



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Domingos Sávio de Oliveira Santos Júnior

Método de *Ensemble* para correção de modelos ARIMA: uma abordagem de sistema híbrido para previsão de séries temporais

Recife

2022

Domingos Sávio de Oliveira Santos Júnior

Método de *Ensemble* para correção de modelos ARIMA: uma abordagem de sistema híbrido para previsão de séries temporais

Trabalho apresentado ao Programa de Pós-graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco, como requisito parcial para obtenção do grau de Doutor em Ciência da Computação

Área de Concentração: Inteligência Computacional

Orientador: Paulo Salgado Gomes de Mattos Neto

Coorientadores: George Darmiton da Cunha Cavalcanti e João Fausto Lorenzato de Oliveira

Recife

2022

Catálogo na fonte
Bibliotecária Nataly Soares Leite Moro, CRB4-1722

S237m Santos Júnior, Domingos Sávio de Oliveira
Método de *Ensemble* para correção de modelos ARIMA: uma abordagem de sistema híbrido para previsão de séries temporais / Domingos Sávio de Oliveira Santos Júnior. – 2022.
80 f.: il., fig., tab.

Orientador: Paulo Salgado Gomes de Mattos Neto.
Tese (Doutorado) – Universidade Federal de Pernambuco. CIn, Ciência da Computação, Recife, 2022.
Inclui referências e anexos.

1. Inteligência computacional. 2. Previsão de séries temporais. 3. ARIMA. 4. Rede neural artificial. 5. Ensemble. I. Mattos Neto, Paulo Salgado Gomes de (orientador). II. Título

006.31

CDD (23. ed.)

UFPE - CCEN 2022 – 83

Domingos Sávio de Oliveira Santos Júnior

“Método de Ensemble para correção de modelos ARIMA: uma abordagem de sistema híbrido para previsão de séries temporais”

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutor em Ciência da Computação. Área de Concentração: Inteligência Computacional

Aprovado em: 17/02/2022.

Orientador: Prof. Dr. Paulo Salgado Gomes de Mattos Neto

BANCA EXAMINADORA

Prof. Dr. Ricardo Bastos Cavalcante Prudêncio
Centro de Informática / UFPE

Prof. Dr. Adriano Lorena Inacio de Oliveira
Centro de Informática / UFPB

Prof. Dr. Manoel Henrique da Nóbrega Marinho
Escola Politécnica de Pernambuco / UPE

Prof. Dr. Paulo Renato Alves Firmino
Centro de Ciências e Tecnologia / UFCA

Prof. Dr. Hugo Valadares Siqueira
Departamento Acadêmico de Engenharia Elétrica / UTFPR

Dedico à minha família

AGRADECIMENTOS

Primeiramente agradeço aos meus orientadores, Paulo, George e Fausto, por toda a parceria, os ensinamentos e a paciência durante esses quatro anos. Em especial ao Paulo, meu orientador desde o mestrado e que sempre esteve disponível para me ajudar, explicar e reexplicar sempre que precisei e, com toda certeza, o pesquisador que almejo ser é espelhado em sua forma de trabalhar.

Agradeço aos meus pais, que me apoiaram incondicionalmente incentivando-me a todo momento, sempre fizeram o possível e o impossível para que minha vida sempre fosse o mais leve e completa.

Agradeço também a minha esposa por toda parceria e apoio nos períodos em que precisei focar nas atividades de pesquisa, ela esteve comigo me apoiando durante todo o processo e acreditando no meu potencial em momentos em que nem eu mesmo acreditava.

RESUMO

Nas últimas décadas Sistemas Híbridos (SH) que utilizam a modelagem residual têm sido amplamente aplicados no contexto de previsão de séries temporais. Esta abordagem utiliza como resultado final a combinação da previsão de um modelo linear com a previsão do resíduo obtida por um modelo de Aprendizagem de Máquina (AM) não linear. Essa série de resíduo representa a diferença entre a saída do modelo linear e valor real da série temporal. Uma vez que normalmente são encontrados padrões lineares e não lineares em séries temporais reais, esta classe de SH tem alcançado resultados empíricos e teóricos promissores em razão da sua arquitetura ser capaz de modelar esses padrões em etapas específicas. Contudo, são identificadas limitações na etapa de modelagem residual, sendo que por conta de sua complexidade, um modelo de AM pode apresentar problemas de má especificação de parâmetros, sobreajuste e subajuste, prejudicando os resultados de todo o SH. Baseado neste problema, este trabalho propõe um método de *ensemble* para previsão residual (*Ensemble method for Residual Forecast* (ERF)). O método ERF é composto por três fases gerais: (i) previsão da série temporal por meio de um modelo linear; (ii) previsão do erro realizada por um *ensemble*; (iii) combinação pela soma das previsões das fases (i) e (ii). A fase (ii) é a principal contribuição desta tese, na qual é proposta uma abordagem homogênea que cria um *ensemble* de modelos de AM diverso e de forma sistemática. O ARIMA é utilizado como modelo linear, já como modelo não linear são avaliados o MLP e SVR. Desta forma, são obtidas duas versões do método proposto. Essas versões são aplicadas em doze séries temporais reais com os respectivos modelos simples (ARIMA, MLP e SVR) e oito sistemas híbridos da literatura. Todos os métodos são avaliados por meio da métrica Raiz do Erro Quadrático Médio e testes estatísticos de Wilcoxon, Friedman e Nemenyi. Com base nessas formas de avaliação, visualiza-se que as abordagens propostas possuem a capacidade de encontrar bons resultados quando aplicadas em diferentes séries temporais.

Palavras-chaves: previsão de séries temporais; ARIMA; rede neural artificial; máquina de vetores de suporte para regressão; sistemas híbridos; modelagem residual; *ensemble*.

