais prático e menos dispendioso.

4.3.2 Métodos Não-paramétricos

Os métodos de 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 25](#) é ilustrado o processo de predição de ST com o emprego dessa abordagem.

Figura 25 – Exemplificação do processo de predição de ST segundo a abordagem global



Fonte: Elaborada pelo autor.

Na Figura 25, 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 predizer 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 26 é exemplificada a aplicação da estratégia de predição multi-etapa à frente, com passos aproximado e atualizado, para calcular as estimativas \hat{y}_k .

Figura 26 – Representação da estratégia de predição multi-etapa à frente para a abordagem global

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}
z_{10}	z_{11}	\hat{z}_{12}	z_{13}	\rightarrow	\hat{z}_{13}	z_{10}	z_{11}	z_{12}	\hat{z}_{13}
z_{11}	\hat{z}_{12}	\hat{z}_{13}	z_{14}	\rightarrow	\hat{z}_{14}	z_{11}	z_{12}	z_{13}	\hat{z}_{14}
\hat{z}_{12}	\hat{z}_{13}	\hat{z}_{14}	z_{15}	\rightarrow	\hat{z}_{15}	z_{12}	z_{13}	z_{14}	\hat{z}_{15}

(a) Passo aproximado

(b) Passo atualizado

Fonte: Elaborada pelo autor.

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 degradação 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).

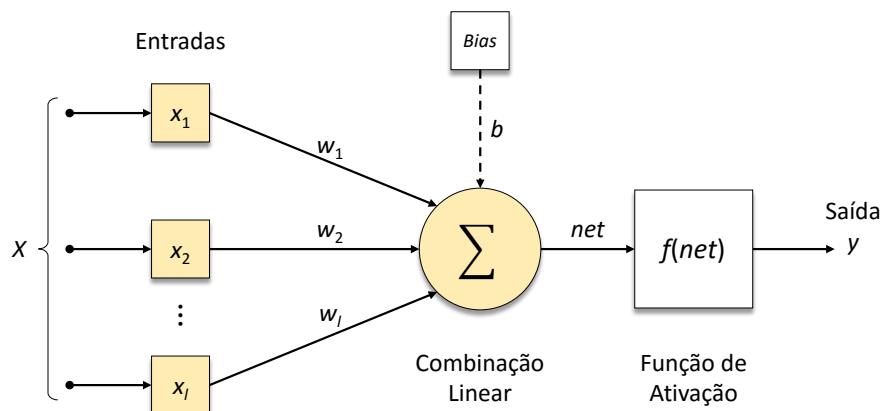
Diferentemente da aproximação global, a abordagem local contempla algoritmos de AM que foram adaptados para incluir, no processo de predição, a informação temporal associada aos dados. Tais métodos particionam a ST original em subsequências cujos valores mais próximos ou mais importantes em relação ao 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*.

4.3.2.1 Redes Neurais Artificiais

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; PATUWO; HU, 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 27, 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; PITTS, 1943).

Figura 27 – Estrutura do *Perceptron*



Fonte: Adaptada de Haykin (2009).

Na Figura 27, 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 4.29). Tal valor é enviado para uma função de ativação f que define a saída y do neurônio.

$$net = \sum_{i=1}^l w_i x_i + b \quad (4.29)$$

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 sigmoidal, 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).

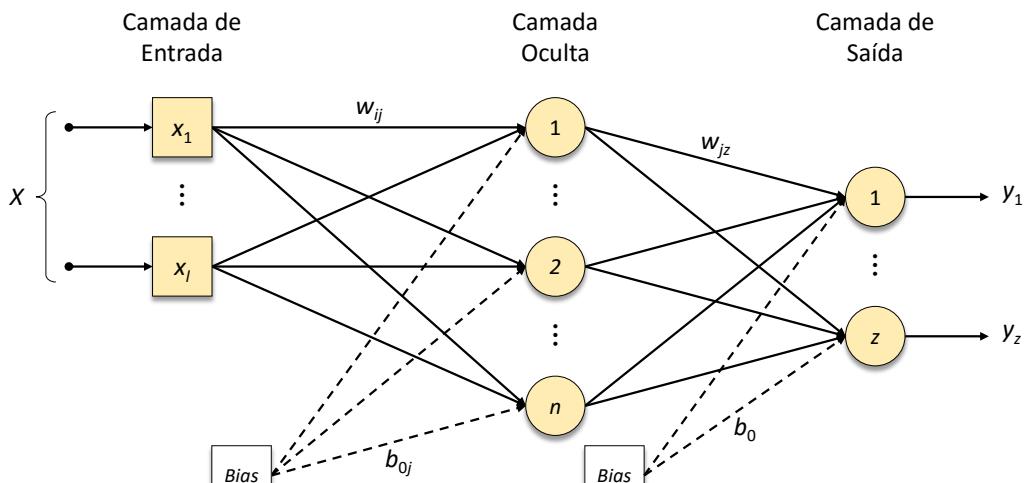
Apesar de ter causado grande impacto nas pesquisas de Inteligência Artificial, o estudo do *Perceptron* foi subitamente interrompido devido às críticas relatadas em Minsky e Papert (1969). No referido trabalho, a partir de uma análise matemática rigorosa, foi demonstrado que o modelo de um neurônio não 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 à 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 realmente retomado após a proposição do algoritmo de aprendizado *Backpropagation* (RUMELHART; HINTON; WILLIAMS, 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 28 é mostrada a estrutura de uma rede *MLP* com três camadas.

Figura 28 – Estrutura de uma rede *MLP* com camada oculta única



Fonte: Elaborada pelo autor.

A saída do modelo esquematizado na Figura 28, porém considerando um único neurônio na camada de saída, pode ser representada equacionalmente do seguinte modo:

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

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 comprehensíveis, 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.

4.3.2.2 Máquinas de Suporte Vetorial

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).

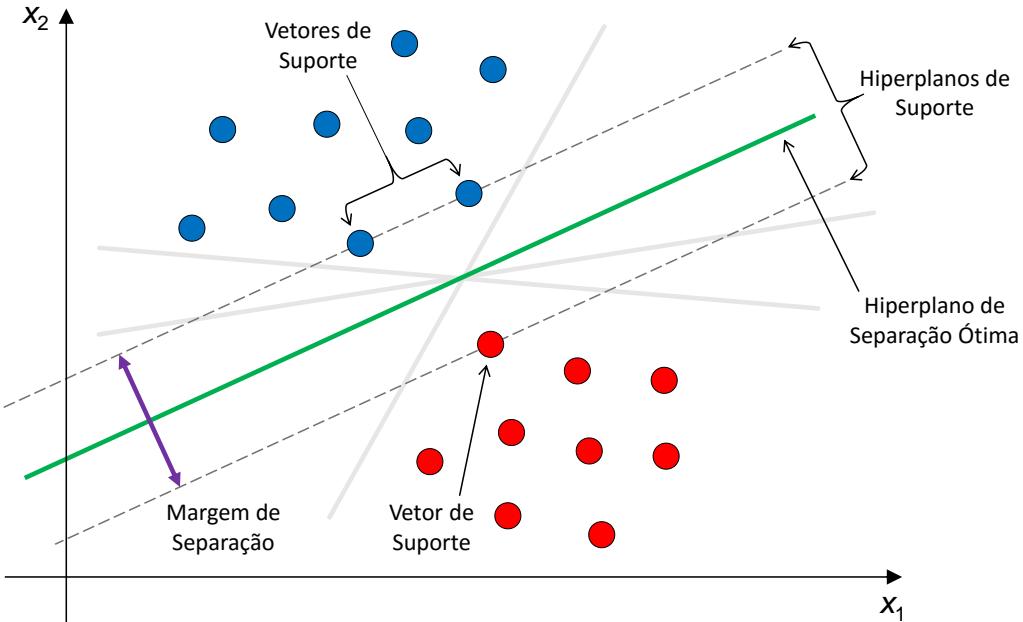
Apesar dos modelos *SVM* apresentarem uma estrutura semelhante a das redes neurais, eles divergem no modo como o aprendizado é conduzido. Enquanto as *ANN* trabalham com a minimização do risco empírico, ou seja, a minimização do erro do modelo induzido sobre os dados de treinamento, as *SVM* são fundamentadas 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; 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 29 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 29 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 de dimensão maior (espaço de características). Tal mapeamento é alcançado mediante ao uso de uma função *kernel* (CRISTIANINI; 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 30. Nessa figura, o mapeamento foi realizado pela função *kernel* $\mathbf{K}(x_i, x_j)$.

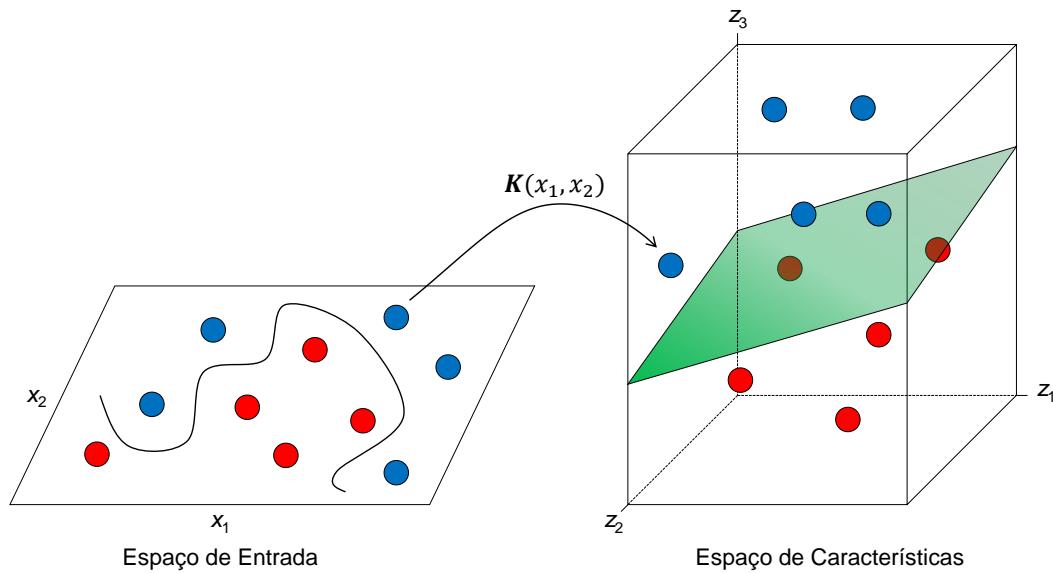
Os *kernels* mais utilizados na prática são os polinomiais, os gaussianos ou Funções de Base Radial (*RBF*) e os sigmoidais. Cada um deles abrange parâmetros que precisam ser estabelecidos pelo usuário. Por exemplo, o modelo *SVM* com *kernel RBF* exige dois parâmetros, C e σ . A constante C é um termo 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; CARVALHO, 2007). O número de funções radiais e os seus respectivos centros são determinados pelos vetores de

Figura 29 – 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



Fonte: Elaborada pelo autor.

Figura 30 – Mapeamento de dados para um espaço de características de mais alta dimensão utilizando como artifício a função *kernel*



Fonte: Adaptada de Kaundal, Kapoor e Raghava (2006).

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, consequentemente, de predição de ST (RISTANOSKI; LIU; BAILEY, 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.

4.3.2.3 *k*-Vizinhos mais Próximos

Algoritmos baseados em similaridade, por exemplo o *k-Nearest Neighbors (kNN)* (FIX; HODGES, 1951), são caracterizados por não construírem um modelo geral que descreve explicitamente o comportamento do conjunto de dados de treino. O modelo é confeccionado pelo simples armazenamento da amostra de dados. A generalização sobre o conjunto de treinamento é efetuada a cada momento em que é solicitado ao algoritmo uma nova classificação. No Capítulo 5 é apresentado, em detalhes, o funcionamento do algoritmo *kNN*, bem como sua respectiva adaptação para a tarefa de predição de ST (PARMEZAN; BATISTA, 2015; FERRERO, 2009).

4.4 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.

4.4.1 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 1](#) transcreve a lógica dessa técnica de amostragem para estabelecer os parâmetros α , β , γ e s do método multiplicativo de Holt-Winters.

Algoritmo 1: 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: P  

1 begin  

2   /* realmax corresponde ao maior número do sistema de ponto
      flutuante */  

3   min_error = realmax;  

4   h' = (max_p + h) ÷ 2;  

5   for s ← 3 : 2 : max_p do  

6     for α ← 0 : 0.25 : 1 do  

7       for β ← 0 : 0.25 : 1 do  

8         for γ ← 0 : 0.25 : 1 do  

9           error = holt_winters(S,s,α,β,γ,h',type = "MULTI");  

10          if error < min_error then  

11            min_error = error;  

12            sbest = s;  

13            αbest = α;  

14            βbest = β;  

15            γbest = γ;  

16          end  

17        end  

18      end  

19    end  

20    P ← {sbest,αbest,βbest,γbest};  

21    return P;  

22 end
  
```

No [Algoritmo 1](#) 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 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.

4.4.2 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 2](#) exemplifica essa técnica para identificar os parâmetros l , \mathbb{C} e σ do algoritmo de regressão *SVM*.

No [Algoritmo 2](#), a ideia por trás da busca pelo conjunto de parâmetros mais adequado é similar à visibilizada no [Algoritmo 1](#), 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 25](#) da página [página 91](#). 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.

4.4.3 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 3](#) exemplifica o uso desse método, intitulado de Box-Jenkins e introduzido na [página 89](#), para determinar os sete parâmetros do modelo *SARIMA* de ordem $(p,d,q) \times (P,D,Q)_s$.

No [Algoritmo 3](#), 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, utilizando a técnica do *scatter plot* ([Capítulo 3](#)). Na décima linha,

Algoritmo 2: *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: P
1 begin
    /* realmax corresponde ao maior número do sistema de ponto
       flutuante */ 
2     min_error = realmax;
3     for l ← 3 : 2 : max_p do
4         // T indica uma tabela atributo-valor
5         T ← generate_data_table(S,l);
6         for C ← 0 : 0.25 : 1 do
7             for σ ← 0.005 : 0.05 : 0.25 do
8                 error = cross_validation(T,kFolds,C,σ,model = "SVM");
9                 if error < min_error then
10                     min_error = error;
11                     l_best = l;
12                     C_best = C;
13                     σ_best = σ;
14                 end
15             end
16         end
17         P ← {l_best,C_best,σ_best};
18         return P;
19 end

```

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 4.28 da página 90).

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 3.

Algoritmo 3: 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: P

1 begin
2     /* realmax corresponde ao maior número do sistema de ponto
       flutuante */ 
3     best_aic ← realmax;
4     // Obtenção do comprimento da subsequência de treino
5     n ← length(S);
6     for p ← 0 : max_ord1[1] do
7         for d ← 0 : max_ord1[2] do
8             for q ← 0 : max_ord1[3] do
9                 for P ← 0 : max_ord2[1] do
10                for D ← 0 : max_ord2[2] do
11                    for Q ← 0 : max_ord2[3] do
12                        if d+D ≤ 1 then
13                            | fit = sarima(S,[p,d,q],[P,D,Q],max_p,δ = TRUE);
14                        else
15                            | fit = sarima(S,[p,d,q],[P,D,Q],max_p,δ = FALSE);
16                        end
17                        /* Computação do Critério de Informação de
                           Akaike */ 
18                        fit_aic = -2 * fit.loglik + (log(n)+1) * length(fit.coef);
19                        if fit_aic < best_aic then
20                            | best_aic ← fit_aic;
21                            | best_fit ← fit;
22                            | best_model ← [p,d,q,P,D,Q,max_p];
23                        end
24                    end
25                end
26            end
27            P ← {best_aic,best_fit,best_model};
28            return P;
29 end

```

4.5 Avaliação da Qualidade de Predição

As abordagens habituais para a seleção de algoritmos em Mineração de Dados, além de empregar conhecimento especialista, envolve procedimentos metódicos de avaliação empírica. Atrelada a essa necessidade de escolha está a questão de que índices considerar para constatar quais métodos apresentaram um desempenho melhor que outros.

No tema de predição de ST, medidas de erro constituem operações fundamentais para quantificar a diferença entre os valores reais observados (z_1, z_2, \dots, z_h) e os valores preditos pelo modelo ajustado ($\hat{z}_1, \hat{z}_2, \dots, \hat{z}_h$). Por outro lado, quando o horizonte de predição h é grande o suficiente, índices de desempenho arrojados, isto é, que oportunizam uma avaliação mais completa da qualidade dos dados projetados, podem ser extraídos a fim de se obter uma estimativa melhor do erro verdadeiro de determinado algoritmo quando aplicado sobre uma ST (HYNDMAN; KOEHLER, 2006).

Entre as medidas extensivamente usadas para calcular os erros de predição está o Erro Médio Percentual Absoluto (*MAPE*) cuja estrutura é denotada pela Equação 4.30 (HYNDMAN; KOEHLER, 2006).

$$MAPE = \frac{1}{h} \sum_{t=1}^h \left| \frac{z_t - \hat{z}_t}{z_t} \right| \times 100 \quad (4.30)$$

O resultado da aplicação da Equação 4.30 é um valor percentual que relaciona o valor predito com o valor real da série. Embora essa medida possibilite a comparação de erros entre sequências de dados com escalas diferentes, ela abrange uma deficiência prática: se uma ST possuir valores zero, ocorrerá uma imprópria divisão por zero. O *MSE*, formalizado pela Equação 4.31 (HYNDMAN; KOEHLER, 2006), não contempla esse problema e, por esse motivo, pode ser elegido como uma alternativa à medida *MAPE*.

$$MSE = \frac{1}{h} \sum_{t=1}^h (z_t - \hat{z}_t)^2 \quad (4.31)$$

Na Equação 4.31, a soma quadrática dos erros de predição é dividida pela quantidade de observações investigadas. Logo, espera-se que a soma de todos os erros computados seja próxima do valor zero. Tanto o *MSE* quanto o *MAPE* são úteis, por exemplo, para comparar dois modelos preditivos. Contudo, tais medidas não indicam se a qualidade das previsões realizadas por um algoritmo em particular é razoável (CLEMENTS; HENDRY, 1993).

A fim de explorar índices que permitam avaliar a qualidade de um preditor relativa aos dados (FAIR, 1986), define-se por meio da Equação 4.32 o coeficiente *U* de Theil (*TU*) (THEIL, 1971). Esse coeficiente 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 .

$$TU = \frac{\sum_{t=1}^h (z_t - \hat{z}_t)^2}{\sum_{t=1}^h (z_t - z_{t-1})^2} \quad (4.32)$$

De acordo com o resultado da [Equação 4.32](#), têm-se que:

- 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;
- Se $TU \leq 0,55$, o algoritmo investigado é confiável para efetuar previsões futuras.

Nota-se que para um modelo de previsão ideal o coeficiente U de Theil tenderá à zero. Ainda, como mencionado, o preditor trivial adotado por esse coeficiente pressupõe que a observação atual é a melhor previsão para o momento seguinte. Obviamente, essa estratégia torna o modelo trivial difícil de ser superado quando tratadas previsões vários passos à frente.

Outro índice de desempenho importante é o designado de *Prediction Of Change In Direction (POCID)*. Essa medida, expressa pela [Equação 4.33](#), mensura a taxa de acerto quanto à tendência da ST.

$$POCID = \frac{\sum_{t=1}^h D_t}{h} \times 100 \quad (4.33)$$

Na [Equação 4.33](#), 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.

A avaliação do desempenho preditivo por meio do emprego das medidas descritas é tão essencial quanto a análise da distribuição gráfica dos erros de previsão. O retrato desses erros ao longo da série precisa compor um padrão aleatório, não uniforme. Esse comportamento é aguardado porque em um modelo ótimo teórico os erros de previsão resultam de flutuações erráticas dos dados, causadas por fatores externos e não previsíveis.

4.6 Considerações Finais

As pesquisas em previsão de ST utilizam desde métodos estatísticos complexos, com uma quantidade considerável de parâmetros, até modelos intuitivos, simples e de fácil implementação. Diante disso, torna-se necessário investigar e ponderar vantagens e desvantagens do uso de cada método nos domínios de aplicação desejados.

Neste capítulo foram descritos alguns dos modelos de predição de ST pertencentes às abordagens paramétrica, que inclui métodos de suavização exponencial e da categoria *ARIMA*, e não-paramétrica, a qual abrange modelos de aproximação global e local.

Como um algoritmo para projeção de dados temporais não condiz, necessariamente, a uma fórmula de avaliação do desempenho preditivo. Foi necessário especificar, além do modelo, uma função de perda capaz de expressar a acurácia de predição. Uma função-perda frequentemente empregada é o Erro Médio Percentual Absoluto, embora em alguns cenários, outras funções-perdas, como o Erro Quadrático Médio, sejam mais apropriadas.

No próximo capítulo é exibida uma adaptação do algoritmo *k-Nearest Neighbors* para predição de ST, bem como os problemas abordados pelo mesmo. Em seguida, é apresentada uma proposta de variação desse método, a qual visa sanar os problemas encontrados por meio da combinação apropriada de invariância à amplitude, invariância à deslocamento, e da invariância de complexidade associada a uma política para evitar casamentos triviais.



ALGORITMO *k-NEAREST NEIGHBORS* PARA PREDIÇÃO DE SÉRIES TEMPORAIS

5.1 Considerações Iniciais

O rápido avanço das pesquisas científicas na área de Mineração de Dados (MD) impulsionou a adaptação dos métodos convencionais de extração de padrões para o contexto de análise de Séries Temporais (ST). Particularmente, para a tarefa de projeção de valores futuros têm sido aplicados, principalmente, métodos de extração de padrões baseados em similaridade (PARMEZAN; BATISTA, 2015; JUNIOR, 2012; FERRERO, 2009), Máquinas de Suporte Vetorial (RISTANOSKI; LIU; BAILEY, 2013; SAPANKEVYCH; SANKAR, 2009) e Redes Neurais Artificiais (CLAVERIA; TORRA, 2014; HUANG *et al.*, 2004).

Neste capítulo são introduzidos os fundamentos da subárea de Aprendizado de Máquina, a qual abrange os estudos de técnicas computacionais para a automação da aquisição de conhecimento, assim como para a organização, estruturação e acesso do conhecimento já existente. Posteriormente, são descritos os algoritmos *k-Nearest Neighbors* (*kNN*) e sua respectiva adaptação, *k-Nearest Neighbors - Time Series Prediction* (*kNN-TSP*), para a predição de dados temporais. Ao final deste capítulo é proposta uma variação do algoritmo *kNN-TSP*, chamada de *kNN-TSP with Invariances* (*kNN-TSPI*), que tem como finalidade contornar os problemas encontrados no método original por meio do uso de três procedimentos para obtenção de invariância à amplitude e deslocamento, invariância à complexidade e tratamento de casamentos triviais.

5.2 Aprendizado de Máquina

Uma das formas mais simples e utilizadas para a representação de dados é realizada por meio de atributos e seus respectivos valores e é denominada tabela atributo-valor. Esse formato consiste na descrição de exemplos por atributos, de modo que cada exemplo equivale a uma

linha e cada atributo a uma coluna, como ilustrado no [Quadro 3](#).

Quadro 3 – Formato atributo-valor

Exemplos	Atributos				Classe (C)
	A_1	A_2	\dots	A_M	
E_1	a_{11}	a_{12}	\dots	a_{1M}	c_1
E_2	a_{21}	a_{22}	\dots	a_{2M}	c_2
E_3	a_{31}	a_{32}	\dots	a_{3M}	c_3
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
E_N	a_{N1}	a_{N2}	\dots	a_{NM}	c_N

Nessa linguagem de descrição de exemplos o tamanho de um conjunto de dados pode ser medido em relação a duas dimensões: número de atributos (M) e número de exemplos (N). Em determinados conjuntos de dados há ainda um atributo especial, denominado classe¹, que representa um conceito a ser aprendido e mapeado pelos modelos construídos por métodos de Aprendizado de Máquina (AM).

O AM é uma subárea da Inteligência Artificial na qual a pesquisa está focalizada no desenvolvimento de métodos computacionais que permitam adquirir, de maneira automática, novos conhecimentos, novas habilidades e novas técnicas de organizar o conhecimento existente ([MITCHELL, 1997](#)).

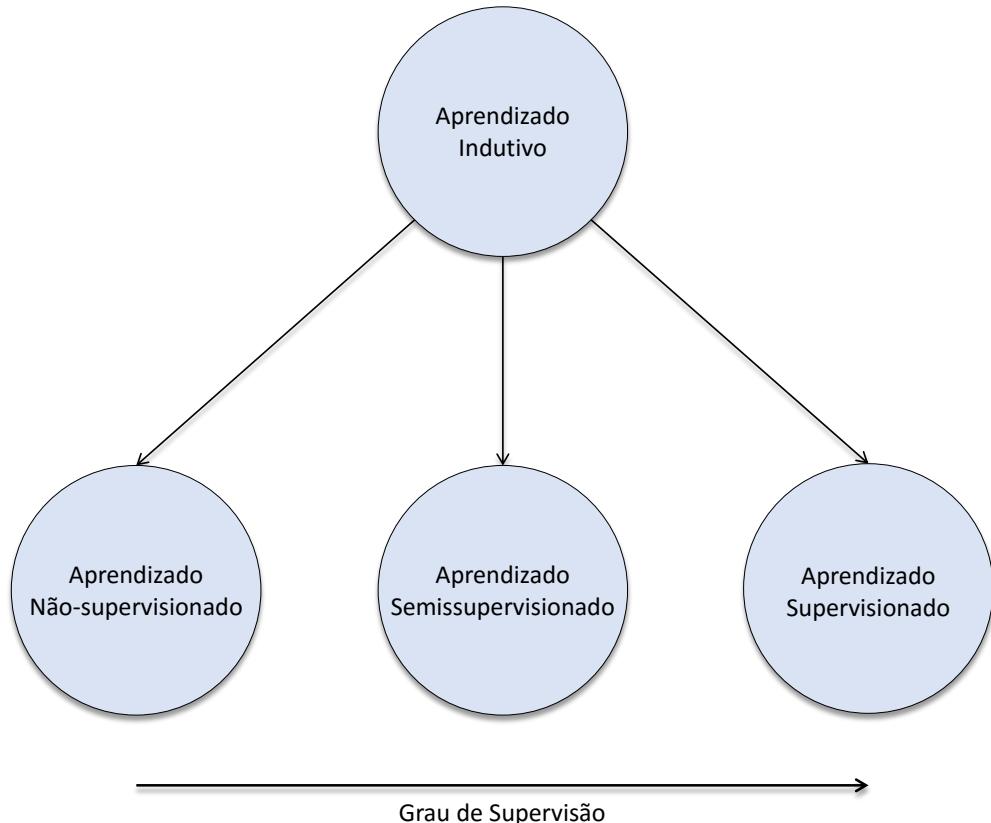
Um sistema de aprendizado consiste em um programa de computador que pode tomar decisões com base na experiência acumulada a partir da solução bem-sucedida de problemas anteriores. Sendo assim, existem diferentes estratégias de aprendizado com a finalidade de extrair conhecimento e predizer eventos futuros, tais como: hábito, instrução, dedução, analogia e indução ([ALPAYDIN, 2004](#)).

A estratégia denominada indução é a mais usada no contexto de AM, tendo em vista que ela permite obter conclusões realizando inferências a partir de um conjunto de exemplos conhecidos, isto é, fatos observados. A inferência indutiva é um dos recursos mais utilizados pelo cérebro humano para derivar conhecimento novo. Porém, deve ser empregada com cautela, pois se o número de exemplos for insuficiente ou não forem representativos, as hipóteses obtidas podem ser de pouco valor. A hierarquia do aprendizado indutivo, em conformidade com o critério do grau de supervisão presente nos dados, pode ser dividida em ([Figura 31](#)): aprendizado não-supervisionado, aprendizado semissupervisionado e aprendizado supervisionado ([MITCHELL, 1997](#)). Na [Figura 31](#), quanto maior a quantidade de exemplos rotulados, maior é o grau de supervisão.

Aprendizado Não-supervisionado: O objetivo reside em analisar os exemplos e determinar se subconjuntos desses exemplos podem ser agrupados de acordo com suas características (atributos), tendo em vista que a classe não está definida para nenhum exemplo;

¹ Os termos classe e rótulo são empregados indistintamente neste trabalho.

Figura 31 – Hierarquia do aprendizado indutivo considerando o grau de supervisão dos dados



Fonte: Adaptada de Braga (2010).

Aprendizado Semissupervisionado: Apenas uma pequena parte do conjunto de exemplos está rotulado. Logo, a ideia é usar essa pequena parcela para auxiliar na determinação das classes de uma quantidade maior de exemplos e permitir que um conjunto de dados rotulados de dimensão superior ao investigado possa ser construído;

Aprendizado Supervisionado: A entrada para um algoritmo de aprendizado supervisionado consiste usualmente de um conjunto de N exemplos (ou casos) de treinamento $\{(a_1, c_1), \dots, (a_N, c_N)\}$ rotulados com os valores c associados a uma função f desconhecida $c = f(a)$, onde os valores a_i são vetores da forma $\langle a_{i1}, a_{i2}, \dots, a_{iM} \rangle$ cujos componentes são valores discretos ou contínuos relacionados aos M atributos $A = \{A_1, A_2, \dots, A_M\}$. Ou seja, a_{ij} denota o valor do atributo A_j para o exemplo i (ALPAYDIN, 2004). Dado esse conjunto de exemplos de treinamento, o algoritmo induz uma hipótese h que deve aproximar a verdadeira função f , tal que dados os valores a de um novo exemplo, h determina o valor c correspondente. Desse modo, cada exemplo é associado a uma classe, que pode ser contínua, sendo nesse caso o processo denominado de regressão, ou discreta, designado de classificação.

O processo de classificação realizado por um algoritmo de AM pode ser dividido em duas fases. Na primeira fase, os exemplos rotulados (exemplos de treinamento) são fornecidos a um

indutor, o qual extrai o conhecimento desses exemplos e gera um classificador representado em uma estrutura interna (modelo). Na segunda fase, o classificador gerado é utilizado para rotular novos exemplos (exemplos de teste). Desse modo, o classificador tem habilidade de determinar tanto a classe dos exemplos usados em sua construção, quanto a classe de novos exemplos (PILA, 2001).

Existem diversos paradigmas de AM propostos na literatura. Dentre esses paradigmas podem ser citados (MITCHELL, 1997):

Paradigma Baseado em Exemplos: Um sistema de aprendizado baseado em exemplos classifica um novo exemplo por meio da análise comparativa com outros exemplos cuja classe é conhecida. Em termos práticos, os sistemas desse paradigma utilizam uma medida de distância ou de similaridade para identificar os exemplos mais similares ao exemplo a ser classificado;

Paradigma Conexionista: Sistemas de aprendizado conexionistas envolvem unidades altamente conectadas. O nome conexionismo é usado para descrever a área de estudo das Redes Neurais Artificiais, as quais são construções matemáticas inspiradas em conexões neurais do sistema nervoso. As Redes Neurais Artificiais constituem uma área de estudo ampla e utilizada em diversas aplicações, incluindo reconhecimento de voz e escrita;

Paradigma Estatístico: Os sistemas desse paradigma empregam modelos estatísticos para encontrar uma aproximação do conceito induzido. Comumente, esses métodos assumem que os dados seguem uma distribuição gaussiana e, a partir disso, usam procedimentos estatísticos para realizar inferências sobre os dados;

Paradigma Simbólico: Os sistemas de aprendizado simbólico visam construir representações simbólicas de conceitos por meio da análise de exemplos e contra-exemplos. Dentre os diversos modos de representação de conceitos, pode-se citar expressões lógicas, redes semânticas, regras de produção e árvores de decisão. Particularmente, a construção de árvores de decisão é uma das formas mais utilizadas para representações simbólicas devido à facilidade de compreensão da sua estrutura quando comparada com outras técnicas, como Redes Neurais Artificiais.

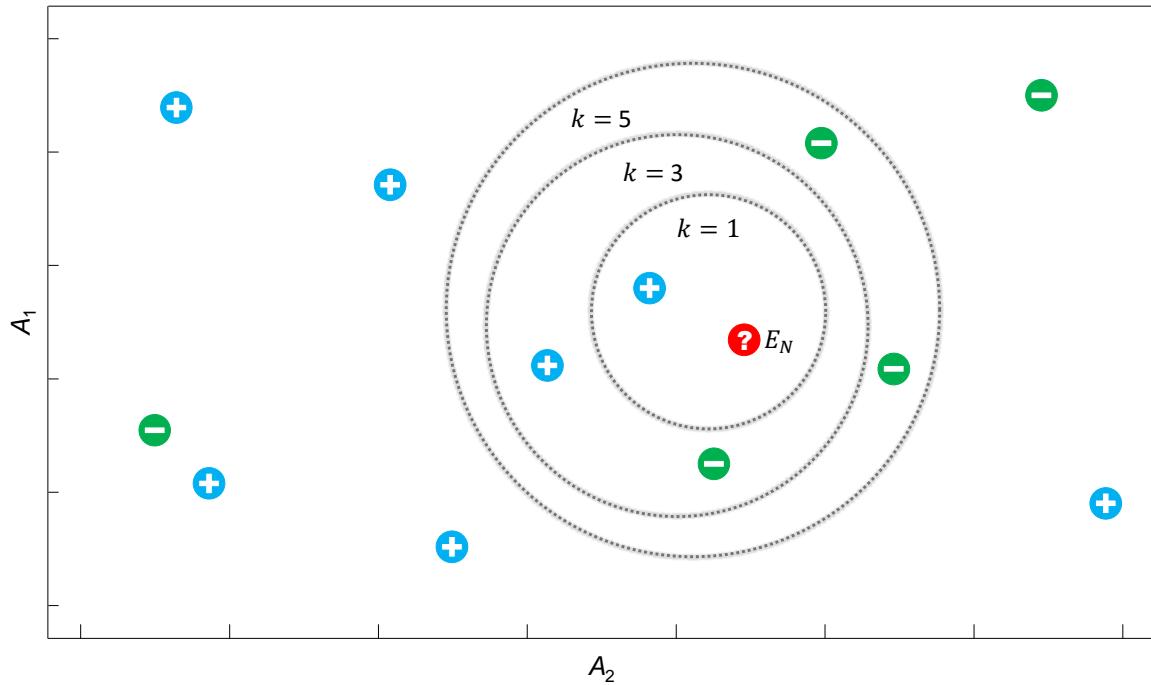
5.3 O Algoritmo kNN

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; HODGES, 1951). Assim, se $k = 1$ o novo exemplo será classificado como pertencente à mesma classe do único exemplo mais próximo de acordo com a medida

de similaridade. Diferentemente, se forem considerados $k > 1$ vizinhos, por exemplo, $k = 3$, a classe predominante dos três exemplos mais próximos será atribuída ao novo exemplo.

Na [Figura 32](#) é esquematizado o funcionamento do algoritmo *kNN* com $k = 1$, 3 e 5 para a classificação de um novo exemplo E_N a partir de um conjunto de treinamento composto por 12 exemplos, sete positivos (+) e cinco negativos (-), descritos pelos atributos A_1 e A_2 .

[Figura 32 – Exemplo da aplicação do algoritmo *kNN* com parâmetro \$k = 1\$, 3 e 5](#)



Fonte: Elaborada pelo autor.

Como pode ser observado na [Figura 32](#), a quantidade de vizinhos próximos considerada para a classificação pode influenciar diretamente no resultado da aplicação do algoritmo, pois para $k = 1$ e 3 o exemplo E_N seria classificado como positivo, enquanto que para $k = 5$ a classificação de E_N seria como negativo. Portanto, verifica-se que o valor do parâmetro k é particular a cada problema e estimá-lo incide em uma tarefa delicada que merece especial atenção.

Outro fator que deve ser levado em consideração quando usado esse algoritmo está associado à definição da medida de similaridade, a qual deve ser escolhida mediante às características dos exemplos ([LAROSE, 2004](#)). Adicionalmente, o *kNN* demanda baixo esforço computacional durante a fase de treinamento, porém o custo para classificar novos exemplos pode ser alto, uma vez que no pior caso o algoritmo utilizará para comparação todos os exemplos contidos no conjunto de treinamento.

O algoritmo *kNN* foi modificado em [Ferrero \(2009\)](#) para lidar com ST. Essa adaptação, disseminada como *kNN-TSP*, permite a predição de valores futuros com base na exploração das observações da série histórica.

5.4 O Algoritmo $kNN-TSP$

A predição de ST apoiada em similaridade visa estimar eventos futuros baseando-se em fatos similares que já aconteceram. Em outras palavras, a partir de 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, o $kNN-TSP$ será discutido com um horizonte unitário ($h = 1$), ou seja, supondo a predição do próximo valor da série.

Como mencionado, a ideia do algoritmo $kNN-TSP$ é bastante simples. Dada 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 predizer $\hat{z}(m, 1)$. 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 (MVA) (Equação 5.1), que tem como finalidade aproximar o valor de $\hat{z}(m, 1)$.

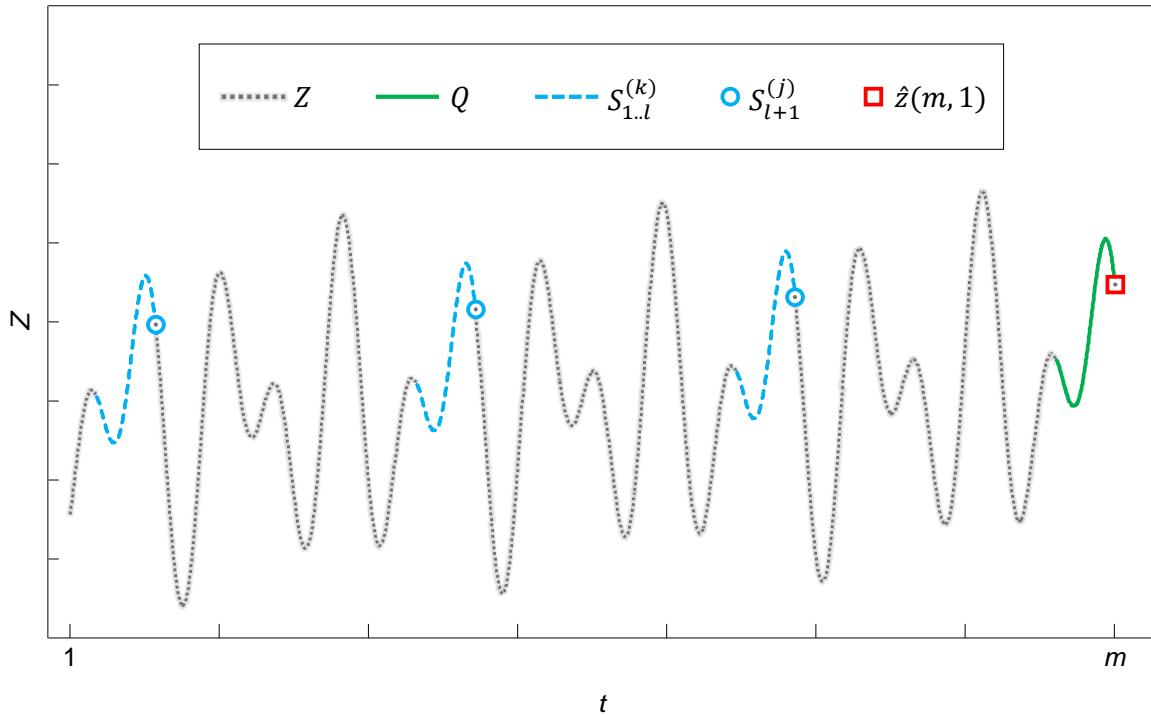
$$f_{\text{MVA}}(S) = \frac{1}{k} \sum_{j=1}^k S_{l+1}^{(j)} \quad (5.1)$$

Na Equação 5.1, 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 simples de combinar observações para se obter projeções futuras, uma vez que a MVA considera que as observações imediatamente seguintes às subsequências de dados mais similares são igualmente prováveis de ocorrer no futuro.

Na Figura 33 é exibido um exemplo de aplicação do algoritmo $kNN-TSP$ com $k = 3$ e $l = 25$. 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.

Cabe enfatizar que o $kNN-TSP$ define uma observação em função das l observações anteriores. Assim, a restrição da dependência a um número limitado de valores passados é válida, haja vista que geralmente o valor atual não é influenciado por observações que aconteceram há muito tempo atrás.

No Algoritmo 4 é apresentado o pseudocódigo do método $kNN-TSP$. Na segunda linha desse algoritmo 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 é armazenada a consulta

Figura 33 – Exemplo da aplicação do algoritmo *kNN-TSP* com parâmetros $k = 3$ e $l = 25$ 

Fonte: Elaborada pelo autor.

de referência Q , ou seja, a subsequência que será usada como parâmetro para a busca por subsequências similares no passado da série. Na quarta linha são computadas as distâncias entre a consulta Q e todas as subsequências de tamanho l geradas. A partir disso, na quinta linha, a busca pelas k subsequências mais parecidas com a consulta de referência Q é realizada. Posteriormente, na sexta linha, são obtidos os valores seguintes de cada uma das k subsequências mais semelhantes. Esses valores são utilizados, na sétima linha, pela função de predição f para o cálculo do valor futuro. Ao final, na oitava linha, o valor estimado é retornado pelo algoritmo.

Embora o [Algoritmo 4](#) tenha dois argumentos de entrada adicionais, a medida de distância d e a função de predição f , eles não são estritamente parâmetros do *kNN-TSP*. Isso porque tanto d quanto f são fixados estaticamente antes do cálculo dos parâmetros k e l . Ainda, pela própria definição dos parâmetros é possível perceber que k e l são muito intuitivos e fáceis de estimar. Por exemplo, o valor de l poderia ser proporcional ao ciclo sazonal da ST, visto que os vizinhos mais próximos seriam mais significativos na projeção dos dados. Além disso, para ajustar esses parâmetros a uma ST específica, poderia ser empregado um método de validação comum, como o de treino e teste ou validação *holdout* ([HAN; KAMBER; PEI, 2011](#)).

Diversas pesquisas foram conduzidas para analisar o desempenho do algoritmo *kNN-TSP* diante de distintas funções de predição ([FERRERO, 2009](#)) e diferentes medidas de distância ([JU-NIOR, 2012](#)). Nesse contexto, é relevante ressaltar que uma das escolhas mais frequentes de medida de similaridade é a Norma L_p ([KULESH; HOLSCHEIDER; KURENNAYA, 2008](#)), sendo usada majoritariamente a distância euclidiana.

Algoritmo 4: $kNN-TSP$

```

// Z representa uma série temporal com m observações
// l refere-se ao tamanho da janela de busca
// k indica a quantidade de vizinhos mais próximos
// d expressa a medida de distância
// f corresponde à função de predição
Input:  $Z, l, k, d, f$ 
Output:  $\hat{z}(m, 1)$ 
1 begin
    /*  $S_{1..l}^{(i)}$  contém a subsequência de comprimento l que começa na
       observação i da série temporal Z */ 
2      $S \leftarrow generate\_subsequences(Z_{1..(m-l)}, l);$ 
    // Obtenção da consulta Q
3      $Q \leftarrow S_{(m-l+1)..m};$ 
    //  $D^{(i)}$  armazena a distância entre Q e  $S^{(i)}$ 
4      $D \leftarrow d(Q, S);$ 
    // Escolha das k subsequências mais semelhantes
5      $P \leftarrow search\_nearest\_neighbors(S, D, k);$ 
    /* Obtenção do valor seguinte  $S_{l+1}^{(k)}$  de cada uma das k subsequências
       mais similares  $\in P$  */
6      $R \leftarrow \{S_{l+1}^{(1)}, \dots, S_{l+1}^{(k)}\};$ 
    // Cálculo da predição
7      $\hat{z}(m, 1) \leftarrow f(R);$ 
8     return  $\hat{z}(m, 1);$ 
9 end

```

5.5 O Algoritmo $kNN-TSPI$

Há três problemas substanciais com o procedimento de busca por similaridade adotado pelo $kNN-TSP$:

1. Falta de invariância à amplitude e deslocamento;
2. Necessidade de invariância à complexidade;
3. Carência da desconsideração de casamentos triviais.

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, neste trabalho foram formuladas técnicas para solucionar esses problemas. As técnicas desenvolvidas, discutidas a seguir, foram conciliadas para originar o algoritmo $kNN-TSPI$.

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

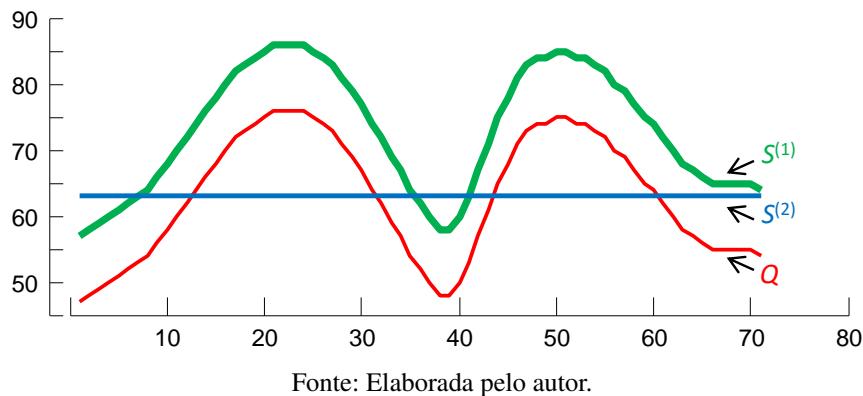
tipo de invariância pode ser obtido por vários métodos, porém uma maneira simples é utilizando a normalização em *z-scores* (ou *z-normalização*), a qual é definida pela [Equação 5.2 \(RAKTHANMANON *et al.*, 2012\)](#).

$$z'_t = \frac{z_t - \mu}{\sigma} \quad (5.2)$$

Na [Equação 5.2](#), 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 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 para aplicações de busca por subsequências em ST ([RAKTHANMANON *et al.*, 2012](#)).

Na [Figura 34](#) 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.

[Figura 34](#) – Exemplo de variância à amplitude e deslocamento: Pequeno deslocamento faz considerar Q mais parecida a $S^{(2)}$ do que a $S^{(1)}$



Fonte: Elaborada pelo autor.

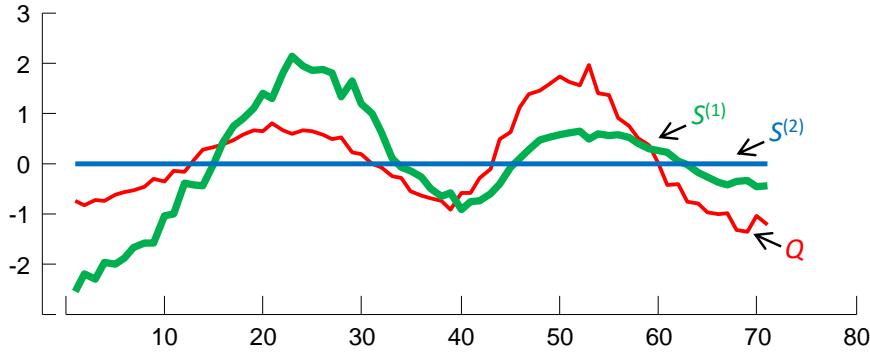
Embora para algumas aplicações a amplitude e o deslocamento sejam atributos importantes para caracterizar a subsequência, na maior parte dos domínios de aplicação, a busca sem invariância à amplitude e deslocamento não leva a resultados significativos. O motivo é que, mesmo em tais domínios, as subsequências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

Na [Figura 34](#) foi retratado um cenário ideal, pois $S^{(1)}$ é uma cópia perfeita de Q . No entanto, as pequenas diferenças de deslocamento já são suficientes 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; KETTERLIN; GANÇARSKI, 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 35.

Figura 35 – Exemplo de variância à complexidade: A subsequência simples $S^{(2)}$ é considerada o melhor casamento para Q , embora $S^{(1)}$ tenha um comportamento parecido



Fonte: Elaborada pelo autor.

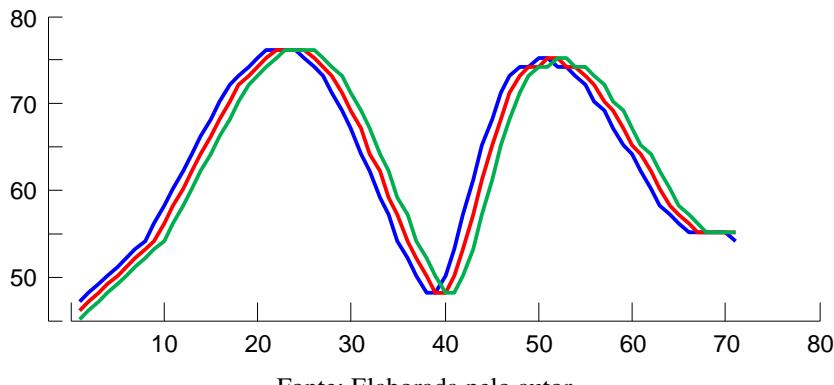
Na Figura 35 foi empregada z -normalização sobre versões das consultas da Figura 34 e introduzidas corrupções nos dados (ruído, diferenças de fase e distorção de amplitude), em pequenas quantidades, de modo que Q e $S^{(1)}$ ainda sejam parecidas. É importante ressaltar que essa configuração é mais realista do que a exibida na Figura 34 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)* (BATISTA *et al.*, 2014). A medida *CID* e sua estimativa de complexidade serão tratadas com maiores detalhes no Capítulo 6.

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 36, na qual a consulta Q é deslocada para a direita e após à esquerda por uma observação.

Figura 36 – 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



Fonte: Elaborada pelo autor.

O exemplo exibido na Figura 36 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 modificar o algoritmo *kNN-TSP* e criar o **Algoritmo 5**: *kNN-TSPI*.

Algoritmo 5: *kNN-TSPI*