ABSTRACT

Hybris systems (HS) that use residual modeling are widely applied in the time series forecasting context. This approach uses as the final forecast, the combination of a linear model forecast and a residual forecast performed by a nonlinear Machine Learning (ML) model. This residual series is the difference between the time series values and linear forecast. Due to real-world time series normally show both linear and nonlinear patterns, this class of SH attained good empirical and theoretical results because its architecture model linear and nonlinear patterns in specific steps. However, are identified some limitations in the residual modeling phase because an ML model can present problems such as overfitting, underfitting, and miss specification of parameters. This problem occurs due to residual series complexity and can affect the final results of the HS negatively. Based on this problem, this work proposes an Ensemble method for Residual Forecast (ERF). The ERF is composed of three general phases: (i) time series forecasting through a linear model; (ii) residual forecasting performed by an ensemble; (iii) sum of the forecasts of (i) and (ii). Phase (ii) is the main contribution of this work, in which is proposed a homogeneous approach that creates a diverse ensemble of ML models in a systematic form. In the experimental analysis, the ARIMA is used as linear model, and as nonlinear model is evaluated the MLP and SVR, so are created two versions of the proposed method. These versions are applied in twelve time series with the single models (ARIMA, MLP e SVR) and seven hybrid systems from the literature. All methods are evaluated through the Root Mean Squared Error and statistical tests of Wilcoxon, Friedman e Nemenyi. Following these evaluation strategies, is observed that the proposed methods found good results in different time series.

Keywords: time series forecasting; ARIMA; artificial neural network; support vector machine for regression; hybrid systems; residual modeling; ensemble.

LISTA DE FIGURAS

Figura 1 – Fases do método proposto	35
Figura 2 – Exemplo de amostra gerada por <i>Random Patches</i> com $m = 3$, $bs = 50\%$, e $fs = 50\%$. Considerando um conjunto de treino de uma série residual (E_t^{tr}) com dezenove pontos e <i>lags</i> iguais a quatro	37
Figura 3 – Série temporal <i>Airlines Passengers</i>	40
Figura 4 – FACP e FAC da série <i>Airlines Passengers</i>	41
Figura 5 – Série temporal <i>Canadian Lynx</i>	41
Figura 6 – FACP e FAC da série <i>Canadian Lynx</i>	42
Figura 7 – Série temporal <i>Colorado River</i>	42
Figura 8 – FACP e FAC da série <i>Colorado River</i>	43
Figura 9 – Série temporal <i>International Business Machines Corporation (IBM)</i>	43
Figura 10 – FACP e FAC da série <i>IBM</i>	44
Figura 11 – Série temporal <i>Lake Erie</i>	45
Figura 12 – FACP e FAC da série <i>Lake Erie</i>	45
Figura 13 – Série temporal <i>Milk</i>	46
Figura 14 – FACP e FAC da série <i>Milk</i>	46
Figura 15 – Série temporal <i>Ozon</i>	47
Figura 16 – FACP e FAC da série <i>Ozon</i>	47
Figura 17 – Série temporal <i>Paper</i>	48
Figura 18 – FACP e FAC da série <i>Paper</i>	48
Figura 19 – Série temporal <i>Pollution</i>	49
Figura 20 – FACP e FAC da série <i>Pollution</i>	49
Figura 21 – Série temporal <i>Star Bright</i>	50
Figura 22 – FACP e FAC da série <i>Star Bright</i>	50
Figura 23 – Série temporal <i>Sunspot</i>	51
Figura 24 – FACP e FAC da série <i>Sunspot</i>	51
Figura 25 – Série temporal <i>Temperature</i>	52
Figura 26 – FACP e FAC da série <i>Temperature</i>	52
Figura 27 – Análise de sensibilidade dos parâmetros do método proposto ERF_{MLP} para as séries temporais (a) <i>Milk</i> , (b) <i>Paper</i> , (c) <i>Pollution</i> e (d) <i>River</i>	59

Figura 28 – Valor real da série temporal (linha preta) com as respectivas previsões, aplicando o <i>Autoregressive Integrated Moving Average</i> (ARIMA) (linha vermelha), ERF_{MLP} (linha azul) e ERF_{SVR} (linha verde)	64
Figura 29 – Classificação do teste de Nemenyi considerando o RMSE	66
Figura 30 – Análise de sensibilidade dos parâmetros do método proposto ERF_{MLP} para as séries temporais (a) <i>Airlines</i> , (b) <i>IBM</i> , (c) <i>Lake</i> e (d) <i>Lynx</i>	77
Figura 31 – Análise de sensibilidade dos parâmetros do método proposto ERF_{MLP} para as séries temporais (a) <i>Ozon</i> , (b) <i>Star</i> , (c) <i>Sunspot</i> e (d) <i>Temp</i>	78

LISTA DE TABELAS

Tabela 1 – Lista de séries temporais aplicadas	39
Tabela 2 – Nomenclatura dos métodos propostos e da literatura aplicados	53
Tabela 3 – Valores dos parâmetros para os modelos lineares e não lineares usados pelo ERF e métodos da literatura.	55
Tabela 4 – Parâmetros do <i>ensemble</i> selecionados baseado no RMSE de validação	58
Tabela 5 – Resultados do RMSE para as séries temporais utilizando os métodos simples e propostos. O melhor e segundo melhor estão destacados com negrito e sublinhado, respectivamente	60
Tabela 6 – Resultados do RMSE para as séries temporais utilizando os Sistemas Híbridos da literatura e propostos. O melhor e segundo melhor estão destacados em negrito e sublinhado, respectivamente	61
Tabela 7 – Teste de hipótese Wilcoxon com 95% de confiança. Os símbolos (+), (-) ou (=) representam que o método proposto ERF_{MLP} é melhor, pior ou equivalente ao outro método, respectivamente.	61
Tabela 8 – Teste de hipótese Wilcoxon com 95% de confiança. Os símbolos (+), (-) ou (=) representam que o método proposto ERF_{SVR} é melhor, pior ou equivalente ao outro método, respectivamente.	62
Tabela 9 – Resultados do <i>Mean Absolute Percentage Error</i> - MAPE (média das 30 execuções) para as séries temporais utilizando os métodos literatura e propostos no conjunto de teste. O melhor e segundo melhor estão destacados em negrito e sublinhado, respectivamente. Os valores reais e preditos são normalizados entre 0,1 e 0,9, devido que existem séries que possuem valores iguais a 0	79
Tabela 10 – Resultados do <i>Mean Absolute Error</i> - MAE (média das 30 execuções) para as séries temporais utilizando os métodos literatura e propostos no conjunto de teste. O melhor e segundo melhor estão destacados em negrito e sublinhado, respectivamente	80