```

// Z representa uma série temporal com m observações
// l refere-se ao tamanho da janela de busca
// k indica a quantidade de vizinhos mais próximos
Input: Z,l,k
Output:  $\hat{z}(m,1)$ 
1 begin
    /*  $S_{1..l}^{(i)}$  contém a subsequência de comprimento l que começa na
       observação i da série temporal Z */ 
2    $S \leftarrow \text{generate\_subsequences}(Z_{1..(m-l)}, l);$ 
    /*  $S^{(i)}$ ' é a z-normalização da subsequência  $S^{(i)}$ . A média e o desvio
       padrão usado na normalização são também alocados para reúso */
3    $[S', \sigma, \mu] \leftarrow z\_scores(S);$ 
    // Obtenção da consulta Q
4    $Q \leftarrow Z_{(m-l+1)..m};$ 
    /* D(i) armazena a distância com invariância à complexidade entre a
       consulta Q e S(i), ambas z-normalizadas */ 
5    $D \leftarrow CID(z\_scores(Q), S');$ 
    /* Escolha das k subsequências mais semelhantes (desconsideração
       de casamentos triviais) */ 
6    $P \leftarrow \text{search\_nearest\_neighbors}(S, D, k);$ 
    /* Obtenção do valor seguinte  $S_{l+1}^{(k)}$  de cada uma das k subsequências
       mais similares  $\in P$  */ 
7    $R \leftarrow \{S_{l+1}^{(1)}, \dots, S_{l+1}^{(k)}\};$ 
    /* Normalização em z-scores de cada valor  $S_{l+1}^{(k)} \in R$  utilizando a
       média e o desvio padrão da sua respectiva subsequência  $S_{1..l}^{(k)}$  */
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 à Equação 5.3, das z-normalizações para os
       valores da consulta Q */ 
9    $R \leftarrow \text{map\_query\_values}(Q, R');$ 
    // Cálculo da predição
10   $\hat{z}(m,1) \leftarrow f(R);$ 
11  return  $\hat{z}(m,1);$ 
12 end

```

Na segunda linha do **Algoritmo 5** 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 à Equação 5.3 para o espaço de valores da consulta Q e utilizados, na décima linha, pela função de predição f (Equação 5.1) para o cálculo do valor futuro. Ao final, na décima primeira linha, o valor estimado é retornado pelo algoritmo.

$$R^{(k)} = \sigma(Q) \times R^{(k)'} + \mu(Q) \quad (5.3)$$

Na Equação 5.3, σ e μ correspondem, nessa ordem, ao desvio padrão e a média da consulta de referência Q . Essa equação comprehende a função inversa da z -normalização e permite mapear, para o espaço 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 $S_{1..l}^{(k)}$.

Por fim, é importante frisar que a aproximação do desconhecido valor verdadeiro $\hat{z}(m, 1)$ é realizada no espaço real das observações usando a função de predição MVA (Equação 5.1), assim como no $kNN-TSP$. Todavia, outras funções de predição poderiam ser empregues, por exemplo a mediana, a *Distance Weighted* (*DW*) e a Média de Valores Relativos (*MVR*).

A função de predição *DW*, definida pela Equação 5.4 (HAN; KAMBER; PEI, 2011), trabalha com a ponderação das observações subsequentes $S_{l+1}^{(k)}$ pelas distâncias entre a consulta de referência Q e as subsequências de dados $S_{1..l}^{(k)}$.

$$f_{DW}(S) = \frac{\sum_{j=1}^k w_j S_{l+1}^{(j)}}{\sum_{j=1}^k w_j} \quad (5.4)$$

Os pesos w_j da Equação 5.4 podem ser determinados por meio distintas relações matemáticas, porém uma maneira usual é dada pela seguinte expressão:

$$w_j = \frac{1}{d(Q, S_{1..l}^{(j)})^2}$$

Uma atenção especial deve ser conferida ao valor de w_j , pois quando os dados da consulta Q forem iguais aos dados da subsequência corrente, o resultado da medida de distância será igual a zero. Logo, é preciso estabelecer uma regra que permita, em situações como a descrita, atribuir um valor unitário à variável w_j .

Em relação à MVR, expressa conforme a Equação 5.5 (FERRERO, 2009), o cálculo do valor futuro é firmado pela soma do último valor z_m amostrado da ST com a média das diferenças entre o último valor $S_l^{(k)}$ de cada subsequência mais parecida e o valor $S_{l+1}^{(k)}$ que a sucede.

$$f_{\text{MVR}}(S) = z_m + \frac{\sum_{j=1}^k (S_{l+1}^{(j)} - S_l^{(j)})}{k} \quad (5.5)$$

A utilização da MVR é mais pertinente na ausência de técnicas para obtenção de invariância à amplitude e deslocamento, pois ela projeta valores levando em consideração padrões em níveis distintos de tendência.

5.6 Considerações Finais

Métodos provenientes da área de MD têm sido estudados e modificados para dar suporte à tarefa de predição de valores em ST. Dentre os algoritmos frequentemente explorados encontra-se o kNN , cuja adaptação $kNN-TSP$ permite localizar padrões (subsequências similares) em dados temporais e usar os padrões encontrados para predizer eventos futuros.

Neste capítulo foram explanados os principais conceitos da subárea de AM, tal como o algoritmo de classificação, do paradigma baseado em exemplos, kNN . Em seguida, foi descrito o funcionamento do método baseado em similaridade para predição de ST $kNN-TSP$. Complementarmente, foram discutidos os três problemas encontrados no referido algoritmo, bem como as respectivas técnicas formuladas para solucionar esses problemas. Todas as técnicas desenvolvidas foram combinadas para criar o algoritmo $kNN-TSPI$, proposto neste trabalho.

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.

No próximo capítulo são resumidos alguns dos procedimentos que corroboram para a obtenção de invariâncias às distorções conhecidas em dados temporais. Adicionalmente, são abordadas as medidas de similaridade, incluindo as distâncias invariantes à complexidade, analisadas neste trabalho. O entendimento dessas propriedades é imprescindível para compreender o quanto elas influenciam no resultado da aplicação dos métodos de predição por similaridade.



SIMILARIDADE ENTRE SÉRIES TEMPORAIS

6.1 Considerações Iniciais

A similaridade pode ser entendida como uma estimativa do grau de semelhança entre dois objetos (XU; WUNSCH, 2009). Na predição de Séries Temporais (ST) via modelos de Aprendizado de Máquina (AM), quantificar quão parecidas são duas subsequências de dados para decidir se elas pertencem a um mesmo espaço de características é uma tarefa altamente subjetiva e que sofre influência de distintos fatores (PARMEZAN; BATISTA, 2015). De acordo com Hetland (2004), entre os possíveis fatores estão o domínio de aplicação e o método escolhido para o cálculo da similaridade. O estudo formalizado em Batista *et al.* (2014) vai além, isto é, demonstra que o desempenho de uma medida de semelhança está relacionado à capacidade dessa medida em capturar corretamente as invariâncias requeridas pelo domínio de aplicação.

Neste capítulo são especificadas as distorções conhecidas em ST, bem como as técnicas empregadas para se obter invariância a esses efeitos indesejáveis. As principais medidas de similaridade, incluindo distâncias invariantes à complexidade, são descritas em termos de alinhamento linear e não-linear. O modo como essas medidas estabelecem a semelhança entre sequências de dados, sua interpretação geométrica e o custo computacional também são discutidos.

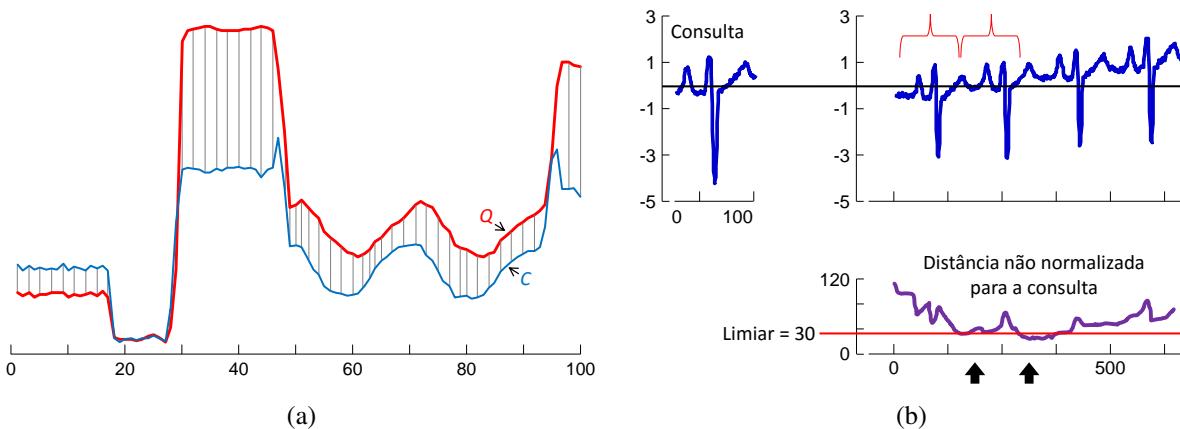
6.2 Invariâncias às Distorções Conhecidas em Dados Temporais

Comumente dados temporais possuem distorções inoportunas. Esses efeitos indesejáveis podem fazer com que medidas de distância não consigam capturar adequadamente a similaridade entre ST. Em outras palavras, quando os dados apresentam distorções no eixo temporal, as medidas existentes para mensurar semelhança entre pares de objetos acabam associando distâncias demasiadamente grandes à objetos similares (BATISTA *et al.*, 2014). A seguir são explicados os

efeitos indesejáveis em dados temporais e as técnicas que permitem desprezar as informações tendenciosas decorrentes dessas distorções.

Invariância à Amplitude e Deslocamento: Duas ST medidas em escalas distintas, como temperaturas em Celsius e Fahrenheit, não podem ser consideradas semelhantes mesmo possuindo uma forma parecida. Nesse contexto, para estimar a verdadeira similaridade entre essas sequências de dados é preciso fazer com que as amplitudes delas sejam as mesmas. Por exemplo, na Figura 37a as ST possuem um formato similar, mas apresentam uma grande distância euclidiana devido à diferença de amplitudes. Analogamente, mesmo que duas ST possuam amplitudes idênticas, elas podem ter deslocamentos diferentes. Como pode ser observado na Figura 37b, uma pequena mudança de deslocamento pode rapidamente dominar a medida de distância e acarretar em falsos negativos, por exemplo, batidas cardíacas não detectadas. As invariâncias à amplitude e deslocamento podem ser obtidas por meio da normalização em *z-scores* ou pela normalização linear dos dados (FALOUTSOS; RANGANATHAN; MANOLOPOULOS, 1994);

Figura 37 – Exemplificação da necessidade de invariância à amplitude e deslocamento: (a) Séries não similares devido à diferença de amplitude; (b) quando um objeto de consulta de batimento cardíaco é comparado com uma sequência de batimentos, os dois primeiros batimentos apresentam-se similares, porém a mudança de deslocamento faz com que os batimentos subsequentes não sejam detectados

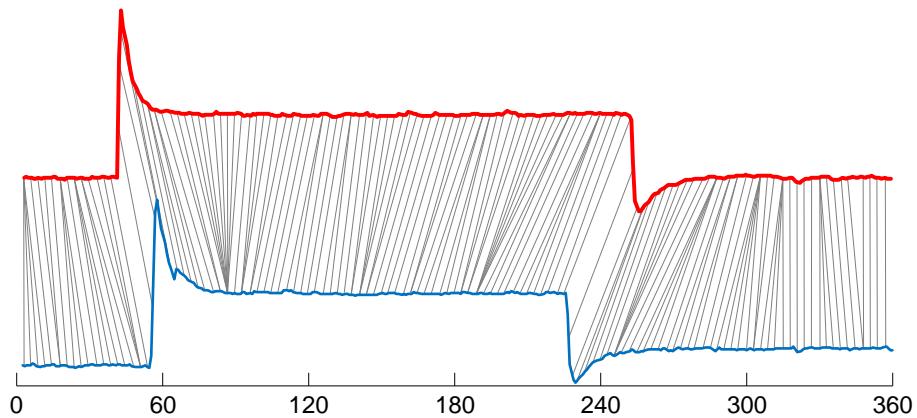


Fonte: Adaptada de Batista *et al.* (2014).

Invariância à Escala Local: Essa invariância é imprescindível para a maioria dos sinais de origem biológica, incluindo a captura de movimentos, reconhecimento de escrita e eletrocardiogramas. Na Figura 38 são mostradas duas ST, cada uma representando o comportamento de um inseto. Tais sequências de dados somente casam quando a invariância à escala local é considerada. À vista disso, trabalhos recentes sugerem que uma técnica proposta há mais de quatro décadas, *Dynamic Time Warping* (DTW), funciona apropriadamente (DING *et al.*, 2008);

Invariância à Escala Uniforme: A obtenção desse tipo de invariância é caracterizada pelo re-dimensionamento global de um conjunto de dados. Por exemplo, na Figura 39 são exibidas

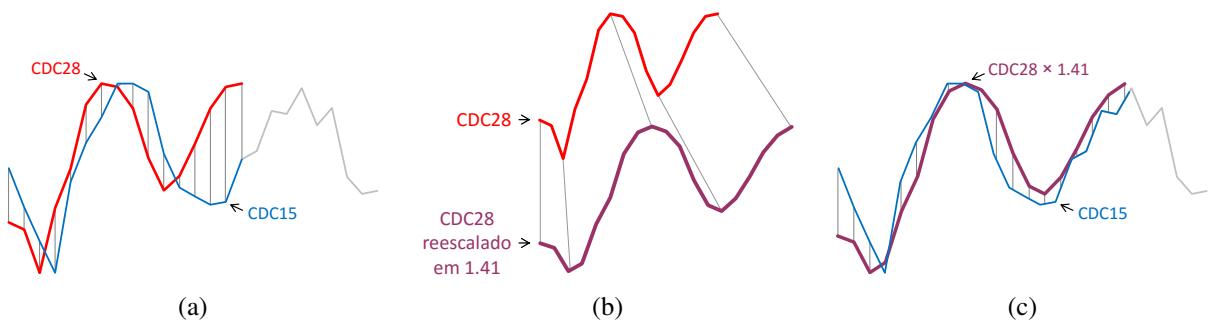
Figura 38 – Exemplificação de invariância à escala local: Duas ST representando comportamentos de insetos apresentam-se similares quando há invariância à escala local (alinhamento não-linear calculado pela medida DTW)



Fonte: Adaptada de [Batista et al. \(2014\)](#).

duas ST de expressão gênica de levedura de dois genes conhecidamente relacionados ([LI; YAN; YUAN, 2002](#)). O alinhamento da sequência mais curta com o prefixo da sequência mais longa resulta em um casamento ruim. Porém, um casamento consideravelmente melhor pode ser conseguido se a sequência mais curta for globalmente reescalada por um fator de 1,41. A principal dificuldade em se criar essa invariância está atrelada ao fator de escala, o qual normalmente não é conhecido *a priori*. Portanto, para determinar esse fator é preciso testar todas as possibilidades dentro de um determinado intervalo ([KEOGH, 2003](#));

Figura 39 – Exemplificação de invariância à escala uniforme: (a) A expressão completa de um gene CDC28 casa mal com o prefixo de um gene correlacionado CDC15; (b) se a expressão for reescalada por um fator 1,41, (c) ela se torna mais similar ao gene CDC15

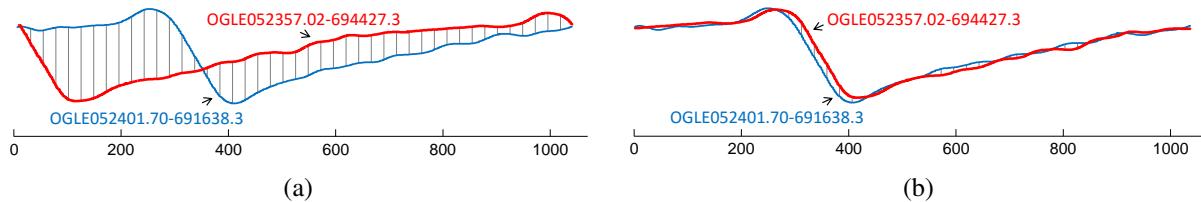


Fonte: Adaptada de [Batista et al. \(2014\)](#).

Invariância à Fase: A adoção de invariância à fase é fundamental para casar ST periódicas, como curvas de luz estrelar e batimentos cardíacos. Além disso, ela também pode ser usada para casar formas bidimensionais que foram convertidas para ST unidimensionais por meio de um truque de representação ([KEOGH et al., 2009](#)). Pesquisadores sugerem

que a invariância à fase pode ser obtida por meio da utilização de alinhamentos cardinais para os quais todas as séries são alinhadas. No entanto, trabalhos recentes expuseram que essa abordagem pode não fornecer bons resultados (**ZUNIC; ROSIN; KOPANJA, 2006**), de modo que a única maneira conhecida para garantir essa invariância é testando todos os alinhamentos possíveis, como ilustrado na **Figura 40**:

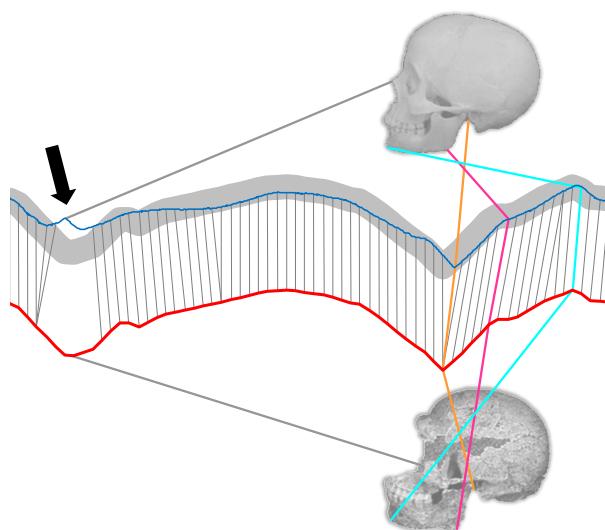
Figura 40 – Exemplificação de invariância à fase: (a) Duas curvas de luz estrelar fora de fase; (b) mantendo uma série em posição fixa e testando todos os deslocamentos circulares da outra, é possível obter invariância à fase



Fonte: Adaptada de [Batista et al. \(2014\)](#).

Invariância à Oclusão: A distorção amenizada por essa invariância ocorre em domínios nos quais as ST podem apresentar uma pequena subsequência faltante. Por exemplo, na **Figura 41**, a imagem de um crânio antigo casa quase perfeitamente com a imagem de um crânio moderno, apesar do crânio antigo ter a região do nariz faltante. Ainda nessa figura, é interessante observar que foi realizada uma transformação de representação entre uma imagem de crânio bidimensional e uma ST unidimensional;

Figura 41 – Exemplificação de invariância à oclusão: A invariância à oclusão pode ser obtida com a recusa do casamento entre subsequências de uma ST. Neste exemplo, a distância se torna robusta a falta da região do nariz no crânio antigo

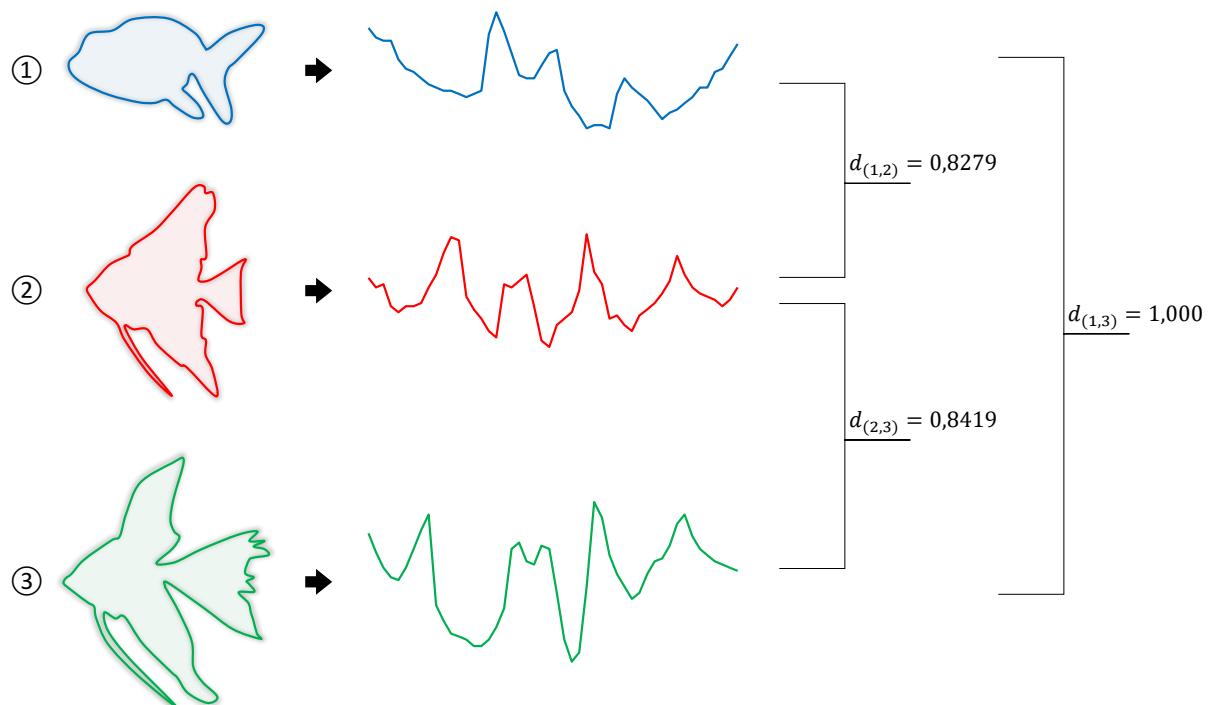


Fonte: Adaptada de [Batista et al. \(2014\)](#).

Invariância à Complexidade: O problema tratado por essa invariância reside no fato de que pares de objetos complexos, mesmo aqueles que subjetivamente são similares, tendem a

estar mais separados em termos de distância do que pares de objetos simples. Esse fato introduz erros na busca de similaridade, uma vez que sequências de dados complexas podem ser consideradas mais similares a outras mais simples. Na Figura 42 esse problema é retratado empregando três ST geradas computacionalmente, a partir do contorno de imagens de peixes, pela técnica de varredura radial (BATISTA; CAMPANA; KEOGH, 2010). Por meio dessa técnica são extraídas as distâncias entre o ponto central do objeto de interesse e os pontos consecutivos de seu contorno para compor uma sequência de distâncias. Na referida figura, o ponto de referência inicial para o cálculo das distâncias foi a boca dos peixes e adotou-se o sentido horário. A distância euclidiana entre as ST construídas demonstra que, embora as formas dos objetos 2 e 3 sejam visualmente semelhantes, o objeto 2 é mais parecido com o objeto 1 ($d_{(1,2)} = 0,8279$ e $d_{(2,3)} = 0,8419$). Uma solução para essa inconsistência está no uso de distâncias com invariância à complexidade.

Figura 42 – Exemplificação da necessidade de invariância à complexidade: Em termos de distância euclidiana, mesmo os objetos 2 e 3 sendo similares ao olho humano, o objeto 2 é mais parecido com o objeto 1



Fonte: Elaborada pelo autor.

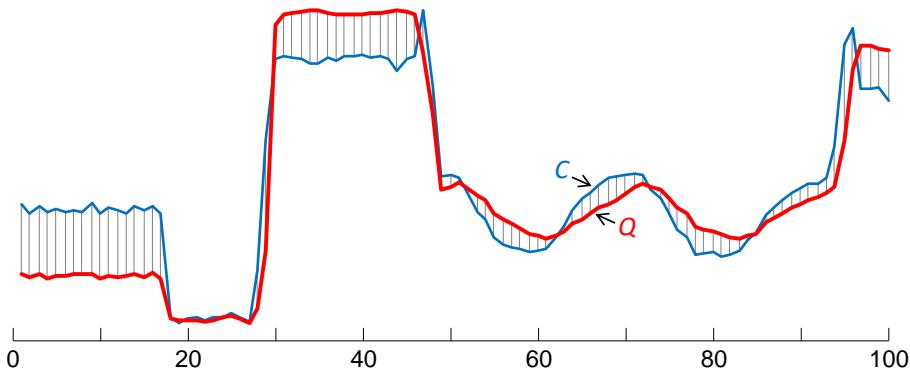
As medidas de distância invariantes à complexidade, inicialmente desenvolvidas para dar suporte à tarefa de classificação em Mineração de Dados Temporais (MDT), são relatadas na Seção 6.4.

6.3 Medidas de Distância

O empenho realizado por parte dos pesquisadores para a exploração do conceito de semelhança entre pares de objetos culminou na proposição de uma quantidade significativa de estimativas de distâncias (BATISTA *et al.*, 2014; DING *et al.*, 2008; CHA, 2007; DEZA; DEZA, 2006). Estudadas majoritariamente nas áreas da Estatística, Ciências de Computação, Física, Química e Biologia, essas medidas podem ser facilmente adaptadas, ou mesmo aplicadas diretamente, para auxiliar na comparação de ST (GIUSTI; BATISTA, 2013).

Como cada estimativa de distância contém particularidades específicas, a escolha de uma em detrimento de outras é frequentemente conduzida de maneira empírica sobre um problema particular. O alinhamento linear, esquematizado na Figura 43, é uma propriedade que destoa da assertiva exposta por ser comum à grande parte dessas medidas.

Figura 43 – Representação gráfica do alinhamento realizado, entre duas ST, pela distância euclidiana



Fonte: Adaptada de Batista *et al.* (2014).

Na Figura 43, a estimativa de semelhança entre as sequências $Q = (q_1, q_2, \dots, q_l)$ e $C = (c_1, c_2, \dots, c_l)$, ambas de comprimento l , é resultado das diferenças ao quadrado entre o pares de observações alinhadas no eixo do tempo. Esse alinhamento foi obtido utilizando a distância euclidiana, a qual é derivada da Equação 6.1 (AGGARWAL; HINNEBURG; KEIM, 2001).

$$L_p(Q, C) = \left(\sum_{i=1}^l |q_i - c_i|^p \right)^{\frac{1}{p}} \quad (6.1)$$

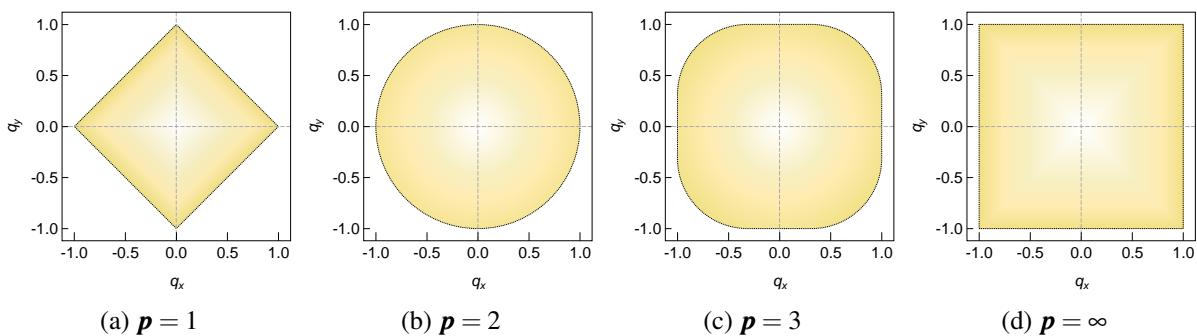
A Equação 6.1, tipicamente conhecida como Norma L_p ou métrica de Minkowski, trabalha com dois vetores l -dimensionais, que neste trabalho indicam ST, e uma constante $p \geq 0$. O custo computacional para calcular a distância de Minkowski está atrelado aos valores de l e p , de modo que quanto maior essas constantes, maior a quantidade de operações a serem computadas. O valor assumido por p dá origem à várias medidas de distância com comportamentos e nomes específicos. Por exemplo, $p = 1$ corresponde à distância de Manhattan ou *City Block*, enquanto $p = 2$ refere-se à distância euclidiana.

No [Quadro 4](#), algumas das mais relevantes estimativas de distância que obedecem ao alinhamento linear são elencadas. Nesse quadro, as medidas foram organizadas em seis categorias conforme suas semelhanças sintáticas e semânticas ([CHA, 2007](#)).

A categoria L_p contempla as métricas de distância que, devido à simplicidade de interpretação e codificação, são normalmente escolhidas para guiar a busca por similaridade em aplicações do mundo real. Uma desvantagem das medidas dessa categoria é o fato delas estarem suscetíveis à maldição da dimensionalidade, a qual significa uma diminuição do poder de discernimento das métricas de distância à medida que a quantidade de dimensões dos dados cresce. Adicionalmente, essas métricas são sensíveis à *outliers* e a pequenas distorções no eixo temporal da série ([VLACHOS; GUNOPULOS; DAS, 2004](#)).

A distância L_1 define um espaço geométrico no qual os pontos possuem o mesmo valor da soma das diferenças absolutas de cada ponto; a L_2 estabelece um espaço geométrico em forma de circunferência, no qual todos os pontos estão equidistantes em relação ao centro; a distância L_3 determina um espaço geométrico quadrado de cantos arredondados; e a L_∞ delimita um espaço quadricular, onde a distância entre dois pontos é a maior de suas diferenças. A variação dos espaços geométricos¹ dessas medidas pode ser visibilizada na [Figura 44](#).

Figura 44 – Efeito da variação de p na Norma L_p



Fonte: Elaborada pelo autor.

As métricas categorizadas como \hat{L}_1 são versões ponderadas da medida Manhattan. Por exemplo, na distância Canberra as diferenças absolutas entre os valores próximos à origem mostram-se maiores quando comparados aos valores mais distantes da origem. As métricas Kulczynski e Soergel incluem em suas estruturas a soma entre os valores máximos e mínimos, respectivamente. Já a Sørensen, usada em Ecologia, é similar à Canberra e seu resultado condiz com $L_1 \div 2$. A Lorentzian, além de trabalhar com a diferença absoluta, aplica o logaritmo natural. Nessa medida, o valor 1 é adicionado para garantir não-negatividade e evitar logaritmo de zero.

A categoria \hat{L}_2 abrange distâncias que fazem utilização das diferenças quadradas entre pares de observações, assim como realizado pela distância euclidiana. A estrutura da métrica Clark assemelha-se a da Canberra e retorna valores no intervalo de $[0, 1]$. Ainda, é conveniente

¹ Circunferências de raio unitário, isto é, conjunto de valores q para os quais $(|q_x|^p + |q_y|^p)^{\frac{1}{p}} = 1$.

Quadro 4 – Medidas de distância entre ST

Categoría	Medida de Distância	Equação
L_p	L_1 ou Manhattan	$d_1(Q, C) = \sum_{i=1}^l q_i - c_i $
	L_2 ou Euclidiana	$d_2(Q, C) = \sqrt{\sum_{i=1}^l (q_i - c_i)^2}$
	L_3 ou Métrica L_3	$d_3(Q, C) = \sqrt[3]{\sum_{i=1}^l q_i - c_i ^3}$
	L_∞ ou Chebychev	$d_4(Q, C) = \max_i q_i - c_i $
\hat{L}_1	Canberra	$d_5(Q, C) = \sum_{i=1}^l \frac{ q_i - c_i }{ q_i + c_i }$
	Kulczynski	$d_6(Q, C) = \frac{\sum_{i=1}^l q_i - c_i }{\sum_{i=1}^l \min(q_i, c_i)}$
	Lorentzian	$d_7(Q, C) = \sum_{i=1}^l \log(1 + q_i - c_i)$
	Sørensen	$d_8(Q, C) = \frac{\sum_{i=1}^l q_i - c_i }{\sum_{i=1}^l (q_i + c_i)}$
	Soergel	$d_9(Q, C) = \frac{\sum_{i=1}^l q_i - c_i }{\sum_{i=1}^l \max(q_i, c_i)}$
\hat{L}_2	Clark	$d_{10}(Q, C) = \sqrt{\sum_{i=1}^l \left(\frac{ q_i - c_i }{q_i + c_i} \right)^2}$
	Neyman	$d_{11}(Q, C) = \sum_{i=1}^l \frac{(q_i - c_i)^2}{q_i}$
	Pearson	$d_{12}(Q, C) = \sum_{i=1}^l \frac{(q_i - c_i)^2}{c_i}$
	Quadrática	$d_{13}(Q, C) = \sum_{i=1}^l \frac{(q_i - c_i)^2}{q_i + c_i}$
	Quadrática Aditiva Simétrica	$d_{14}(Q, C) = \sum_{i=1}^l \frac{(q_i - c_i)^2 (q_i + c_i)}{q_i c_i}$
Produto Interno	Correlação	$d_{15}(Q, C) = 1 - \frac{m \sum_{i=1}^l q_i c_i - \sum_{i=1}^l q_i \sum_{i=1}^l c_i}{\sqrt{m \sum_{i=1}^l q_i^2 - (\sum_{i=1}^l q_i)^2} \sqrt{m \sum_{i=1}^l c_i^2 - (\sum_{i=1}^l c_i)^2}}$
	Cosseno	$d_{16}(Q, C) = 1 - \frac{\sum_{i=1}^l q_i c_i}{\sqrt{\sum_{i=1}^l q_i^2} \sqrt{\sum_{i=1}^l c_i^2}}$
	Geodésica	$d_{17}(Q, C) = \arccos \left(\frac{\sum_{i=1}^l q_i c_i}{\sqrt{\sum_{i=1}^l q_i^2} \sqrt{\sum_{i=1}^l c_i^2}} \right)$
	Jaccard	$d_{18}(Q, C) = \frac{\sum_{i=1}^l (q_i - c_i)^2}{\sum_{i=1}^l q_i^2 \sum_{i=1}^l c_i^2 \sum_{i=1}^l q_i c_i}$
Informação	Jeffreys	$d_{19}(Q, C) = \sum_{i=1}^l (q_i - c_i) \log \left(\frac{q_i}{c_i} \right)$
	Topsøe	$d_{20}(Q, C) = \sum_{i=1}^l \left(q_i \log \left(\frac{2q_i}{q_i + c_i} \right) + c_i \log \left(\frac{2c_i}{q_i + c_i} \right) \right)$
Híbrida	Average Distance (L_1, L_∞)	$d_{21}(Q, C) = \frac{\sum_{i=1}^l q_i - c_i + \max_i q_i - c_i }{2}$

salientar que as medidas Neyman e Pearson, assim como a Quadrática e Quadrática Aditiva Simétrica, são bastante correlatas.

As distâncias pertencentes à categoria Produto Interno usam explicitamente em suas definições a multiplicação escalar entre dois vetores, isto é, entre duas sequências de dados. No campo da Estatística, a Correlação e o Cosseno são duas medidas de similaridade que podem ser transformadas em distâncias subtraindo seus resultados do valor 1. A Correlação quantifica o grau da relação linear entre duas sequências e retorna um valor entre 0 e 1. O valor 0 indica a ausência de relação linear e o valor 1 demonstra uma relação linear positiva perfeita. Quanto mais próximo o resultado estiver de 1, mais forte será a associação linear entre as duas sequências de dados. O cosseno é parecido com a Correlação, retornando valores no intervalo de [0, 1]. Essa medida de similaridade mensura o ângulo entre dois vetores num espaço vetorial. Portanto, quanto mais próximo de 1 for o resultado, mais semelhantes são os dois vetores. Outra métrica dessa categoria é a Geodésica, que consiste de uma generalização da noção de linha reta para espaços curvos. Essa distância é procedente da Ciência Geodésia, a qual se ocupa da determinação da forma, das dimensões e do campo de gravidade da Terra. Entre as demais variações da soma das diferenças quadradas divida pelo produto interno normalizado, encontra-se a distância Jaccard.

A categoria Informação é composta pelas métricas Jeffreys e Topsøe, as quais tem origem na área de Teoria da Comunicação. Em resumo, essas medidas são baseadas na aproximação da entropia da informação ([HAN; KAMBER; PEI, 2011](#)), isto é, na estimativa da incerteza de uma distribuição de probabilidade.

A *Average Distance*, por sua vez, é a única estimativa de distância categorizada como Híbrida. Essa medida é resultado da média entre as métricas Manhattan (L_1) e Chebychev (L_∞).

Segundo [Cha \(2007\)](#), problemas técnicos precisarão ser tratados durante a implementação de algumas das equações pontuadas no [Quadro 4](#). Essas distâncias, juntamente com suas advertências de codificação, são apresentadas no [Quadro 5](#). Nesse quadro, os problemas evidenciados podem ser solucionados da seguinte maneira: considera-se zero o resultado das operações $0 \div 0$ e $0\log(0)$. Nos casos de divisão por zero e aplicação do logaritmo de zero, o 0 pode ser substituído por um valor ϵ extremamente pequeno e positivo.

Quadro 5 – Problemas técnicos encontrados na implementação de algumas medidas de distância

Problema	Medidas de Distância
$0 \div 0$	Clark, Quadrática e Quadrática Aditiva Simétrica
Divisão por zero	Kulczynski, Pearson, Neyman e Jeffreys
$0\log(0)$	Topsøe
Logaritmo de zero	Jeffreys

Fonte: [Cha \(2007\)](#).

No que diz respeito ao custo computacional, todas as medidas exibidas no [Quadro 4](#) apresentam complexidade de tempo linear, isto é, $O(m)$, sendo m o tamanho da ST. Contudo, a

adição de termos no denominador dessas distâncias faz com que as operações a serem efetuadas aumente em quantidade constante.

6.4 Medidas de Distância Invariantes à Complexidade

Os fenômenos predominantemente investigados em AM exibem ST com diferentes morfologias. Nesse contexto, a aplicação de distâncias com invariância à complexidade, tais como *Complexity-Invariant Distance* (BATISTA *et al.*, 2014) e *Dynamic Time Warping - Delta* (DTW-D) (CHEN; KEOGH; BATISTA, 2013), permite que a informação morfológica das sequências de dados seja estimada e utilizada como um fator de correção para as medidas de distâncias existentes. Como resultado dessa ponderação, têm-se casamentos mais precisos e pertinentes.

6.4.1 Complexity-Invariant Distance

Definida a partir da distância euclidiana e em conformidade com a Equação 6.2, a medida *CID* considera a morfologia das sequências sendo comparadas e atribui distâncias maiores à sequências com complexidades diferentes (BATISTA *et al.*, 2014).

$$CID(Q, C) = ED(Q, C) \times CF(Q, C) \quad (6.2)$$

Na Equação 6.2, Q e C são duas ST, ED corresponde à distância euclidiana e CF é um fator de correção de complexidade calculado pela Equação 6.3, 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.

$$CF(Q, C) = \frac{\max(CE(Q), CE(C))}{\min(CE(Q), CE(C))} \quad (6.3)$$

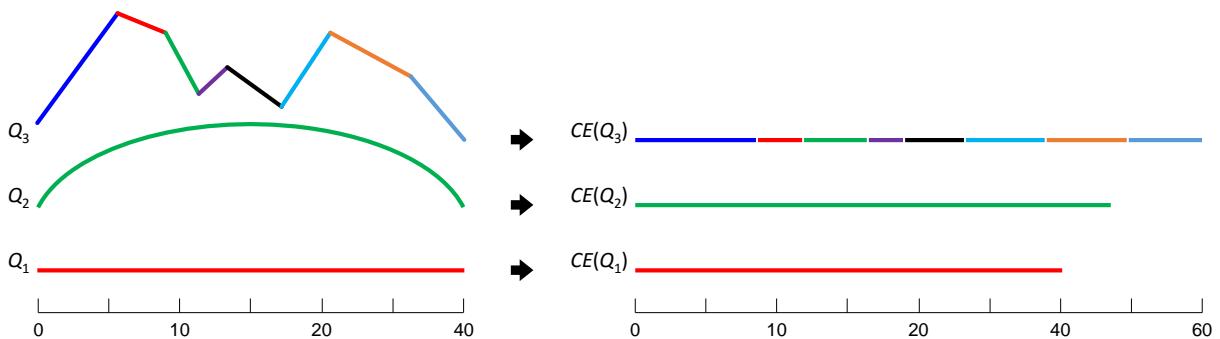
A estimativa de complexidade usada pela medida *CID* é bastante simples. Ela é apoiada na intuição de que uma ST pode ser “esticada” até que se torne um segmento de reta. Consequentemente, uma sequência complexa resultaria em um segmento de reta mais longo que uma sequência simples. Na Figura 45, essa ideia é sintetizada com um exemplo.

A estimativa de complexidade ilustrada na Figura 45 pode ser determinada por meio do emprego da Equação 6.4.

$$CE(S) = \sqrt{\sum_{i=1}^{l-1} (s_i - s_{i+1})^2} \quad (6.4)$$

Em uma análise da Equação 6.4 fica evidente a simplicidade da estimativa de complexidade, bem como das suas limitações. Essa medida deve ser aplicada para sequências com

Figura 45 – 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



Fonte: Adaptada de [Batista et al. \(2014\)](#).

o mesmo número de observações, uma vez que somente as diferenças entre os valores das observações são levadas em consideração, ignorando-se as diferenças de tempo em que as observações ocorreram. Tal limitação não restringe a utilização da complexidade quando aliada a uma métrica de distância como a euclidiana, pois essa medida também requer que as ST possuam o mesmo número de observações. Porém, ela restringe o uso de outras distâncias mais flexíveis, como a *Dynamic Time Warping (DTW)*. Ainda, a [Equação 6.4](#) requer que as sequências a serem comparadas estejam previamente normalizadas em amplitude e deslocamento. Embora essa normalização seja padrão para a maioria dos domínios de aplicação, uma estimativa de complexidade que não requeira invariância à amplitude e deslocamento pode ser bastante útil em outros campos de estudo ([BATISTA et al., 2014](#)).

No tema de processamento digital de sinais consta uma quantidade considerável de estimativas de complexidade que são aplicáveis à ST. A seguir são descritas cinco dessas medidas, as quais fundamentam-se em conceitos dos temas de teoria da informação, em especial entropia, teoria do caos e aproximação da complexidade de Kolmogorov.

Diferença Absoluta: Essa medida de complexidade é análoga à Diferença Quadrática ([Equação 6.4](#)) utilizada pela CID. No entanto, ao invés de serem computadas as diferenças quadradas, são calculadas as diferenças absolutas entre os pares de observações consecutivas. Formalmente, tem-se que $CE(S) = \sum_{i=1}^{l-1} |s_i - s_{i+1}|$;

Compressão: Essa medida baseia-se na complexidade de Kolmogorov que, apesar de compor uma teoria não-computável, pode ser aproximada por algoritmos de compressão, tal como o Lempel-Ziv-Welch. Inicialmente, cada ST é convertida para símbolos usando a representação *Symbolic Aggregate approXimation (SAX)* ([LIN et al., 2003](#)). Em seguida, as sequências discretizadas são comprimidas por meio de um utilitário de compactação de arquivos. A estimativa de complexidade de uma ST equivale ao tamanho em *bytes* do arquivo compactado;