LISTA DE ABREVIATURAS E SIGLAS

AIC	<i>Akaike's Information Criterion</i>
AM	<i>Aprendizagem de Máquina</i>
AR	<i>Autoregressive</i>
ARIMA	<i>Autoregressive Integrated Moving Average</i>
ARMA	<i>Autoregressive Moving Average</i>
DBN	<i>Deep Belief Network</i>
DReF	<i>Dynamic Residual Forecasting</i>
ERF	<i>Ensemble method for Residual Forecast</i>
ESN	<i>Echo State Network</i>
FAC	<i>Função de Autocorrelação</i>
FACP	<i>Função de Autocorrelação Parcial</i>
IBM	<i>International Business Machines Corporation</i>
KPSS	<i>Kwiatkowski Phillips Schmidt Shin</i>
LSTM	<i>Long Short-Term Memory network</i>
MA	<i>Moving Average</i>
MLP	<i>Multilayer Perceptron</i>
PSO	<i>Particle Swarm Optimization</i>
RBF	<i>Radial Basis Function network</i>
RMSE	<i>Root Mean Square Error</i>
SH	<i>Sistemas Híbridos</i>
SKU	<i>Stock Keeping Unit</i>
SVR	<i>Support Vector machine for Regression</i>
TSDL	<i>Time Series Data Library</i>

SUMÁRIO

1	INTRODUÇÃO	14
1.1	CONTEXTO E MOTIVAÇÃO	14
1.2	OBJETIVO	17
1.3	ESTRUTURA DA TESE	17
2	TRABALHOS RELACIONADOS	19
2.1	SÉRIES TEMPORAIS	19
2.2	PREVISÃO DE SÉRIES TEMPORAIS	21
2.2.1	<i>Autoregressive Integrated Moving Average</i>	22
2.2.2	Rede Neural Perceptron de Multicamadas	24
2.2.3	Máquinas de Vetores de Suporte para Regressão	24
2.3	COMBINAÇÃO DE PREDITORES	26
2.3.1	<i>Ensemble</i>	27
2.3.1.1	<i>Geração</i>	28
2.3.1.2	<i>Seleção</i>	29
2.3.1.3	<i>Agregação</i>	30
2.3.2	Sistemas Híbridos Residuais	30
2.4	RESUMO DO CAPÍTULO	34
3	MÉTODO PROPOSTO	35
3.1	FASE DE TREINAMENTO	36
3.2	FASE DE TESTE	37
3.3	RESUMO DO CAPÍTULO	38
4	METODOLOGIA	39
4.1	BASE DE DADOS	39
4.1.1	<i>Airlines Passengers</i>	40
4.1.2	<i>Canadian Lynx</i>	41
4.1.3	<i>Colorado River</i>	42
4.1.4	IBM	43
4.1.5	<i>Lake Erie</i>	44
4.1.6	<i>Milk</i>	45
4.1.7	<i>Ozon</i>	46

4.1.8	<i>Paper</i>	47
4.1.9	<i>Pollution</i>	48
4.1.10	<i>Star Bright</i>	49
4.1.11	<i>Sunspot</i>	50
4.1.12	<i>Temperature</i>	51
4.2	MÉTODOS APLICADOS	52
4.3	RESUMO DO CAPÍTULO	56
5	ANÁLISE EXPERIMENTAL	57
5.1	ANÁLISE DE SENSIBILIDADE	57
5.2	RESULTADOS EXPERIMENTAIS	60
5.3	ANÁLISE FINAL	65
5.4	RESUMO DO CAPÍTULO	66
6	CONCLUSÃO	68
6.1	CONSIDERAÇÕES FINAIS	68
6.2	TRABALHOS FUTUROS	69
	REFERÊNCIAS	71
	ANEXO A – ANÁLISE DE SENSIBILIDADE	77
	ANEXO B – RESULTADOS DE MÉTRICAS ADICIONAIS	79

1 INTRODUÇÃO

Este capítulo apresenta a introdução deste trabalho, compreendendo os seguintes elementos: o contexto e a motivação da abordagem proposta na tarefa de previsão de séries temporais, objetivos de pesquisa, e a organização dos próximos capítulos.

1.1 CONTEXTO E MOTIVAÇÃO

O contexto da previsão de séries temporais tem atraído a atenção da comunidade científica nas últimas décadas devido a sua importância em muitas aplicações do conhecimento, tais como: mercado financeiro, economia, meteorologia e mercado varejista (CHATFIELD, 2003; BOX et al., 2015).