Edges: Essa medida utiliza a quantidade de arestas, a qual também pode ser interpretada como o número de mudanças de tendência ou o número de vezes que a derivada primeira muda de sinal, como uma estimativa de complexidade para sequências de dados;

Zero-crossings: Essa medida expressa a complexidade de uma dada ST a partir da taxa de cruzamentos por zero, ou seja, do número de vezes que o sinal muda sua amplitude de um valor positivo para negativo ou vice-versa. Na área de análise de discurso, essa estimativa é amplamente empregada para detecção de presença e/ou ausência de voz, além de sons isolados (RABINER; SCHAFER, 1978);

Entropia de Permutação: Essa medida equivale à entropia da informação de um conjunto de padrões (BANDT; POMPE, 2002). Tais padrões são obtidos por meio da geração de todas as permutações de números naturais entre 0 e $n - 1$, sendo n um valor de parâmetro geralmente escolhido no intervalo discreto de [3, 7]. Por exemplo, para $n = 3$, são válidas as permutações $\{[0, 1, 2], [1, 0, 2], \dots, [2, 1, 0]\}$. O padrão [0, 1, 2] deve ser interpretado como uma subsequência, extraída de uma ST e composta por três observações, onde a segunda observação é maior do que a primeira e a terceira observação é maior do que a segunda; o padrão [1, 0, 2] deve ser visibilizado como uma subsequência de três observações cuja segunda observação é menor do que a primeira e a terceira observação é maior do que a primeira; e assim por diante. A probabilidade de cada padrão investigado é computada por intermédio de uma janela deslizante de tamanho n , a qual percorre toda a ST contando as ocorrências de cada padrão. A estimativa de complexidade é resultado da entropia do conjunto de padrões identificados na ST.

Mesmo havendo uma grande disponibilidade de estimativas de complexidade, apenas uma parcela pode ser usada para a comparação de pequenas sequências (ou subsequências). As cinco medidas descritas fazem parte desse grupo seletivo e, em virtude disso, podem ser aplicadas à tarefa de projeção de valores. O entendimento desse fato é importante porque as subsequências comparadas na predição baseada em similaridade são geralmente curtas, variando de acordo com o número de observações que constituem uma estação sazonal na ST (PARMEZAN; BATISTA, 2015).

6.4.2 Dynamic Time Warping - Delta

A DTW-D, que foi originalmente proposta para aprimorar o desempenho da classificação de exemplos em AM semissupervisionado, pode ser considerada uma medida de distância invariante à complexidade por aproximar ST globalmente similares por meio da utilização da Equação 6.5 (CHEN; KEOGH; BATISTA, 2013).

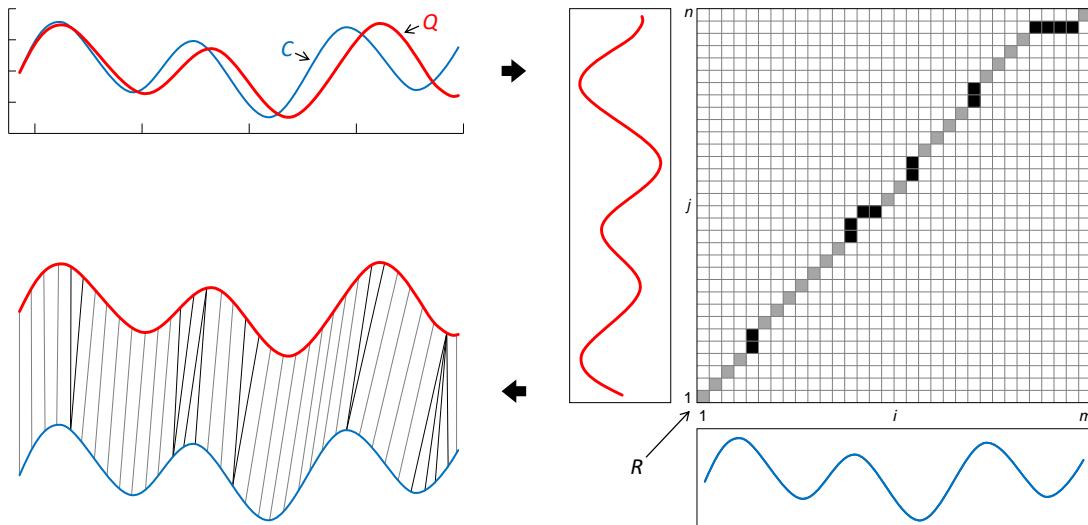
$$DTWD(Q, C) = \frac{DTW(Q, C)}{ED(Q, C) + \epsilon} \quad (6.5)$$

Na [Equação 6.5](#), o resultado da técnica *DTW* é dividido pela distância euclidiana acrescida de um valor ε (extremamente pequeno e positivo), o qual visa evitar a divisão por zero.

A *DTW* busca alinhar, da maneira mais adequada possível, os valores das sequências a serem comparadas. Isso permite que duas ST visualmente semelhantes, mas que estejam distorcidas no eixo temporal, possam ser alinhadas para posterior comparação ponto-a-ponto ([KEOGH; RATANAMAHATANA, 2005](#)).

Sejam duas sequências $Q = (q_1, q_2, \dots, q_n)$ e $C = (c_1, c_2, \dots, c_m)$, de comprimentos n e m respectivamente. Para compará-las usando a técnica *DTW*, constrói-se uma matriz de tamanho $n \times m$ na qual o elemento índice (i, j) contém a distância d , tipicamente euclidiana, entre as observações q_i e c_j . Em termos práticos, cada célula (i, j) da matriz corresponde ao alinhamento entre as observações nela representadas, assim como mostrado na [Figura 46](#).

Figura 46 – Esquematização da matriz de distâncias acumuladas, com rota de ajuste traçado, decorrente da aplicação da medida *DTW*



Fonte: Adaptada de [Keogh e Ratanamahatana \(2005\)](#).

Um percurso de ajuste $R = (r_1, \dots, r_t)$, onde $\max(n, m) \leq t < m + n - 1$, é um conjunto de elementos contíguos da matriz que define um mapeamento entre Q e C . Há muitos possíveis caminhos de ajuste, porém o percurso a ser escolhido é aquele que minimiza o custo da deformação, ou seja, cuja distância acumulada ao longo do caminho é mínima. A [Equação 6.6](#) estabelece o percurso ótimo, onde r_i indica o i -ésimo elemento do caminho de ajuste.

$$DTW(Q, C) = \min_R \left(\sum_{i=1}^t r_i \right) \quad (6.6)$$

Outras três restrições devem ser obedecidas durante a construção do referido percurso:

- 1. Restrição de Fronteira:** Iniciar e terminar em células diagonalmente opostas da matriz, ou seja, $r_1 = (1, 1)$ e $r_t = (n, m)$;

- 2. Restrição de Continuidade:** Os casamentos precisam ser realizados em etapas de uma unidade. Isso significa que um casamento nunca salta uma ou mais observações;
- 3. Restrição de Monotonicidade:** A ordem relativa das observações necessita ser preservada, de modo que a sequência de dados não poderá “voltar” no caminho dos elementos da matriz.

As distâncias acumuladas podem ser mensuradas aplicando um algoritmo de programação dinâmica, o qual implementa a seguinte relação de recorrência:

$$DTW(i, j) = d(q_i, c_j) + \min \begin{cases} DTW(i - 1, j) \\ DTW(i, j - 1) \\ DTW(i - 1, j - 1) \end{cases} \quad (6.7)$$

Na [Equação 6.7](#), soma-se ao resultado acumulado na célula atual a menor distância de suas três adjacentes, sendo essas a célula à esquerda, superior ou diagonal superior direita. A quantidade de iterações necessária para a realização desses cálculos concede à *DTW* uma complexidade de $O(n \cdot m)$, entretanto, é possível reduzir esse custo computacional de tempo utilizando uma janela de *warping*. Essa janela, que pode ser codificada de diferentes formas, limita o quanto a rota de ajuste pode se afastar da diagonal principal da matriz ([SAKOE; CHIBA, 1978](#); [ITAKURA, 1975](#)).

6.5 Considerações Finais

Medidas de similaridade são geralmente apoiadas em cálculos de distância. Do ponto de vista matemático, distância compreende um grau quantitativo de quão longínquo são dois objetos. Na área de MDT, as medidas de distância têm sido empregadas com sucesso na resolução de diversas tarefas de reconhecimento de padrões, tais como classificação, agrupamento, recuperação de informação e, mais recentemente, predição de ST.

Observa-se em algumas aplicações reais a predominância do uso de certas medidas de distância, por exemplo a Manhattan e a euclidiana. Isso acontece por que elas exprimem a noção de similaridade de maneira simples e, em geral, permitem alcançar resultados aceitáveis com os métodos propostos. Apesar das muitas evidências de que a análise de ST empregando as medidas supracitadas gera bons resultados ([JUNIOR, 2012](#); [DING *et al.*, 2008](#); [CHA, 2007](#); [DEZA; DEZA, 2006](#)), a constatação do seu desempenho em grandes quantidades de dados e na presença de distorções continua sendo um desafio.

No próximo capítulo é apresentada uma sucessão de experimentos que objetiva o entendimento dos distintos aspectos inerentes à predição de ST por similaridade, como invariâncias às distorções conhecidas em dados temporais, medidas de distância e funções de predição. Além

disso, o algoritmo proposto neste trabalho é comparado com outros dois métodos baseados em similaridade.



AVALIAÇÃO EXPERIMENTAL I: EXPLORANDO AS PROPRIEDADES DA PREDIÇÃO POR SIMILARIDADE

7.1 Considerações Iniciais

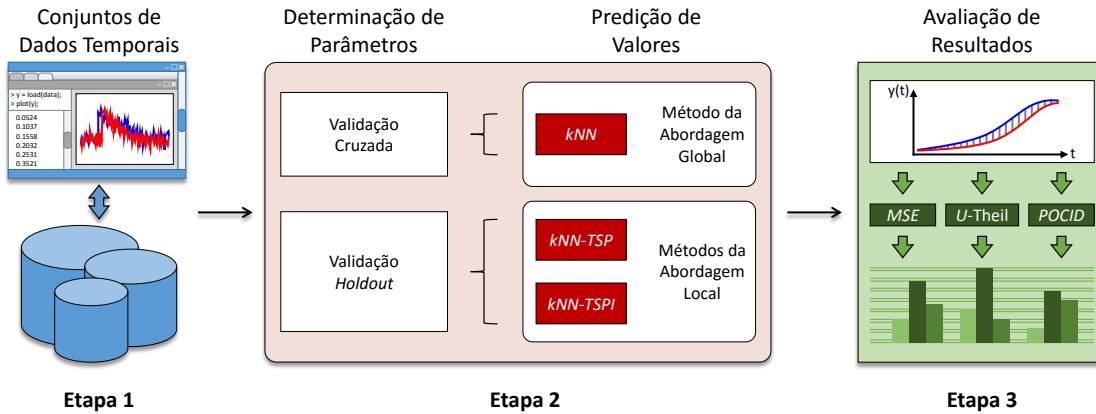
Pesquisas empíricas são necessárias para ancorar e comprovar, por meio de experimentos ou observação, aquilo formalizado conceitualmente. Nesse tipo de pesquisa, as hipóteses são construídas a partir das premissas e as variáveis que possivelmente exercem influência no comportamento do fenômeno de interesse 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 55 conjuntos de dados reais, neste capítulo são mostrados o protocolo, assim como os resultados e sua discussão, de uma sequência de testes computacionais. Esses experimentos foram realizados no intuito de explorar as propriedades intrínsecas à predição baseada em similaridade, como invariâncias às distorções em dados temporais, medidas de distância, medidas de complexidade aplicadas à *CID* e funções de predição. Adicionalmente, o *k-Nearest Neighbors - Time Series Prediction with Invariances (kNN-TSPI)* foi confrontado com outros dois métodos de predição de ST por similaridade: *kNN*, aplicado segundo a abordagem global; e *kNN-TSP*, executado de acordo com a abordagem local.

7.2 Configuração Experimental

O protocolo para avaliação do desempenho preditivo dos algoritmos baseados em similaridade foi organizado em três etapas, as quais são ilustradas na Figura 47.

Figura 47 – Configuração experimental I



Fonte: Elaborada pelo autor.

Na Etapa 1, a partir do resultado da revisão sistemática especificada no [Capítulo 2](#), foram selecionados 55 conjuntos de dados reais. Esses conjuntos são descritos no [Apêndice B](#), porém um resumo de suas particularidades e configurações é apresentado no [Quadro 6](#). Nesse quadro, para cada conjunto de dados real, são exibidos o tipo de aquisição dos dados, as datas de início e término do período de coleta, o tamanho da ST (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.

Na Etapa 2, o método kNN , que é aplicado conforme a abordagem global, teve seus parâmetros determinados por meio de validação cruzada em dez partições com minimização do Erro Quadrático Médio (MSE) ([Algoritmo 2](#) da [página 101](#)). Já os parâmetros dos algoritmos $kNN-TSP$ e $kNN-TSPI$, os quais são executados segundo a abordagem local, foram definidos usando validação *holdout* com minimização da medida MSE ([Algoritmo 1](#) da [página 99](#)).

Os métodos guiados por similaridade possuem dois parâmetros, a cardinalidade do conjunto de subsequências similares (k) e a quantidade de observações utilizadas como referência na busca por subsequências (l). A fim de realizar uma comparação justa, foram testados valores de k no intervalo de 1 a 9 em incrementos de 2, e l no intervalo de 3 até max_p também em incrementos de 2. Como max_p é limite superior para o número de observações que constituem um período sazonal, l será proporcional ao ciclo sazonal da série.

Após a identificação dos melhores parâmetros, os modelos preditivos foram construídos e ajustados aos dados de treinamento. Cada modelo produzido foi, em seguida, extrapolado h períodos à frente de acordo com as duas estratégias de projeção relatadas no [Capítulo 4](#) e intituladas de:

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

Quadro 6 – Sumário de características e de configurações dos conjuntos de dados reais

ID	Conjunto de Dados	Aquisição	Início	Término	<i>m</i>	<i>max_p</i>	<i>h</i>
01.A	Fortaleza	Anual	1849	1997	149	6	7
02.A	Manchas	Anual	1749	1942	176	11	12
03.D	Atmosfera: • Temperatura	Diária	01-jan-1997	31-dez-1997	365	7	31
04.D	• Umidade Relativa do Ar					7	31
05.D	Banespa	Diária	03-jan-1995	27-dez-2000	1499	7	88
06.D	CEMIG	Diária	03-jan-1995	27-dez-2000	1499	7	88
07.D	IBV	Diária	03-jan-1995	27-dez-2000	1499	7	88
08.D	<i>Patient Demand</i>	Diária	01-jan-2007	31-mar-2009	821	7	90
09.D	Petrobras	Diária	03-jan-1995	27-dez-2000	1499	7	88
10.D	Poluição: • PM10	Diária	01-jan-1997	31-dez-1997	365	7	31
11.D	• SO2					7	31
12.D	• CO					7	31
13.D	• O3					7	31
14.D	• NO2					7	31
15.D	<i>Star</i>	Diária	1922	1924	600	7	25
16.D	<i>Stock Market:</i> • Amsterdam	Diária	06-jan-1986	31-dez-1997	3128	7	92
17.D	• Frankfurt					7	92
18.D	• London					7	92
19.D	• Hong Kong					7	92
20.D	• Japan					7	92
21.D	• Singapore					7	92
22.M	• New York					7	92
23.M	Bebida	Mensal	jan-1985	jul-2000	187	12	7
24.M	<i>CBE:</i> • Chocolate	Mensal	jan-1958	dez-1990	396	12	24
25.M	• Beer					12	24
26.M	• Electricity Production					12	24
27.M	<i>Chicken</i>	Mensal	jan-1999	jul-2014	187	12	7
28.M	Consumo	Mensal	jan-1984	out-1996	154	12	10
29.M	Darwin	Mensal	1882	1998	1400	12	36
30.M	Dow Jones	Mensal	jan-1950	maio-2003	641	12	29
31.M	Energia	Mensal	jan-1968	set-1979	141	12	9
32.M	Global	Mensal	jan-1856	dez-2005	1800	12	36
33.M	ICV	Mensal	jan-1970	jun-1980	126	12	6
34.M	IPI	Mensal	jan-1985	jul-2000	187	12	7
35.M	<i>Latex</i>	Mensal	jan-1998	jul-2014	199	12	7
36.M	Lavras	Mensal	jan-1966	dez-1997	384	12	12
37.M	Maine	Mensal	jan-1996	ago-2006	128	12	8
38.M	MPrime	Mensal	jan-1949	nov-2007	707	12	23
39.M	OSVisit	Mensal	1977	1995	228	12	12
40.M	Ozônio	Mensal	jan-1956	dez-1970	180	12	12
41.M	PFI	Mensal	jan-1991	jul-2000	115	12	7
42.M	<i>Reservoir</i>	Mensal	jan-1909	dez-1980	864	12	24
43.M	<i>STemp</i>	Mensal	jan-1850	dez-2007	1896	12	36
44.M	Temperatura: • Cananéia	Mensal	jan-1976	dez-1985	120	12	12
45.M	• Ubatuba					12	12
46.M	<i>USA</i>	Mensal	jan-1996	out-2006	130	12	6
47.M	<i>Wine:</i> • <i>Fortified White</i>	Mensal	jan-1980	jul-1995	187	12	19
48.M	• <i>Dry White</i>					12	19
49.M	• <i>Sweet White</i>					12	19
50.M	• <i>Red</i>					12	19
51.M	• <i>Rose</i>					12	19
52.M	• <i>Sparkling</i>					12	19
53.S	<i>ECG:</i> • A	Intervalos de 0,5s	out-1996	out-1996	1800	60	120
54.S	• B					60	120
55.S	<i>Laser</i>	Intervalos de 1s	1991	1991	1000	8	100

Na Etapa 3, os dados projetados foram comparados com os dados de teste, em termos de erro preditivo, por meio do uso da medida *MSE* (Equação 4.31 da página 103) e do coeficiente *U* de Theil (*TU*) (Equação 4.32 da página 104). O índice de desempenho *Prediction Of Change In Direction (POCID)* (Equação 4.33 da página 104), que mensura a taxa de acerto quanto à tendência do horizonte de predição, também foi calculado. A partir dos valores dessas medidas foi possível confrontar o *kNN-TSPI* com os algoritmos *kNN* e *kNN-TSP*. 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 Nemenyi¹.

Todos os métodos utilizados na execução do protocolo experimental exposto, incluindo os algoritmos para estimação de parâmetros, foram implementados utilizando a linguagem de programação e o ambiente computacional MATLAB², o qual possui considerável compatibilidade com o, gratuito e de código aberto, ambiente e linguagem de programação GNU Octave³.

7.3 Resultados e Discussão

Os resultados empíricos obtidos, mostrados e discutidos a seguir, estão organizados do seguinte modo:

1. Impacto do uso de invariâncias às distorções em ST no desempenho do algoritmo *kNN-TSP*;
2. Estudo comparativo da relação custo-benefício dos métodos *kNN*, *kNN-TSP* e *kNN-TSPI*;
3. Análise do emprego de distintas medidas de similaridade na predição de ST utilizando o algoritmo *kNN-TSPI*;
4. Investigação do uso de medidas de complexidade aplicadas à medidas de distância invariáveis à complexidade no processo de busca adotado pelo método *kNN-TSPI*;
5. Avaliação da influência de diversas funções de predição na qualidade das projeções realizadas pelo algoritmo *kNN-TSPI*.

Os valores dos índices *MSE*, *TU* e *POCID*, tratados nesta seção, podem, juntamente com os parâmetros utilizados para alcançar tais resultados, ser encontrados no *ICMC-USP Time Series Prediction Repository* (PARMEZAN; BATISTA, 2014).

É importante frisar que, para cada grupo de experimentos, as configurações investigadas foram comparadas, a partir dos três índices de avaliação de desempenho supracitados, usando o teste estatístico de Friedman, com nível de significância de 5%, seguido do pós-teste de Nemenyi.

¹ Testes estatísticos realizados utilizando *KEEL Software Tool* para Windows, <<http://www.keel.es>>.

² <<http://www.mathworks.com>>.

³ <<http://www.gnu.org/software/octave>>.

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.

De maneira complementar e devido às particularidades dos índices *TU* e *POCID*, estes tiveram seus valores sumarizados em gráficos de barras totalmente empilhadas e gráficos de barras com desvios padrão, respectivamente.

7.3.1 Invariâncias às Distorções em Séries Temporais

Como mencionado, dados temporais podem apresentar distorções indesejáveis. Essas distorções fazem com que medidas de distância não consigam capturar adequadamente a similaridade entre as ST, associando distâncias demasiadamente grandes à objetos similares. No Quadro 7 são descritos alguns problemas conhecidos em dados temporais, bem como as técnicas pesquisadas neste trabalho para se obter invariância a esses efeitos indesejáveis.

Quadro 7 – Problemas em dados temporais e técnicas invariantes a essas adversidades

ID	Problema	Técnica Invariante
D	Deslocamento	Normalização de deslocamento por meio da subtração de cada observação pela média aritmética dos valores observados
AD	Amplitude e Deslocamento	Normalização de amplitude e deslocamento assemelha-se à obtenção de invariância à deslocamento. No entanto, após a subtração da média, é realizada a divisão pelo desvio padrão das observações. Esse procedimento é conhecido como normalização em <i>z-scores</i> ou <i>z-normalização</i>
EL	Escala Local	<i>Dynamic Time Warping (DTW)</i>
C	Complexidade	<i>Complexity-Invariant Distance (CID)</i>
CT	Casamentos Triviais	Exclusão de casamentos triviais por checagem iterativa

Com o objetivo de averiguar o uso de invariâncias na predição de ST por similaridade, as técnicas elencadas no Quadro 7 foram combinadas alternadamente e acopladas ao algoritmo *kNN-TSP*. Tais composições totalizaram 1870 configurações (1 modelo preditivo \times 17 tipos de invariância \times 2 estratégias de projeção \times 55 conjuntos de dados). As distintas combinações, além de comparadas entre si, foram confrontadas com os resultados computados pelo *kNN-TSP* sem invariâncias (Original).

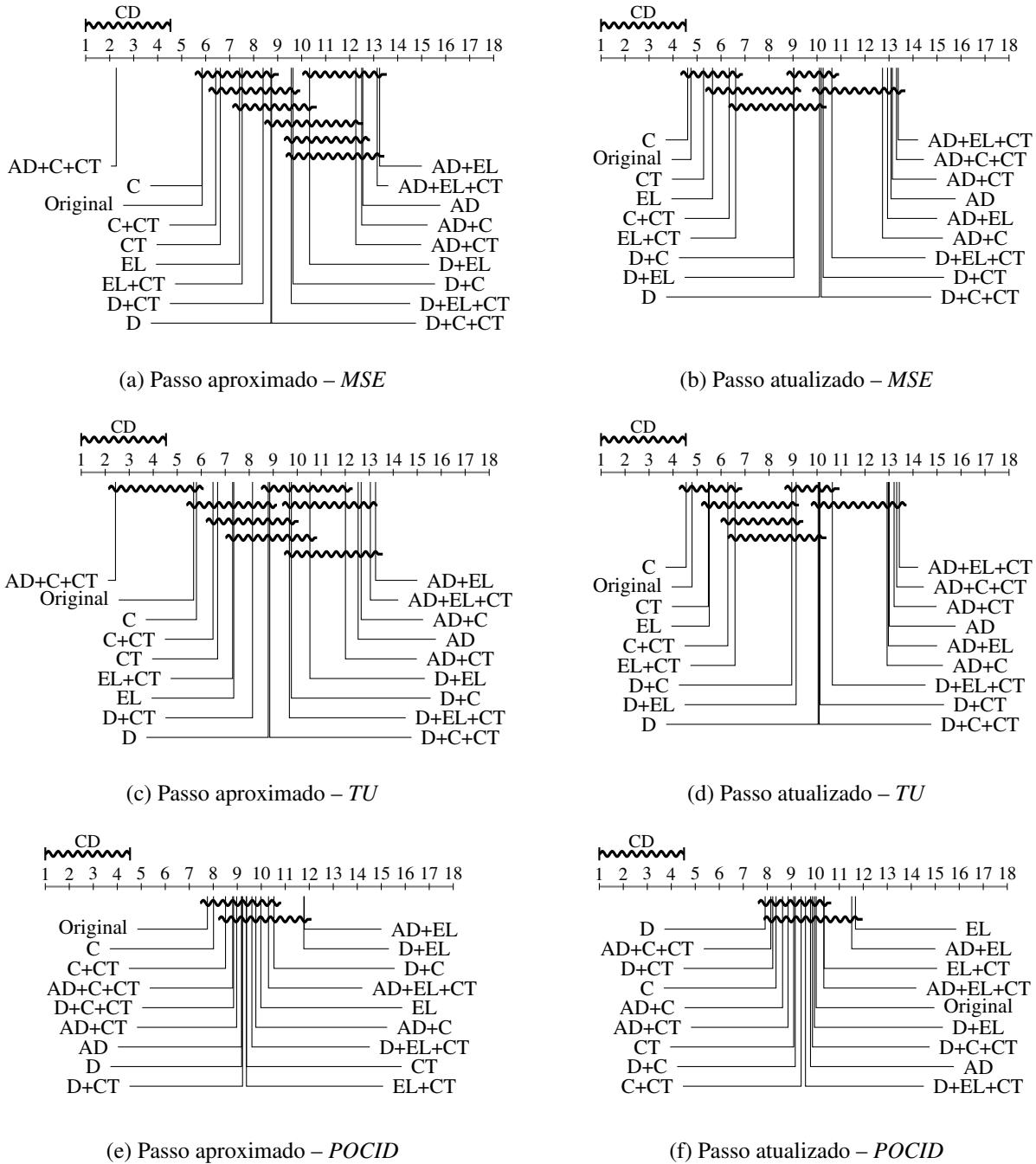
Na Figura 48 são mostrados os diagramas de distância crítica com relação aos valores das medidas *MSE*, *TU* e *POCID* decorrentes das 1870 configurações. Nessa figura, considerando a estratégia de predição multi-etapa à frente com passo aproximado, o *kNN-TSP* invariante à Am-

plitude e Deslocamento, Complexidade e com eliminação de Casamentos Triviais (AD+C+CT) exibiu o melhor resultado segundo os índices *MSE* (Figura 48a) e *TU* (Figura 48c). Especificamente para a medida *MSE* (Figura 48a), a configuração supracitada foi estatisticamente melhor que o *kNN-TSP* em sua versão Original e com as demais combinações de técnicas invariantes. É interessante observar que o emprego da *z*-normalização juntamente com a *DTW* no *kNN-TSP* (AD+EL) acarretou, em conformidade com os três índices de desempenho (Figuras 48a, 48c e 48e), nos piores resultados.

No que diz respeito à estratégia de predição multi-etapa à frente com passo atualizado, o *kNN-TSP* invariante à Complexidade (C) demonstrou o melhor resultado considerando as medidas *MSE* (Figura 48b) e *TU* (Figura 48d). Entretanto, ele não foi estatisticamente melhor que o *kNN-TSP* Original e com as seguintes técnicas invariantes: Casamentos Triviais (CT); Escala Local (EL); Complexidade e Casamentos Triviais (C+CT); e Escala Local e Casamentos Triviais (EL+CT). Os piores resultados de *MSE* (Figura 48b) e *TU* (Figura 48d) foram obtidos pelo *kNN-TSP* invariante à Amplitude e Deslocamento, Escala Local e com exclusão de Casamentos Triviais (AD+EL+CT). Nessa mesma linha de raciocínio, o *kNN-TSP* invariante à Amplitude e Deslocamento, Complexidade e com tratamento de Casamentos Triviais (AD+C+CT) apresentou o segundo pior resultado (Figuras 48b e 48d). Isso pode ter acontecido porque, ao contrário do funcionamento do algoritmo *kNN-TSPI*, a referida configuração não garante que os valores fornecidos para a função de predição tenham a mesma distribuição que as suas respectivas subsequências de origem. Ainda assim, ela assumiu a segunda posição no *ranking* gerado a partir das taxas de acerto *POCID* (Figura 48f), mas não expôs diferenças estatisticamente significativas em relação ao *kNN-TSP* invariante à Escala Local (EL).

Na Figura 49, as estatísticas procedentes da aplicação do coeficiente *TU* indicam que, para a predição multi-etapa à frente com passo aproximado (Figura 49a), o *kNN-TSP* sem técnicas invariantes (Original) promoveu, para 30 conjuntos de dados, os menores valores de *TU*. Em contraste, o *kNN-TSP* invariante à Complexidade (C) mostrou que para 26 conjuntos de dados (18 + 8) sua utilização é preferível em relação ao modelo trivial ou ingênuo (*TU* < 1), o qual pressupõe que o valor atual constitui a melhor predição para o período seguinte. Adicionalmente, essa mesma configuração revelou-se confiável para projetar os valores de 18 dos 26 conjuntos de dados (*TU* ≤ 0,55). As demais comparações entre valores de *TU* indicam que as composições D, D+C, D+EL+CT, D+C+CT, EL, EL+CT, CT+C e CT foram promissoras, em média, para 22 conjuntos de dados, dos quais 11 são tidos como confiáveis. Na Figura 49b, que trata da estratégia de predição multi-etapa à frente com passo atualizado, o *kNN-TSP* invariante à Complexidade (C) exibiu os menores valores de *TU*, seguido das versões sem técnicas invariantes (Original) e com eliminação de Casamentos Triviais (CT). Por outro lado, o *kNN-TSP* invariante à Amplitude, Deslocamento e Escala Local (AD+EL) não conseguiu superar o desempenho do modelo ingênuo em 41 do total de 55 conjuntos de dados (*TU* > 1).

Figura 48 – Diagramas de distância crítica para os valores dos índices *MSE*, *TU* e *POCID* provenientes do *kNN-TSP*, usando distintas técnicas invariantes, sobre ST reais

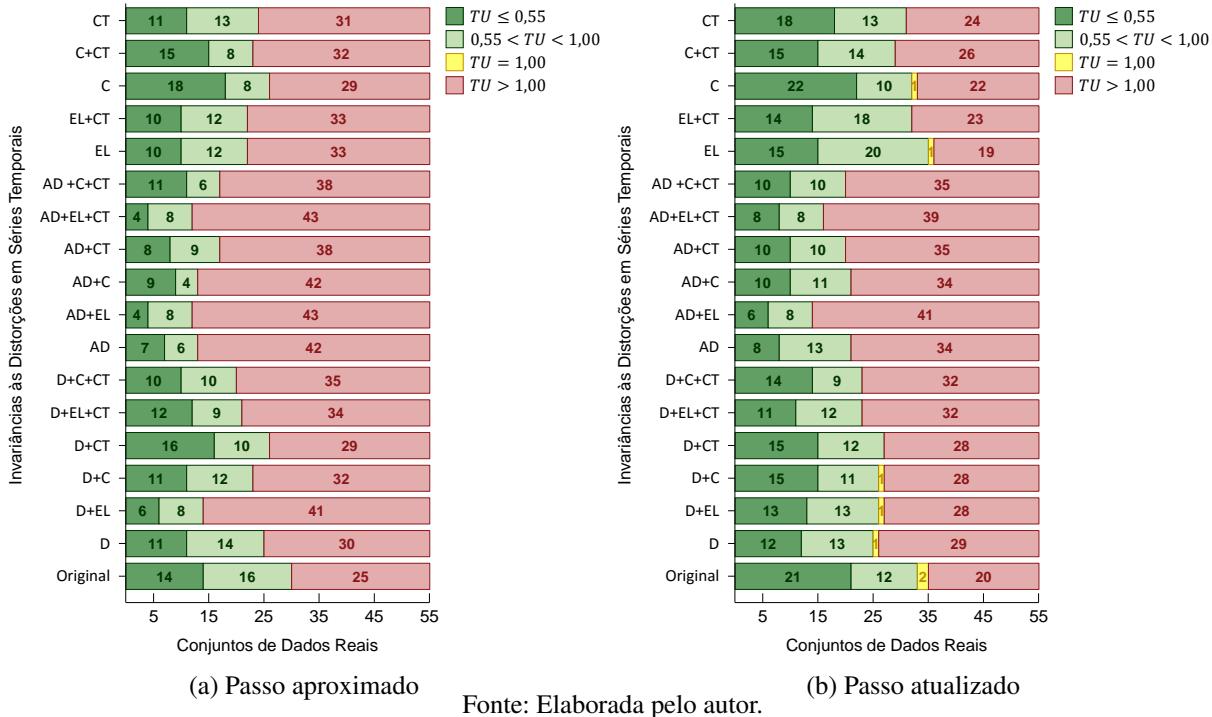


Fonte: Elaborada pelo autor.

Na Figura 50, para cada combinação de técnicas invariantes e estratégia de projeção, é apresentada a média dos valores do índice *POCID* com a indicação dos desvios padrão.

Observa-se na Figura 50 que, a partir da estratégia de predição com passo aproximado, o *kNN-TSP* Original implicou na maior taxa média de acerto, isto é, 59,31% com desvio padrão de 20,96%. Já o *kNN-TSP* invariante à Complexidade (C) emplacou um acerto médio de 58,42%,

Figura 49 – Desempenho em ST reais do *kNN-TSP*, empregando diferentes técnicas invariantes, para quatro faixas de valores do coeficiente *TU*



Fonte: Elaborada pelo autor.

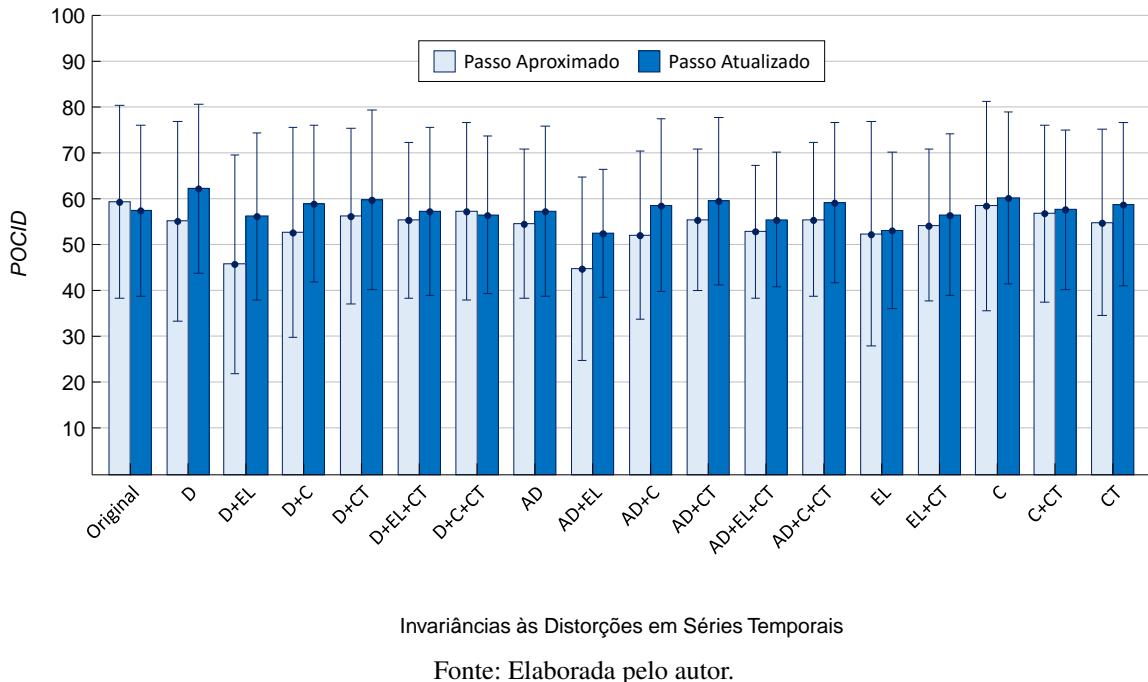
com desvio padrão de 22,78%, sobre as tendências dos horizontes projetados. A menor taxa média de acerto, ou seja, 52,82% com desvio padrão de 14,44%, foi procedente do *kNN-TSP* invariante à Amplitude e Deslocamento, Escala Local e com exclusão de Casamentos Triviais (AD+EL+CT). No que se refere à estratégia de predição com passo atualizado, as maiores taxas de acerto foram obtidas pelo *kNN-TSP* invariante à Deslocamento (D) (62,19% com desvio padrão de 18,42%) e à Complexidade (C) (60,19% com desvio padrão de 18,69%). Diferentemente, o *kNN-TSP* invariante à Amplitude, Deslocamento e Escala Local (AD+EL) proporcionou os menores valores de *POCID* (acerto médio de 52,48% com desvio padrão de 13,98%).

7.3.2 Métodos Baseados em Similaridade

O algoritmo *kNN* é aplicado conforme a abordagem global, na qual assume-se que os dados de entrada são independentes e identicamente distribuídos. Já os métodos *kNN-TSP* e *kNN-TSPI* foram formulados para considerar em suas modelagens as características temporais associadas aos dados. No intuito de evidenciar a relação custo-benefício desses algoritmos, eles foram confrontados entre si em termos de desempenho preditivo. Tais comparações totalizaram 330 configurações (3 modelos preditivos × 2 estratégias de projeção × 55 conjuntos de dados).

Na Figura 51 são exibidos os diagramas de distância crítica com relação aos valores das medidas *MSE*, *TU* e *POCID* decorrentes das 330 configurações. Nessa figura, avaliando a estratégia de predição multi-etapa à frente com passo aproximado, o algoritmo *kNN-TSPI* exibiu,

Figura 50 – Médias e desvios padrão das taxas de acerto *POCID* obtidas pelo *kNN-TSP*, utilizando distintas técnicas invariantes, em ST reais



Invariâncias às Distorções em Séries Temporais

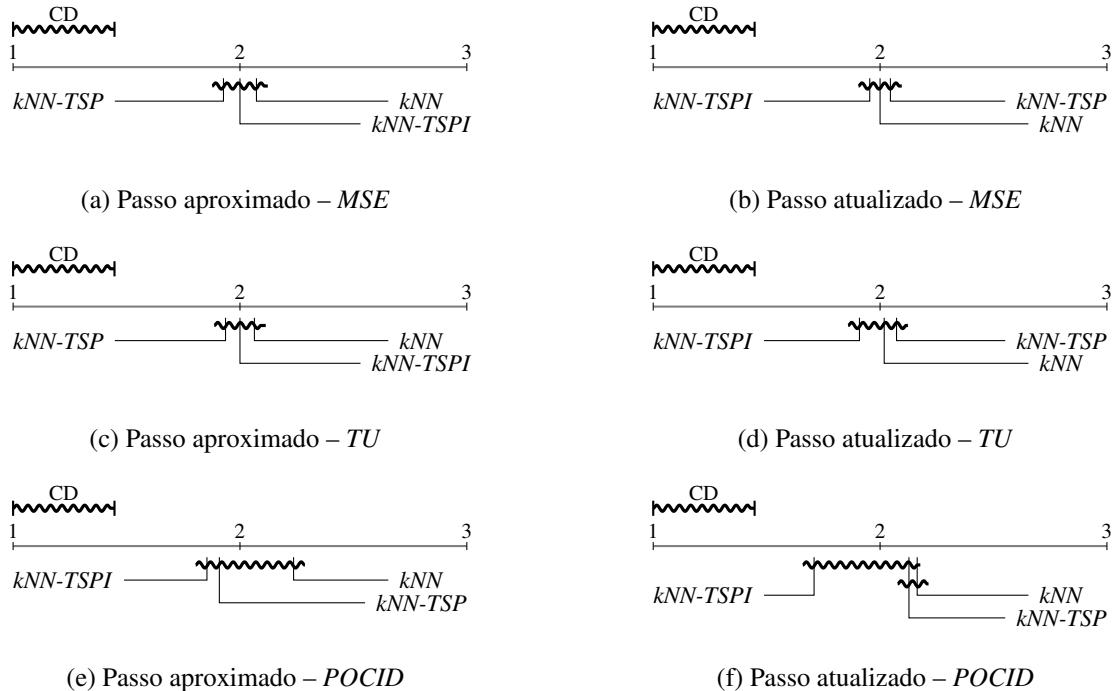
Fonte: Elaborada pelo autor.

sem diferenças estatisticamente significativas, o segundo melhor resultado de acordo com os índices *MSE* (Figura 51a) e *TU* (Figura 51c). Todavia, essa mesma configuração apresentou, em termos de *POCID* (Figura 51e), as melhores taxas de acerto sobre as tendências dos horizontes projetados. Analisando a estratégia de predição multi-etapa à frente com passo atualizado, o *kNN-TSPI* acarretou, sem diferenças estatisticamente significativas e segundo as três medidas de desempenho (Figuras 51b, 51d, 51f), nos melhores resultados.

As informações sintetizadas nos diagramas das Figuras 51c e 51d são semelhantes às retratadas nos gráficos da Figura 52. Nesses gráficos, para cada método baseado em similaridade, as quatro faixas de valores do coeficiente *TU* mostraram-se coerentes para com os resultados esperados. Notoriamente, o *kNN-TSPI* com passo aproximado deve, em geral, prover um desempenho inferior a sua versão com passo atualizado. Além disso, os algoritmos *kNN* e *kNN-TSP* podem, na maioria dos casos, manter-se próximos ou até mesmo superar o erro preditivo do *kNN-TSPI*. Esse fato acontece porque, no algoritmo *kNN-TSPI*, a busca por similaridade é conduzida de maneira a evitar escolhas errôneas das subsequências mais similares. Nesse sentido, como pode ser visibilizado na Figura 53, o *kNN-TSPI* frequentemente apresentará uma taxa de acerto maior sobre a tendência dos horizontes de predição em relação aos métodos *kNN* e *kNN-TSP*.

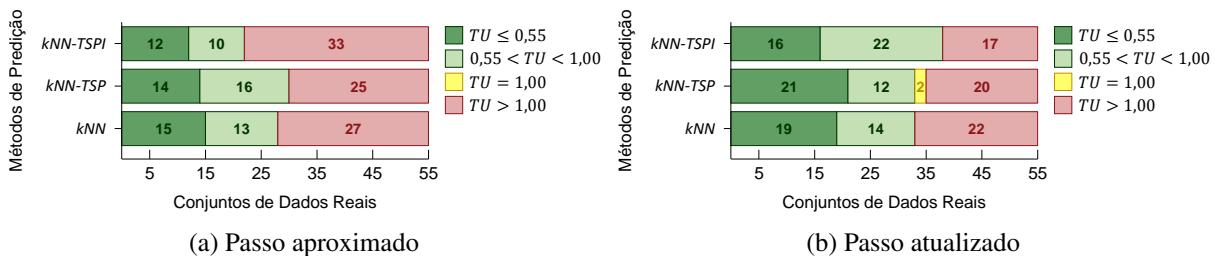
Observa-se na Figura 53 que, para ambas as estratégias de predição, os algoritmos *kNN-TSPI* e *kNN* expuseram, nessa ordem, as melhores e piores taxas médias de acerto quanto às tendências dos horizontes projetados.

Figura 51 – Diagramas de distância crítica para os valores dos índices MSE , TU e $POCID$ provenientes dos métodos baseados em similaridade sobre ST reais



Fonte: Elaborada pelo autor.

Figura 52 – Desempenho em ST reais dos métodos baseados em similaridade para quatro faixas de valores do coeficiente TU

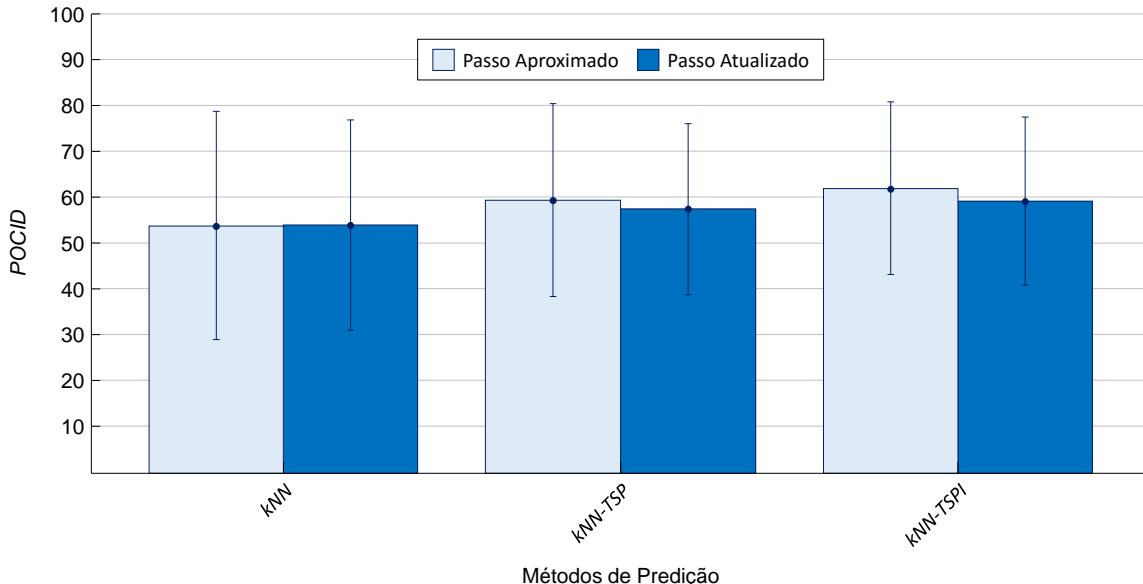


Fonte: Elaborada pelo autor.

7.3.3 Medidas de Distância

O algoritmo $kNN-TSPI$, proposto neste trabalho, difere do método original por envolver três técnicas para obtenção de invariância à amplitude e deslocamento, invariância à complexidade e tratamento de casamentos triviais. A fim de verificar a influência que a medida de similaridade exerce sobre o desempenho do algoritmo $kNN-TSPI$, foram utilizadas e confrontadas 25 medidas de similaridade. Essas combinações totalizaram 2750 configurações (1 modelo preditivo \times 25 medidas de distância \times 2 estratégias de projeção \times 55 conjuntos de dados).

Figura 53 – Médias e desvios padrão das taxas de acerto *POCID* obtidas pelos métodos baseados em similaridade em ST reais



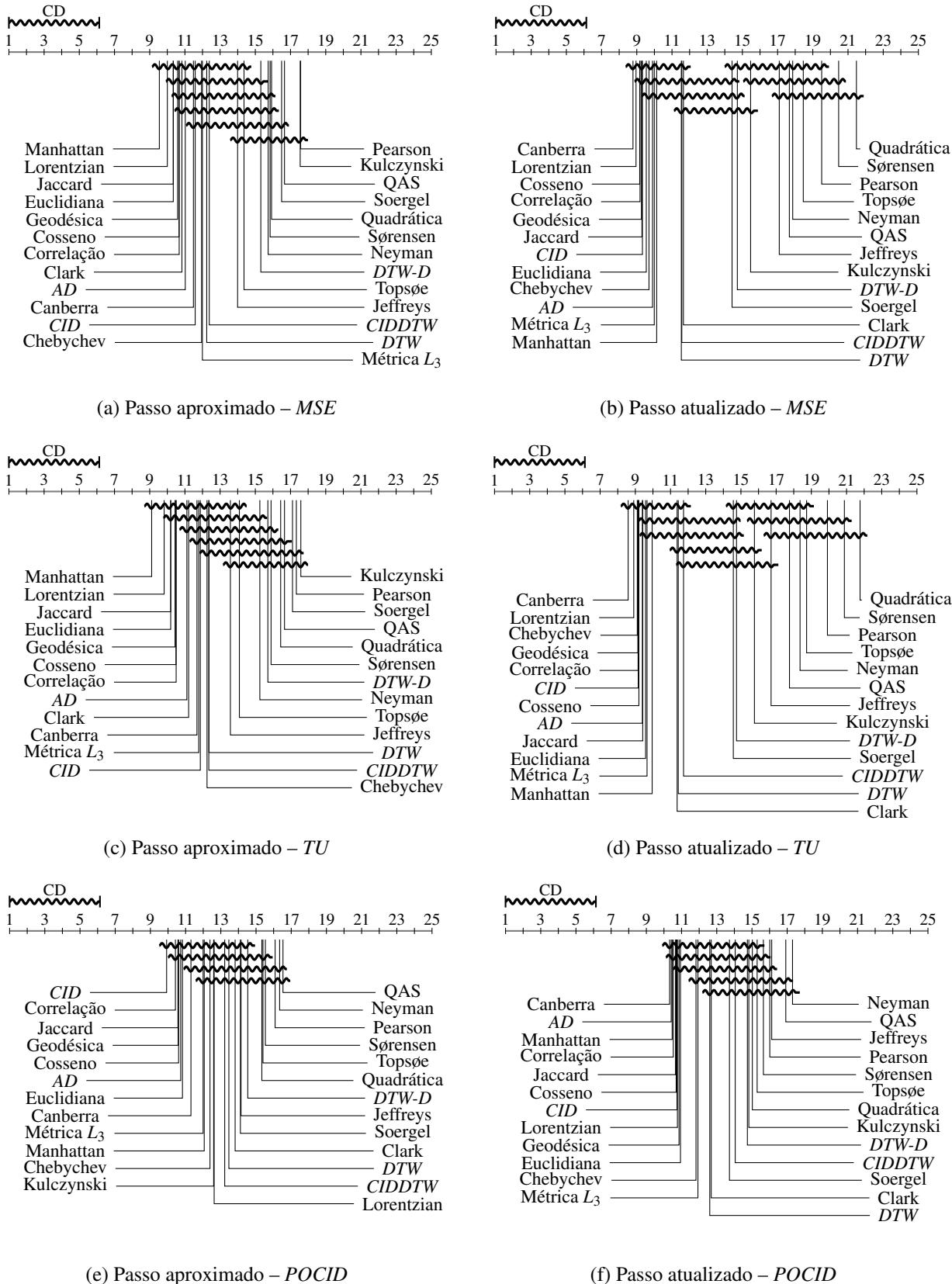
Fonte: Elaborada pelo autor.

Na Figura 54 são mostrados os diagramas de distância crítica com relação aos valores das medidas *MSE*, *TU* e *POCID* provenientes das 2750 configurações. Nessa figura, analisando a estratégia de predição multi-etapa à frente com passo aproximado, a distância Manhattan foi a que forneceu, em média, os melhores resultados de *MSE* (Figura 54a) e *TU* (Figura 54c). Contudo, ela não apresentou diferenças estatisticamente significativas quando comparada com as medidas usualmente difundidas na literatura, tais como Euclidiana, Cosseno, Geodésica, *Average Distance* (*AD*), *CID*, *DTW*, *CIDDTW* e *DTW-D*. Em contraste, as distâncias Kulczynski, Pearson, Soergel e Quadrática Aditiva Simétrica (*QAS*) demonstraram os piores resultados de *MSE* (Figura 54a) e *TU* (Figura 54c). Isso significa que essas medidas têm pouco poder discriminativo. Em termos de *POCID* (Figura 54e), a *CID* foi a distância mais promissora no quesito predizer com precisão a tendência dos horizontes futuros.

Considerando a estratégia multi-etapa à frente com passo atualizado, a distância Quadrática exibiu, segundo os índices *MSE* (Figura 54b) e *TU* (Figura 54d), os piores resultados. Em oposição, a medida Canberra acarretou, em conformidade com os três índices de desempenho (Figuras 54b, 54d e 54f), nos melhores resultados.

Na Figura 55, os valores do coeficiente *TU* indicam que, para a predição multi-etapa à frente com passo aproximado (Figura 55a), a utilização das métricas da Norma L_p (Manhattan, Euclidiana, Métrica L_3 e Chebychev), da categoria Produto Interno (Correlação, Cosseno, Geodésica e Jaccard) e das distâncias Canberra, Lorentizian, Clark, *AD*, *DTW* e *CID* foram apropriadas em, aproximadamente, 23 conjuntos de dados ($TU < 1$), sendo que destes, 12 apresentaram uma modelagem confiável para a realização de projeções futuras ($TU \leq 0,55$). Os resultados do uso dessas mesmas distâncias e da *CIDDTW*, porém adotando a estratégia de predição multi-etapa

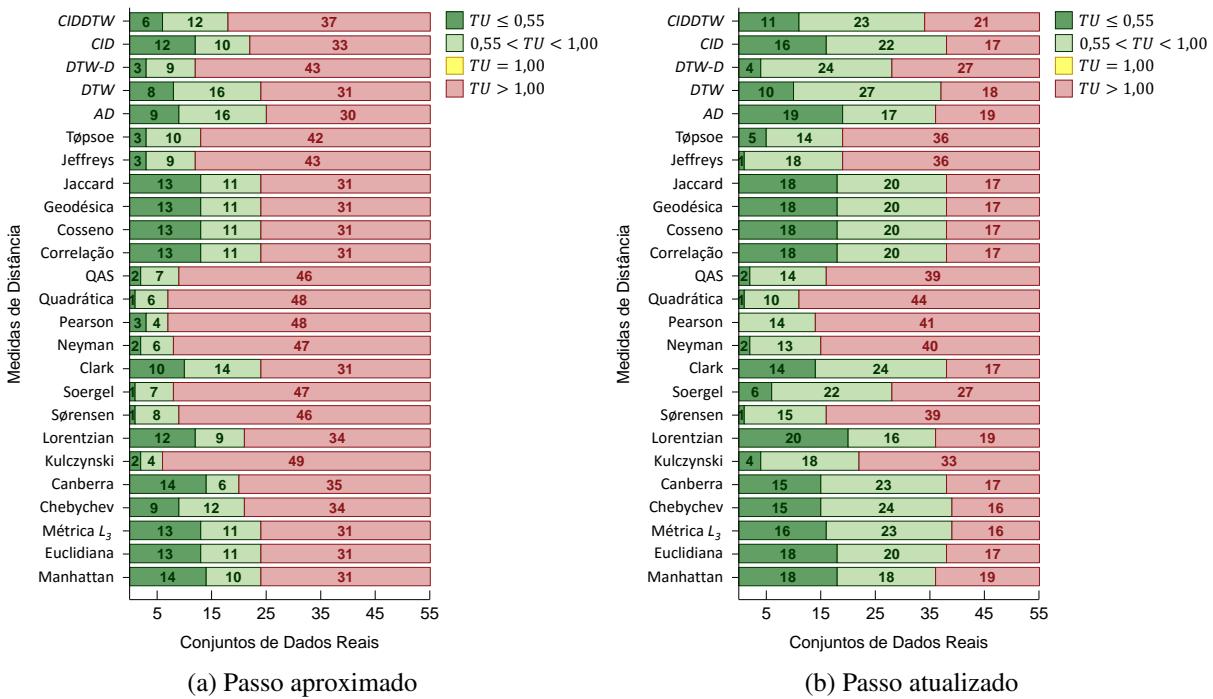
Figura 54 – Diagramas de distância crítica para os valores dos índices *MSE*, *TU* e *POCID* provenientes do *kNN-TSPI*, usando diferentes medidas distâncias, sobre ST reais



Fonte: Elaborada pelo autor.

à frente com passo atualizado (Figura 55b), foram adequadas, em média, para 37 conjuntos de dados ($TU < 1$), dos quais 16 acarretaram em modelos preditivos confiáveis ($TU \leq 0,55$). É importante ressaltar que, para ambas as estratégias de predição, os piores resultados foram alcançados por meio da aplicação das métricas Sørensen, Neyman, Pearson, Quadrática e QAS.

Figura 55 – Desempenho em ST reais do *kNN-TSPI*, empregando distintas medidas de distância, para quatro faixas de valores do coeficiente TU



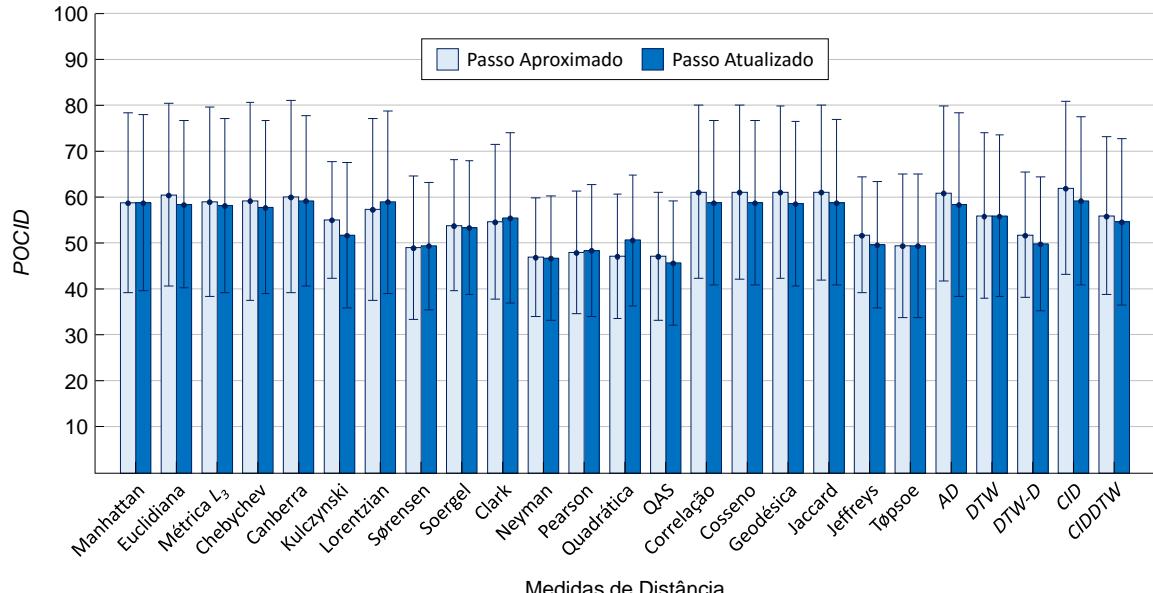
Fonte: Elaborada pelo autor.

As médias dos valores do índice *POCID* com a indicação dos desvios padrão, para cada medida de distância e estratégia de predição, são mostradas graficamente na Figura 56. Usando a predição multi-etapa à frente com passo aproximado, a medida *CID* proporcionou a melhor taxa média de acerto (61,86% com desvio padrão de 18,83%), enquanto a distância Neyman produziu o pior índice médio de desempenho (46,94% com desvio padrão de 12,88%). Já para a predição multi-etapa à frente com passo atualizado, a melhor taxa média de acerto foi obtida pelo emprego da *CID* (59,11% com desvio padrão de 18,34%) e a pior foi decorrente do uso da distância QAS (45,63% com desvio padrão de 13,58%). Portanto, independentemente da estratégia de predição adotada, a medida *CID* proporcionou as melhores taxas de acerto quanto à tendência dos horizontes projetados.

7.3.4 *Medidas de Complexidade Aplicadas à CID*

Como tratado no Capítulo 6, a *CID* utiliza uma estimativa de complexidade baseada na intuição de que uma ST pode ser “esticada” até que se torne um segmento de reta. Por conseguinte, uma série complexa resultaria em um segmento de reta mais longo do que uma

Figura 56 – Médias e desvios padrão das taxas de acerto *POCID* obtidas pelo *kNN-TSPI*, utilizando diferentes medidas de distâncias, em ST reais



Fonte: Elaborada pelo autor.

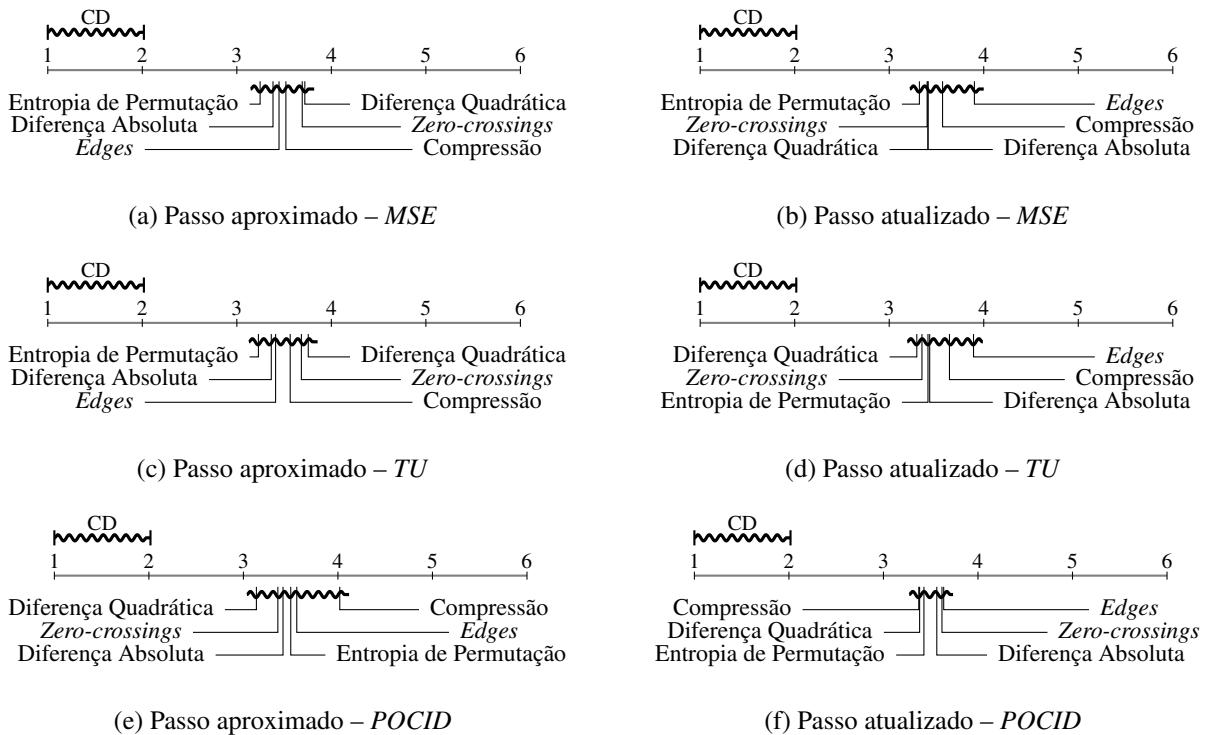
ST simples. Por meio do uso dessa estimativa de complexidade, denominada de Diferença Quadrática, é possível atribuir maiores distâncias à subsequências com diferentes complexidades.

Neste trabalho foram pesquisadas, além da estimativa de complexidade usada originalmente pela *CID* (Diferença Quadrática), cinco diferentes medidas de complexidade: Diferença Absoluta, Compressão, *Edges*, *Zero-crossings* e Entropia de Permutação. Essas estimativas de complexidade foram aplicadas à *CID*, no intuito de avaliar o impacto dessas combinações no processo de busca por similaridade adotado pelo *kNN-TSPI* e, consequentemente, na qualidade do desempenho preditivo do mesmo. Os experimentos derivados dessas aplicações totalizaram 660 configurações (1 modelo preditivo \times 6 medidas de complexidade \times 2 estratégias de projeção \times 55 conjuntos de dados).

Na Figura 57 são exibidos os diagramas de distância crítica com relação aos valores das medidas *MSE*, *TU* e *POCID* decorrentes das 660 configurações. Nessa figura, examinando a estratégia de predição multi-etapa à frente com passo aproximado, a Entropia de Permutação foi a estimativa de complexidade que forneceu, em média, os melhores resultados de *MSE* (Figura 57a) e *TU* (Figura 57c). Contudo, ela não apresentou diferenças estatisticamente significativas quando comparada com as outras medidas de complexidade. A Diferença Quadrática, embora tenha demonstrado, sem diferenças significativas, os piores resultados de *MSE* (Figura 57a) e *TU* (Figura 57c), obteve as melhores taxas de acerto *POCID* (Figura 57e).

Em relação à estratégia de predição multi-etapa à frente com passo atualizado, a estimativa de complexidade *Edges* exibiu, em conformidade com os três índices de desempenho (Figuras 57b, 57d, 57f), os piores resultados. Isso pode significar que essa medida tem pouco

Figura 57 – Diagramas de distância crítica para os valores dos índices *MSE*, *TU* e *POCID* provenientes do *kNN-TSPI*, usando distintas estimativas de complexidade, sobre ST reais



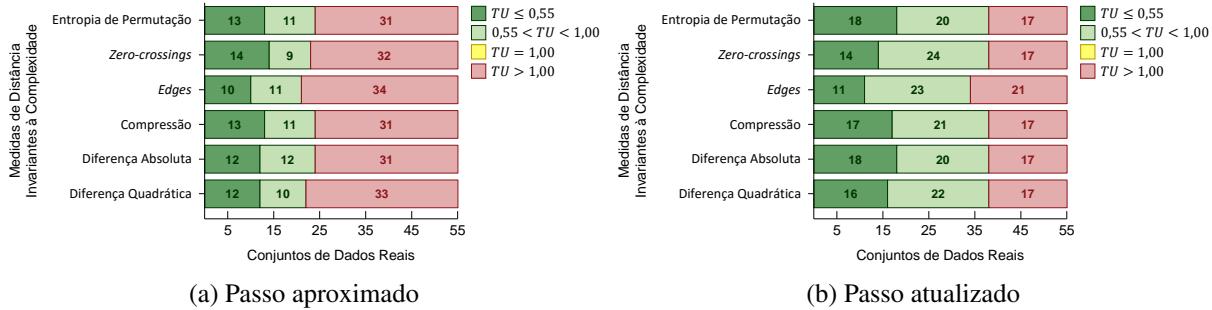
Fonte: Elaborada pelo autor.

poder discriminativo em comparação com as demais. É importante observar que a estimativa de complexidade *Zero-crossings* (Figura 57b) é competitiva com a Diferença Quadrática (Figura 57d), embora esta última seja mais simples, tanto conceitualmente quanto na prática. Os valores de *POCID* (Figura 57f) evidenciam que as medidas Compressão, Diferença Quadrática e Entropia de Permutação foram as mais promissoras para predizer a tendência dos horizontes projetados.

Nos gráficos da Figura 58, os valores do coeficiente *TU* indicam que, dentre todas as estimativas de complexidade investigadas, a medida *Edges* acarretou nos piores resultados. Especificamente, para predição multi-etapa à frente com passo aproximado (Figura 58a), a utilização do *kNN-TSPI* empregando a medida supracitada não foi preferível comparado ao modelo trivial em 34 do total de 55 conjuntos de dados ($TU > 1$). Aplicando a estratégia de predição multi-etapa à frente com passo atualizado (Figura 58b), a medida *Edges* apresentou-se adequada ($TU < 1$) para modelar e predizer 34 (11 + 23) conjuntos de dados, dos quais apenas 11 acarretaram em modelos tidos como confiáveis para a projeção de valores futuros.

Nota-se na Figura 59 que, para ambas as estratégias de predição, a média e o desvio padrão dos valores de *POCID* estão distribuídos uniformemente entre as seis estimativas de complexidade. Esse fato demonstra que o uso de qualquer uma dessas estimativas de com-

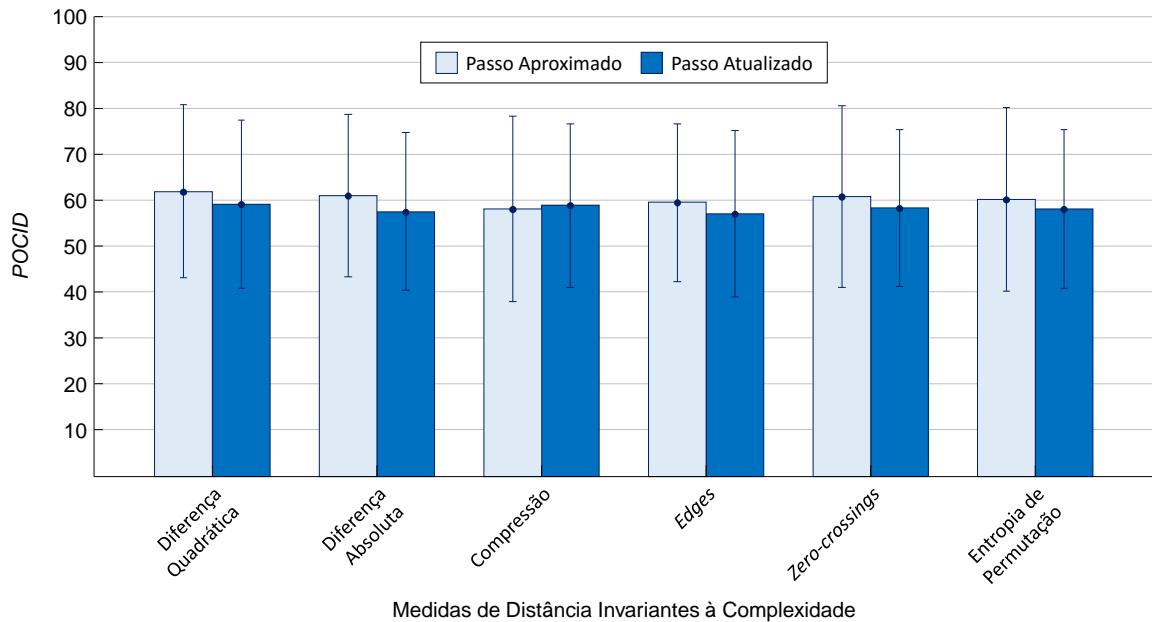
Figura 58 – Desempenho em ST reais do *kNN-TSPI*, empregando diferentes estimativas de complexidade, para quatro faixas de valores do coeficiente *TU*



Fonte: Elaborada pelo autor.

plexidade pelo *kNN-TSPI* implicou em uma taxa média de acerto sobre as tendências futuras de, aproximadamente, 59,21% com desvio padrão de 18,27%. Considerando todas as medidas investigadas, a estimativa de complexidade Diferença Quadrática exibiu o melhor resultado, isto é, emplacou uma taxa média de acerto de 61,86% com desvio padrão de 18,83%, na predição via passo aproximado, e 59,11% com desvio padrão de 18,34%, na predição utilizando passo atualizado.

Figura 59 – Médias e desvios padrão das taxas de acerto *POCID* obtidas pelo *kNN-TSPI*, utilizando distintas estimativas de complexidade, em ST reais



Fonte: Elaborada pelo autor.

7.3.5 Funções de Predição

A função de predição utilizada em conjunto com o algoritmo *kNN-TSPI* deve possibilitar a projeção de dados que apresentem variação de amplitude ao longo do tempo. Desse modo,

foram analisadas quatro funções de predição: Mediana; Média de Valores Absolutos (MVA); Média de Valores Relativos (MVR); *Index Weighted* (*IW*), a qual consiste na aplicação de pesos ponderados pelo índice temporal da subsequência a ser comparada; e *Distance Weighted* (*DW*), que é uma média ponderada pelos valores das distâncias (d) calculadas. Para este último caso, foram investigadas seis variações de pesos, as quais são listadas no [Quadro 8](#). As combinações realizadas resultaram em 1100 configurações (1 modelo preditivo \times 10 funções de predição \times 2 estratégias de projeção \times 55 conjuntos de dados).

Quadro 8 – Pesos considerados na aplicação da função de predição *DW*

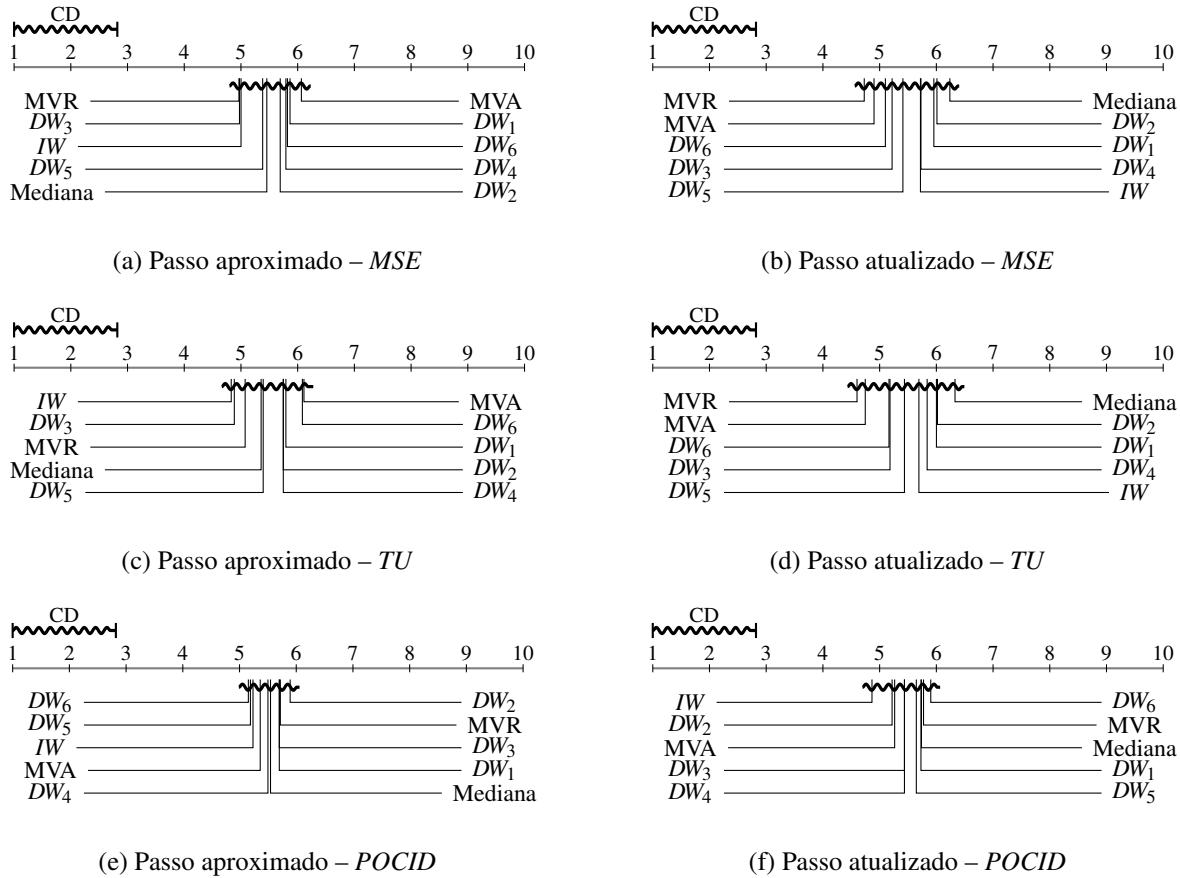
ID	Peso (w)
1	$w = 1/d(Q, S)$
2	$w = 1/d(Q, S)^2$
3	$w = \exp(-d(Q, S)^2)$
4	$w = \exp(-d(Q, S)^2/(2\sigma^2))$ para $\sigma = 0,5$
5	$w = \exp(-d(Q, S)/(2\sigma^2))$ para $\sigma = 0,5$
6	$w = \exp(-d(Q, S)/\sigma)$ para $\sigma = 0,5$

Na [Figura 60](#) são mostrados os diagramas de distância crítica com relação aos valores das medidas *MSE*, *TU* e *POCID* decorrentes das 1100 configurações. Nessa figura, analisando ambas as estratégias de predição, a função MVR proporcionou ao *kNN-TSPI* os menores erros de projeção. Diferentemente, a Mediana foi a função de predição que acarretou nos piores resultados. É interessante observar que a DW_3 , DW_5 , a DW_6 e a *IW* se apresentaram, ponderando as três medidas de desempenho, como as funções mais estáveis para a predição de ST.

Na [Figura 61](#), as estatísticas provenientes da utilização do coeficiente *TU* indicam que, para a predição multi-etapa à frente com passo aproximado ([Figura 61a](#)), a função MVR em conjunto com o *kNN-TSPI* promoveu, para 24 (14 + 10) do total de 55 conjuntos de dados, os menores valores de *TU*. Considerando a predição multi-etapa à frente com passo atualizado ([Figura 61b](#)), as funções MVA, DW_3 , DW_5 e DW_6 registraram, para 38 do total de 55 conjuntos de dados, os melhores resultados ($TU < 1$).

Como pode ser observado na [Figura 62](#), para ambas as estratégias de projeção, a média e o desvio padrão dos valores de *POCID* estão distribuídos uniformemente entre as 10 funções de predição. Isso significa que o uso de qualquer uma dessas funções de projeção pelo *kNN-TSPI* acarretou em uma taxa média de acerto sobre as tendências futuras de, aproximadamente, 60,05% com desvio padrão de 19,19%. Particularmente, na predição via passo aproximado, o uso da função de predição *IW* acarretou nos melhores resultados (taxa média de acerto de 62,70% com desvio padrão de 18,44%). Em contraste, na predição conduzida por passo atualizado, os melhores resultados foram obtidos por meio da aplicação da função de predição DW_2 (taxa média de acerto de 60,32% com desvio padrão de 18,58%).

Figura 60 – Diagramas de distância crítica para os valores dos índices *MSE*, *TU* e *POCID* provenientes do *kNN-TSPI* usando, diferentes funções de predição, sobre ST reais



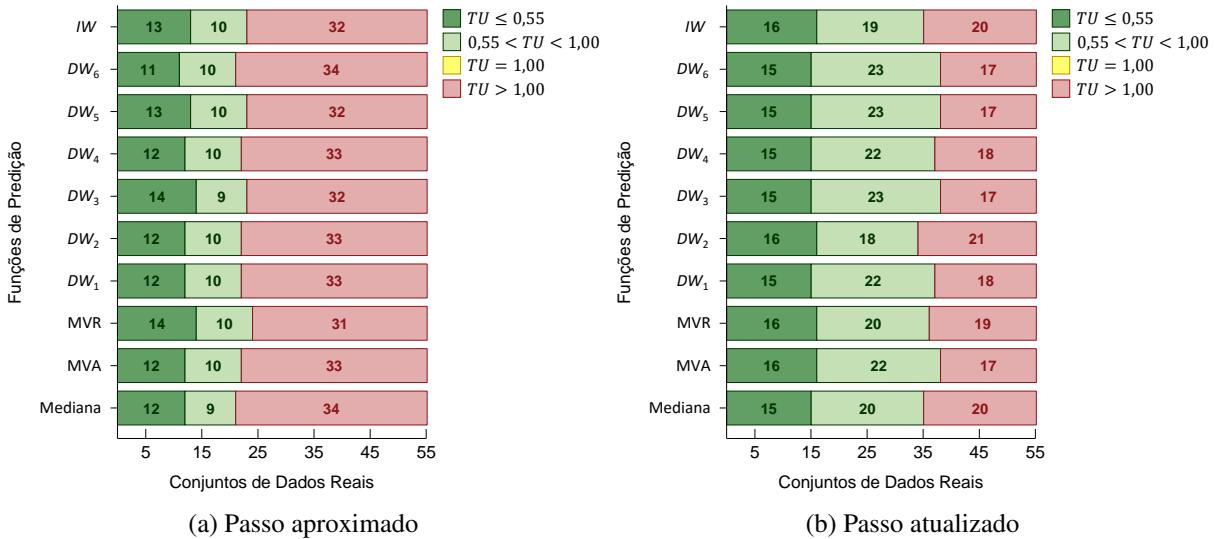
Fonte: Elaborada pelo autor.

7.4 Considerações Finais

Neste capítulo, a partir de 55 conjuntos de dados reais, foram realizados diversos experimentos computacionais a respeito das questões que exercem impacto na qualidade da predição de ST por similaridade. As propriedades pesquisadas foram, sobretudo, invariâncias às distorções em dados temporais, medidas de distância, medidas de complexidade aplicadas à *CID* e funções de predição. Complementarmente, o *kNN-TSPI* foi comparado com outros dois métodos de predição de ST por similaridade: *kNN* e *kNN-TSP*.

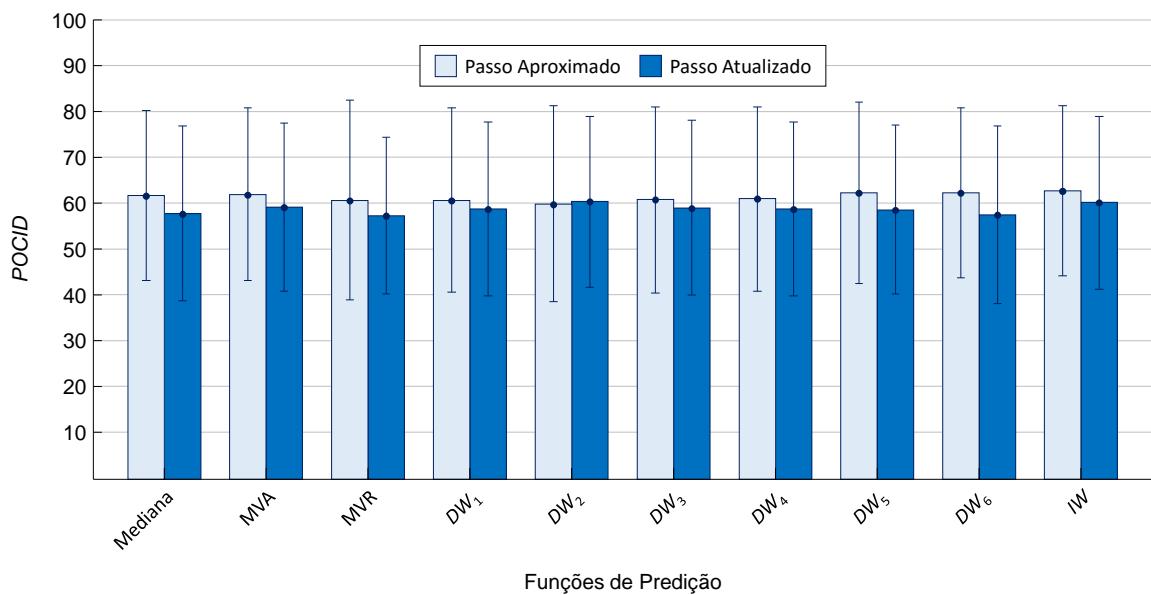
O *kNN-TSP* com invariância à amplitude e deslocamento, complexidade e tratamento de casamentos triviais apresentou os resultados mais promissores. Essas três técnicas invariantes são as utilizadas pelo *kNN-TSPI*, o qual ainda assegura que os valores a serem usados pela função de predição tenham a mesma distribuição que suas respectivas subsequências de origem. Tal certificação permite um melhor acerto quanto à tendência do horizonte futuro. Adicionalmente, verificou-se que a adoção de invariância à complexidade é um fator determinante para o bom desempenho da predição de ST por similaridade.

Figura 61 – Desempenho em ST reais do *kNN-TSPI*, empregando distintas funções de predição, para quatro faixas de valores do coeficiente *TU*



Fonte: Elaborada pelo autor.

Figura 62 – Médias e desvios padrão das taxas de acerto *POCID* obtidas pelo *kNN-TSPI*, utilizando diferentes funções de predição, em ST reais



Fonte: Elaborada pelo autor.

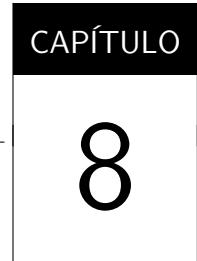
Quando confrontado com os métodos *kNN* e *kNN-TSP*, o algoritmo *kNN-TSPI* apresentou, em geral, os melhores resultados. Embora não tenha sido verificada diferenças estatisticamente significativas entre essas comparações, o *kNN-TSPI* é mais robusto contra escolhas errôneas das subsequências mais similares e, por esse motivo, obteve as melhores taxas de acerto quanto à tendência dos horizontes projetados.

Em relação as medidas de distâncias pesquisadas, a Manhattan providenciou ao *kNN-TSPI* com passo aproximado os menores erros de predição. Esse fato também foi observado para a Canberra em relação ao *kNN-TSPI* com passo atualizado. Tanto a distância Manhattan quanto à Canberra não exibiram, para ambas as estratégias de predição, diferenças estatisticamente significativas quando confrontadas com as métricas usualmente empregadas na literatura, a exemplo: Euclidiana, Cosseno, Geodésica, *CID*, *DTW*, *CIDDTW* e *DTW-D*. Independentemente da estratégia de predição averiguada, a *CID* providenciou ao *kNN-TSPI* as melhores taxas de acerto quanto à tendência dos horizontes projetados.

No que diz respeitos às estimativas de complexidade, as medidas Diferença Quadrática e Diferença Absoluta não apresentaram diferenças estatisticamente significativas quando comparadas às demais estimativas. Pela simplicidade de codificação e por acarretar nas melhores taxas de acerto sobre as tendências futuras, recomenda-se a utilização da Diferença Quadrática cujo desempenho foi previamente verificado em [Parmezan e Batista \(2015\)](#).

Os experimentos envolvendo as funções de predição indicaram, sem diferenças estatisticamente significativas, a eficiência da MVR. No entanto, foi possível notar que a DW_3 , DW_5 , DW_6 e *IW* foram as funções mais estáveis para a predição de valores em ST. Nesse contexto, a *IW* demonstra ser uma boa candidata à função de predição, visto que atribui pesos maiores para as observações mais recentes.

No próximo capítulo são exibidos os resultados que demonstram que é possível alcançar bons desempenhos na predição de ST usando o método *kNN-TSPI*. Esses resultados tornam verdadeira a hipótese levantada nesta dissertação de mestrado, ou seja, que os métodos de predição de ST baseados em similaridade provém resultados competitivos em relação aos obtidos com a aplicação de métodos estatísticos estado-da-arte.



AVALIAÇÃO EXPERIMENTAL II: COMPARANDO O ALGORITMO *kNN-TSPI* COM MÉTODOS TRADICIONAIS DA LITERATURA

8.1 Considerações Iniciais

A escolha do algoritmo mais promissor para explicar e/ou predizer um determinado fenômeno reside em uma das atividades mais críticas do processo de Mineração de Dados Temporais (MDT). Como diferentes modelos são capazes de descrever uma mesma sequência de dados, argumenta-se que nenhum algoritmo de modelagem pode ser considerado o melhor independentemente do problema em questão. À vista disso, uma das alternativas é testar os modelos disponíveis a fim de selecionar aquele que apresenta o melhor ajuste aos dados. Todavia, além de exigir vasto conhecimento técnico, esses experimentos podem consumir um longo período de tempo. É prática comum, portanto, guiar-se em estudos empíricos consistentes reportados na bibliografia relacionada.

Os resultados da revisão sistemática conduzida neste trabalho permitiram constatar uma carência de publicações envolvendo avaliações experimentais robustas que viabilizem o confronto entre métodos estatísticos e de Aprendizado de Máquina (AM) para predição de Séries Temporais (ST). Uma análise minuciosa acerca do desempenho desses algoritmos em conjuntos de dados projetados para *benchmarking*, de origem artificial (sintético) e real, poderia ratificar as vantagens e desvantagens do uso de cada método.

Neste capítulo, a partir de 95 conjuntos de dados (40 ST sintéticas e 55 ST reais), o algoritmo *k-Nearest Neighbors - Time Series Prediction with Invariances* (*kNN-TSPI*) é comparado com nove métodos amplamente difundidos na literatura. Essa extensa pesquisa

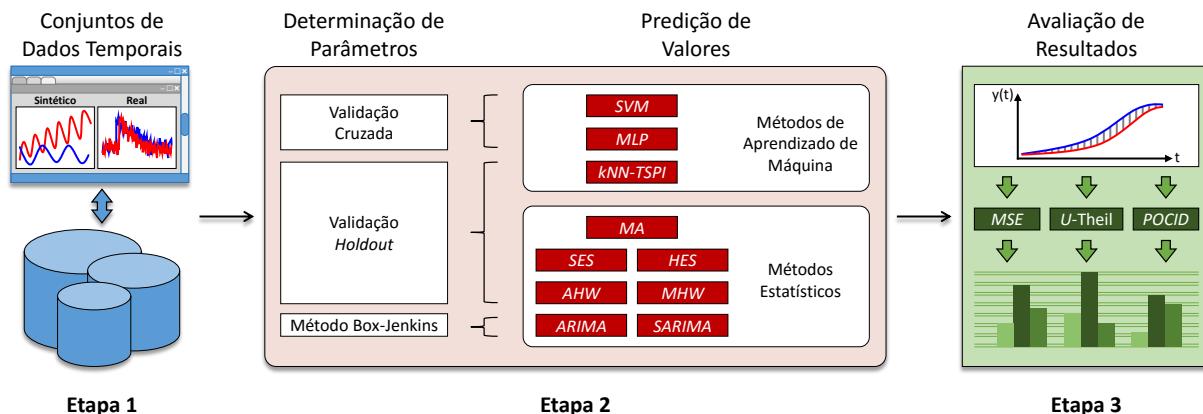
empírica, além de comprovar a hipótese levantada nesta dissertação de mestrado, foi planejada para demonstrar, de maneira compreensível e replicável, a eficiência e efetividade dos algoritmos estado-da-arte destinados à construção de modelos preditivos.

8.2 Configuração Experimental

Na intenção de certificar que o *kNN-TSPI* pode fornecer resultados tão precisos quanto outros métodos estatísticos e de AM, o algoritmo proposto foi confrontado com outros nove métodos: Máquinas de Suporte Vetorial (*SVM*), *Multilayer Perceptron* (*MLP*), Médias Móveis (*MA*), Suavização Exponencial Simples (*SES*), Suavização Exponencial de Holt (*HES*), Holt-Winters Aditivo (*AHW*), Holt-Winters Multiplicativo (*MHW*), Autorregressivo Integrado de Médias Móveis (*ARIMA*) e Autorregressivo Integrado de Médias Móveis Sazonal (*SARIMA*). Grande parte desses algoritmos são empregados em aplicações do mundo real, sendo os modelos *ARIMA* e *SARIMA* o estado-da-arte para a modelagem e a predição de ST.

O protocolo experimental foi organizado em três etapas, como ilustrado na Figura 63. Essa configuração é análoga à retratada na Figura 47, exceto pela adição de 40 conjuntos de dados sintéticos e pelo modo como os parâmetros dos algoritmos foram identificados.

Figura 63 – Configuração experimental II



Fonte: Elaborada pelo autor.

Na Etapa 1, além das 55 ST reais listadas no Quadro 6 da página 141, foram utilizadas 40 sequências de dados geradas computacionalmente a partir de propriedades previamente estabelecidas. Uma descrição dos 40 conjuntos de dados sintéticos agrupados, segundo seu processo originário, nas categorias determinística, estocástica e caótica pode ser visibilizada no Apêndice B, porém um resumo de suas particularidades e configurações é mostrado no Quadro 9. Nesse quadro, para cada conjunto de dados construído, 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 série de dados ($h = m \times 5\%$).

Quadro 9 – Sumário de características e de configurações dos conjuntos de dados sintéticos

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

Na Etapa 2, os parâmetros de ajuste foram estimados em conformidade com as características de cada método investigado. 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 2 da página 101). Já os parâmetros dos modelos paramétricos

ARIMA e *SARIMA* foram definidos usando o método Box-Jenkins com minimização do Critério de Informação de Akaike (*AIC*) e máxima verossimilhança ([Algoritmo 3 da página 102](#)). Os métodos restantes tiveram seus parâmetros estabelecidos empregando validação *holdout* com minimização do *MSE* ([Algoritmo 1 da página 99](#)).

Os algoritmos de predição adotados na Etapa 2, em conjunto com seus parâmetros, são listados no [Quadro 10](#). Nesse quadro, na terceira coluna, são indicados os intervalos de variação numérica, estipulados para cada parâmetro, que foram testados pelas técnicas de estimação.

Quadro 10 – Algoritmos utilizados e intervalos de variação numérica definidos para os seus parâmetros

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

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 introduzidas no [Capítulo 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, em termos de erro preditivo, por meio da utilização da medida *MSE* ([Equação 4.31 da página 103](#)) e do coeficiente *U* de Theil (*TU*) ([Equação 4.32 da página 104](#)). O índice de desempenho *Prediction Of Change In Direction (POCID)* ([Equação 4.33 da página 104](#)), que mensura a taxa de acerto

quanto à tendência do horizonte de predição, também foi computado. A partir dos valores dessas medidas foi possível confrontar o algoritmo proposto neste trabalho com os métodos estado-da-arte. 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 Nemenyi¹.

A execução do protocolo experimental exposto abrangeu o uso dos seguintes ambientes computacionais e linguagens de alto nível: (1) MATLAB² e GNU Octave³, assim como seus respectivos pacotes de funções relacionados ao tema de predição de ST; (2) linguagem R⁴ aliada ao pacote *Forecast*⁵, ambos integrados ao ambiente de desenvolvimento RStudio⁶; e (3) linguagem Java (DEITEL; DEITEL, 2012), no ambiente *Eclipse*⁷, com a biblioteca Weka (WITTEN; FRANK; HALL, 2011).

8.3 Resultados e Discussão

Os resultados experimentais alcançados, apresentados e discutidos a seguir, estão organizados da seguinte maneira:

1. Estudo comparativo entre o *kNN-TSPI* e os modelos preditivos clássicos aplicados à dados temporais sintéticos;
2. Confronto do algoritmo *kNN-TSPI* com os métodos convencionais para projeção de valores sobre dados temporais reais;
3. Análise do desempenho preditivo do *kNN-TSPI* em relação aos modelos de predição tradicionais aplicados à dados temporais sintéticos e reais.

Os valores dos índices *MSE*, *TU* e *POCID*, tratados nesta seção, podem, juntamente com os parâmetros empregados para obter tais resultados, ser encontrados no *ICMC-USP Time Series Prediction Repository* (PARMEZAN; BATISTA, 2014).

Como mencionado, os algoritmos de projeção foram comparados, por meio das três medidas de avaliação de desempenho supracitadas, empregando o teste estatístico de Friedman, com nível de significância de 5%, seguido do pós-teste de Nemenyi. Os resultados dessa validação estatística foram esboçados em diagramas de Diferença Crítica (CD) (DEMŠAR, 2006), os quais determinam que quanto mais próximo de 1 está o valor indicado para uma configuração, melhor

¹ Testes estatísticos realizados utilizando *KEEL Software Tool* para Windows, <<http://www.keel.es>>.

² <<http://www.mathworks.com/products/matlab>>.

³ <<https://www.gnu.org/software/octave>>.

⁴ <<https://www.r-project.org>>.

⁵ <<https://cran.r-project.org/web/packages/forecast>>.

⁶ <<https://www.rstudio.com/>>.

⁷ <<http://www.eclipse.org>>.

é o seu desempenho. Adicionalmente, os valores dos índices *TU* e *POCID* foram representados, nessa ordem, em gráficos de barras totalmente empilhadas e gráficos de barras com desvios padrão.

8.3.1 Séries Sintéticas

Nessa coleção de experimentos foram utilizadas 40 séries sintéticas, das quais 17 são determinísticas, 15 são estocásticas e oito são caóticas. Os resultados, além de examinados separadamente para cada um dos grupos de ST, foram discorridos considerando todos os conjuntos de dados. Particularmente, o uso de dados sintéticos foi escolhido para possibilitar a execução de testes computacionais em um ambiente relativamente controlado, o que permite analisar e compreender o desempenho dos algoritmos de predição em ST com características distintas.

8.3.1.1 Séries Determinísticas

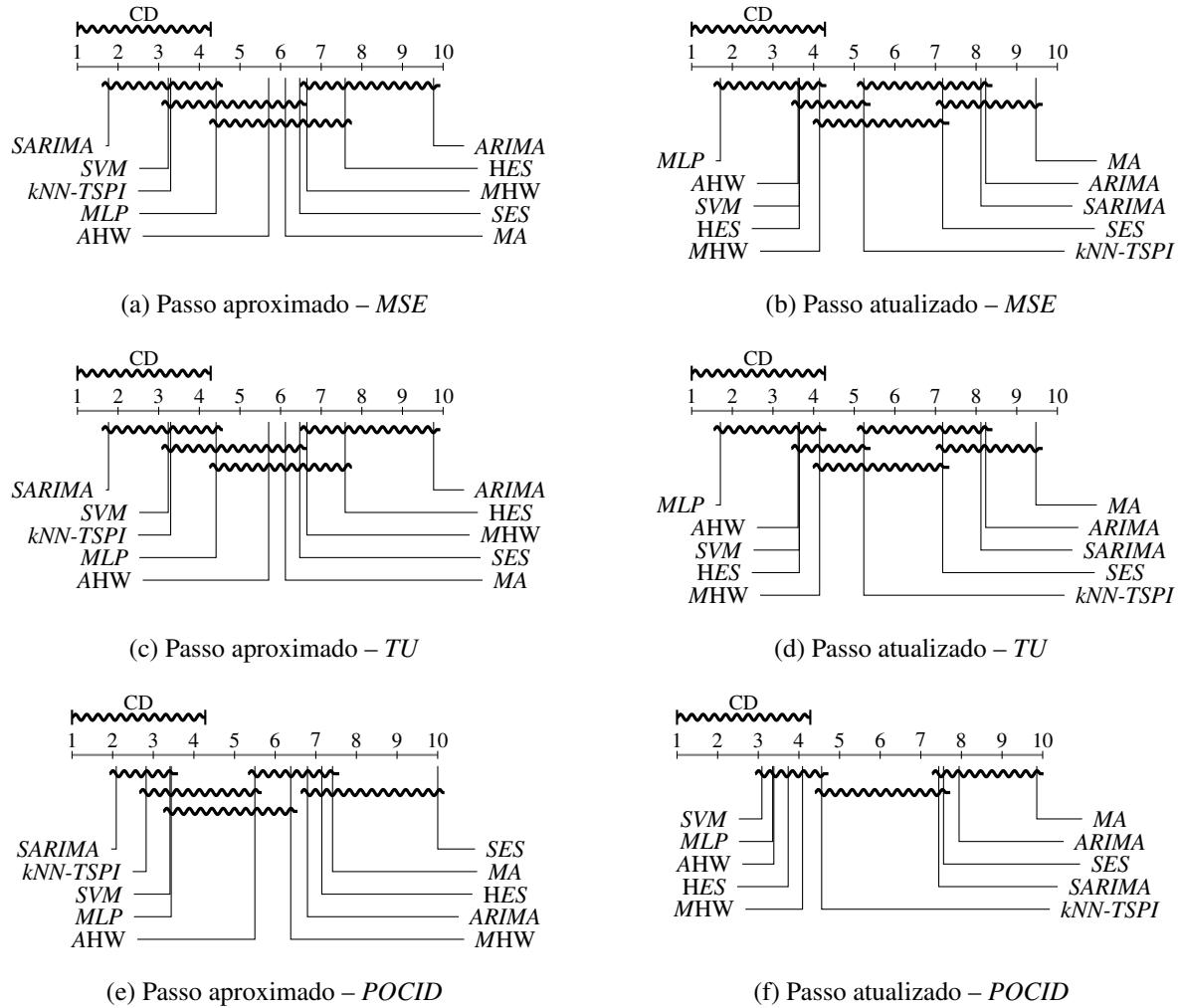
Os experimentos conduzidos sobre as sequências de dados determinísticas envolveram 340 configurações (10 modelos preditivos \times 2 estratégias de predição \times 17 conjuntos de dados). Na [Figura 64](#) são mostrados os diagramas de distância crítica com relação aos valores das medidas *MSE*, *TU* e *POCID* procedentes das configurações pesquisadas.

Em conformidade com as escalas representadas na [Figura 64](#), as quais indicam o *ranking* médio de desempenho dos algoritmos sobre os conjuntos de dados, o *kNN-TSPI* com passo aproximado ficou em terceiro lugar segundo as medidas *MSE* ([Figura 64a](#)) e *TU* ([Figura 64c](#)), perdendo sem diferenças estatisticamente significativas para os modelos *SARIMA* e *SVM*. Já para o índice *POCID* ([Figura 64e](#)), 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 64b](#)) e *TU* ([Figura 64d](#)). No entanto, quando analisadas as taxas de acerto quanto à tendência dos horizontes projetados ([Figura 64f](#)), essa configuração ocupou a segunda posição no *ranking* médio, não exibindo diferenças estatisticamente significativas em comparação com o modelo *SVM*.

Na [Figura 65](#), as quatro faixas de valores do coeficiente *TU* demonstram que, para a predição multi-etapa à frente com passo aproximado ([Figura 65a](#)), as projeções realizadas pelo *SARIMA* foram confiáveis em 14 do total de 17 conjuntos de dados ($TU \leq 0,55$). Ainda nesse cenário, a projeção por similaridade com invariâncias alcançou o segundo melhor desempenho, de modo que para 14 conjuntos de dados ($12 + 2$) sua utilização foi preferível em relação à predição trivial ou ingênua ($TU < 1$). Como esperado, o algoritmo *SES* não conseguiu vencer o modelo ingênuo para nenhum conjunto de dados. Isso aconteceu porque a predição com passo aproximado não é suportada por ele.

Figura 64 – Diagramas de distância crítica para os valores dos índices *MSE*, *TU* e *POCID* provenientes dos métodos de predição sobre ST determinísticas



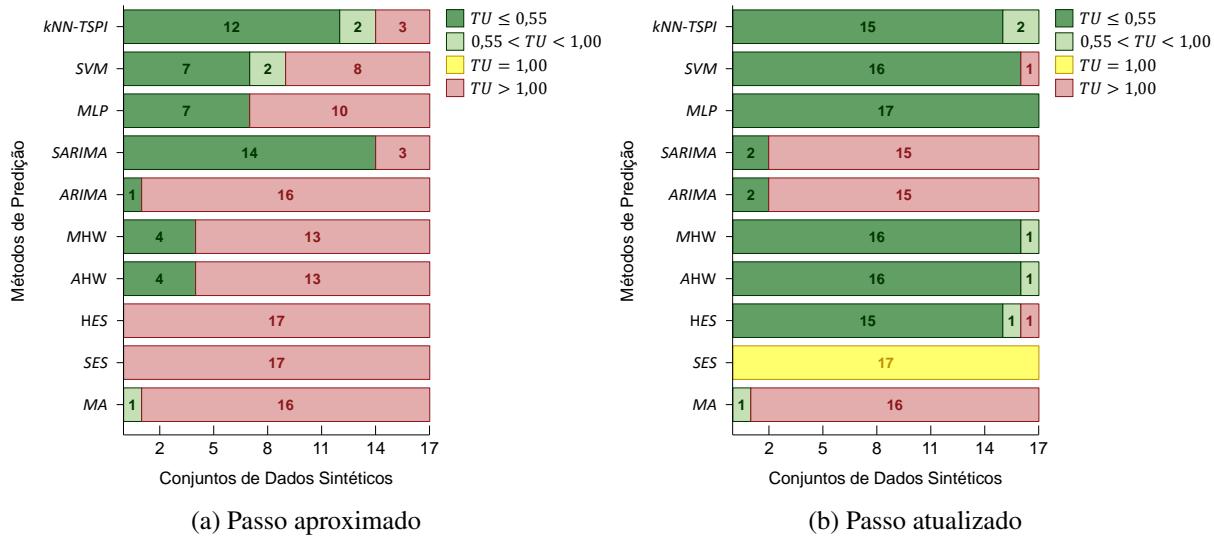
Fonte: Elaborada pelo autor.

Quanto aos resultados da estratégia de predição multi-etapa à frente com passo atualizado (Figura 65b), o *MLP* mostrou-se confiável para projetar os valores dos 17 conjuntos de dados ($TU \leq 0,55$). Já o *kNN-TSPI* foi o algoritmo que apresentou o segundo melhor desempenho, ou seja, seu uso foi preferível ($TU < 1$) em relação ao modelo ingênuo para 17 conjuntos de dados ($15 + 2$), dos quais 15 proporcionaram modelagens confiáveis para a projeção de valores futuros ($TU \leq 0,55$).

Analizando simultaneamente as duas estratégias de predição da Figura 65, o desempenho do modelo não-paramétrico *kNN-TSPI* foi, em geral, mais estável que os outros métodos para predizer ST determinísticas.

Na Figura 66, para cada algoritmo e estratégia de projeção, é esboçada a média dos valores da medida *POCID* com a indicação dos desvios padrão. Nessa figura, o modelo *SARIMA* com passo aproximado emplacou uma taxa média de acerto de 98,29%, com desvio padrão de

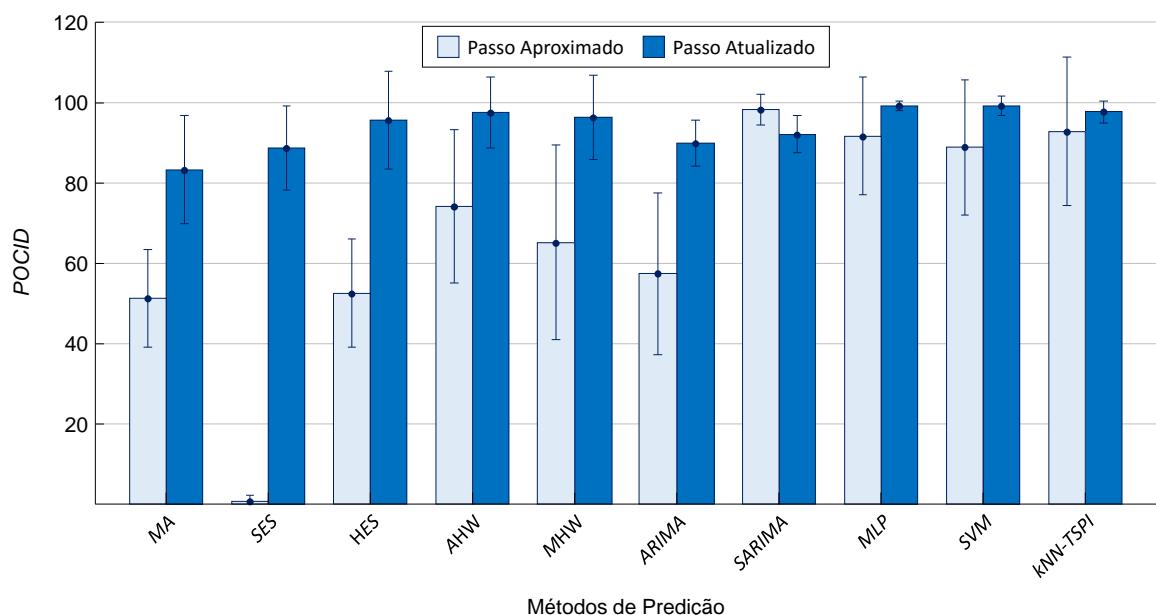
Figura 65 – Desempenho dos métodos de predição para quatro faixas de valores do coeficiente *TU* em ST determinísticas



Fonte: Elaborada pelo autor.

3,83%, sobre as tendências dos horizontes de predição. Diferentemente, o *kNN-TSPI* com passo aproximado implicou em uma taxa média de certo de 92,99% com desvio padrão de 18,44%. Considerando a predição multi-etapa à frente com passo atualizado, o *kNN-TSPI* resultou na terceira melhor taxa média de acerto (97,83% com desvio padrão de 2,76%), perdendo para os modelos *MLP* (99,39% com desvio padrão de 1,24%) e *SVM* (99,36% com desvio padrão de 2,42%).

Figura 66 – Médias e desvios padrão das taxas de acerto *POCID* obtidas pelos métodos de predição em ST determinísticas

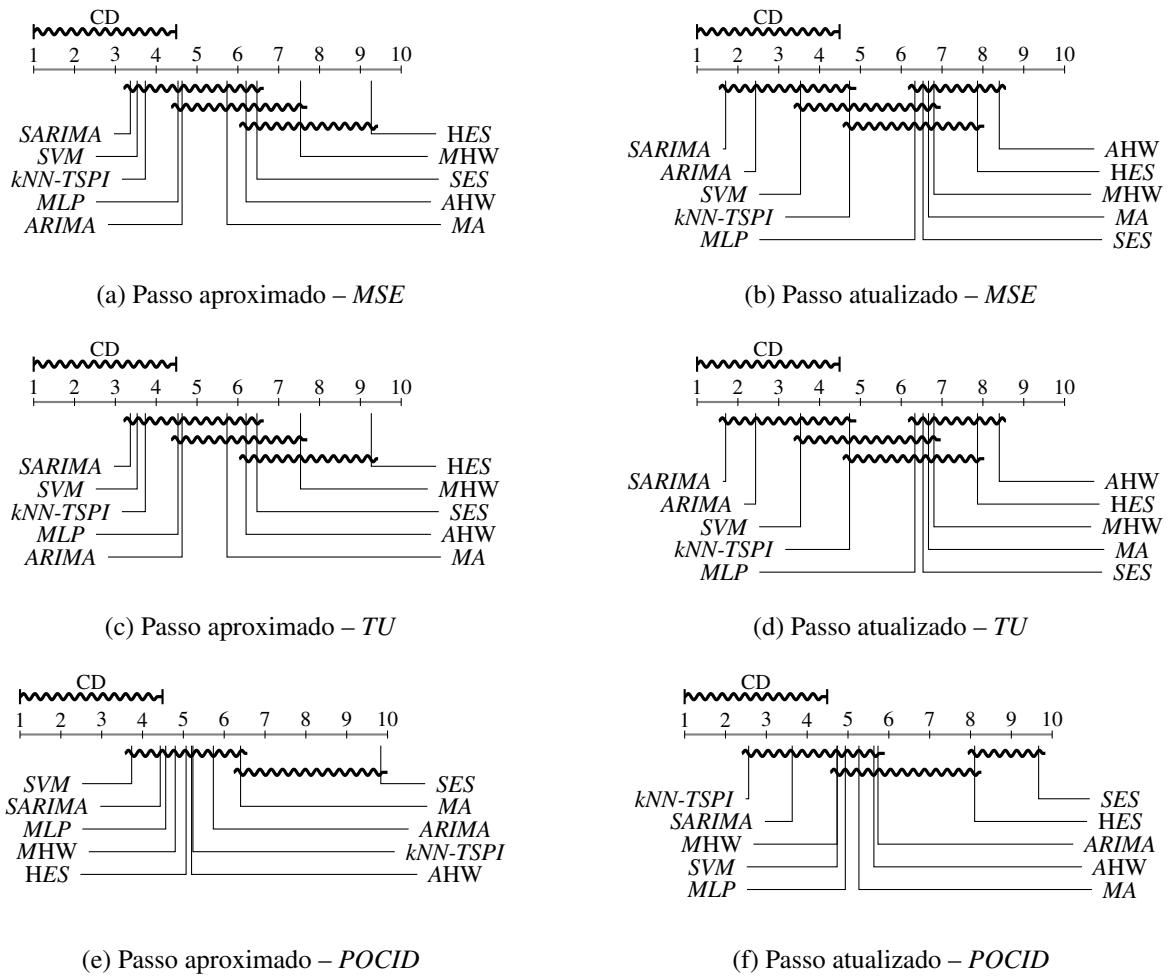


Fonte: Elaborada pelo autor.

8.3.1.2 Séries Estocásticas

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 67.

Figura 67 – Diagramas de distância crítica para os valores dos índices *MSE*, *TU* e *POCID* provenientes dos métodos de predição sobre ST estocásticas



Fonte: Elaborada pelo autor.

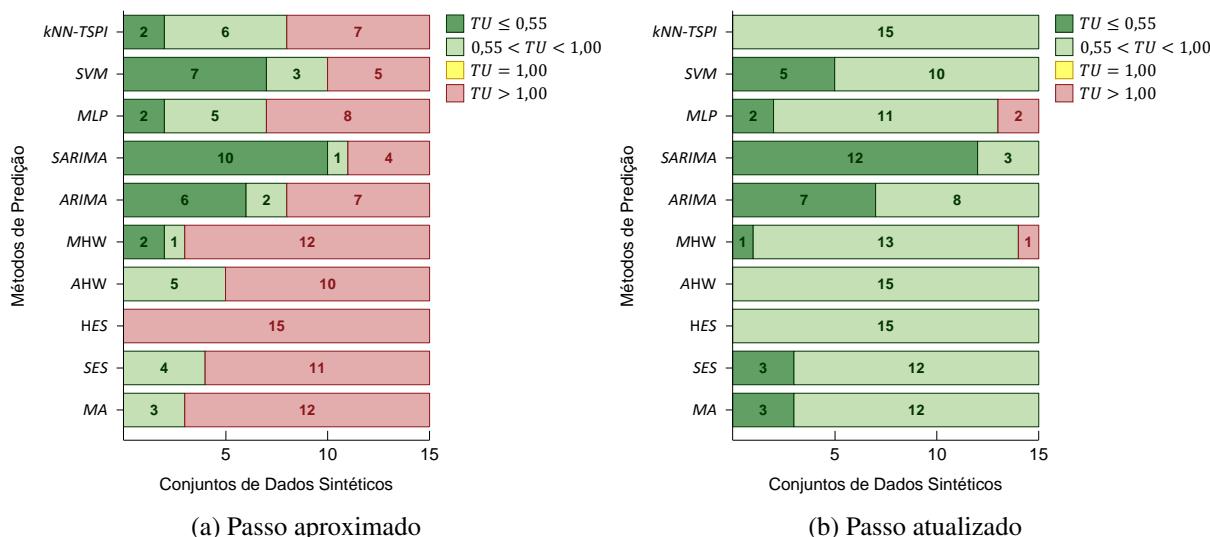
Na Figura 67, o algoritmo *kNN-TSPI* com passo aproximado ocupou a terceira posição no ranking dos valores de *MSE* (Figura 67a) e *TU* (Figura 67c). Tal configuração perdeu, nessa mesma perspectiva e por uma pequena margem de diferença, para os modelos *SARIMA* e *SVM*. Particularmente, o *SVM* com passo aproximado apresentou, em média, as melhores taxas de acerto *POCID* (Figura 67e), 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* (Fi-

gura 67b) quanto pelo coeficiente *TU* (Figura 67d). 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*.

Na Figura 68, as faixas de valores decorrentes da aplicação do coeficiente *TU* evidenciam que o *SARIMA*, com passos aproximado e atualizado, incidiu no método mais promissor para predizer ST com comportamento estocástico. Esse fato pode ser explicado pela própria estrutura do *SARIMA*, a qual frequentemente inclui um procedimento de Médias Móveis que abrange estimativas do fator de inovação (ruído branco) que não pode ser explicado pelo modelo.

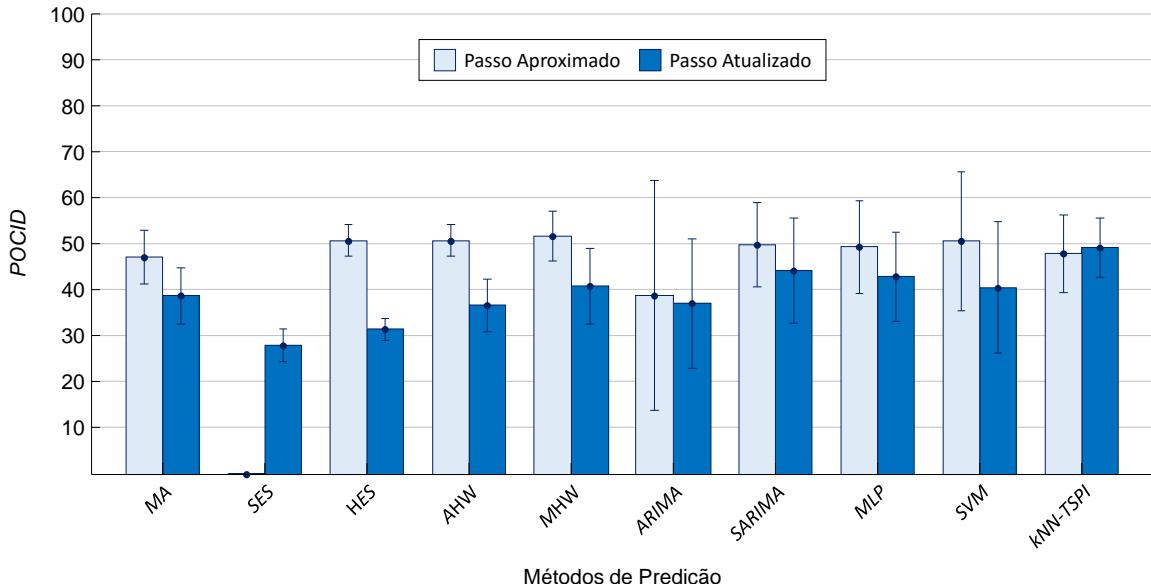
Figura 68 – Desempenho dos métodos de predição para quatro faixas de valores do coeficiente *TU* em ST estocásticas



Fonte: Elaborada pelo autor.