Métodos estatísticos e de Aprendizagem de Máquina (AM) são amplamente aplicados na área de previsão de séries temporais (BOX et al., 2015; ZHANG; PATUWO; HU, 1998). Os modelos Autorregressivos (*Autoregressive* (AR)), Média Móveis (*Moving Average* (MA)) e Autorregressivos Integrado de Média Móveis (ARIMA) são exemplos de métodos estatísticos amplamente aplicados na literatura (BOX et al., 2015). Estas abordagens possuem metodologias bem definidas para identificação do melhor modelo, assumindo que a série temporal é composta por estruturas lineares (BOX et al., 2015; HYNDMAN; KHANDAKAR et al., 2007). Desta forma, estes modelos podem apresentar acurácia reduzida em séries temporais do mundo real, dado que estas séries normalmente apresentam ambos os padrões lineares e não lineares (ZHANG, 2003; SANTOS JÚNIOR; de OLIVEIRA; de MATTOS NETO, 2019).

Alternativamente, métodos de AM são estratégias baseadas em dados que podem modelar padrões não lineares da série temporal, Rede Neural Perceptron de Multicamadas (*Multilayer Perceptron* (MLP)) (ZHANG; PATUWO; HU, 1998) e Máquinas de Vetores de Suporte para Regressão (*Support Vector machine for Regression* (SVR)) (VAPNIK et al., 1997) são exemplos de abordagens deste tipo. Contudo, modelos de aprendizagem de máquina podem não capturar igualmente bem os padrões lineares e não lineares devido a problemas como má especificação de parâmetros, sobreajuste (*overfitting*) e subajuste (*underfitting*) (ZHANG, 2003; TASKAYA-TEMIZEL; CASEY, 2005; de MATTOS NETO et al., 2020).

Muitas abordagens têm combinado métodos lineares e não lineares com o objetivo de resolver as limitações mencionadas acima (ZHANG, 2003; PAI; LIN, 2005; KHASHEI; BIJARI,

2011; SANTOS JÚNIOR; de OLIVEIRA; de MATTOS NETO, 2019; de MATTOS NETO et al., 2020; PUROHIT et al., 2021; XU et al., 2021). Entre essas, Sistemas Híbridos (SH) que combinam modelos estatísticos com de AM por meio de modelagem da série residual¹ têm alcançado resultados teóricos e empíricos promissores (ZHANG, 2003; KHASHEI; BIJARI, 2011; SANTOS JÚNIOR; de OLIVEIRA; de MATTOS NETO, 2019; de MATTOS NETO et al., 2020). Esta classe de SH combina a previsão da série temporal realizada pelo modelo estatístico com a previsão do resíduo executada pelo modelo de AM. Esta combinação é normalmente feita por meio de soma (ZHANG, 2003; PANIGRAHI; BEHERA, 2017; PUROHIT et al., 2021). O objetivo principal é a melhora dos resultados finais em termos de acurácia, criando um sistema capaz de modelar, em etapas separadas, os padrões lineares e não lineares.

Um dos maiores desafios na criação um SH preciso é a fase de modelagem residual. O treinamento do método de AM pode ser uma tarefa complexa, dado que a série de resíduo pode apresentar flutuações aleatórias, heterocedasticidade² e ruído (TASKAYA-TEMIZEL; CASEY, 2005).

Estas características podem enviesar o modelo, reduzindo sua capacidade de generalização. Portanto a má especificação, *underfitting* ou *overfitting* do modelo de AM continua sendo um problema na fase de previsão residual, prejudicando o desempenho final do SH. Além disso, a utilização de um único modelo não linear pode não ser o suficiente para capturar flutuações variáveis no tempo (padrões locais) devido à heterocedasticidade.

Neste contexto, um *ensemble* de modelos de AM pode ser utilizado para lidar com os problemas mencionados acima. Esta abordagem combina previsões de diversos modelos com o objetivo de criar uma única saída final (SHEN et al., 2021). Esta estratégia é amplamente aplicada na literatura e, quando comparada a um modelo simples, consegue lidar melhor com os impactos do *overfitting* e *underfitting* (BREIMAN, 1996; REN; ZHANG; SUGANTHAN, 2016). Em relação à decomposição do viés-variância (*bias-variance*) do erro (KEIJZER; BABOVIC, 2000), abordagens de *ensemble* possui o objetivo de reduzir a variância do erro por meio da combinação de diferentes previsões. Contudo, é necessário existir modelos acurados e diversos no *ensemble*, obtendo uma combinação de pontos fortes de diferentes previsores (BROWN et al., 2005; MENDES-MOREIRA et al., 2012).

Considerando os desafios exibidos no contexto de previsão residual e as capacidades das

¹ A série de resíduo é obtida a partir da diferença entre o valor real da série temporal e a respectiva previsão.

² Séries com heterocedasticidade exibem em seus dados históricos períodos com diferentes níveis de volatilidade ou variabilidade (BOX et al., 2015).

abordagens de *ensemble*, este trabalho propõe um novo SH que combina um modelo estatístico linear com um *ensemble* de modelos de AM por meio da modelagem residual. Esta abordagem é chamada de Método de *Ensemble* para Previsão Residual (*Ensemble method for Residual Forecast* (ERF)).

O método ERF é composto de três fases: (a) modelagem da série temporal por meio um modelo estatístico linear, a partir da qual é obtido o resíduo, (b) modelagem residual feita por um *ensemble* de modelos de AM e (c) combinação por meio da soma das previsões de (a) e (b). O método proposto aplica um *ensemble* homogêneo, no qual cada modelo (utilizando a mesma arquitetura) é treinado com uma amostra diferente dos dados. A abordagem chamada de *Random Patches* (LOUPPE; GEURTS, 2012a) foi escolhida como estratégia para geração dos modelos do *ensemble*. Esta foi selecionada para tal tarefa, pois combina as estratégias de Bagging (BREIMAN, 1996) e *Random Subspace* (HO, 1998) buscando maximizar a diversidade no *ensemble*, como também reduzir os requisitos de memória necessários para criação de um modelo de AM (BREIMAN, 1996; LOUPPE; GEURTS, 2012a; HO, 1998).