As médias e os desvios padrão das taxas de acerto *POCID* são esquematizados, para cada algoritmo e estratégia de predição, na Figura 69. Nessa figura, examinando a estratégia de predição multi-etapa à frente com passo aproximado, as maiores taxas de acerto foram atingidas pelos métodos *AHW* (acerto médio de 50,73% com desvio padrão de 3,51%) e *MHW* (acerto médio de 51,71% com desvio padrão de 5,39%). O *kNN-TSPI* exibiu o sétimo melhor resultado, isto é, uma taxa média de acerto equivalente à 47,83% com desvio padrão de 8,46%. Por outro lado, quando empregada a estratégia de predição multi-etapa à frente com passo atualizado, o algoritmo baseado em similaridade com invariâncias demonstrou a maior taxa média de acerto quanto à tendência dos horizontes projetados (49,20% com desvio padrão de 6,47%). Os piores valores de *POCID* foram procedentes da aplicação dos modelos *SES* (acerto médio de 27,94% com desvio padrão de 3,52%), *HES* (taxa média de acerto equivalente à 31,47% com desvio padrão de 2,41%) e *ARIMA* (acerto médio de 37,02% com desvio padrão de 14,09%).

Figura 69 – Médias e desvios padrão das taxas de acerto *POCID* obtidas pelos métodos de predição em ST estocásticas



Fonte: Elaborada pelo autor.

É relevante frisar que o *ARIMA* e o *SARIMA* costumam apresentar um desempenho melhor que os algoritmos de suavização exponencial quando a ST é relativamente longa e “bem comportada”. Se a série é muito irregular, os resultados dos modelos da categoria *ARIMA* são, normalmente, inferiores aos obtidos pelos métodos de suavização exponencial.

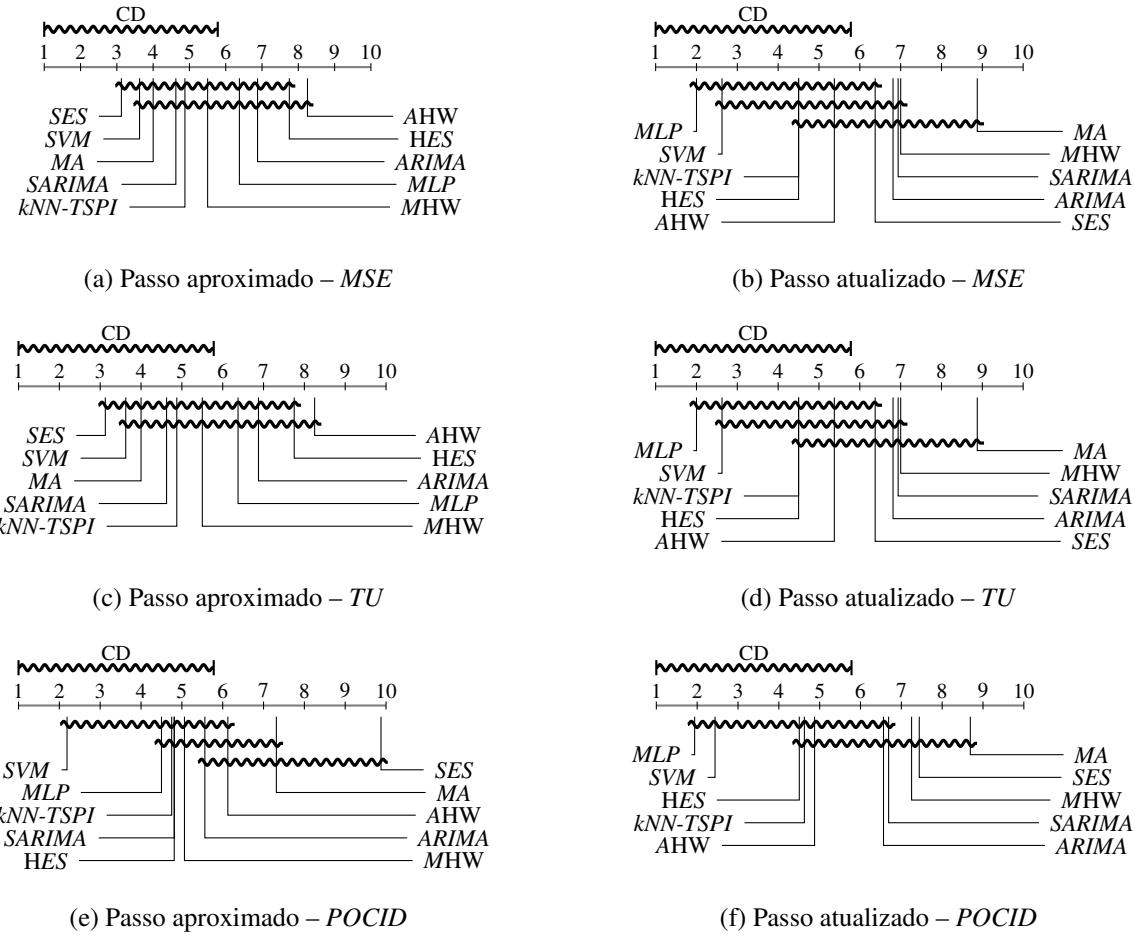
8.3.1.3 Séries Caóticas

Os experimentos conduzidos sobre as sequências de dados caóticas contemplaram 160 configurações (10 modelos preditivos \times 2 estratégias de predição \times 8 conjuntos de dados). Conforme retratado nos diagramas de distância crítica da Figura 70, 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 70a) e *TU* (Figura 70c). 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 70e).

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 70b, 70d e 70f), 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 70, o desempenho do algoritmo não-paramétrico *SVM* foi, em geral, mais estável que os outros métodos para predizer ST caóticas.

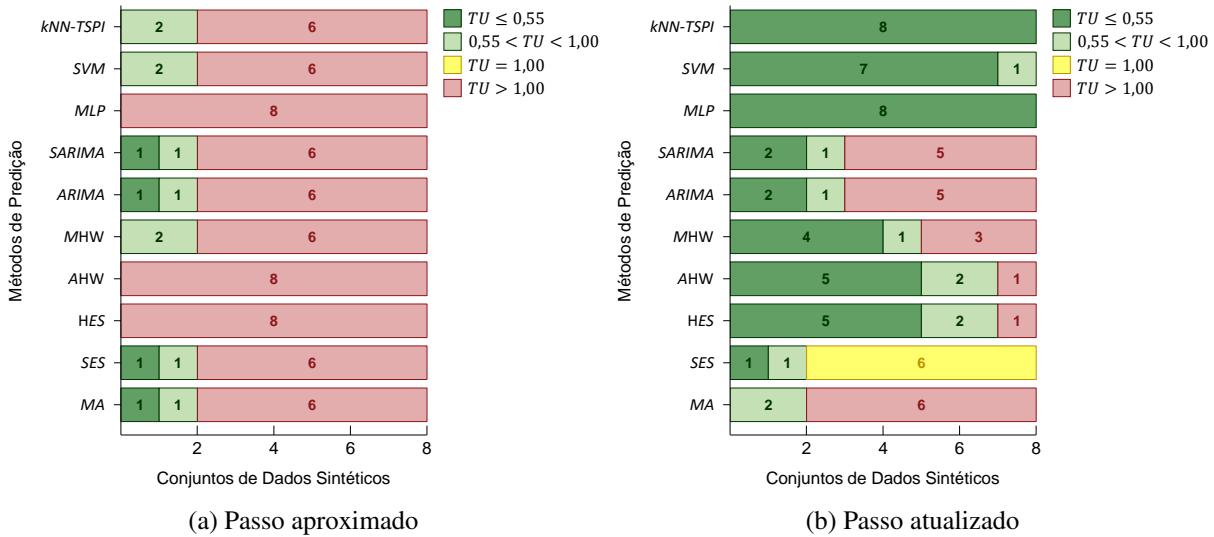
Figura 70 – Diagramas de distância crítica para os valores dos índices *MSE*, *TU* e *POCID* provenientes dos métodos de predição sobre ST caóticas



Fonte: Elaborada pelo autor.

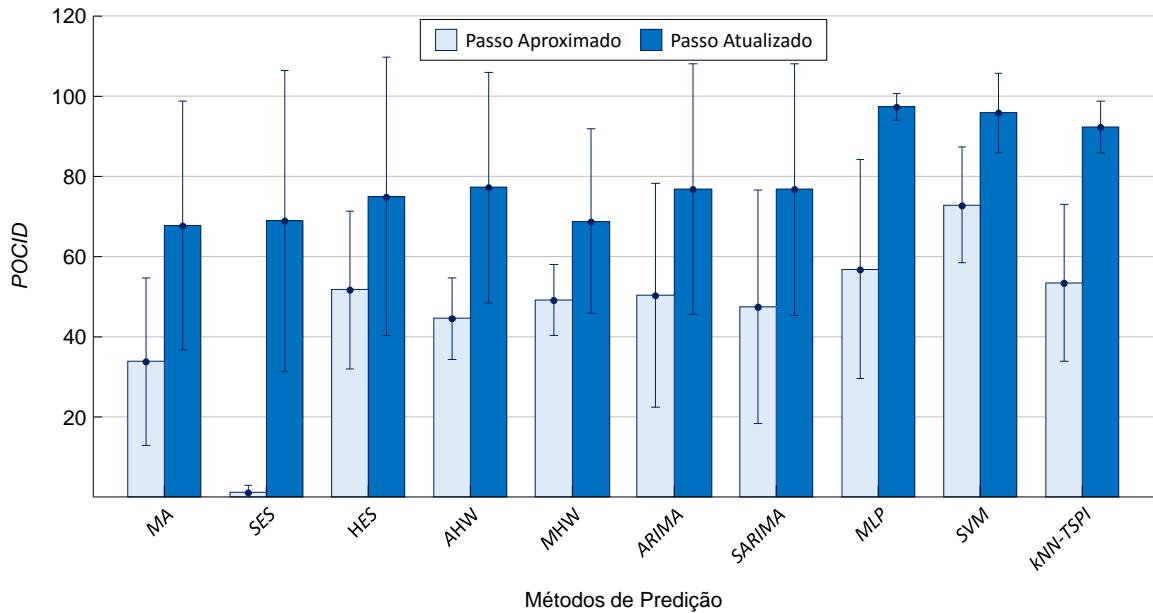
Na [Figura 71](#), as estatísticas decorrentes da aplicação do coeficiente *TU* demonstram que, utilizando a estratégia de predição multi-etapa à frente com passo aproximado ([Figura 71a](#)), os modelos preditivos foram convenientes para predizer, aproximadamente, dois total de oito conjuntos de dados ($TU < 1$). Nesse cenário, os algoritmos *HES*, *AHW* e *MLP* obtiveram, para todos os conjuntos de dados, desempenho inferior ao do modelo trivial ($TU > 1$). Tais resultados reforçam a dificuldade em se predizer ST caóticas, especialmente quando a estratégia de projeção empregada oportuniza a propagação do erro no horizonte de predição.

Adotando a estratégia de predição multi-etapa à frente com passo atualizado ([Figura 71b](#)), o uso dos métodos *MLP*, *SVM* e *kNN-TSPI* foi, em geral, confiável para modelar sete do total de oito conjunto de dados ($TU \leq 0,55$). Esse fato evidencia que, para ST caóticas, os algoritmos de predição mais promissores são os oriundos da subárea de Aprendizado de Máquina. Como esses modelos são não-paramétricos, ou seja, eles não assumem que os dados seguem uma distribuição específica, era esperado que os mesmos atingissem um desempenho superior ao dos métodos estatísticos.

Figura 71 – Desempenho dos métodos de predição para quatro faixas de valores do coeficiente TU em ST caóticas

Fonte: Elaborada pelo autor.

Na [Figura 72](#), para cada algoritmo e estratégia de predição, é apresentada a média dos valores da medida $POCID$ com a indicação dos desvios padrão.

Figura 72 – Médias e desvios padrão das taxas de acerto $POCID$ obtidas pelos métodos de predição em ST caóticas

Fonte: Elaborada pelo autor.

Nessa figura, analisando a estratégia de predição multi-etapa à frente com passo aproximado, o *SVM* exibiu os melhores valores de $POCID$ (acerto médio de 72,90% com desvio padrão de 14,47%), enquanto que os métodos *MA* (acerto médio de 33,76% com desvio padrão de 20,85%) e *HES* (acerto médio de 0,98% com desvio padrão de 1,62%) implicaram nas piores taxas de certo sobre as tendências dos horizontes projetados. Quanto à estratégia de predição

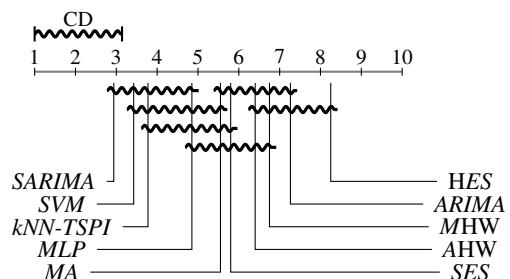
multi-etapa à frente com passo atualizado, o algoritmo *MLP* registrou as melhores taxas de acerto *POCID* (acerto médio de 97,41% com desvio padrão de 3,30%). Já o modelo *MA* manteve o pior desempenho global (acerto médio de 67,78% com desvio padrão de 31,16%).

Em síntese, os valores de *POCID* retratados graficamente na Figura 72 refletem as conclusões adquiridas por meio da execução do teste estatístico de Friedman com pós-teste de Nemenyi (Figura 70), as quais sugerem o *SVM* como candidato à modelo mais apropriado para predizer ST caóticas.

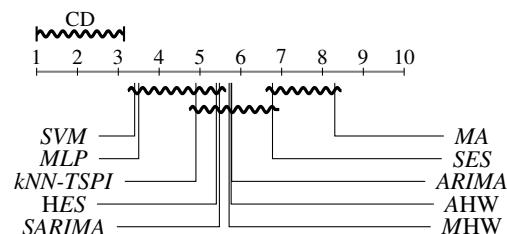
8.3.1.4 Comparação Geral

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 73.

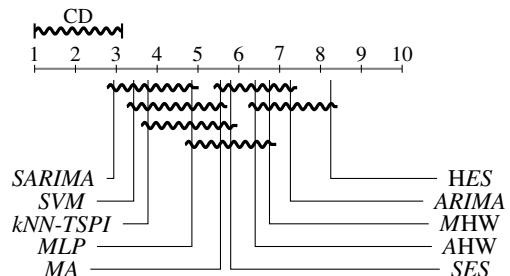
Figura 73 – Diagramas de distância crítica para os valores dos índices *MSE*, *TU* e *POCID* provenientes dos métodos de predição sobre ST sintéticas



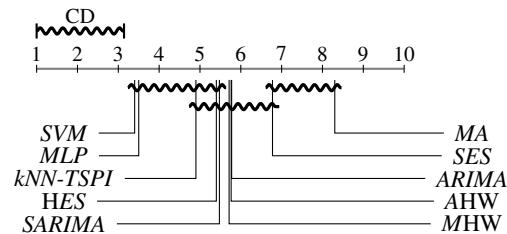
(a) Passo aproximado – *MSE*



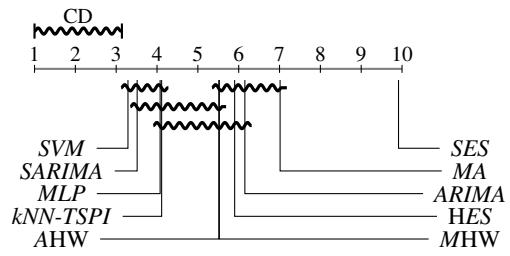
(b) Passo atualizado – *MSE*



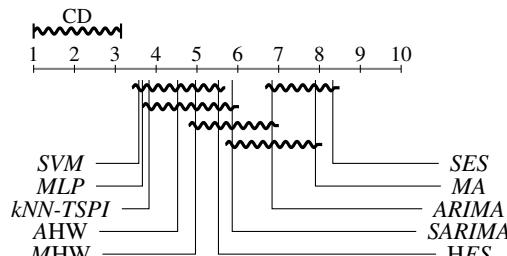
(c) Passo aproximado – *TU*



(d) Passo atualizado – *TU*



(e) Passo aproximado – *POCID*



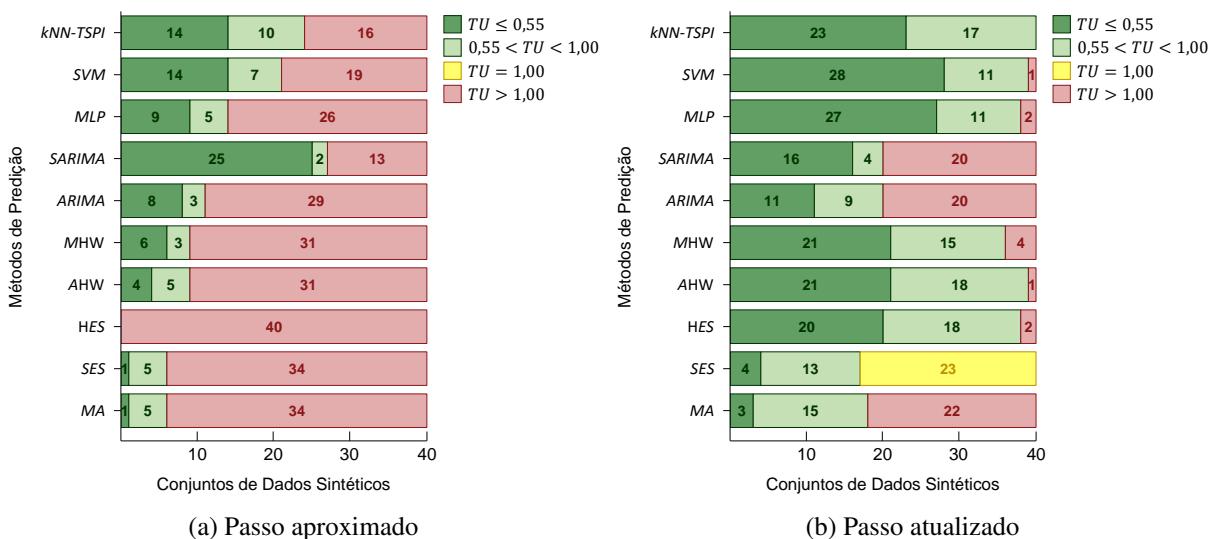
(f) Passo atualizado – *POCID*

Fonte: Elaborada pelo autor.

Na [Figura 73](#), o modelo *SARIMA* com passo aproximado apresentou os melhores resultados utilizando as medidas *MSE* ([Figura 73a](#)) e *TU* ([Figura 73c](#)). Por outro lado, quando analisadas as taxas de acerto sobre as tendências dos horizontes projetados ([Figura 73e](#)), 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, a primeira, segunda e terceira posições nos *rankings* derivados dos índices *MSE* ([Figura 73b](#)), *TU* ([Figura 73d](#)) e *POCID* ([Figura 73f](#)).

Examinando simultaneamente as duas estratégias de projeção da [Figura 73](#), 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*. Esse fato fica ainda mais evidente na [Figura 74](#), a qual aborda as faixas de valores do coeficiente *TU*.

Figura 74 – Desempenho dos métodos de predição para quatro faixas de valores do coeficiente *TU* em ST sintéticas



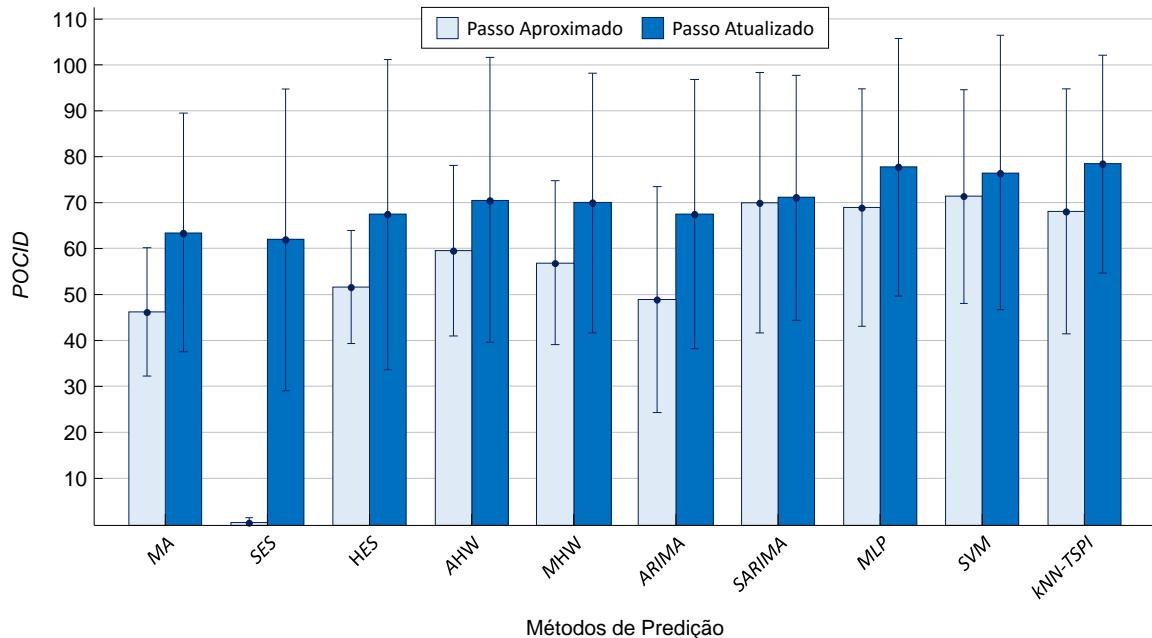
Fonte: Elaborada pelo autor.

Na [Figura 74](#), analisando os resultados computados por meio da aplicação da estratégia de predição multi-etapa à frente com passo aproximado ([Figura 74a](#)), o uso do *SARIMA* foi adequado em 27 ($25 + 2$) conjuntos de dados ($TU < 1$), sendo que destes, 25 viabilizaram uma modelagem confiável para a realização de projeções futuras ($TU \leq 0,55$). Em contraste, o algoritmo *HES* não conseguiu superar o desempenho do método ingênuo para nenhum dos 40 conjuntos de dados. A utilização dos modelos *MLP*, *SVM* e *kNN-TSPI*, porém adotando a estratégia de predição multi-etapa à frente com passo atualizado ([Figura 74b](#)), foram apropriadas, em média, para 39 conjuntos de dados ($TU < 1$), dos quais 26 acarretaram em modelos preditivos

confiáveis ($TU \leq 0,55$).

As médias e os desvios padrão das taxas de acerto *POCID* são esquematizados, para cada algoritmo e estratégia de predição, na [Figura 75](#).

Figura 75 – Médias e desvios padrão das taxas de acerto *POCID* obtidas pelos métodos de predição em ST sintéticas



Fonte: Elaborada pelo autor.

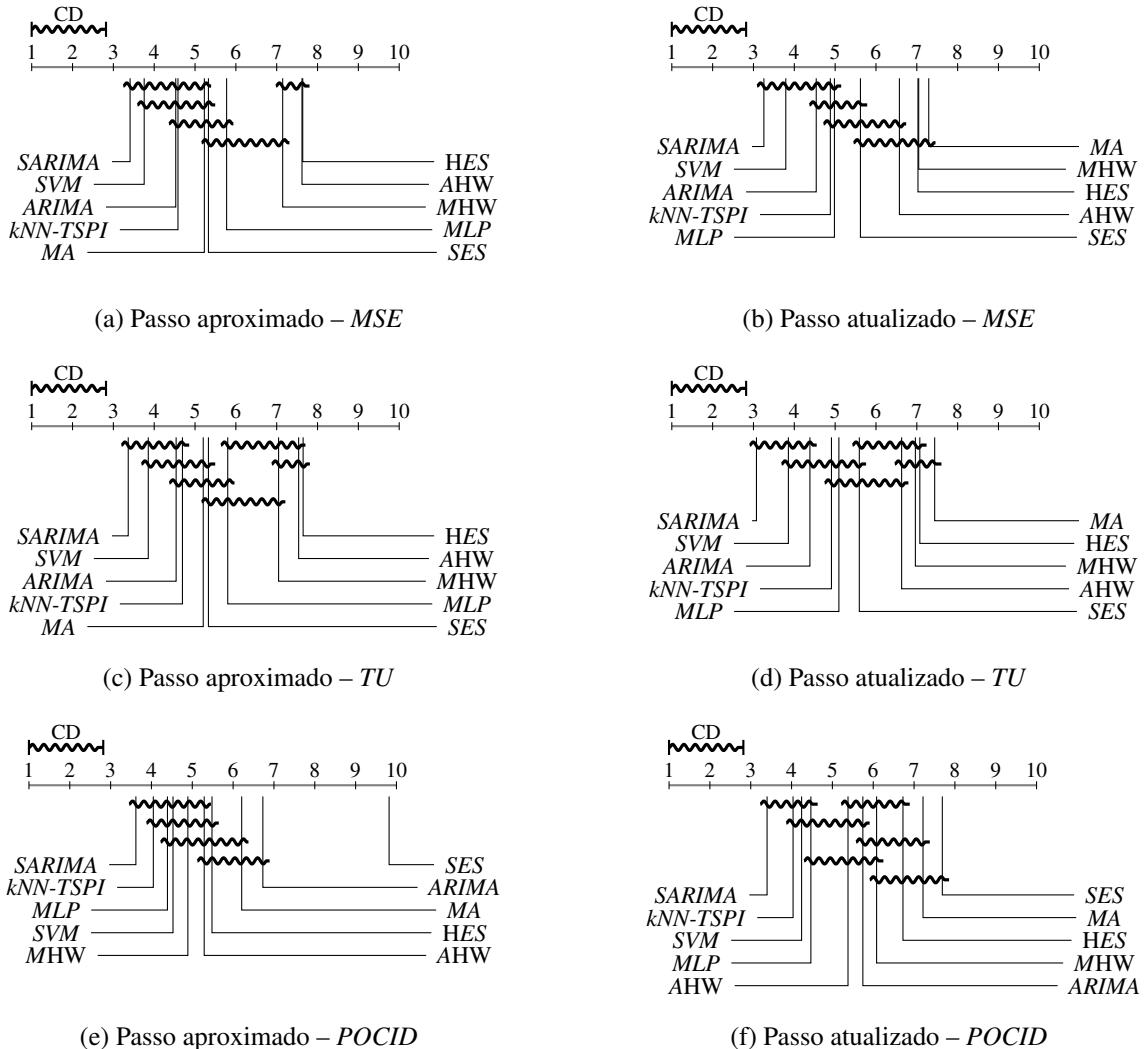
Como pode ser observado na [Figura 75](#), independentemente da estratégia de projeção adotada, os métodos de Aprendizado de Máquina (*MLP*, *SVM* e *kNN-TSPI*) conseguiram acertar com aproximadamente 73,55% de precisão, e desvio padrão de 26,19%, as tendências dos horizontes de predição.

8.3.2 Séries Reais

Os experimentos realizados sobre as sequências de dados reais totalizaram 1100 configurações (10 modelos preditivos \times 2 estratégias de predição \times 55 conjuntos de dados). Na [Figura 76](#) é apresentado o diagrama de distância crítica com relação aos valores das medidas *MSE*, *TU* e *POCID* decorrentes das 1100 configurações pesquisadas.

Na [Figura 76](#), considerando ambas as estratégias de projeção, o *SARIMA* acarretou, segundo as três medidas de desempenho, nos melhores resultados. Diferentemente, os modelos baseados na técnica de suavização exponencial simples exibiram os piores desempenhos. Não foi possível verificar diferenças estatisticamente significativas entre o algoritmo *kNN-TSPI* e o *SARIMA*. Isso permite aceitar como verdadeira a hipótese de que métodos guiados por similaridade proporcionam resultados competitivos quando comparados com os métodos estatísticos estado-da-arte na literatura.

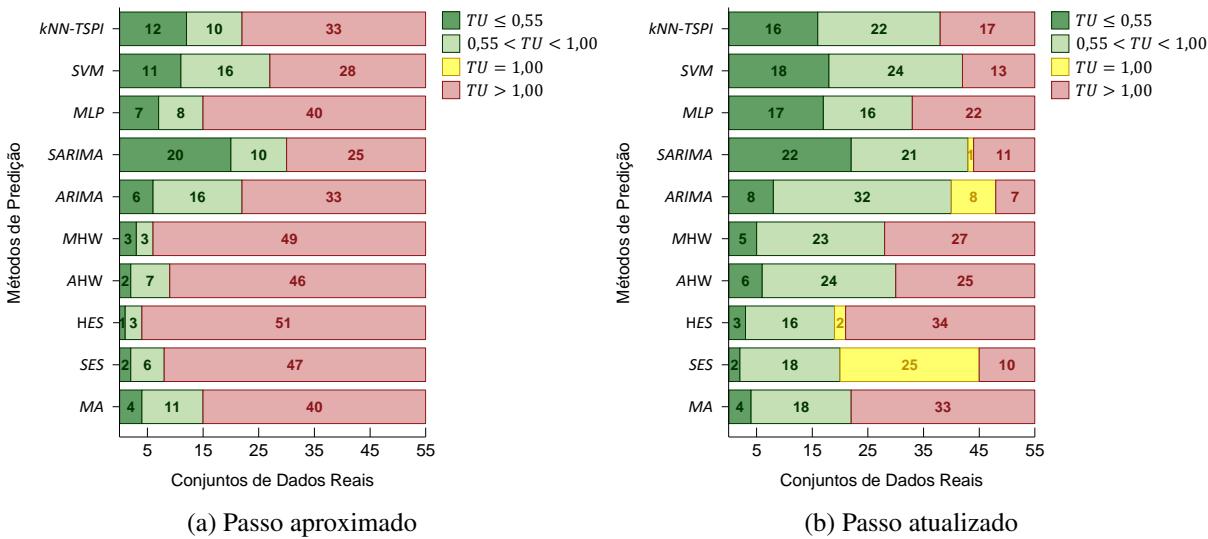
Figura 76 – Diagramas de distância crítica para os valores dos índices MSE , TU e $POCID$ provenientes dos métodos de predição sobre ST reais



Fonte: Elaborada pelo autor.

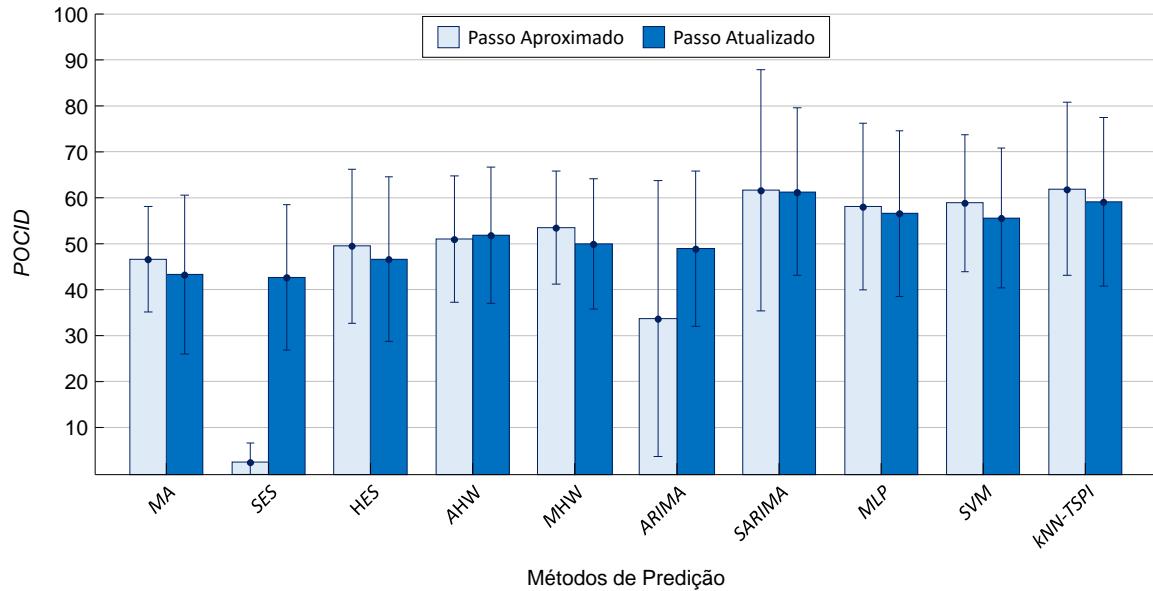
As quatro faixas de valores do coeficiente TU indicadas na Figura 77 demostram que, utilizando a estratégia de predição multi-etapa à frente com passo aproximado (Figura 77a), o modelo *SARIMA* foi adequado para predizer 30 ($20 + 10$) do total de 55 conjuntos de dados ($TU < 1$). O algoritmo *SVM* foi apropriado para modelar 27 ($11 + 16$) ST, enquanto que a utilização dos métodos *ARIMA* e *kNN-TSPI* foi preferível em relação ao modelo trivial para 22 ($12 + 10$) do total de 55 conjuntos de dados.

Quanto à estratégia de predição multi-etapa à frente com passo atualizado (Figura 77b), o uso do *SARIMA* foi conveniente para predizer 43 ($22 + 21$) conjuntos de dados ($TU < 1$), sendo que destes, 22 oportunizaram uma modelagem confiável para a realização de projeções futuras ($TU \leq 0,55$). Ainda nesse cenário, o *SVM* e o *kNN-TSPI* foram propícios ($TU < 1$) para predizer os valores de 42 ($18 + 24$) e 38 ($16 + 22$) conjunto de dados, respectivamente.

Figura 77 – Desempenho dos métodos de predição para quatro faixas de valores do coeficiente *TU* em ST reais

Fonte: Elaborada pelo autor.

Na [Figura 78](#), para cada algoritmo e estratégia de predição, é apresentada a média dos valores da medida *POCID* com a indicação dos desvios padrão.

Figura 78 – Médias e desvios padrão das taxas de acerto *POCID* obtidas pelos métodos de predição em ST reais

Fonte: Elaborada pelo autor.

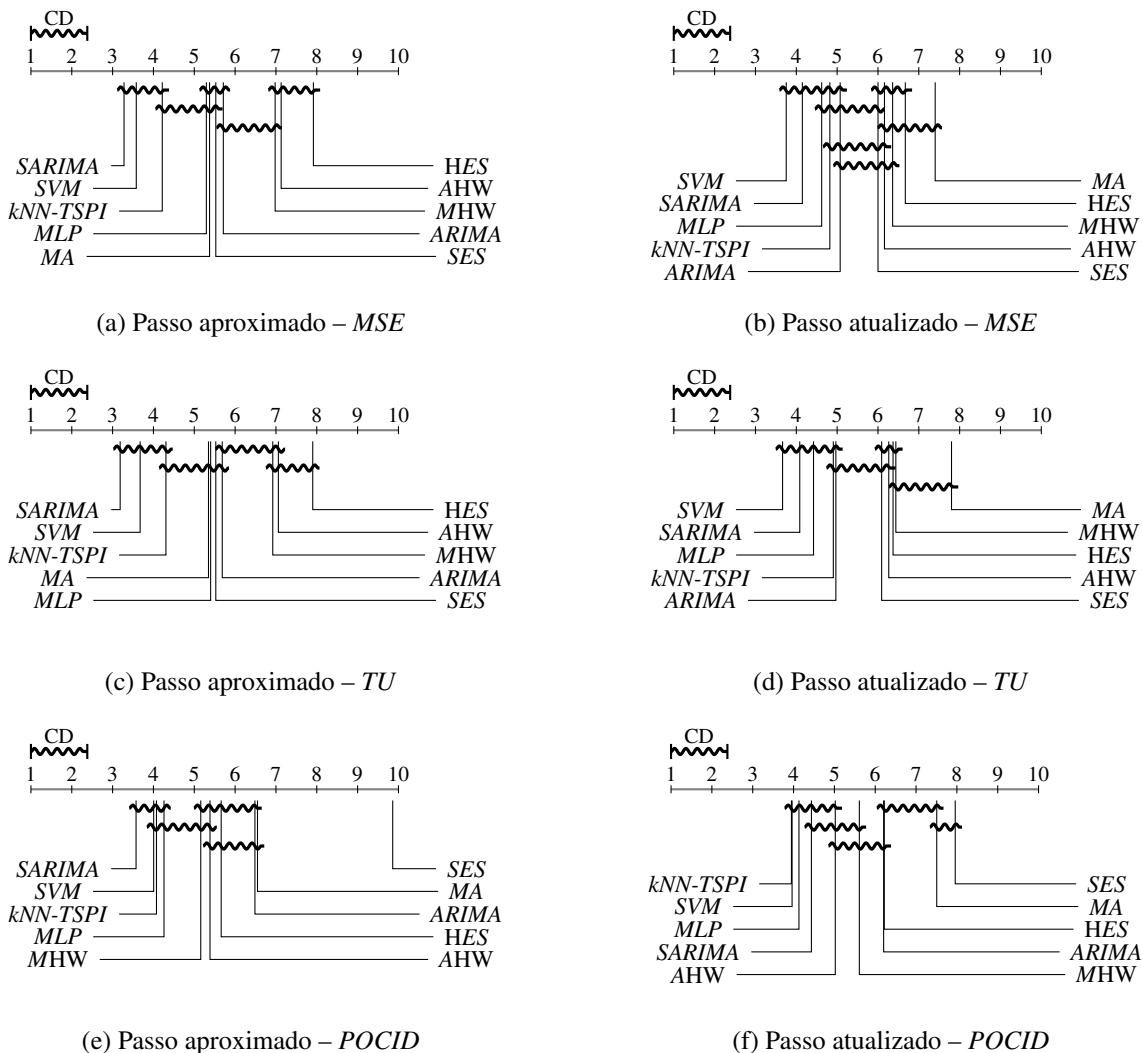
Na [Figura 78](#), a partir do emprego da estratégia de predição multi-etapa à frente com passo aproximado, o *kNN-TSPI* obteve uma taxa média de acerto sobre as tendências dos horizontes de predição equivalente à 61,86%, com desvio padrão de 18,83%, ultrapassando o desempenho dos métodos *SARIMA* (acerto médio de 61,64% com desvio padrão de 26,29%) e *MLP* (acerto médio de 58,07% com desvio padrão de 18,04%). Examinando a estratégia de predição multi-etapa à frente com passo atualizado, os maiores valores de *POCID* foram

provenientes da aplicação dos algoritmos *SARIMA* (taxa média de acerto de 61,27% com desvio padrão de 18,20%) e *kNN-TSPI* (taxa média de acerto de 59,11% com desvio padrão de 18,34%).

8.3.3 Comparação Geral

Os experimentos conduzidos sobre as sequências de dados sintéticas e reais envolveram 1900 configurações (10 modelos preditivos \times 2 estratégias de predição \times 95 conjuntos de dados). Nos diagramas de distância crítica da [Figura 79](#) é possível notar o quanto competitivo são os métodos de Aprendizado de Máquina, em especial o algoritmo proposto neste trabalho, em relação aos modelos estado-da-arte *ARIMA* e *SARIMA*.

[Figura 79](#) – Diagramas de distância crítica para os valores dos índices *MSE*, *TU* e *POCID* provenientes dos métodos de predição sobre ST sintéticas e reais

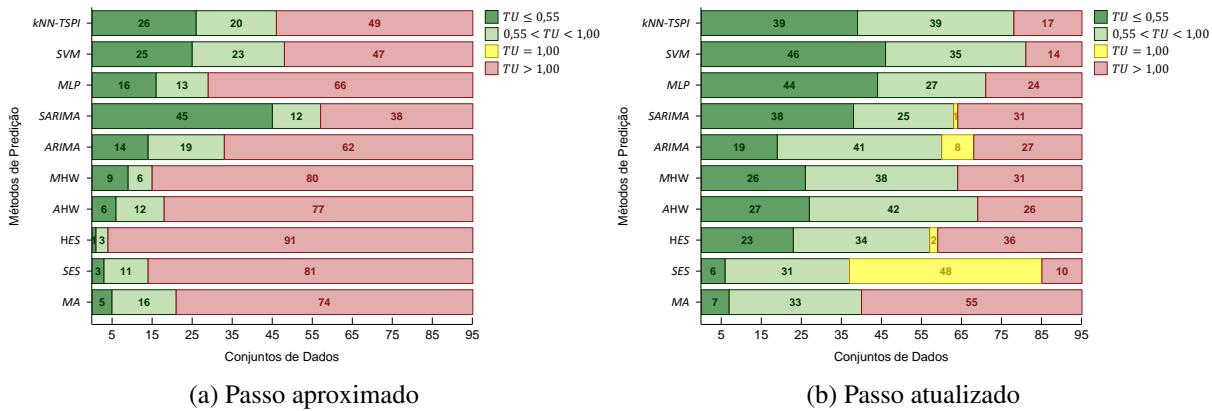


Fonte: Elaborada pelo autor.

Na [Figura 80](#), as estatísticas decorrentes da aplicação do coeficiente *TU* demonstram que, por meio da estratégia de predição multi-etapa à frente com passo aproximado ([Figura 80a](#)), o

SARIMA foi preferível ao modelo ingênuo em 57 ($45 + 12$) do total de 95 conjuntos de dados ($TU < 1$). O desempenho dos métodos *SVM* e *kNN-TSPI* foram melhores ($TU < 1$) que o modelo trivial em 48 ($25 + 23$) e 46 ($26 + 20$) conjuntos de dados, respectivamente. Analisando a estratégia de predição multi-etapa à frente com passo atualizado, o *SVM* e o *kNN-TSPI* foram os algoritmos que demonstraram os menores erros de projeção para, aproximadamente, 80 conjuntos de dados.

Figura 80 – Desempenho dos métodos de predição para quatro faixas de valores do coeficiente TU em ST sintéticas e reais



Fonte: Elaborada pelo autor.

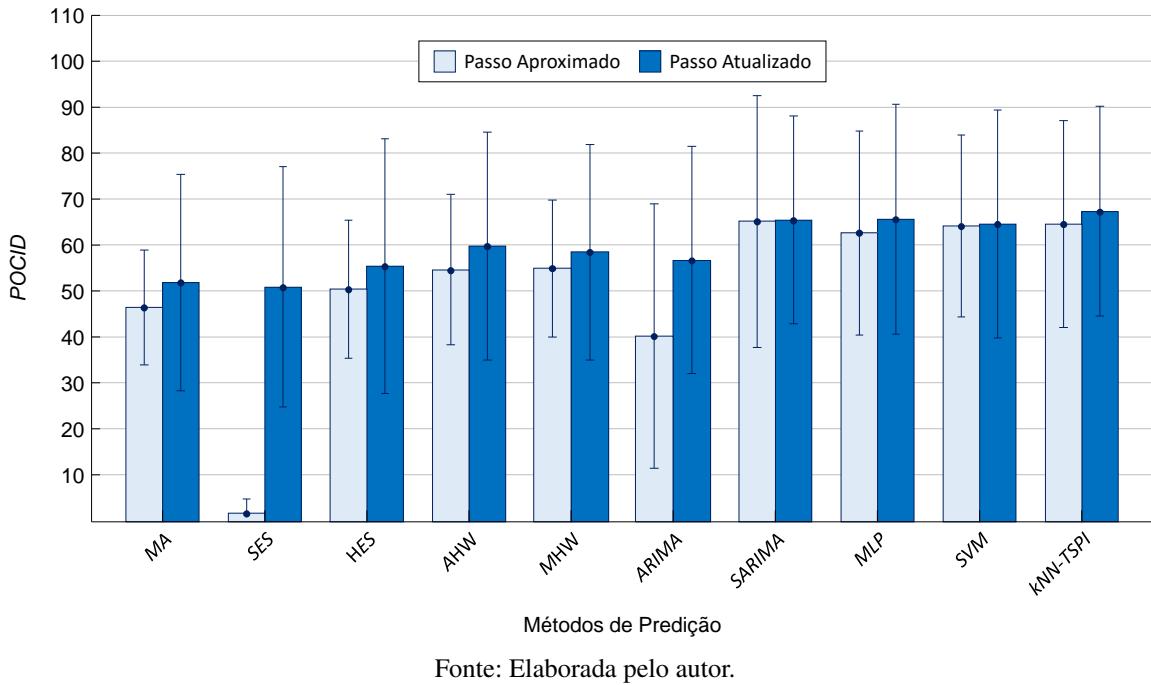
Na [Figura 81](#), a partir da representação gráfica das médias e dos desvios padrão dos valores de *POCID*, verifica-se que as taxas de acerto sobre as tendências dos horizontes de predição estão levemente distribuídas entre os modelos *SARIMA*, *MLP*, *SVM* e *kNN-TSPI*. Cada um desses algoritmos, independentemente da estratégia de predição empregada, emplacou uma taxa média de acerto de, aproximadamente, 64,89% com desvio padrão de 23,33%.

8.4 Considerações Finais

Neste capítulo foram relatados os resultados de uma ampla análise experimental envolvendo a aplicação de dez métodos de predição ST sobre 95 conjuntos de dados, dos quais 40 são de origem artificial (sintética) e 55 são provenientes de eventos do mundo real. Os conjuntos de dados sintéticos construídos neste trabalho foram agrupados, conforme seu processo originário, em três categorias: determinística, estocástica e caótica.

Para as ST determinísticas, o modelo *SARIMA* forneceu os melhores resultados quando usada 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.

Figura 81 – Médias e desvios padrão das taxas de acerto *POCID* obtidas pelos métodos de predição em ST sintéticas e reais



Fonte: Elaborada pelo autor.

Considerando as ST estocásticas, o *SARIMA* com passos aproximado e atualizado demonstrou ser o modelo de predição mais promissor. Esse fato pode ser esclarecido tendo em vista a estrutura do *SARIMA*, 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 de Aprendizado de Máquina *MLP*, *SVM* e *kNN-TSPI* obtiveram os melhores desempenhos. Como esses modelos são não-paramétricos, isto é, não fazem suposições acerca da distribuição dos dados, era esperado que eles apresentassem um desempenho superior ao dos métodos estatísticos.

Analizando 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 à 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*.

Examinando os 55 conjuntos de dados reais, o *kNN-TSPI* expôs o quarto melhor resultado em termos de desempenho preditivo e ficou em segundo lugar quando avaliada as taxas de acerto sobre tendência dos horizontes projetados. Não foram verificadas diferenças estatisticamente significativas entre o algoritmo *kNN-TSPI* e os modelos *SVM* e *SARIMA*.

Os resultados decorrentes da comparação geral, isto é, considerando os 95 conjuntos de dados (40 ST sintéticas e 55 ST reais), indicaram que os métodos *SARIMA*, *MLP*, *SVM* e *kNN-TSPI* são os mais promissores para a predição de valores em dados temporais. Esses

resultados estão em conformidade com a hipótese inicial levantada nesta dissertação de mestrado, isto é, a de que o algoritmo *kNN-TSPI* é capaz de prover resultados comparáveis com os métodos estado-da-arte. O algoritmo baseado em similaridade com invariâncias, proposto neste trabalho, ainda contempla as vantagens de ser fácil de implementar e requerer a estimação de apenas dois parâmetros.

No próximo capítulo são apresentadas as conclusões desta dissertação de mestrado, bem como as principais limitações e trabalhos futuros.



CONCLUSÃO

Um dos principais objetivos dos modelos de predição é a redução da incerteza futura, sobretudo devido à volatilidade de determinados fenômenos. Esses modelos evoluíram no decorrer dos anos, passando de simples técnicas de regressão para métodos robustos provenientes da Estatística e da Inteligência Artificial.

Nesta dissertação de mestrado foram pesquisadas adaptações de algoritmos provenientes da subárea de Aprendizado de Máquina (AM) para o problema da predição de Séries Temporais (ST). Especificamente, o interesse se concentrou em métodos baseados em similaridade que buscam, 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 usam os valores seguintes dessas subsequências como entrada para uma função de predição, a qual realiza o cálculo do valor futuro.

A motivação e justificativa para investigar algoritmos por similaridade incidiu no fato de que pesquisas empíricas têm demonstrado que variações desses métodos proporcionam resultados muito competitivos, frequentemente superando métodos mais complexos.

Por meio de uma revisão sistemática, seguida de uma meta-análise da literatura, foi possível detectar uma importante lacuna de pesquisa. Essa lacuna refere-se à comparação objetiva e subjetiva entre modelos estatísticos e de AM para a predição de ST. Nesse contexto, os objetivos desta dissertação de mestrado abrangeu duas temáticas principais: (1) a exploração das propriedades inerentes à predição baseada em similaridade e (2) a comparação do algoritmo *k-Nearest Neighbors - Time Series Prediction with Invariances* (*kNN-TSPI*), aqui proposto, com os métodos estatísticos e de Aprendizado de Máquina considerados o estado-da-arte na literatura. Todos esses algoritmos foram executados utilizando a estratégia de predição multi-etapa à frente com passos aproximado e atualizado.

O algoritmo *kNN-TSPI* difere da versão convencional pela incorporação de três técnicas para obtenção de invariância à amplitude e deslocamento, invariância à complexidade e trata-

mento de casamentos triviais. Como demonstrado, o uso simultâneo dessas técnicas proporciona uma melhor correspondência entre as subsequências de dados e a consulta de referência.

A primeira avaliação experimental conduzida contemplou a aplicação dos algoritmos *kNN*, *kNN-TSP* e *kNN-TSPI* sobre 55 conjuntos de dados reais. Especificamente para o *kNN-TSPI*, foram investigadas diferentes invariâncias às distorções em dados temporais, medidas de distância, medidas de complexidade aplicadas à *CID* e funções de predição. Os resultados gerados empiricamente foram apresentados e discutidos considerando os seguintes aspectos:

1. Impacto do uso de invariâncias às distorções em ST no desempenho do algoritmo *kNN-TSP*

kNN-TSP – Foram pesquisadas cinco técnicas para obtenção de invariância aos seguintes problemas indesejáveis em dados temporais: deslocamento, amplitude e deslocamento, escala local, complexidade e casamentos triviais. Essas técnicas foram combinadas alternadamente e acopladas ao algoritmo *kNN-TSP*, totalizando 1870 configurações (1 modelo preditivo \times 17 tipos de invariância \times 2 estratégias de projeção \times 55 conjuntos de dados). Os resultados dessas configurações, além de comparados entre si, foram também confrontados com os resultados obtidos pelo *kNN-TSP* sem técnicas invariantes (Original). O método *kNN-TSP* com invariância à amplitude e deslocamento, complexidade e tratamento de casamentos triviais exibiu, em média, os melhores resultados, porém não foi estatisticamente melhor que o *kNN-TSP* em sua versão original;

2. Estudo comparativo da relação custo-benefício dos métodos *kNN*, *kNN-TSP* e *kNN-TSPI*

***TSP* –** No intuito de firmar a relação custo-benefício dos algoritmos baseados em similaridade pesquisados, estes foram confrontados em termos de desempenho preditivo. Tais comparações totalizaram 330 configurações (3 modelos preditivos \times 2 estratégias de projeção \times 55 conjuntos de dados). O algoritmo *kNN-TSPI* registrou, em geral, os melhores resultados;

3. Análise do emprego de distintas medidas de similaridade na predição de ST utilizando o algoritmo *kNN-TSPI*

***TSP* –** A fim de verificar a influência que a medida de similaridade exerce sobre o desempenho do algoritmo *kNN-TSPI*, foram utilizadas e comparadas 25 medidas de similaridade, totalizando 2750 configurações (1 modelo preditivo \times 25 medidas de distância \times 2 estratégias de projeção \times 55 conjuntos de dados). As métricas Manhattan e Canberra forneceram, em média, os menores erros de predição, mas não apresentaram diferenças estatisticamente significativas quando confrontadas com as distâncias comumente aplicadas na literatura, tais como Euclidiana, Cosseno, Geodésica, *Complexity Invariant Distance (CID)*, *Dynamic Time Warping (DTW)*, *CIDDTW* e *DTW-D*;

4. Investigação do uso de medidas de complexidade aplicadas à medidas de distância invariantes à complexidade no processo de busca adotado pelo método *kNN-TSPI* –

Neste trabalho foram pesquisadas, além da estimativa de complexidade usada pela *CID* em sua formalização convencional, cinco diferentes medidas de complexidade: Diferença

Absoluta, Compressão, *Edges*, *Zero-crossings* e Entropia de Permutação. Essas estimativas de complexidade foram aplicadas à medida *CID*, na intenção de verificar o impacto dessas combinações no processo de busca por similaridade adotado pelo *kNN-TSPI*. As referidas combinações resultaram em 660 configurações (1 modelo preditivo \times 6 medidas de complexidade \times 2 estratégias de projeção \times 55 conjuntos de dados). Pela simplicidade de codificação e por acarretar nas melhores taxas de acerto sobre as tendências futuras, recomenda-se a utilização da Diferença Quadrática cujo desempenho foi previamente verificado em [Parmezan e Batista \(2015\)](#);

5. Avaliação da influência de diversas funções de predição na qualidade das projeções realizadas pelo algoritmo *kNN-TSPI* – Foram analisadas quatro funções de predição: Mediana; Média de Valores Absolutos (MVA); Média de Valores Relativos (MVR); *Index Weighted (IW)* e *Distance Weighted (DW)*. Para este último caso, foram investigadas seis variações de pesos. Nota-se que as combinações realizadas resultaram em 1100 configurações (1 modelo preditivo \times 10 funções de predição \times 2 estratégias de projeção \times 55 conjuntos de dados). A MVR apresentou, em geral, o melhor desempenho. Apesar disso, a *IW* demonstrou ser uma boa candidata à função de predição, visto que atribui pesos maiores para as observações mais recentes;

Ao longo da execução dos experimentos supracitados houve a construção de um portal *Web* ([PARMEZAN; BATISTA, 2014](#)), intitulado de ICMC-USP *Time Series Prediction Repository*, que concede acesso aos materiais produzidos e também aos usados no decorrer das atividades contempladas por esta pesquisa. Atualmente, o repositório mantém 100 conjuntos de dados temporais (40 ST sintéticas e 60 ST reais) de acesso público e destinados às comunidades de Estatística e AM.

No segundo conjunto de experimentos, considerando dados sintéticos e reais, foi realizada uma das mais extensas e imparciais avaliações experimentais já idealizadas no tema de predição de ST. Essa avaliação foi conduzida com o propósito de evidenciar em que situações um algoritmo de predição supera o outro e quais aspectos dos dados têm maior influência no desempenho desses algoritmos. Para tanto, o *kNN-TSPI* foi confrontado com outros nove algoritmos: Máquinas de Suporte Vetorial (*SVM*), *Multilayer Perceptron (MLP)*, Médias Móveis (*MA*), Suavização Exponencial Simples (*SES*), Suavização Exponencial de Holt (*HES*), Holt-Winters Aditivo (*AHW*), Holt-Winters Multiplicativo (*MHW*), Autorregressivo Integrado de Médias Móveis (*ARIMA*) e Autorregressivo Integrado de Médias Móveis Sazonal (*SARIMA*). Os resultados desses testes computacionais foram mostrados e discutidos sob as seguintes perspectivas:

1. Estudo comparativo entre o *kNN-TSPI* e os modelos preditivos clássicos aplicados à dados temporais sintéticos – Os experimentos realizados nessa análise totalizaram 800 configurações (10 modelos preditivos \times 2 estratégias de predição \times 40 conjuntos de dados). Os menores erros de projeção foram fornecidos pelo modelo *SARIMA*, seguido do

algoritmo *SVM* e *kNN-TSPI*. O *kNN-TSPI* foi o mais adequado para predizer ST determinísticas, enquanto o modelo *SARIMA* foi o mais promissor para predizer ST estocásticas. Já para as ST caóticas, os algoritmos mais apropriados foram os de Aprendizado de Máquina, ou seja, *MLP*, *SVM* e *kNN-TSPI*;

- 2. Confronto do algoritmo *kNN-TSPI* com os métodos convencionais para projeção de valores sobre dados temporais reais** – Os testes computacionais conduzidos nessa avaliação resultaram em 1100 configurações (10 modelos preditivos \times 2 estratégias de predição \times 55 conjuntos de dados). O *kNN-TSPI* expôs o quarto melhor resultado em termos de desempenho preditivo e ficou em segundo lugar quando avaliada as taxas de acerto sobre a tendência dos horizontes projetados. Não foram verificadas diferenças estatisticamente significativas entre o algoritmo *kNN-TSPI* e os modelos *SVM* e *SARIMA*;
- 3. Análise do desempenho preditivo do *kNN-TSPI* em relação aos modelos de predição tradicionais aplicados à dados temporais sintéticos e reais** – Os experimentos realizados nesse estudo totalizaram 1900 configurações (95 conjuntos de dados \times 10 modelos de predição \times 2 estratégias de predição) e os resultados indicaram que os métodos *SARIMA*, *MLP*, *SVM* e *kNN-TSPI* são os mais propícios para a predição de ST.

Considerando os resultados da avaliação experimental com séries artificiais e reais, pode-se concluir que o *kNN-TPI* é competitivo com os métodos *SARIMA* e *SVM*. Por mais esses modelos caixa-preta tenham atingido uma precisão superior a do método baseado em similaridade, o algoritmo com invariâncias proposto é consideravelmente 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.

9.1 Limitações

No decorrer do desenvolvimento deste trabalho foram detectadas as seguintes limitações:

- A medida de erro *MAPE* possui um problema de divisão por zero e isso faz com que seu resultado fique, em determinadas situações, distantes dos valores de outras medidas de erro;
- Não foi utilizada uma medida multicritério de avaliação da qualidade de predição que leve em consideração os diversos índices de desempenho com características distintas;
- A implementação do *kNN-TSPI* se restringe à dados univariados.

Os resultados das previsões produzidas neste trabalho foram, inicialmente, analisados por meio do emprego da medida *MAPE*. No entanto, verificou-se que esta apresenta uma deficiência

prática: se uma ST possuir valores zero, ocorrerá uma imprópria divisão por zero. Quando comparados os valores de *MAPE* com os resultados de outras medidas de erro, estes mostraram-se razoavelmente discrepantes. Por esse motivo, os resultados das projeções passaram a ser verificados de acordo com o Erro Quadrático Médio (*MSE*) como uma alternativa à medida *MAPE*.

Durante a revisão sistemática foi observado que, ao contrário do que acontece em trabalhos na área de predição de curíssimo prazo, onde o uso do índice *MAPE* predomina, não há preferência entre os pesquisadores sobre qual medida de erro escolher. Isso sugere que mais de um índice deve ser levado em conta na avaliação de diferentes métodos de predição. Neste trabalho, a medida *MSE*, o coeficiente *U* de Theil (*TU*) e a taxa de acerto *Prediction of Change in Direction (POCID)* foram adotados para avaliação do desempenho dos modelos preditivos.

É importante destacar que a utilização de mais de uma medida de desempenho acarreta em análises extensas e, em alguns casos, de difícil interpretação. Portanto, a utilização de uma medida multicritério de desempenho para a avaliação conjunta dos índices individuais torna-se bastante útil.

O algoritmo *kNN-TSPI* implementado neste trabalho está preparado para lidar com dados univariados. Sendo assim, o algoritmo considera como entrada apenas os dados das ST, sem considerar possíveis variáveis explanatórias.

9.2 Trabalhos Futuros