Desta forma, esta tese é baseada nas hipóteses de que o método proposto ERF possui as capacidades de:

- Melhorar o poder de generalização na fase de previsão residual por meio da redução da variância do erro;
- Reduzir o risco de selecionar um modelo inapropriado para previsão residual;
- Expandir o espaço de funções representáveis, aumentando a acurácia na previsão da série do resíduo.

O modelo ARIMA é aplicado para a modelagem linear e a MLP e SVR são usadas na modelagem da série do resíduo. Estes modelos foram selecionados por serem amplamente utilizados na literatura de SH (ZHANG, 2003; PAI; LIN, 2005; KHASHEI; BIJARI, 2011; SANTOS JÚNIOR; de OLIVEIRA; de MATTOS NETO, 2019). A seleção dos modelos supracitados busca facilitar a comparação com os modelos da literatura e deixar clara a contribuição do método proposto. Portanto, duas versões do método ERF são avaliadas usando na fase (b) um *ensemble* de MLPs ou SVRs, nomeados como ERF_{MLP} ou ERF_{SVR} , respectivamente.

Os métodos propostos são avaliados em doze séries temporais reais utilizando a métrica de raiz do erro médio quadrático (*Root Mean Square Error* (RMSE)), e os testes estatísticos de Wilcoxon, Friedman e Nemenyi. Nesse contexto os métodos propostos são comparados aos

modelos simples ARIMA, MLP e SVR e oito SHs da literatura (ZHANG, 2003; PAI; LIN, 2005; KHASHEI; BIJARI, 2011; WANG et al., 2013; YAN; MA, 2016; XU et al., 2019; XU et al., 2021; de HOLANDA; de OLIVEIRA, 2020).

1.2 OBJETIVO

O objetivo geral deste trabalho é propor um sistema híbrido que combine um modelo linear estatístico com um *ensemble* por meio da modelagem residual. Esta abordagem tem o objetivo de criar um sistema híbrido que consiga corrigir as previsões do modelo linear quando aplicado em diferentes problemas de previsão de séries temporais. Dado este objetivo geral, os objetivos específicos são:

- Desenvolver estratégia de geração de *ensemble* aplicada ao contexto de previsão residual;
- Propor a arquitetura para treinamento e teste do sistema híbrido;
- Procurar e selecionar os parâmetros do *ensemble* por meio de um método sistemático;
- Aplicar o método proposto em séries temporais reais com diferentes características;
- Analisar o comportamento da variação dos parâmetros do método proposto;
- Realizar avaliação de resultados do método proposto juntamente as abordagens da literatura por meio de métrica de desempenho e testes estatísticos.

1.3 ESTRUTURA DA TESE

A estrutura desta tese é composta de seis capítulos, conforme descritos a seguir:

Capítulo 2 – Trabalhos relacionados: Neste capítulo são exibidos conceitos sobre séries temporais e previsão de séries temporais. Em seguida é detalhada a implementação dos modelos ARIMA, MLP e SVR (modelos simples). Por fim, são exibidas as abordagens de combinação preditores da literatura que serviram de base para a construção do método proposto.

Capítulo 3 – Método proposto: Este capítulo expõe detalhes sobre a estratégia para o treinamento e teste da abordagem proposta. Desta forma são apresentados fluxogramas e exemplos com o principal objetivo de facilitar o entendimento e a replicabilidade.

Capítulo 4 – Metodologia: Este capítulo exibe a metodologia aplicada para o desenvolvimento dos resultados. Portanto, são exibidas as bases de dados, métrica de avaliação e estratégia para seleção de parâmetros dos modelos.

Capítulo 5 – Análise Experimental: Neste capítulo o método proposto é avaliado juntamente às abordagens da literatura seguindo o escopo experimental definido no capítulo de metodologia. Estes resultados são mensurados por meio de uma métrica de avaliação e dos testes estatísticos de Wilcoxon, Friedman e Nemenyi.

Capítulo 6 – Conclusão: Este capítulo apresenta a conclusão desta tese, e expõe pontos importantes deste trabalho e contribuições alcançadas levando em consideração os resultados obtidos. Por fim, são elencadas novas pesquisas que podem ser desenvolvidas a partir desta pesquisa.

2 TRABALHOS RELACIONADOS

Este capítulo aborda pontos relacionados à previsão de séries temporais, explanando sobre conceitos básicos e modelos que são amplamente aplicados nesta área. Por fim, são exibidas estratégias para a combinação de preditores. Todos esses tópicos abordados servem de base para o método proposto nesta tese. Desta forma este capítulo é dividido em seções de Séries Temporais (Seção 2.1), Previsão de Séries Temporais (Seção 2.2) e Combinação de Preditores (Seção 2.3).

2.1 SÉRIES TEMPORAIS

Uma série temporal pode ser considerada como um conjunto de variáveis que estão sequencialmente expostas no tempo (CHATFIELD, 2003). Em termos formais, os dados da série (z_t) estão expostos ao longo do tempo (t) (Equação 2.1), no qual N representa o tamanho da amostra (COCHRANE, 1997).

$$z_0, z_1, \dots, z_N \text{ ou } z_t, t = 0, 1, \dots, N. \quad (2.1)$$

Séries temporais são encontradas nas mais diversas áreas de estudo, alguns exemplo são:

1. quantidade semanal de acidentes de trânsito na avenida Agamenon Magalhães em Recife;
2. valores mensais de temperatura da cidade de Petrolina;
3. registro de marés no porto de Suape.

Os Itens 1 e 2 são exemplos de séries discretas que possuem observações divididas em intervalos iguais em um determinado espaço de tempo. O Item 3 é um exemplo de série contínua que não se apresenta em intervalos fixos. Contudo, é possível aproximar uma série contínua em discreta por meio de amostragem (MORETTIN; TOLOI, 1981). Neste procedimento, é necessário capturar intervalos de tempo iguais Δ , utilizando a mesma unidade temporal (por exemplo: hora, dia ou mês), obtendo assim uma série discreta com N pontos observados no intervalo $[0, T]$, no qual $N = \frac{T}{\Delta}$.

Normalmente supõe-se que uma série temporal é afetada por um ou mais componentes, os quais podem ser caracterizados por (ADHIKARI; AGRAWAL, 2013; COWPERTWAIT; METCALFE, 2009):

- Sazonalidade: são flutuações que ocorrem dentro de intervalo de tempo, ou de forma mais genérica, são padrões que se repetem dentro de um período fixo. Períodos climáticos, condições meteorológicas e costumes culturais, são exemplos de fatores que podem causar variações sazonais;
- Tendência: é uma mudança na série temporal que não se apresenta de forma periódica, cuja forma mais simples apresenta um aumento ou diminuição linear da série;
- Ciclos: são variações que não se repetem de forma fixa (HYNDMAN; KHANDAKAR et al., 2007). Por exemplo, em séries temporais financeiras são exibidos os ciclos de prosperidade, declínio, depressão e recuperação;
- Irregularidades: também conhecidas como variações aleatórias, as quais são influenciadas por causas desconhecidas e não se repetem em um determinado padrão.

Existe também outro conceito importante em séries temporais, chamado de estacionariedade. Este conceito representa que uma série temporal está em equilíbrio estatístico (BOX et al., 2015), podendo se apresentar de forma estrita ou fraca. Estacionariedade é dita como estrita quando seu comportamento probabilístico da série temporal permanece o mesmo durante o tempo (SHUMWAY; STOFFER; STOFFER, 2000). Na estacionariedade fraca por sua vez, possui propriedades estocásticas como média e variância estáveis ou constantes ao longo do tempo (BOX et al., 2015). Desta forma uma estacionariedade estrita pode se apresentar como fraca, mas o contrário não é possível (SHUMWAY; STOFFER; STOFFER, 2000). No decorrer deste trabalho vamos utilizar o termo estacionariedade para representar o tipo fraca, caso o processo seja estritamente estacionário, usaremos o nome completo.

Séries temporais do mundo real podem apresentar algum grau de não-estacionariedade, por exemplo, séries financeiras normalmente são não-estacionárias, as quais apresentam no caso mais simples flutuações ao redor de uma reta com inclinação positiva ou negativa (MORETTIN; TOLOI, 1981). Existem modelos que supõem que a série temporal seja estacionária, por exemplo o ARIMA, descrito na Seção 2.2.1. Desta forma, para utilizar esses modelos em séries temporais não-estacionárias, é necessário transformar a série em estacionária. Para isso, podem-se utilizar diferenças sucessivas nos dados originais até obter uma série estacionária. A primeira diferença de $z(t)$ é definida por:

$$\Delta z(z) = z(t) - z(t - 1); \quad (2.2)$$

a segunda diferença:

$$\Delta^2 z(t) = \Delta[\Delta z(t)] = \Delta[z(t) - z(t-1)]. \quad (2.3)$$

Normalmente são necessárias até duas diferenças para obter-se uma série estacionária (MORETTIN; TOLOI, 1981), se fazendo necessário, no fim do processo de modelagem fazer o processo inverso da diferenciação para a escala original.

2.2 PREVISÃO DE SÉRIES TEMPORAIS

Modelos podem ser utilizados para realizar estimações do valor futuro (previsão) da série temporal z_{t+h} , no qual h representa o horizonte de tempo da estimação desejada. Um modelo deve possuir a capacidade de identificar o comportamento da série baseando-se em lm dados passados (*lags*) e captando as relações estatísticas do processo ou o fenômeno que a gera (MORETTIN; TOLOI, 1981). Esses dados passados podem ser oriundos da própria série ($z_{t+1}, z_{t-2}, \dots, z_{t-lm}$), ou de séries temporais relacionadas ($z_{1t-1}, z_{2t-1}, z_{1t-2}, z_{2t-2}, \dots, z_{1t-lm}, z_{1t-lm}$), obtendo assim uma modelagem univariada ou multivariada, respectivamente (BOX et al., 2015).

É possível definir a quantidade de *lags* (lm) por meio da análise da *Função de Autocorrelação* (FAC) e *Função de Autocorrelação Parcial* (FACP). Estas medidas estatísticas refletem o quanto uma observação de uma série temporal é relacionada com seus valores anteriores (ADHIKARI; AGRAWAL, 2013).

A FAC é uma função $\rho(\cdot)$ definida pela Equação 2.5 (BROCKWELL et al., 2016), na qual γ é a função de covariância para o *lag* h , μ a média da série e N o total de pontos da série.

$$\gamma(h) = N^{-1} \sum_{t=1}^{N-|h|} (z_{t+|h|} - \mu)(z_t - \mu), \quad (2.4)$$

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}. \quad (2.5)$$

A FACP é uma função $\alpha(\cdot)$, na qual $\alpha(0) = 1$ e para $h \geq 1$ $\alpha(h) = \phi_{hh}$, ϕ_{hh} é o último componente de:

$$\phi_h = \Gamma_h^{-1} \gamma_h, \quad (2.6)$$

$$\Gamma_h = [\gamma(i-j)]_{i,j=1}^h, \gamma_h = [\gamma(1), \gamma(2), \dots, \gamma(h)]'. \quad (2.7)$$

Na literatura existem dois grupos de modelos que são popularmente utilizados para realizar a previsão de séries temporais: modelos lineares (BOX et al., 2015) e não lineares (ZHANG; PATUWO; HU, 1998).

Os modelos lineares são bastantes populares por conta de sua simples implementação e capacidade de interpretação. Os modelos AR, MA e ARIMA são exemplos de modelos lineares amplamente aplicados na literatura (BOX et al., 2015).