Durante o desenvolvimento deste trabalho foram encontradas diversas questões de interesse que podem contribuir no seguimento deste trabalho. Algumas dessas questões, que podem estender a pesquisa realizada, são:

- A implementação do algoritmo *kNN-TSPI*, proposto neste trabalho, está limitada à dados univariados. Como mencionado no Capítulo 4, a adição de variáveis explanatórias pode auxiliar na exatidão de predição de dados temporais. Portanto, a adaptação do método por similaridade com invariâncias para dados multivariados é de grande interesse;
- A utilização de comitês de preditores para compor as previsões do algoritmo *kNN-TSPI* com o objetivo de melhorar a projeção final;
- A extensão dos métodos pesquisados para tratar dupla (diária e semanal) e tripla (diária, semanal e mensal) sazonalidade. Os métodos de predição investigados neste trabalho consideram uma única sazonalidade, ou seja, a sazonalidade que faz referência ao período de aquisição dos dados. No entanto, a consideração de outros períodos sazonais que estão inerentes ou que englobam essa sazonalidade pode acarretar na melhoria do desempenho dos algoritmos;

- O desenvolvimento de uma medida multicritério para a avaliação de distintos índices individuais de desempenho;
- A proposição de uma medida de erro que leve em consideração a disposição das observações no horizonte de predição;
- A paralelização do algoritmo *kNN-TSPI*, de maneira que seja possível tomar proveito de computadores de vários núcleos, ou mesmo, de *clusters* de computadores.

REFERÊNCIAS

- AGGARWAL, C. C.; HINNEBURG, A.; KEIM, D. A. On the surprising behavior of distance metrics in high dimensional space. In: **Database Theory**. London, United Kingdom: Springer Berlin Heidelberg, 2001, (Lecture Notes in Computer Science, v. 1973). p. 420–434. Citado na página 128.
- AHMED, N. K.; ATIYA, A. F.; GAYAR, N. E.; EL-SHISHINY, H. An empirical comparison of machine learning models for time series forecasting. **Econometric Reviews**, v. 29, n. 5-6, p. 594–621, 2010. Citado 2 vezes nas páginas 44 e 48.
- ALPAYDIN, E. **Introduction to machine learning**. Cambridge, United States of America: MIT Press, 2004. Citado 2 vezes nas páginas 108 e 109.
- AMAT, C.; MICHALSKI, T. K.; STOLTZ, G. Fundamentals and exchange rate forecastability with machine learning methods. **HEC Paris Research Paper No. ECO/SCD-2014-1049**, 2014. Citado na página 73.
- ANDRAWIS, R. R.; ATIYA, A. F.; EL-SHISHINY, H. Forecast combinations of computational intelligence and linear models for the NN5 time series forecasting competition. **International Journal of Forecasting**, v. 27, n. 3, p. 672–688, 2011. Citado na página 44.
- BACHE, K.; LICHMAN, M. **UCI machine learning repository**. 2013. 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 47.
- BANDT, C.; POMPE, B. Permutation entropy: A natural complexity measure for time series. **Physical Review Letters**, American Physical Society, Washington, United States of America, v. 88, p. 1–4, 2002. Citado na página 134.
- BARROSO, L. C.; BARROSO, M. M. A.; FILHO, F. F. C.; CARVALHO, M. L. B.; MAIA, M. L. **Cálculo numérico com aplicações**. 2. ed. São Paulo, Brasil: Harbra, 1987. Citado 2 vezes nas páginas 55 e 56.
- BATISTA, G. E. A. P. A.; CAMPANA, B.; KEOGH, E. J. Classification of live moths combining texture, color and shape primitives. In: . Washington, United States of America: IEEE, 2010. p. 903–906. Citado na página 127.
- BATISTA, G. E. A. P. A.; KEOGH, E. J.; TATAW, O. M.; SOUZA, V. M. A. CID: An efficient complexity-invariant distance for time series. **Data Mining and Knowledge Discovery**, Springer, United States of America, v. 28, n. 3, p. 634–669, 2014. Citado 11 vezes nas páginas 38, 39, 116, 117, 123, 124, 125, 126, 128, 132 e 133.
- BLOOMFIELD, P. **Fourier analysis of time series: An introduction**. 2. ed. New York, United States of America: Wiley, 2000. (Wiley Series in Probability and Statistics). Citado na página 204.

BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C.; LJUNG, G. M. **Time series analysis: Forecasting and control.** 5. ed. New Jersey, United States of America: Wiley, 2015. (Wiley Series in Probability and Statistics). Citado 6 vezes nas páginas 36, 47, 76, 86, 89 e 204.

BRAGA, I. A. **Aprendizado semissupervisionado multidescrição em classificação de textos.** 2010. Dissertação de Mestrado, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, Brasil. Citado na página 109.

BROCKLEBANK, J.; DICKEY, D. **SAS for forecasting time series.** 2. ed. Cary, United States of America: SAS Institute, 2003. Citado na página 60.

BROCKWELL, P. J.; DAVIS, R. A. **Introduction to time series forecasting.** 2. ed. New York, United States of America: Springer, 2002. Citado na página 52.

BUFFA, E. S.; SARIN, R. K. **Modern production/operations management.** New York, United States of America: Wiley, 1987. Citado na página 61.

CHA, S. H. Comprehensive survey on distance/similarity measures between probability density functions. **International Journal of Mathematical Models and Methods in Applied Sciences**, v. 1, n. 4, p. 300–307, 2007. Citado 4 vezes nas páginas 128, 129, 131 e 136.

CHANDOLA, V.; CHEBOLI, D.; KUMAR, V. **Detecting anomalies in a time series database.** Minneapolis, United States of America, 2009. Citado na página 38.

CHANG, Y.-W.; LIAO, M.-Y. A comparison of time series models for forecasting outbound air travel demand. **Journal of Aeronautics, Astronautics and Aviation**, Aeronautical and Astronautical Society of the Republic of China, v. 42, n. 2, p. 73–78, 2010. Citado na página 73.

CHATFIELD, C. **The analysis of time series: An introduction.** Boca Raton, United States of America: Taylor & Francis, 2013. (Chapman & Hall/CRC Texts in Statistical Science). Citado 7 vezes nas páginas 35, 36, 52, 61, 69, 75 e 87.

CHATTERJEE, S.; HADI, A. S. **Regression analysis by example.** 5. ed. New York, United States of America: Wiley, 2012. Citado 2 vezes nas páginas 55 e 56.

CHEN, Y.; KEOGH, E. J.; BATISTA, G. E. A. P. A. DTW-D: Time series semi-supervised learning from a single example. In: **International Conference on Knowledge Discovery and Data Mining.** Chicago, United States of America: ACM, 2013. p. 383–391. Citado 2 vezes nas páginas 132 e 134.

CHRISTODOULOU, N. S. An algorithm using Runge-Kutta methods of orders 4 and 5 for systems of ODEs. **International Journal of Numerical Methods and Applications**, Pushpa Publishing House, v. 2, n. 1, p. 47–57, 2009. Citado na página 212.

CLAVERIA, O.; TORRA, S. Forecasting tourism demand to Catalonia: Neural networks vs. time series models. **Economic Modelling**, Elsevier, United States of America, v. 36, n. C, p. 220–228, 2014. Citado 4 vezes nas páginas 36, 49, 73 e 107.

CLEMENTS, M. P.; HENDRY, D. F. On the limitations of comparing mean square forecast errors. **Journal of Forecasting**, Wiley, v. 12, n. 8, p. 617–637, 1993. Citado na página 103.

CORTEZ, P. Sensitivity analysis for time lag selection to forecast seasonal time series using neural networks and support vector machines. In: **International Joint Conference on Neural Networks.** Barcelona, Spain: IEEE, 2010. p. 3694–3701. Citado na página 48.

- COWPERTWAIT, P. S. P.; METCALFE, A. V. **Introductory time series with R**. New York, United States of America: Springer, 2009. Citado 6 vezes nas páginas 35, 53, 67, 88, 89 e 90.
- CRISTIANINI, N.; SHawe-Taylor, J. **An introduction to support vector machines and other kernel-based learning methods**. Cambridge, United Kingdom: Cambridge University Press, 2000. Citado na página 96.
- CRONE, S. F.; HIBON, M.; NIKOLOPOULOS, K. Advances in forecasting with neural networks? Empirical evidence from the NN3 competition on time series prediction. **International Journal of Forecasting**, v. 27, n. 3, p. 635–660, 2011. Citado na página 44.
- CYBENKO, G. Approximation by superpositions of a sigmoidal function. **Mathematics of Control, Signals and Systems**, Springer, v. 2, n. 4, p. 303–314, 1989. Citado na página 95.
- DEITEL, P. J.; DEITEL, H. M. **Java: How to Program**. 9. ed. Massachusetts, United States of America: Prentice Hall, 2012. Citado na página 163.
- DEMŠAR, J. Statistical comparisons of classifiers over multiple data sets. **Journal of Machine Learning Research**, JMLR.org, v. 7, p. 1–30, 2006. Citado 2 vezes nas páginas 143 e 163.
- DEZA, E.; DEZA, M. M. **Dictionary of distances**. Amsterdam, Netherlands: Elsevier, 2006. Citado 2 vezes nas páginas 128 e 136.
- DING, H.; TRAJCEVSKI, G.; SCHEUERMANN, P.; WANG, X.; KEOGH, E. Querying and mining of time series data: Experimental comparison of representations and distance measures. In: **International Conference on Very Large Data Bases**. Auckland, New Zealand: VLDB Endowment, 2008. v. 1, n. 2, p. 1542–1552. Citado 4 vezes nas páginas 37, 124, 128 e 136.
- EHLERS, R. S. **Análise de séries temporais**. 2009. 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 2 vezes nas páginas 54 e 61.
- EISENCRAFT, M. **Sistemas de comunicação utilizando sinais caóticos**. 2001. Dissertação de Mestrado, Escola Politécnica, Universidade de São Paulo, São Paulo, Brasil. Citado na página 214.
- FAIR, R. Handbook of econometrics. In: **ScienceDirect**. 1. ed. North Holland: Elsevier, 1986. v. 3, cap. Evaluating the predictive accuracy of models, p. 1979–1995. Citado na página 103.
- FALOUTSOS, C.; RANGANATHAN, M.; MANOLOPOULOS, Y. Fast subsequence matching in time-series databases. In: **Proceedings of the 1994 ACM SIGMOD International Conference on Management of Data**. Minneapolis, United States of America: ACM, 1994. p. 419–429. Citado na página 124.
- FERRERO, C. A. **Algoritmo kNN para previsão de dados temporais: Funções de previsão e critérios de seleção de vizinhos próximos aplicados à variáveis ambientais em limnologia**. 2009. Dissertação de Mestrado, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, Brasil. Citado 6 vezes nas páginas 37, 98, 107, 111, 113 e 120.
- FIX, E.; HODGES, J. L. **Discriminatory analysis, nonparametric discrimination, consistency properties**. Randolph Field, United States of America, 1951. Citado 2 vezes nas páginas 98 e 110.

FRUNZETE, M.; POPESCU, A.; BARBOT, J. P. Dynamical discrete-time Rössler map with variable delay. In: **Computational Science and Its Applications**. Banff, Canada: Springer International Publishing, 2015, (Lecture Notes in Computer Science, v. 9155). p. 431–446. Citado na página [213](#).

FU, T. A review on time series data mining. **Engineering Applications of Artificial Intelligence**, Tarrytown, United States of America, v. 24, n. 1, p. 164–181, 2011. Citado 3 vezes nas páginas [35](#), [37](#) e [51](#).

GARDNER, E. S. Exponential smoothing: The state of the art. **Journal of forecasting**, Wiley Online Library, v. 4, n. 1, p. 1–28, 1985. Citado 5 vezes nas páginas [76](#), [77](#), [78](#), [79](#) e [81](#).

GIUSTI, R.; BATISTA, G. E. A. P. A. An empirical comparison of dissimilarity measures for time series classification. In: **Brazilian Conference on Intelligent Systems**. Fortaleza, Brasil: IEEE, 2013. p. 82–88. Citado na página [128](#).

GLASS, L.; MACKEY, M. C. **Dos relógios ao caos: Os ritmos da vida**. São Paulo, Brasil: Editora da Universidade de São Paulo, 1997. Citado na página [212](#).

GOOIJER, J. G. D.; HYNDMAN, R. J. 25 years of time series forecasting. **International Journal of Forecasting**, v. 22, n. 3, p. 443–473, 2006. Citado 3 vezes nas páginas [36](#), [38](#) e [44](#).

GUNN, S. R. *et al.* **Support vector machines for classification and regression**. Southampton, United Kingdom, 1998. Faculty of Engineering, Science and Mathematics, University of Southampton. Citado na página [96](#).

HAN, J.; KAMBER, M.; PEI, J. **Data mining: Concepts and techniques**. 3. ed. San Francisco, United States of America: Morgan Kaufmann, 2011. Citado 5 vezes nas páginas [37](#), [69](#), [113](#), [119](#) e [131](#).

HAYKIN, S. S. **Neural networks and learning machines**. 3. ed. Upper Saddle River, United States of America: Prentice Hall, 2009. Hardcover. Citado 3 vezes nas páginas [93](#), [94](#) e [95](#).

HETLAND, M. L. A survey of recent methods for efficient retrieval of similar time sequences. In: **Data Mining in Time Series Databases**. Danvers, United States of America: World Scientific, 2004, (Series in Machine Perception and Artificial Intelligence, v. 57). cap. 2, p. 27–49. Citado na página [123](#).

HOLT, C. C. Forecasting seasonals and trends by exponentially weighted moving averages. **International journal of forecasting**, Elsevier, v. 20, n. 1, p. 5–10, 2004. Citado na página [81](#).

HUANG, W.; LAI, K. K.; NAKAMORI, Y.; WANG, S. Forecasting foreign exchange rates with artificial neural networks: A review. **International Journal of Information Technology and Decision Making**, v. 3, n. 1, p. 145–165, 2004. Citado na página [107](#).

HYNDMAN, R. J.; KOEHLER, A. B. Another look at measures of forecast accuracy. **International Journal of Forecasting**, v. 22, n. 4, p. 679–688, 2006. Citado na página [103](#).

HYNDMAN, R. J.; KOEHLER, A. B.; SNYDER, R. D.; GROSE, S. A state space framework for automatic forecasting using exponential smoothing methods. **International Journal of Forecasting**, Elsevier, v. 18, n. 3, p. 439–454, 2002. Citado 2 vezes nas páginas [78](#) e [79](#).

- ISLAM, M.; SIVAKUMAR, B. Characterization and prediction of runoff dynamics: A nonlinear dynamical view. **Advances in Water Resources**, Elsevier, v. 25, n. 2, p. 179–190, 2002. Citado 2 vezes nas páginas [75](#) e [91](#).
- ITAKURA, F. Minimum prediction residual principle applied to speech recognition. **IEEE Transactions on Acoustics, Speech and Signal Processing**, IEEE, v. 23, n. 1, p. 67–72, 1975. Citado na página [136](#).
- JUNIOR, J. A. **Estudo da influência de diversas medidas de similaridade na previsão de séries temporais utilizando o algoritmo kNN-TSP**. 2012. Dissertação de Mestrado em Engenharia de Sistemas Dinâmicos e Energéticos, Universidade Estadual do Oeste do Paraná, Foz do Iguaçu, Brasil. Citado 4 vezes nas páginas [37](#), [107](#), [113](#) e [136](#).
- KANDANANOND, K. A comparison of various forecasting methods for autocorrelated time series. **International Journal of Engineering Business Management**, Croatia, European Union, v. 4, p. 1–6, 2012. Citado 2 vezes nas páginas [36](#) e [49](#).
- KAUNDAL, R.; KAPOOR, A. S.; RAGHAVA, G. P. S. Machine learning techniques in disease forecasting: a case study on rice blast prediction. **BMC Bioinformatics**, v. 7, n. 1, p. 1–16, 2006. Citado na página [97](#).
- KEOGH, E. Efficiently finding arbitrarily scaled patterns in massive time series databases. In: **Knowledge Discovery in Databases**. Cavtat, Croatia: Springer Berlin Heidelberg, 2003, (Lecture Notes in Computer Science, v. 2838). p. 253–265. Citado na página [125](#).
- KEOGH, E.; RATANAMAHAATANA, C. A. Exact indexing of dynamic time warping. **Knowledge and Information Systems**, Springer, London, v. 7, n. 3, p. 358–386, 2005. Citado na página [135](#).
- KEOGH, E.; WEI, L.; XI, X.; VLACHOS, M.; LEE, S.-H.; PROTOPAPAS, P. Supporting exact indexing of arbitrarily rotated shapes and periodic time series under euclidean and warping distance measures. **The VLDB Journal**, Secaucus, United States of America, v. 18, n. 3, p. 611–630, 2009. Citado na página [125](#).
- KIRCHGÄSSNER, G.; WOLTERS, J.; HASSLER, U. **Introduction to modern time series analysis**. 2. ed. New York, United States of America: Springer, 2013. Citado na página [67](#).
- KITCHENHAM, B. A. **Guidelines for performing systematic literature reviews in software engineering**. United Kingdom, 2007. Evidence-based Software Engineering, EBSE-2007-01. Citado na página [43](#).
- KULESH, M.; HOLTSCHNEIDER, M.; KURENNAYA, K. Adaptive metrics in the nearest neighbours method. **Physica D: Nonlinear Phenomena**, v. 237, n. 3, p. 283–291, 2008. Citado 3 vezes nas páginas [113](#), [205](#) e [207](#).
- LAROSE, D. T. **Discovering knowledge in data: An introduction to data mining**. Hoboken, United States of America: Wiley, 2004. Citado na página [111](#).
- LAROSE, D. T.; LAROSE, C. D. **Discovering knowledge in data: An introduction to data mining**. New Jersey, United States of America: Wiley, 2014. (Wiley Series on Methods and Applications in Data Mining). Citado na página [35](#).

LEMKE, C.; GABRYS, B. Meta-learning for time series forecasting and forecast combination. **Neurocomputing**, Tarrytown, United States of America, v. 73, n. 10-12, p. 2006–2016, 2010. Citado na página [44](#).

LI, K. C.; YAN, M.; YUAN, S. S. A simple statistical model for depicting the cdc15-synchronized yeast cell-cycle regulated gene expression data. **Statistica Sinica**, v. 12, n. 1, p. 141–158, 2002. Citado na página [125](#).

LIN, J.; KEOGH, E.; LONARDI, S.; CHIU, B. A symbolic representation of time series, with implications for streaming algorithms. In: **Workshop on Research Issues in Data Mining and Knowledge Discovery**. New York, United States of America: ACM, 2003. p. 2–11. Citado na página [133](#).

LIPPmann, R. P. An introduction to computing with neural nets. **IEEE ASSP Magazine**, v. 4, n. 2, p. 4–22, 1987. Citado na página [94](#).

LORENA, A. C.; CARVALHO, A. C. P. L. F. de. Uma introdução às support vector machines. **Revista de Informática Teórica e Aplicada**, v. 14, n. 2, p. 43–67, 2007. Citado na página [96](#).

LORENZ, E. N. Deterministic nonperiodic flow. **Journal of the Atmospheric Sciences**, American Meteorological Society, v. 20, n. 2, p. 130–141, 1963. Citado na página [213](#).

MAIMON, O.; ROKACH, L. (Ed.). **Data mining and knowledge discovery handbook**. 2. ed. Secaucus, United States of America: Springer, 2010. Citado 2 vezes nas páginas [35](#) e [70](#).

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The bulletin of mathematical biophysics**, Kluwer Academic Publishers, v. 5, n. 4, p. 115–133, 1943. Citado 2 vezes nas páginas [93](#) e [94](#).

MCSHARRY, P. E.; CLIFFORD, G. D.; TARASSENKO, L.; SMITH, L. A. A dynamical model for generating synthetic electrocardiogram signals. **Transactions on Biomedical Engineering**, IEEE, v. 50, n. 3, p. 289–294, 2003. Citado na página [215](#).

MINSKY, M.; PAPERT, S. **Perceptrons: an introduction to computational geometry**. Cambridge, United States of America: MIT Press, 1969. Citado na página [94](#).

MITCHELL, T. M. **Machine learning**. Ohio, United States of America: McGraw-Hill, 1997. Citado 2 vezes nas páginas [108](#) e [110](#).

MONTGOMERY, D. **Introduction to statistical quality control**. 6. ed. New York, United States of America: Wiley, 2009. Citado na página [208](#).

MONTGOMERY, D. C.; JENNINGS, C. L.; KULAHCI, M. **Introduction to time series analysis and forecasting**. 2. ed. New Jersey, United States of America: Wiley, 2015. (Wiley Series in Probability and Statistics). Citado 7 vezes nas páginas [35](#), [76](#), [78](#), [79](#), [81](#), [86](#) e [87](#).

MORÉ, J. J. Numerical analysis. In: WATSON, G. A. (Ed.). **Proceedings of the Biennial Conference Held at Dundee**. Heidelberg, Germany: Springer, 1978, (Lecture Notes in Mathematics, v. 630). cap. The Levenberg-Marquardt algorithm: Implementation and theory, p. 105–116. Citado na página [90](#).

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de séries temporais**. 2. ed. São Paulo, Brasil: Blucher, 2006. Citado 10 vezes nas páginas [52](#), [54](#), [60](#), [75](#), [77](#), [79](#), [81](#), [83](#), [89](#) e [90](#).

- MUEEN, A.; KEOGH, E.; ZHU, Q.; CASH, S. Exact discovery of time series motifs. In: **SIAM International Conference on Data Mining**. Sparks, United States of America: Society Industrial Applied Mathematics, 2009. Citado na página 117.
- PARMEZAN, A. R. S.; BATISTA, G. E. A. P. A. **ICMC-USP time series prediction repository**. 2014. Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, Brasil. Disponível em: <http://sites.labic.icmc.usp.br/icmc_tspr>. Citado 7 vezes nas páginas 37, 41, 52, 142, 163, 183 e 203.
- _____. A study of the use of complexity measures in the similarity search process adopted by kNN algorithm for time series prediction. In: **International Conference on Machine Learning and Applications**. Miami, United States of America: IEEE, 2015. p. 45–51. Citado 8 vezes nas páginas 36, 40, 98, 107, 123, 134, 158 e 183.
- PETITJEAN, F.; KETTERLIN, A.; GANÇARSKI, P. A global averaging method for dynamic time warping, with applications to clustering. **Pattern Recognition**, v. 44, n. 3, p. 678–693, 2011. Citado na página 116.
- PILA, A. D. Seleção de atributos relevantes para aprendizado de máquina utilizando a abordagem de rough sets. Dissertação de Mestrado, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, Brasil. 2001. Citado na página 110.
- PLATT, J. C. Using analytic QP and sparseness to speed training of support vector machines. In: **Advances in Neural Information Processing Systems**. Cambridge, United States of America: MIT Press, 1999. p. 557–563. Citado na página 98.
- PRATI, R. C.; BATISTA, G. E. A. P. A. A complexity-invariant measure based on fractal dimension for time series classification. **International Journal of Natural Computing Research**, v. 3, n. 3, p. 59–73, 2012. Citado na página 38.
- PYLE, D. **Data preparation for data mining**. California, United States of America: Morgan Kaufmann, 1999. Citado na página 71.
- RABINER, L. R.; SCHAFER, R. W. **Digital processing of speech signals**. Englewood Cliffs, United States of America: Prentice Hall, 1978. Citado na página 134.
- RAKTHANMANON, T.; CAMPANA, B.; MUEEN, A.; BATISTA, G.; WESTOVER, B.; ZHU, Q.; ZAKARIA, J.; KEOGH, E. Searching and mining trillions of time series subsequences under dynamic time warping. In: **International Conference on Knowledge Discovery and Data Mining**. Beijing, China: ACM, 2012. p. 262–270. Citado na página 115.
- REINSEL, G. C. **Elements of multivariate time series analysis**. 2. ed. New York, United States of America: Springer, 2003. (Springer Series in Statistics). Citado na página 74.
- RISTANOSKI, G.; LIU, W.; BAILEY, J. A time-dependent enhanced support vector machine for time series regression. In: **International Conference on Knowledge Discovery and Data Mining**. Chicago, United States of America: ACM, 2013. p. 946–954. Citado 3 vezes nas páginas 49, 98 e 107.
- ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. **Psychological Review**, v. 65, n. 6, p. 386–408, 1958. Citado na página 93.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Parallel distributed processing: explorations in the microstructure of cognition. In: . Cambridge, United States of America: MIT Press, 1986. v. 1, cap. Learning internal representations by error propagation, p. 318–362. Citado na página 94.

SAKOE, H.; CHIBA, S. Dynamic programming algorithm optimization for spoken word recognition. **IEEE Transactions on Acoustics, Speech and Signal Processing**, IEEE, v. 26, n. 1, p. 43–49, 1978. Citado na página 136.

SAPANKEVYCH, N. I.; SANKAR, R. Time series prediction using support vector machines: A survey. **Computational Intelligence Magazine**, IEEE, Piscataway, United States of America, v. 4, n. 2, p. 24–38, 2009. Citado 2 vezes nas páginas 36 e 107.

SCHILLER, J.; SPIEGEL, M.; SRINIVASAN, R. **Schaum's outline of probability and statistics**. New York, United States of America: McGraw-Hill, 2012. Citado 2 vezes nas páginas 57 e 63.

SHAWE-TAYLOR, J.; BARTLETT, P. L.; WILLIAMSON, R. C.; ANTHONY, M. Structural risk minimization over data-dependent hierarchies. IEEE, v. 44, n. 5, p. 1926–1940, 1998. Citado na página 96.

SHUMWAY, R. H.; STOFFER, D. S. **Time series analysis and its applications: With R examples**. 3. ed. New York, United States of America: Springer Science & Business Media, 2011. (Springer Texts in Statistics). Citado na página 87.

SORJAMAA, A.; HAO, J.; REYHANI, N.; JI, Y.; LENDASSE, A. Methodology for long-term prediction of time series. **Neurocomputing**, v. 70, p. 2861–2869, 2007. Citado na página 73.

STROGATZ, S. H. **Nonlinear dynamics and chaos: With applications to physics, biology, chemistry, and engineering**. 2. ed. Boulder, United States of America: Westview Press, 2014. (Studies in Nonlinearity). Citado na página 210.

TAHERSIMA, H.; TAHERSIMA, F.; SOHANI, A.; JAFAR, M.; SALEH, K. Prediction of Lorenz chaotic time series via genetic algorithm. In: **International Conference on Computational Intelligence for Measurement Systems and Applications**. Taranto, Italia: IEEE, 2010. p. 13–17. Citado na página 212.

TAIEB, S. B.; BONTEMPI, G.; ATIYA, A. F.; SORJAMAA, A. A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition. **Expert Systems with Applications**, v. 39, n. 8, p. 7067–7083, 2012. Citado na página 44.

THEIL, H. **Principles of econometrics**. New York, United States of America: Wiley, 1971. Citado na página 103.

VAPNIK, V. N. **The nature of statistical learning theory**. 2. ed. New York, United States of America: Springer Science & Business Media, 1999. (Information Science and Statistics). Citado na página 96.

VERHULST, P. F. Recherches mathématiques sur la loi d'accroissement de la population. **Nouveaux mémoires de l'Académie Royale des Sciences et Belles-Lettres de Bruxelles**, v. 18, p. 14–54, 1845. Citado na página 210.

- VERPLANCKE, T.; LOOY, S. V.; STEURBAUT, K.; BENOIT, D.; TURCK, F. D.; MOOR, G. D.; DECRUYENAERE, J. A novel time series analysis approach for prediction of dialysis in critically ill patients using echo-state networks. **BMC Medical Informatics and Decision Making**, v. 10, p. 4, 2010. Citado na página 73.
- VLACHOS, M.; GUNOPULOS, D.; DAS, G. Indexing time-series under conditions of noise. In: **Data Mining in Time Series Databases**. Danvers, United States of America: World Scientific, 2004. p. 67–100. Citado na página 129.
- WINTERS, P. R. Forecasting sales by exponentially weighted moving averages. **Management Science**, INFORMS, v. 6, n. 3, p. 324–342, 1960. Citado 2 vezes nas páginas 81 e 83.
- WITTEN, I. H.; FRANK, E.; HALL, M. A. **Data mining: Practical machine learning tools and techniques**. 3. ed. San Francisco, United States of America: Morgan Kaufmann, 2011. (The Morgan Kaufmann Series in Data Management Systems). Citado 3 vezes nas páginas 35, 47 e 163.
- XU, R.; WUNSCH, D. **Clustering**. New Jersey, United States of America: Wiley, 2009. (IEEE Series on Computational Intelligence). Citado na página 123.
- ZHANG, G.; PATUWO, B. E.; HU, M. Y. Forecasting with artificial neural networks: the state of the art. **International Journal of Forecasting**, v. 14, n. 1, p. 35–62, 1998. Citado na página 93.
- ZHANG, X.; ZHANG, T.; YOUNG, A. A.; LI, X. Applications and comparisons of four time series models in epidemiological surveillance data. **PLOS ONE**, Public Library of Science, v. 9, n. 2, 2014. Citado na página 50.
- ZHU, Q.; BATISTA, G. E. A. P. A.; RAKTHANMANON, T.; KEOGH, E. J. A novel approximation to dynamic time warping allows anytime clustering of massive time series datasets. In: **SIAM International Conference on Data Mining**. Anaheim, United States of America: Society Industrial Applied Mathematics, 2012. p. 999–1010. Citado na página 38.
- ZUNIC, J.; ROSIN, P. L.; KOPANJA, L. Shape orientability. In: **Computer Vision**. Hyderabad, India: Springer Berlin Heidelberg, 2006, (Lecture Notes in Computer Science, v. 3852). p. 11–20. Citado na página 126.

TRABALHOS SELECIONADOS NA REVISÃO SISTEMÁTICA

Neste apêndice estão localizados os quadros que contemplam as informações de publicação dos trabalhos selecionados na revisão sistemática.

Quadro 11 – Sumário de informações dos trabalhos identificados

ID	Item de Interesse	Informação Extraída
1	Título Autores Evento/Revista Ano	A comparison of time series models for forecasting outbound air travel demand Yu-Wei Chang; Meng-Yuan Liao Journal of Aeronautics, Astronautics and Aviation 2010
2	Título Autores Evento/Revista Ano	An artificial neural network (p, d, q) model for timeseries forecasting Mehdi Khashei; Mehdi Bijari Expert Systems with Applications 2010
3	Título Autores Evento/Revista Ano	An empirical comparison of machine learning models for time series forecasting Nesreen K. Ahmed; Amir F. Atiya; Neamat El Gayar; Hisham El-Shishiny Econometric Reviews 2010
4	Título Autores Evento/Revista Ano	An experimental study of fitness function and time series forecasting using artificial neural networks Aranildo R. Lima Junior; David A. Silva; Paulo S. Mattos Neto; Tiago A. E. Ferreira Genetic and Evolutionary Computation Conference 2010
5	Título Autores Evento/Revista Ano	Combination of long term and short term forecasts, with application to tourism demand forecasting Robert R. Andrawis; Amir F. Atiya; Hisham El-Shishiny International Journal of Forecasting 2010
Continua na página seguinte.		

Quadro 2 – Sumário de informações dos trabalhos identificados

Continuação da página anterior.		
ID	Item de Interesse	Informação Extraída
6	Título	Conditionally dependent strategies for multiple-step-ahead prediction in local learning
	Autores	Gianluca Bontempi; Souhaib Ben Taieb
	Evento/Revista	International Journal of Forecasting
	Ano	2010
7	Título	Financial time series forecasting with machine learning techniques: A survey
	Autores	Bjoern Krollner; Bruce Vanstone; Gavin Finnie
	Evento/Revista	European Symposium on Artificial Neural Networks
	Ano	2010
8	Título	Integrated time series forecasting approaches using moving average, grey prediction, support vector regression and bagging for NNFC
	Autores	Chihli Hung; Xin-Yi Huang; Hao-Kai Lin; Yen-Hsu Hou
	Evento/Revista	International Joint Conference on Neural Networks
	Ano	2010
9	Título	Meta-learning for time series forecasting and forecast combination
	Autores	Christiane Lemke; Bogdan Gabrys
	Evento/Revista	Neurocomputing
	Ano	2010
10	Título	Multiple-output modeling for multi-step-ahead time series forecasting
	Autores	Souhaib Ben Taieb; Antti Sorjamaa; Gianluca Bontempi
	Evento/Revista	Neurocomputing
	Ano	2010
11	Título	Performance of radial basis function and support vector machine in time series forecasting
	Autores	Mazlina Mamat; Salina Abdul Samad
	Evento/Revista	International Conference on Intelligent and Advanced Systems
	Ano	2010
12	Título	Prediction of daily patient numbers for a regional emergency medical center using time series analysis
	Autores	Hye Jin Kam; Jin Ok Sung; Rae Woong Park
	Evento/Revista	Healthcare Informatics Research
	Ano	2010
13	Título	Sensitivity analysis for time lag selection to forecast seasonal time series using neural networks and support vector machines
	Autores	Paulo Cortez
	Evento/Revista	International Joint Conference on Neural Networks
	Ano	2010
14	Título	Advances in forecasting with neural networks? Empirical evidence from the NN3 competition on time series prediction
	Autores	Sven F. Crone; Michèle Hibon; Konstantinos Nikolopoulos
	Evento/Revista	International Journal of Forecasting
	Ano	2011
15	Título	Forecast combinations of computational intelligence and linear models for the NN5 time series forecasting competition
	Autores	Robert R. Andrawis; Amir F. Atiya; Hisham El-Shishiny
	Evento/Revista	International Journal of Forecasting
	Ano	2011

Continua na página seguinte.

Quadro 2 – Sumário de informações dos trabalhos identificados

			Continuação da página anterior.
ID	Item de Interesse	Informação Extraída	
16	Título Autores Evento/Revista Ano	F orecasting seasonal time series with computational intelligence: contribution of a combination of distinct methods M. Štěpnička; J. Peralta; P. Cortez; L. Vavříčková; G. Gutierrez Conference of the European Society for Fuzzy Logic and Technology 2011	
17	Título Autores Evento/Revista Ano	Hybrid models for future event prediction Giuseppe Amodeo; Roi Blanco; Ulf Brefeld Conference on Information and Knowledge Management 2011	
18	Título Autores Evento/Revista Ano	Recursive multi-step time series forecasting by perturbing data Souhaib Ben Taieb; Gianluca Bontempi International Conference on Data Mining 2011	
19	Título Autores Evento/Revista Ano	Time series forecasting by using seasonal autoregressive integrated moving average subset multiplicative or additive model Suhartono Journal of Mathematics and Statistics 2011	
20	Título Autores Evento/Revista Ano	A comparative study on the forecast of fresh food sales using logistic regression, moving average and BPNN methods Wan-I Lee; Cheng-Wu Chen; Kung-Hsing Chen; Tsung-Hao Chen; Chia-Chi Liu Journal of Marine Science and Technology 2012	
21	Título Autores Evento/Revista Ano	A comparison of various forecasting methods for autocorrelated time series Karin Kandananond International Journal of Engineering Business Management 2012	
22	Título Autores Evento/Revista Ano	A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition Souhaib Ben Taieb; Gianluca Bontempi; Amir F. Atiya; Antti Sorjamaa Expert Systems with Applications 2012	
23	Título Autores Evento/Revista Ano	Bayesian model for time series with trend, autoregression and outliers Pitsanu Tongkhow; Nantachai Kantanantha International Conference on ICT and Knowledge Engineering 2012	
24	Título Autores Evento/Revista Ano	Comparative analysis of machine learning techniques in sale forecasting Suresh Kumar Sharma; Vinod Sharma International Journal of Computer Applications 2012	
25	Título Autores Evento/Revista Ano	Fuzzy time series and SARIMA model for forecasting tourist arrivals to Bali Maria Elena; Muhamad Hisyam Lee; Suhartono H.; Hossein I.; Nur Haizum Abd Rahman; Nur Arina Bazilah Jurnal Teknologi 2012	
Continua na página seguinte.			

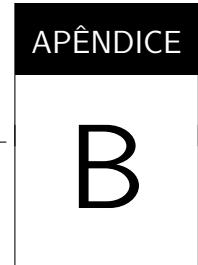
Quadro 2 – Sumário de informações dos trabalhos identificados

Continuação da página anterior.		
ID	Item de Interesse	Informação Extraída
26	Título	A time-dependent enhanced support vector machine for time series regression
	Autores	Goce Ristanoski; Wei Liu; James Bailey
	Evento/Revista	Conference on Knowledge Discovery and Data Mining
	Ano	2013
27	Título	An ensemble model for day-ahead electricity demand time series forecasting
	Autores	Wen Shen; Vahan Babushkin; Zeyar Aung; Wei Lee Woon
	Evento/Revista	International Conference on Future Energy Systems
	Ano	2013
28	Título	Comparative study of four time series methods in forecasting typhoid fever incidence in China
	Autores	Xingyu Zhang; Yuanyuan Liu; Min Yang; Tao Zhang; Alistair A. Young; Xiaosong Li
	Evento/Revista	PLOS ONE
	Ano	2013
29	Título	Design of experiments on neural network's parameters optimization for time series forecasting in stock markets
	Autores	Mu-Yen Chen; Min-Hsuan Fan; Young-Long Chen; Hui-Mei Wei
	Evento/Revista	International Journal on Non-Standard Computing and Artificial Intelligence
	Ano	2013
30	Título	Machine learning methods to forecast temperature in buildings
	Autores	Fernando Mateo; Juan José Carrasco; Abderrahim Sellami; Mónica Millán-Giraldo; Manuel Domínguez; Emilio Soria-Olivas
	Evento/Revista	Expert Systems with Applications
	Ano	2013
31	Título	Machine learning strategies for time series forecasting
	Autores	Gianluca Bontempi; Souhaib Ben Taieb; Yann-Aël Le Borgne
	Evento/Revista	Business Intelligence
	Ano	2013
32	Título	Online learning for time series prediction
	Autores	Oren Anava; Elad Hazan; Shie Mannor; Ohad Shamir
	Evento/Revista	Journal of Machine Learning Research
	Ano	2013
33	Título	SARIMA (Seasonal ARIMA) implementation on time series to forecast the number of Malaria incidence
	Autores	Adhistya Erna Permanasari; Indriana Hidayah; Isna Alfi Bustoni
	Evento/Revista	International Conference on Information Technology and Electrical Engineering
	Ano	2013
34	Título	Time series analysis of household electric consumption with ARIMA and ARMA models
	Autores	Pasapitch Chujai; Nittaya Kerdprasop; Kittisak Kerdprasop
	Evento/Revista	International MultiConference of Engineers and Computer Scientists
	Ano	2013
35	Título	A comparison of machine learning techniques for modeling river flow time series: The case of upper Cauvery river basin
	Autores	Shivshanker Singh Patel; Parthasarathy Ramachandran
	Evento/Revista	Water Resources Management
	Ano	2014

Continua na página seguinte.

Quadro 2 – Sumário de informações dos trabalhos identificados

			Continuação da página anterior.
ID	Item de Interesse	Informação Extraída	
36	Título Autores Evento/Revista Ano	A hybrid forecasting approach for HRG parameters based on output time series Qi Ziyang; Li Qinghua; Yi Guoxing; Fang Haibin Chinese Control and Decision Conference 2014	
37	Título Autores Evento/Revista Ano	A strategy for forecasting option prices using fuzzy time series and least square support vector regression with a bootstrap model C. P. Lee; W. C. Lin; C. C. Yang International Journal of Science and Technology 2014	
38	Título Autores Evento/Revista Ano	Applications and comparisons of four time series models in epidemiological surveillance data Xingyu Zhang; Tao Zhang; Alistair A. Young; Xiaosong Li PLOS ONE 2014	
39	Título Autores Evento/Revista Ano	Employing time-series forecasting to historical medical data: An application towards early prognosis within elderly health monitoring environments Antonis S. Billis; Panagiotis D. Bamidis International Workshop on Artificial Intelligence and Assistive Medicine 2014	
40	Título Autores Evento/Revista Ano	Forecasting tourism demand to Catalonia: Neural networks vs. time series models Oscar Claveria; Salvador Torra Economic Modelling 2014	
41	Título Autores Evento/Revista Ano	Investigation of Empirical Mode Decomposition in Forecasting of Hydrological Time Series Ozgur Kisi; Levent Latifoğlu; Fatma Latifoğlu Water Resources Management 2014	
42	Título Autores Evento/Revista Ano	Time series forecasting using least square support vector machine for Canadian Lynx data Shuhaida Ismail; Ani Shabri Jurnal Teknologi 2014	



ICMC-USP TIME SERIES PREDICTION REPOSITORY

As informações gerenciadas pelo portal *Web ICMC-USP Time Series Prediction Repository* (PARMEZAN; BATISTA, 2014) são esquematizadas na Figura 82.

Figura 82 – Conteúdo administrado pelo ICMC-USP Time Series Prediction Repository

Fonte: Elaborada pelo autor.

O ICMC-USP *Time Series Prediction Repository*, que foi planejado e desenvolvido neste trabalho, tem como objetivo estabelecer um canal de comunicação entre o Laboratório de Inteligência Computacional (LABIC) do ICMC-USP, instituições de ensino e a comunidade acadêmica, de modo a disponibilizar por meio da *Internet*, conjuntos de dados, programas de

computador, resultados de pesquisa, trabalhos relacionados e outras informações relevantes acerca das atividades conduzidas no projeto financiado pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) e vinculado ao processo 2013/10978-8.

Até o momento, o repositório mantém 100 conjuntos de dados temporais (40 ST sintéticas e 60 ST reais) de acesso público e destinados às comunidades de Estatística e Aprendizado de Máquina. Uma descrição sucinta desses conjuntos de dados é apresentada a seguir.

Conjuntos de Dados Sintéticos

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

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 B.1 (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.

$$S_t = \sum_{i=0}^{\infty} a_i \cos\left(\frac{2\pi i t}{T}\right) + \sum_{j=0}^{\infty} b_j \sin\left(\frac{2\pi j t}{T}\right) + c \quad (\text{B.1})$$

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 B.1 com $t \in [1, 790]$, foram confeccionadas 12 ST cujas estruturas são listadas no Quadro 12.

Todas as equações dispostas no Quadro 12 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

Quadro 12 – Padrões temporais gerados a partir da composição de Fourier

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

tendência decrescente. Os padrões temporais dessas séries são mostrados graficamente na Figura 83.

Dependência Sazonal: Esse tipo de padrão temporal, estabelecido pela Equação B.2, possui tendência linear e sazonalidade constante (KULESH; HOLSCHEIDER; KURENNAYA, 2008).

$$S_t = \cos\left(\frac{t}{25}\right)\sin\left(\frac{t}{100}\right) \quad (\text{B.2})$$

Na Figura 84 é ilustrada três ST geradas usando a Equação B.2 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.

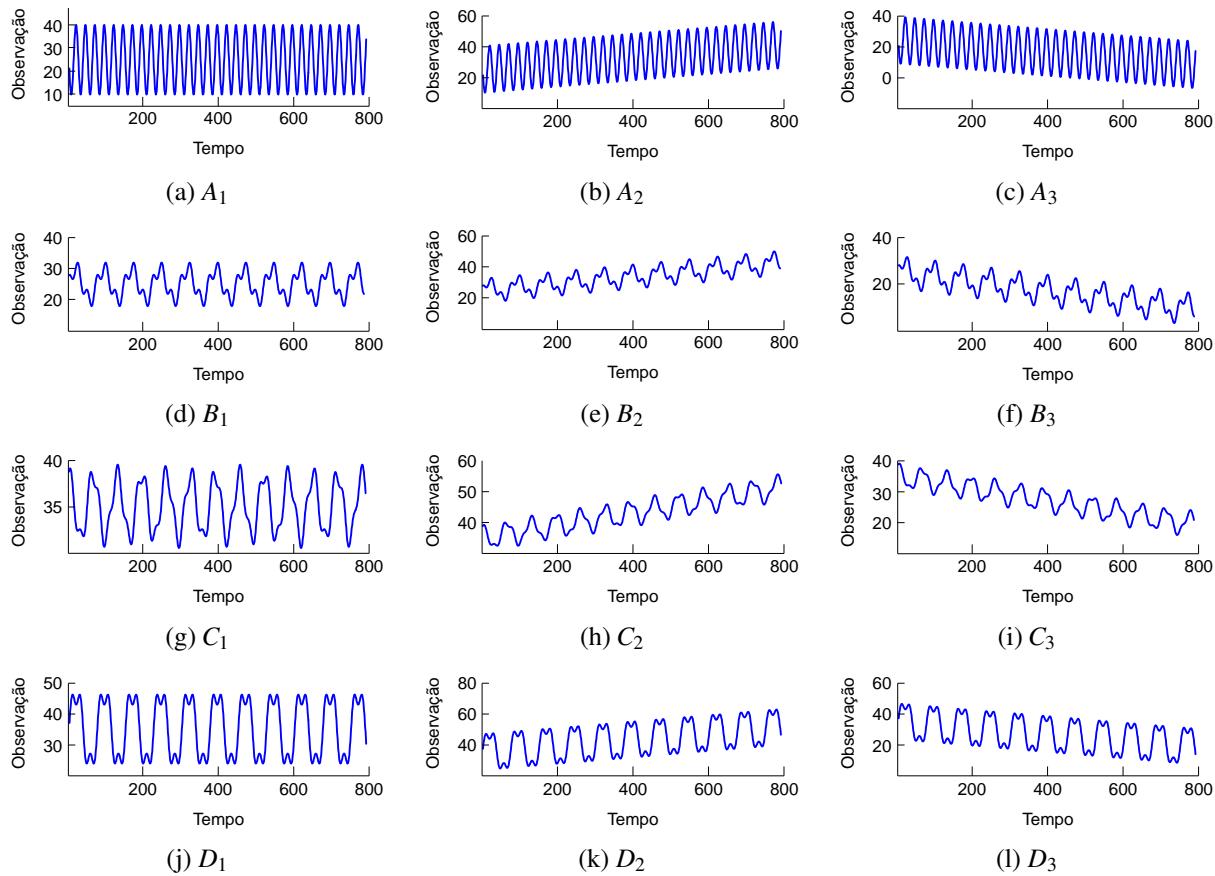
Sazonalidade Multiplicativa: Modelado por meio da Equação B.3, esse tipo de padrão temporal abrange uma tendência não-linear e a amplitude das oscilações sazonais aumenta com o tempo (KULESH; HOLSCHEIDER; KURENNAYA, 2008).

$$S_t = \begin{cases} R_t, & t \in [1, 79] \\ \frac{(S_{t-t_0})^2}{S_{t-2t_0}}, & t \in [80, 590], \quad t_0 = 14 \end{cases} \quad (\text{B.3})$$

Complementarmente à Equação B.3, tem-se que:

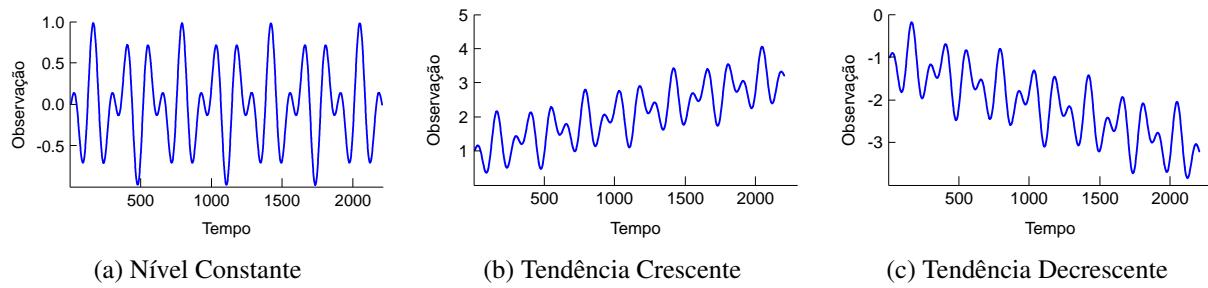
$$R_t = \frac{t}{70000} \left[\cos\left(\frac{9t}{7}\right)\sin\left(\frac{t}{350}\right) + 10 \right]$$

Figura 83 – Séries com sazonalidade aditiva derivadas da composição de Fourier



Fonte: Elaborada pelo autor.

Figura 84 – Séries com dependência sazonal

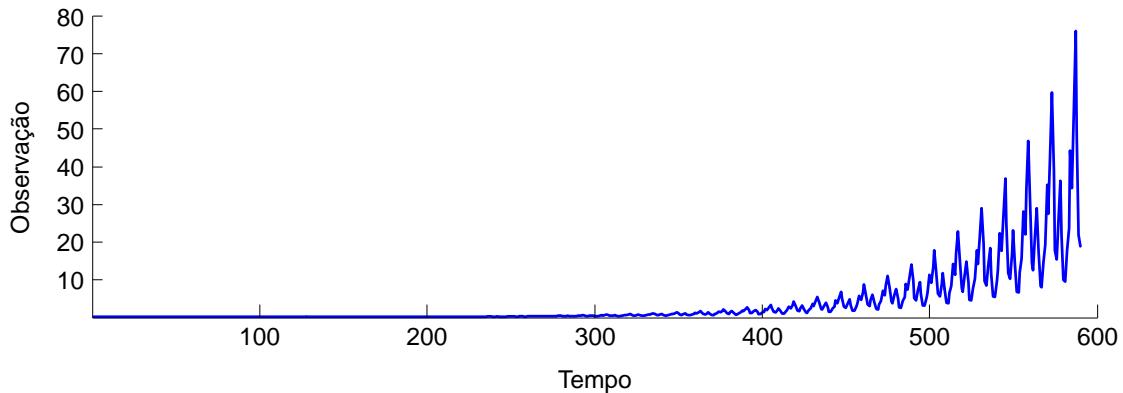


Fonte: Elaborada pelo autor.

O resultado da aplicação das equações supracitadas é esquematizado na [Figura 85](#). Como pode ser observado nessa figura, a série sintética expõe um crescimento levemente curvo e de ordem quadrática.

Alta Frequência: Descritos pela [Equação B.4](#), os dados desse padrão temporal assumem sazonalidade multiplicativa e constante aumento de amplitude ([KULESH; HOLSCHNEIDER](#);

Figura 85 – Série com sazonalidade multiplicativa



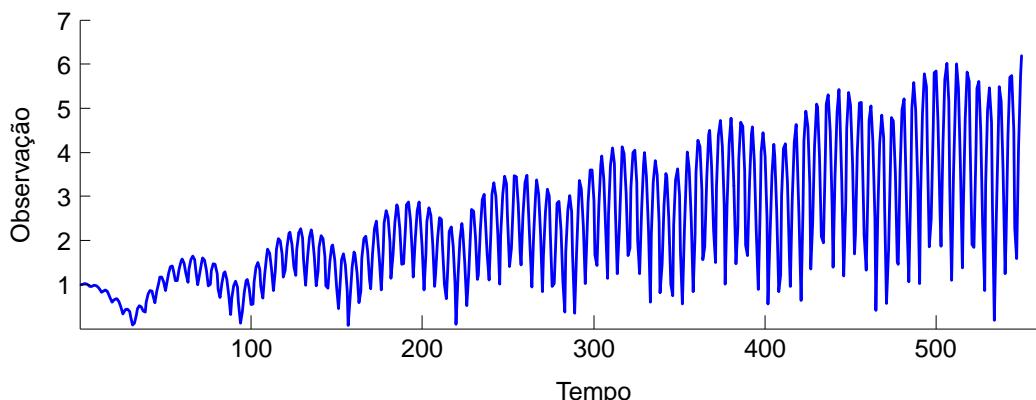
Fonte: Elaborada pelo autor.

KURENNAYA, 2008).

$$S_t = \frac{t}{100} |\sin(t/2)| + |\cos(t/20)|, \quad t \in [0, 550] \quad (\text{B.4})$$

Na [Figura 86](#) é exibida uma série sintética constituída de 550 amostras que retrata o padrão de alta frequência.

Figura 86 – Série de alta frequência com sazonalidade multiplicativa



Fonte: Elaborada pelo autor.

Séries Estocásticas

Séries estocásticas são originárias de eventos aleatórios e, consequentemente, 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 predizer 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 no [Quadro 13](#). Nesse quadro, 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.

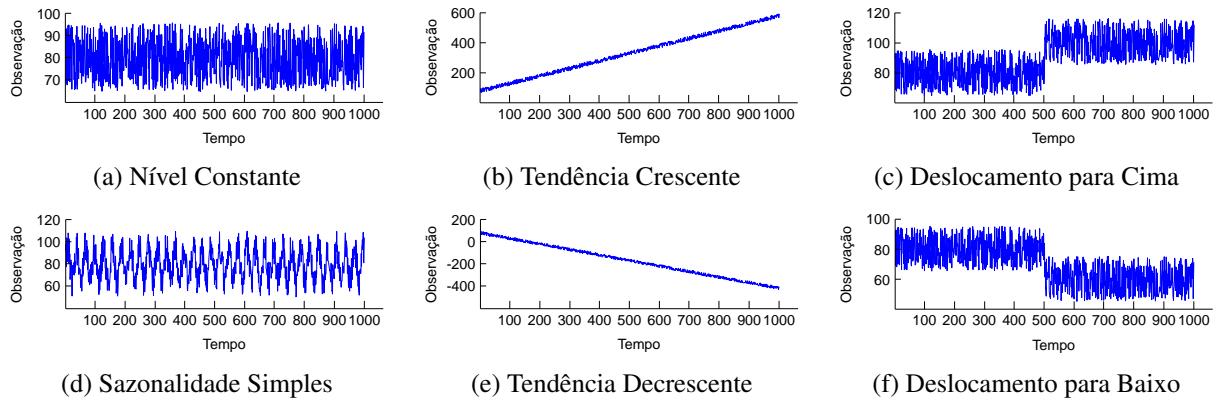
Quadro 13 – Padrões temporais comumente observados em gráficos de controle

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

Nas equações organizadas no [Quadro 13](#), 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 comprehende uma função geradora 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 no [Quadro 13](#), 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 $t = 500$.

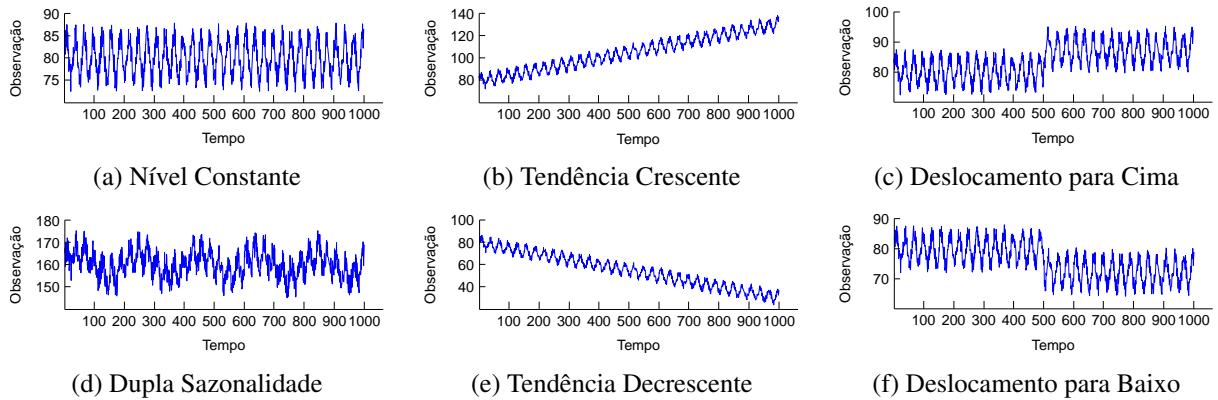
As séries provenientes da aplicação das equações catalogadas como GCA são retratadas na [Figura 87](#), enquanto que as séries geradas usando as equações da categoria GCB são esquematizadas na [Figura 88](#).

Figura 87 – Padrões temporais provenientes da categoria GCA



Fonte: Elaborada pelo autor.

Figura 88 – Padrões temporais provenientes da categoria GCB



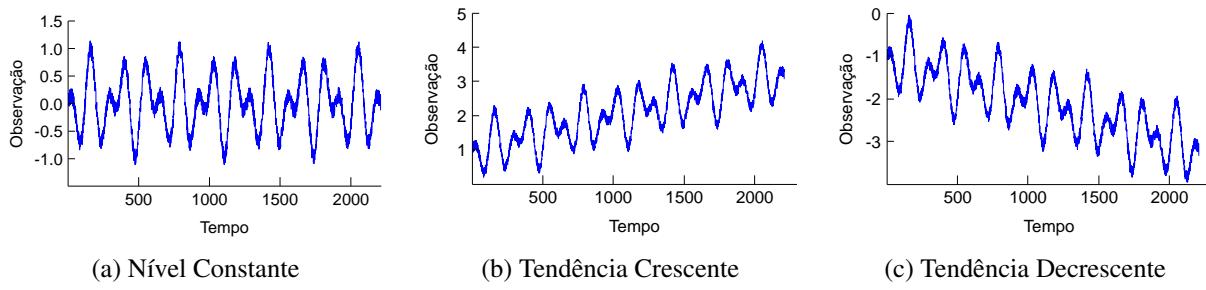
Fonte: Elaborada pelo autor.

Dependência Sazonal e Ruído: Esse padrão temporal é resultado da utilização da [Equação B.2](#) acrescida de uma função r que produz, a cada instante t , um número aleatório (ruído). Na [Figura 89](#), 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 205](#).

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

Figura 89 – Séries com dependência sazonal e ruído



Fonte: Elaborada pelo autor.

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 B.5 \(VERHULST, 1845\)](#), 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.

$$x_{t+1} = rx_t(1 - x_t) \quad (\text{B.5})$$

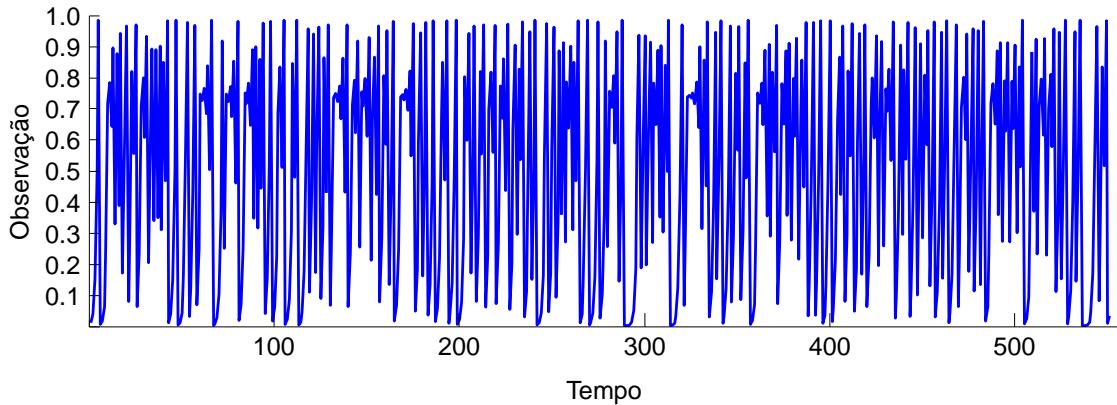
Geometricamente, o mapa logístico pode ser visualizado como uma parábola cuja concavidade é definida segundo o valor de r . Dependendo do valor assumido por esse parâmetro, o comportamento futuro do mapa perde a regularidade e passa a ser altamente sensível às condições iniciais. Por isso, na proposta 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 90](#) é mostrada uma série sintética, constituída por 551 amostras, resultante da aplicação da [Equação B.5](#) com $r = 4$ e $x_1 = 0,01$.

Mapa de Hénon: Consiste de um sistema discreto bidimensional, expresso por meio da [Equação B.6 \(STROGATZ, 2014\)](#), 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 dinâmicos pelas seguintes equações:

$$\begin{cases} x_{t+1} = 1 - ax_t^2 + by_t \\ y_{t+1} = x_t \end{cases} \quad (\text{B.6})$$

Figura 90 – Mapa Logístico

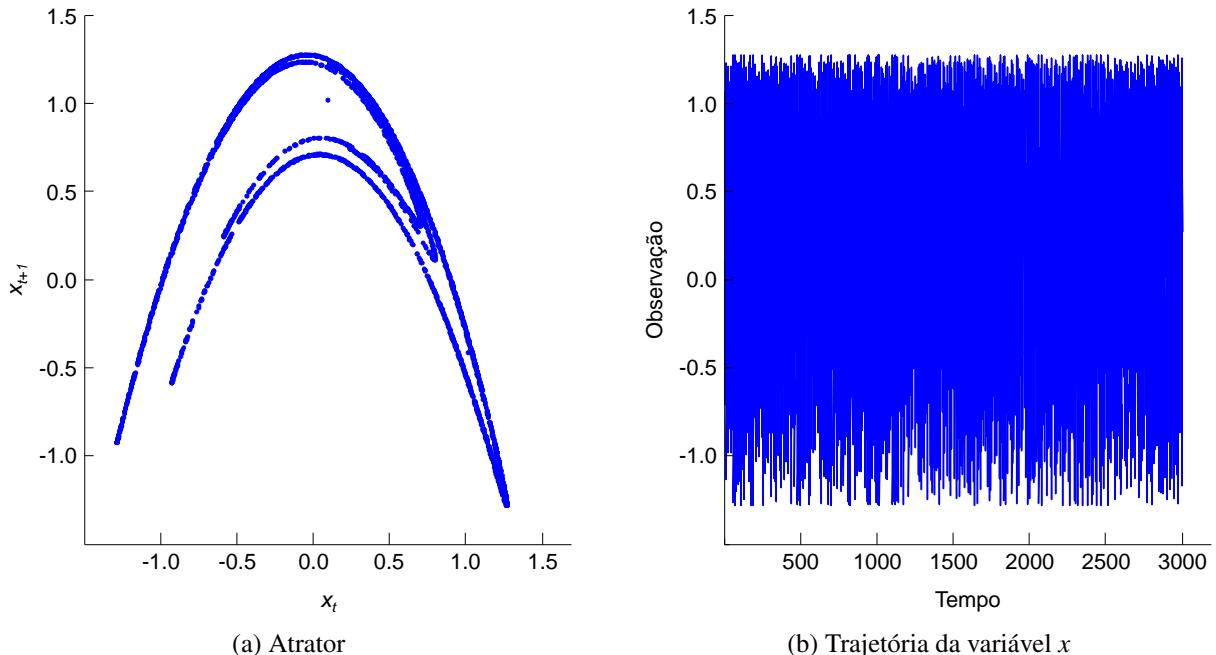


Fonte: Elaborada pelo autor.

Na [Equação B.6](#), 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 91](#) 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 91 – Mapa de Hénon



Fonte: Elaborada pelo autor.

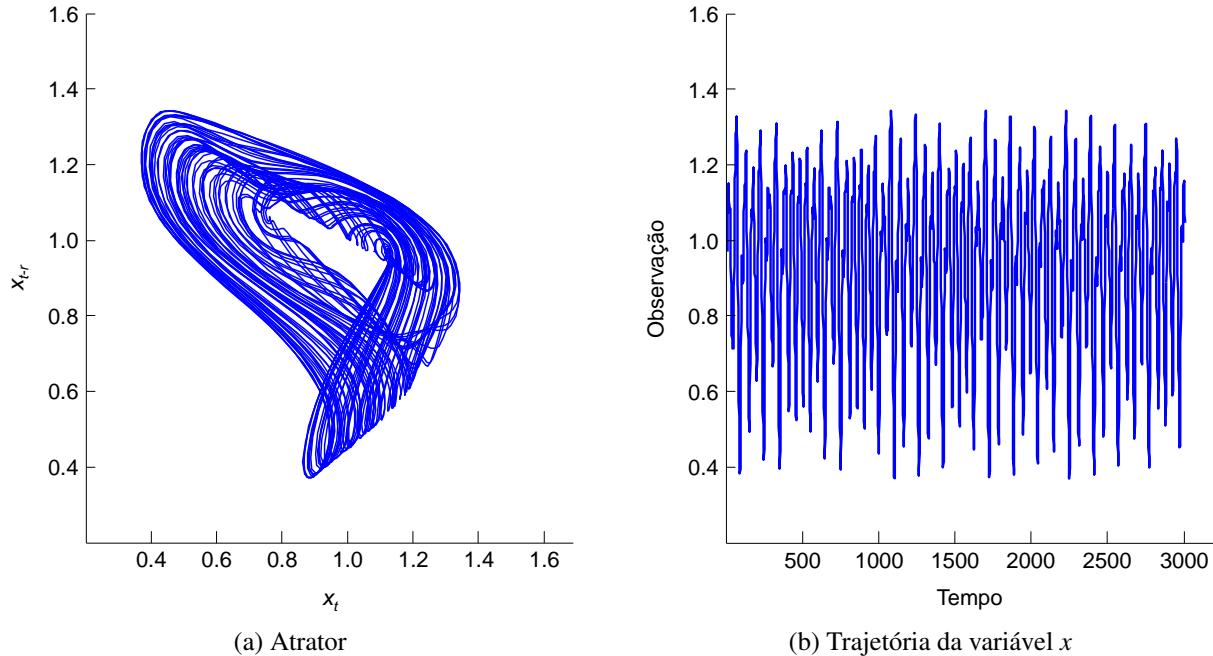
Sistema de Mackey-Glass: Permite a criação de séries caóticas por meio da integração da

Equação B.7 (GLASS; 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.

$$\frac{d(x)}{dt} = \frac{\beta \theta^n x(t - \tau)}{x(t - \tau)^m + \theta^m} - \gamma x \quad (\text{B.7})$$

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 92](#), foi obtida com auxílio do método iterativo de Runge-Kutta de quarta ordem ([CHRISTODOULOU, 2009](#)) para a aproximação numérica da Equação B.7.

Figura 92 – Sistema de Mackey-Glass



Fonte: Elaborada pelo autor.

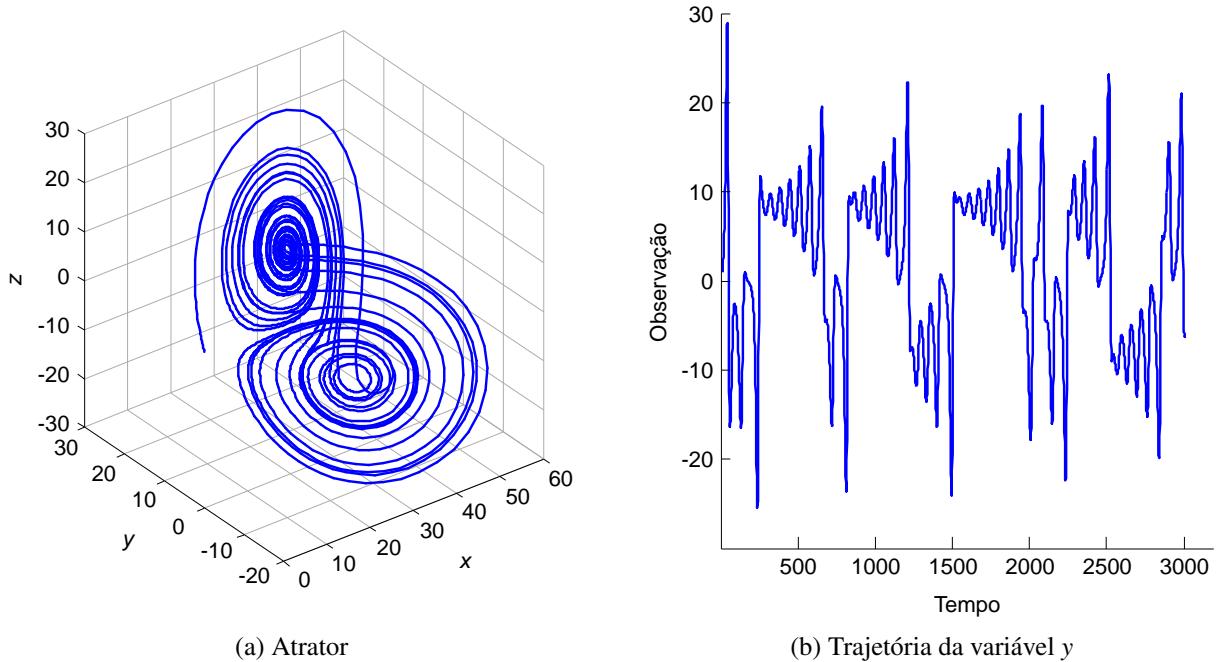
Sistema de Lorenz: Esse sistema foi derivado de um modelo matemático tridimensional relacionado ao estudo do movimento convectivo de fluidos na atmosfera terrestre. O sistema de Lorenz é composto por três equações que podem ser expressas do seguinte modo ([TAHER-SIMA et al., 2010](#)):

$$\begin{cases} x(t+1) = x(t) + d\alpha(y(t) - x(t)) \\ y(t+1) = y(t) + d(x(t)(r - z(t)) - y(t)) \\ z(t+1) = z(t) + d(x(t)y(t) - \beta z(t)) \end{cases} \quad (\text{B.8})$$

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 93](#) 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 93 – Sistema de Lorenz



Fonte: Elaborada pelo autor.

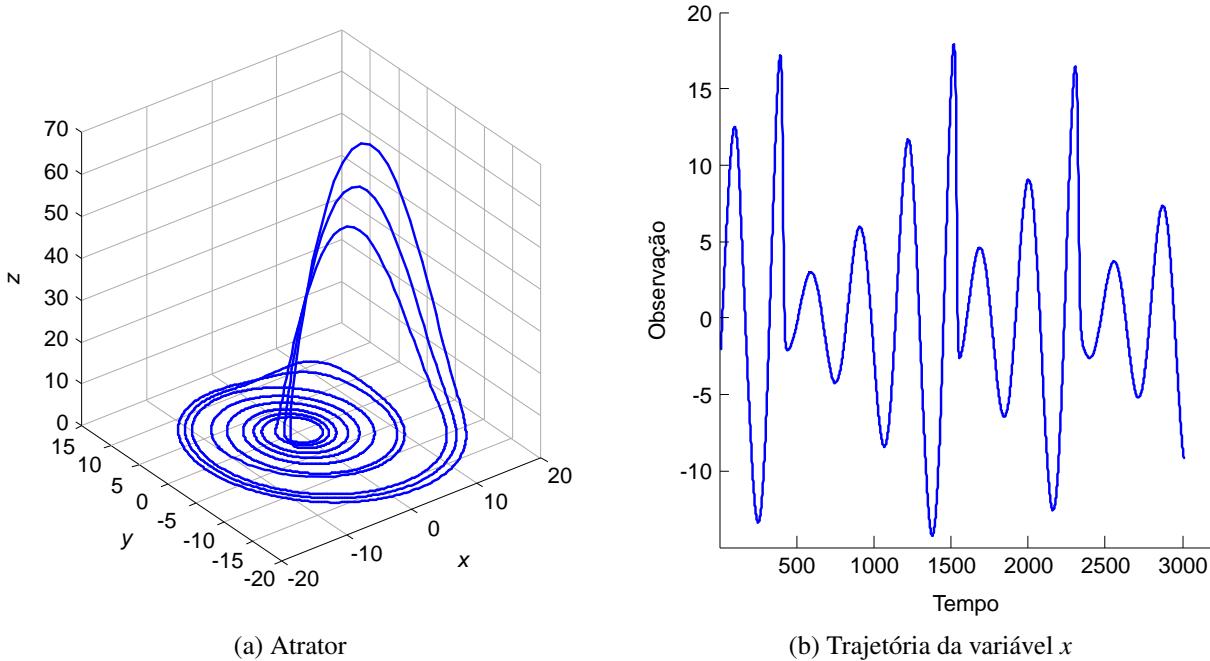
Sistema de Rössler: Formulado com base no modelo de Lorenz, o sistema de Rössler não objetiva a simulação de nenhum fenômeno físico real conhecido e pode ser determinado em conformidade com as seguintes equações ([FRUNZETE; POPESCU; BARBOT, 2015](#)):

$$\begin{cases} x(t+1) = x(t) - d(y(t) + z(t)) \\ y(t+1) = y(t) + d(x(t) + ay(t)) \\ z(t+1) = z(t) + d(b + z(t))(x(t) - c) \end{cases} \quad (\text{B.9})$$

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 94.

Figura 94 – Sistema de Rössler



Fonte: Elaborada pelo autor.

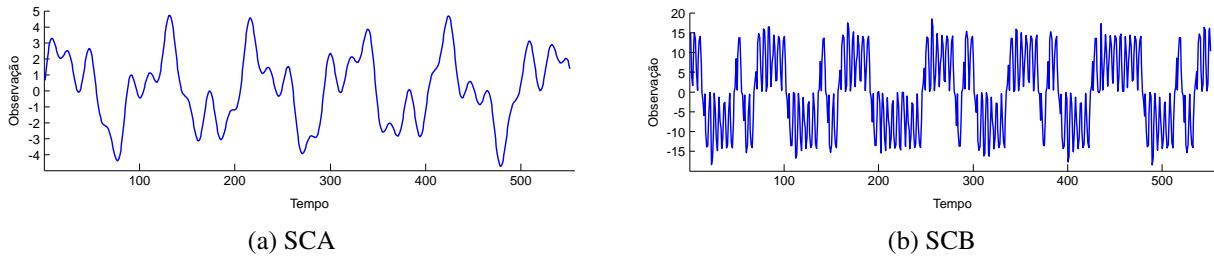
Sinais Caóticos: Esses sinais, além de apresentarem dependência sensível às condições iniciais e parametrizações, são determinísticos e aperiódicos. O estudo dos sinais caóticos é útil, sobretudo, para algumas subáreas da Engenharia de Telecomunicações por abranger espectro de Fourier plano, dificuldade de predição e serem facilmente confundíveis com ruído (EISENCRIFT, 2001). Neste trabalho, o Sinal Caótico A (SCA) foi construído utilizando a Equação B.10, enquanto que o Sinal Caótico B (SCB) foi construído usando a Equação B.11.

$$S = \frac{5}{2} \sin\left(\frac{2\pi t}{100}\right) + \frac{3}{2} \sin\left(\frac{2\pi t}{41}\right) + \sin\left(\frac{2\pi t}{21}\right) \quad (\text{B.10})$$

$$x_t = \frac{x_{t-1}}{2} + 25 \frac{x_{t-1}}{1+x_{t-1}^2} + 7 \cos(1,2t) \quad (\text{B.11})$$

Na Figura 95 são exibidos graficamente os sinais caóticos sintéticos projetados para 550 observações. É importante destacar que a aplicação da Equação B.11 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.

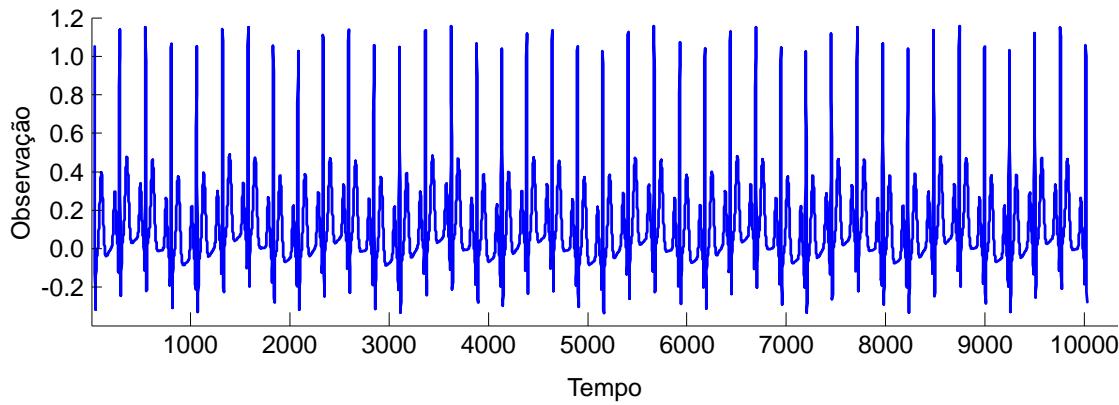
Figura 95 – Sinais Caóticos



Fonte: Elaborada pelo autor.

ECGSYN: Série constituída por 3000 observações e cujos valores foram obtidos a partir de um simulador de eletrocardiograma sustentado por três equações diferenciais ordinárias ([MC SHARRY *et al.*, 2003](#)). O comportamento dessa série é ilustrado na [Figura 96](#).

Figura 96 – Série de valores ECGSYN



Fonte: Elaborada pelo autor.

Conjuntos de Dados Reais

Os conjuntos de dados reais foram selecionados a partir do resultado da revisão sistemática apresentada no [Capítulo 2](#). Esses dados, descritos a seguir, foram considerados com o objetivo futuro de viabilizar a comparação do algoritmo proposto nesta dissertação com métodos desenvolvidos em outros trabalhos que se utilizam das mesmas ST.

Atmosfera: Dados de (i) temperatura, medida em graus Celsius, e (ii) umidade relativa do ar, expressa em percentagem, ao meio dia, na cidade de São Paulo. Observações diárias de 1 de janeiro a 31 de dezembro de 1997;

Banespa: Preços diários das ações, sem direito a voto (PN), do Banco do Estado de São Paulo (Banespa). Observações adquiridas no período de 3 de janeiro de 1995 a 27 de dezembro de 2000;

Beer: Produção trimestral de cerveja, em milhões de barris, nos Estados Unidos da América (EUA). Observações realizadas no período de janeiro de 1975 a dezembro de 1982;

Bebida: Produção física de alimentos e bebidas para indústria. Observações mensais de janeiro de 1985 a julho de 2000;

CBE: Produção mensal de (i) chocolate (toneladas), (ii) cerveja (Ml) e (iii) energia elétrica (milhões de kWh) na Austrália. Observações coletadas no período de janeiro de 1958 a dezembro de 1990 e disponibilizadas pela Agência Nacional de Estatísticas da Austrália;

CEMIG: Preços diários das ações da Companhia Energética de Minas Gerais (CEMIG) no período de 3 de janeiro de 1995 a 27 de dezembro de 2000;

Chicken: Dados mensais, adquiridos em kg, relativos à exportação de frango produzido na Tailândia. Observações mensuradas no período de janeiro de 1999 a julho de 2014 e cedidas pelo Ministério tailandês da Agricultura e Cooperativas por meio do Gabinete da Economia Agrícola em parceria com o Departamento de Alfândegas de Bangkok;

Consumo: Dados referentes às vendas físicas na região metropolitana de São Paulo. Observações mensais de janeiro de 1984 a outubro de 1996;

Darwin: Valores mensais da pressão ao nível do mar nas estações de Darwin, localizada no norte da Austrália, entre os anos de 1882 e 1998. Essa série, por conter importantes padrões climatológicos, tem sido usada em diversos estudos relacionados ao Índice de Oscilação do Sul e, consequentemente, ao fenômeno El Niño;

Dow Jones: Valores mensais de fechamento do Dow Jones *Industrial Average*, o qual constitui o principal índice acionário das bolsas de valores dos EUA. Observações realizadas no período de janeiro de 1950 a maio de 2003;

ECG: Frequência cardíaca, em bastimentos por minuto (bpm), obtida do eletrocardiograma de dois indivíduos (A e B) com níveis diferentes de condicionamento físico. Durante o procedimento, após um eletrodo ser fixado à pele, os indivíduos foram envolvidos em atividades passíveis de comparação. As medições, coletadas em intervalos de 0,5 segundo ao longo de 15 minutos, foram disponibilizadas pelo Centro Médico Beth Israel Deaconess de Boston, EUA, em outubro de 1996;

Energia: Valores mensais do consumo de energia elétrica no estado do Espírito Santo. Observações adquiridas no período de janeiro de 1968 a setembro de 1979;

Fortaleza: Precipitação atmosférica na cidade de Fortaleza, capital do Ceará. Observações anuais coletadas entre os anos de 1849 e 1997;

Global: Aumento mensal da temperatura global no período de janeiro de 1856 a dezembro de 2005. Segundo a Convenção-Quadro das Nações Unidas sobre a Mudança do Clima, a

temperatura média global deverá continuar aumentando no futuro, a menos que as emissões de gases de efeito estufa sejam reduzidas em escala global;

IBV: Índice diário da Bolsa de Valores de São Paulo (BOVESPA) no período de 3 de janeiro de 1995 a 27 de dezembro de 2000;

ICV: Índice do Custo de Vida (ICV) no município de São Paulo. Observações mensais de janeiro de 1970 a junho de 1980;

IPI: Fabricação de produtos alimentares para indústria. Observações mensais de janeiro de 1985 a julho de 2000;

Latex: Dados mensais, adquiridos em kg, relativos à exportação de látex produzido na Tailândia. Observações mensuradas no período de janeiro de 1998 a julho de 2014 e cedidas pelo Ministério tailandês da Agricultura e Cooperativas por meio do Gabinete da Economia Agrícola em parceria com o Departamento de Alfândegas de Bangkok;

Lavras: Precipitação atmosférica no município de Lavras, Minas Gerais. Observações mensais realizadas no período de janeiro de 1966 a dezembro de 1997;

Laser: Dados coletados em laboratório, durante um experimento controlado de física, e que correspondem à intensidade de pulsação de um *far-infrared laser* de amônia (NH_3) sob estado caótico ao longo de 1000 segundos. As medições de intensidade, além de terem sido registradas por um osciloscópio LeCroy, foram utilizadas pela primeira vez no ano de 1991, em uma competição entre métodos de predição promovida pelo Santa Fe *Institute*;

Maine: Taxa desemprego mensal para o estado do Maine considerando relatório produzido no período de janeiro de 1996 a agosto de 2006. Observações calculadas conforme definições estipuladas pelo Ministério do Trabalho dos EUA;

Manchas: Número de manchas solares de Wölfer. Observações anuais entre os anos de 1749 e 1924;

MPrime: Dados provindos da Reserva Federal dos EUA e que refletem a taxa mensal de empréstimo percentual no período de janeiro de 1949 a novembro de 2007. Observações fornecidas pelo Conselho de Governadores do Sistema da Reserva Federal;

OSVisit: Números mensais de visitantes estrangeiros na Nova Zelândia entre os anos de 1977 e 1995;

Ozônio: Valores mensais de concentração de ozônio em Azusa, Califórnia. Observações adquiridas no período de janeiro de 1956 a dezembro de 1970;

Patient Demand: Número de pacientes diários atendidos pelo departamento de emergência hospitalar em um hospital coreano. Observações realizadas no período de 1 de janeiro de 2007 a 31 de março de 2009;

Petrobras: Preços diários das ações, sem direito a voto (PN), da Petrobras. Observações coletadas no período de 3 de janeiro de 1995 a 27 de dezembro de 2000;

PFI: Produção Física Industrial (PFI) referente à indústria geral. Observações mensais de janeiro de 1991 a julho de 2000;

Poluição: Emissão diária dos seguintes poluentes na cidade de São Paulo: (i) material particulado (PM_{10}), (ii) dióxido de enxofre (SO_2), (iii) monóxido de carbono (CO), (iv) ozônio (O_3) e (v) dióxido de nitrogênio (NO_2). Observações adquiridas no período de 1 de janeiro a 31 de dezembro de 1997;

Reservoir: Valores mensais de entrada de água, expressos em m^3/s , no reservatório fonte em Northumberland, Inglaterra. Observações realizadas no período de janeiro de 1909 a dezembro de 1980 e fornecidas pela companhia de abastecimento de água de Northumberland;

São Carlos: Dados referentes ao (i) nível do rio Monjolinho antes e (ii) após o encontro de suas águas com o córrego Tijuco Preto. Observações realizadas, em intervalos de cinco minutos, a partir de uma Rede de Sensores Sem Fio implantada na cidade de São Carlos entre os anos de 2013 e 2014;

Star: Medições do brilho (magnitude) de uma estrela oscilante. Valores coletados diariamente, sempre no mesmo local e à meia-noite, entre os anos de 1922 e 1924;

STemp: Dados referentes à temperatura do hemisfério sul extraídos do banco de dados mantido pela Unidade de Pesquisa Climática da Universidade de East Anglia, Reino Unido. Observações mensais de janeiro de 1850 a dezembro de 2007;

Stock Market: Dados do mercado de ações para sete cidades: (i) Amsterdã, (ii) Frankfurt, (iii) Londres, (iv) Hong Kong, (v) Japão, (vi) Singapura e (vii) Nova York. Observações diárias de 6 janeiro de 1986 a 31 de dezembro de 1997;

Super Bowl: Preço dos ingressos para o Super Bowl, o jogo do campeonato da Liga Nacional de Futebol Americano (NFL). Observações anuais entre os anos de 1985 e 2006;

Temperatura: Valores médios de temperatura, em graus Celsius, nos municípios de (i) Cananéia e (ii) Ubatuba, São Paulo. Observações mensais de janeiro de 1976 a dezembro de 1985;

Truck: Número médio diário de defeitos de fabricação em caminhões nos EUA. Os dados, coletados ao longo de seis semanas e meia, foram disponibilizados gratuitamente no ano de 1994;

USA: Taxa mensal de desemprego nos EUA. Observações adquiridas no período de janeiro de 1996 a outubro de 2006;

Wine: Vendas mensais de vinho australiano, em milhares de litros, de acordo com as categorias:
(i) fortificado branco, (ii) branco seco, (iii) branco doce, (iv) tinto, (v) rosé e (vi) espumante.
Valores obtidos no período de janeiro de 1980 a julho de 1995.