Os modelos não lineares, por sua vez, têm sido bastante explorados nas últimas décadas, principalmente as abordagens de AM. A MLP e SVR capazes de realizar uma modelagem não linear (ZHANG; PATUWO; HU, 1998; VAPNIK et al., 1997). Estes não necessitam que a série seja estacionária, como também são considerados aproximadores universais de funções (ZHANG; PATUWO; HU, 1998). Nas próximas subseções serão abordados com mais detalhes os modelos aplicados nesse trabalho: ARIMA, MLP e SVR.

2.2.1 *Autoregressive Integrated Moving Average*

O ARIMA combina operadores lineares para a previsão de séries temporais. Este modelo possui capacidade adaptativa para representar diferentes problemas por meio de um modelo parcimonioso e apresenta a habilidade de extensão, permitindo a modelagem com a presença de eventos externos. Além de que o ARIMA possui uma metodologia consolidada por meio de um *framework* coeso para criação de um modelo (LIU et al., 1992).

O modelo ARIMA é baseado em dois modelos: AR (Equação 2.8) e MA (Equação 2.9). A junção das duas equações supracitadas resulta no modelo Auto-regressivo de Médias Móveis (*Autoregressive Moving Average* (ARMA)), definido pela Equação 2.10.

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_p z_{t-p} + \epsilon_t, \quad (2.8)$$

$$z_t = \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q} + \epsilon_t, \quad (2.9)$$

$$z_t = \theta_0 + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_p z_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q} + \epsilon_t, \quad (2.10)$$

em que z_t é o valor atual no tempo t , $\phi_1, \phi_2, \dots, \phi_p$ e $\theta_1, \theta_2, \dots, \theta_q$ são parâmetros do modelo. No qual p e q são inteiros que se referem às ordens do modelo. Para os erros aleatórios,

(ϵ_t) assume-se que são independente e identicamente distribuídos com média zero e variância constante δ^2 . Os modelos AR, MA e ARMA são adequados apenas na previsão de séries temporais estacionárias.

O modelo ARIMA, por sua vez, é uma extensão do ARMA e possui a capacidade de realizar diferenças sucessivas nos dados (integração ordem d), removendo a tendência e tornando uma série não-estacionária em estacionária.

Para a construção do modelo ARIMA, pode-se utilizar a metodologia desenvolvida por Hyndman, Khandakar et al. (2007), que busca minimizar o critério de informação de Akaike (*Akaike's Information Criterion* (AIC)) por meio da Equação 2.11:

$$AIC = -2\log(L) + 2(p + q + k), \quad (2.11)$$

em que k será igual a 1 caso exista diferenciação (d) e 0 caso o contrário, enquanto L é o valor da função de densidade conjunta dos erros do modelo ajustado, diante dos seus resíduos. O AIC é uma métrica de avaliação estatística que busca maximizar o \log da probabilidade (L) e penalizar a utilização de parâmetros com altos valores (BOZDOGAN, 2000). Verifica-se que a utilização do AIC para definir a diferenciação (d) pode não ser o suficiente, pois a escolha do (d) por AIC tende a diferenças excessivas (HYNDMAN; KHANDAKAR et al., 2007). Por este motivo, a diferenciação (d) é definida a partir do teste de *Kwiatkowski Phillips Schmidt Shin* (KPSS) de raiz unitária. Esse teste possui a hipótese nula de que a série seja estacionária e a alternativa uma raiz unitária.

Então, são utilizadas as seguintes fases para identificação do modelo ARIMA:

1. Número de diferenças (d) é definido por meio de sucessivos testes de hipótese KPSS. Em seguida, é realizada uma busca exaustiva para selecionar o modelo que possua o menor AIC;
2. São consideradas 30 variações, em que p e q são variados ± 1 a partir do modelo atual. Sempre que for encontrado um AIC melhor, este será o novo modelo atual. O processo é terminado quando não encontrar nenhum modelo próximo ao desempenho do atual com menor AIC.

2.2.2 Rede Neural Perceptron de Multicamadas

A MLP possui vários neurônios artificiais distribuídos entre as camadas de entrada, intermediária(s) e saída. A estrutura da MLP segue a alimentação da esquerda para direita (*feedforward*), estando parcialmente ou totalmente conectada com a camada posterior (HAYKIN, 2004).

O relacionamento entre a saída (z_t) e entradas ($z_{t-1}, z_{t-2}, \dots, z_{t-lm}$) possui a representação matemática mostrada na equação 2.12. Esta equação exibe uma rede neural com uma camada escondida e um neurônio de saída, em que $\alpha_j (j = 0, 1, 2, \dots, s)$ e $\beta_{ij} (i = 0, 1, 2, \dots, lm; j = 1, 2, \dots, s)$ são pesos, lm o número de entradas e s o número de camadas escondidas. A função de ativação ($g(net)$) utilizada neste trabalho é a sigmoide logística (Equação 2.13).

$$z_t = \alpha_0 + \sum_{j=1}^s \alpha_j g \left(\beta_{0j} + \sum_{i=1}^{lm} \beta_{ij} z_{t-i} \right) + \varepsilon_t. \quad (2.12)$$

$$g(x) = \frac{1}{1 + \exp(-x)}. \quad (2.13)$$

Utiliza-se o gradiente descendente para buscar no espaço de hipóteses dos possíveis vetores de pesos que melhor representem o espaço de exemplos de treino, sendo esta a base para o algoritmo *backpropagation* (MITCHELL, 1997). Interações de aprendizado (uma época) consistem em duas fases (ENGELBRECHT, 2007):

1. *Feedforward*: Calcular os valores de saída da rede neural para cada padrão de treino;
2. *Backward propagation*: Propagar o sinal do erro de volta da saída para a entrada. Os pesos são ajustados como funções do sinal de erro de retro propagação.

No processo de treinamento podem existir muitas condições de parada, por exemplo: alcançar o erro médio quadrático desejado, atingir o número de interações máximas ou obter certa quantidade de época sem melhoras. A condição de parada não se limita à utilização de apenas uma das formas citadas, podendo ser utilizadas em conjunto.

2.2.3 Máquinas de Vetores de Suporte para Regressão

As SVRs são modelos que utilizam a teoria de aprendizado estatístico desenvolvido por Vapnik. Esta é uma ferramenta universal para resolver problemas de estimação de funções

multidimensionais, que possui o objetivo de buscar regras de decisão (VAPNIK et al., 1997). As SVRs têm grande aplicação em todos os domínios do reconhecimento de padrões, inclusive a previsão de séries temporais. Este modelo ainda possui capacidade de boa generalização e ausência de mínimos locais (CAO; TAY, 2003).

A SVR é baseada nos princípios de minimização do risco estrutural. Adicionalmente, os modelos geram uma função de regressão, aplicando um conjunto de altas dimensões de funções lineares (PAI; LIN, 2005).

Considera-se o conjunto de treino $[(\hat{z}_1, z_1), \dots, (\hat{z}_N, z_N)]$, em que \hat{z}_t é o vetor de características ($\hat{z}_t = [z_{t-1}, z_{t-2}, \dots, z_{t-lm}]$) e z_t é a saída desejada. A forma utilizada para máquinas de vetores de suporte realizar regressão é mostrada nas equações a seguir, em que $C > 0$ e $\varepsilon > 0$ são parâmetros fornecidos. A Equação 2.14 representa a forma padrão para tal realização, em que se busca a minimização da função (CHANG; LIN, 2011).

$$\min(w, b, \xi, \xi^*) = \frac{1}{2}w^T w + C \sum_{t=1}^N \xi_t + C \sum_{t=1}^N \xi_t^*, \quad (2.14)$$

sujeito a

$$\begin{aligned} w^T \nu(\hat{z}_t) + b - z_t &\leq \varepsilon + \xi_t, \\ z_t - w^T \nu(\hat{z}_t) - d &\leq \varepsilon + \xi_t^*, \\ \xi_t, \xi_t^* &\geq 0, t = 1, \dots, N. \end{aligned} \quad (2.15)$$

O problema dual é

$$\min(\alpha, \alpha^*) = \frac{1}{2}(\alpha - \alpha^*)^T Q(\alpha - \alpha^*) + \varepsilon \sum_{t=1}^N (\alpha_t + \alpha_t^*) + \sum_{t=1}^N z_t(\alpha_t - \alpha_t^*), \quad (2.16)$$

sujeito a

$$\begin{aligned} e^T(\alpha_t - \alpha_t^*) &= 0 \\ 0 &\leq \alpha_t, \alpha_t^* \leq C, t = 1, \dots, N, \end{aligned} \quad (2.17)$$

em que $Q_{tj} = K(\hat{z}_t, \hat{z}_j) = \nu(\hat{z}_t)^T \nu(\hat{z}_j)$.

Após realizado o problema de minimização da Equação 2.16, pode-se utilizar a função da Equação 2.18 para realizar aproximações, em que $K(\hat{z}_t, \hat{z})$ é chamada de função de *kernel*, cujo valor é igual ao produto interno entre as duas variáveis \hat{z}_t e \hat{z} no espaço de características $\nu(\hat{z}_t)$ e $\nu(\hat{z})$. A função de *kernel* utilizada neste trabalho é a função de base radial (Equação 2.19).

$$f(y, \alpha, \alpha^*) = \sum_{t=1}^N (-\alpha_t + \alpha_t^*) K(\hat{z}_t, \hat{z}) + b. \quad (2.18)$$

$$K(x_i, x_j) = \exp(-\gamma|x_i - x_j|^2). \quad (2.19)$$

2.3 COMBINAÇÃO DE PREDITORES

A combinação de preditores (modelos) possui o principal objetivo de agregar diferentes previsões, sendo que esta tende a ser melhor em termos de acurácia que as previsões dos modelos separados (modelos simples)¹. Essa melhora está relacionada aos aspectos positivos que uma abordagem de combinação traz ao processo de modelagem, podendo ser listadas como:

1. Simplificação do processo de criação dos modelos e seleção de parâmetros. Selecionar parâmetros pode ser algo custoso em termos de tempo de execução. Podem existir modelos que, além de um processo de treinamento demorado, tenham muitas possíveis combinações de parâmetros, podendo tornar este processo inviável dependendo da aplicação. Nestes casos, utilizar a combinação de modelos diferentes pode reduzir a necessidade de uma busca extensiva nos parâmetros, mantendo bons resultados em termos de acurácia e em um tempo hábil (WANG et al., 2018; SILVA et al., 2018)
2. Melhor generalização quando aplicado em um conjunto futuro de dados. Há a possibilidade de selecionar um modelo que obteve bons resultados na validação e esse bom comportamento não se repetir em dados futuros da série temporal em questão. Abordagens de combinação também trazem efeitos positivos nesses casos, pois eliminam a necessidade de escolher um único modelo, como também a agregação de diferentes previsões criar um mecanismo de complementariedade entre elas (WINKLER; MAKRIDAKIS, 1983).

Um ponto comum de ambos os aspectos citados é que para criar uma boa abordagem de combinação é necessário que existam modelos diferentes (diversidade). É comum na literatura de previsão de séries temporais que se criem abordagens de combinação por meio de métodos de *Ensemble* (WANG et al., 2013; KOURENTZES; BARROW; CRONE, 2014; SILVA et al., 2020) e Sistemas Híbridos Residuais (ZHANG, 2003; PANIGRAHI; BEHERA, 2017; de MATTOS NETO et al., 2021a). Estas abordagens, por sua vez, utilizam mecanismos diferentes para obter diversidade.

¹ Modelos aplicados de forma individual, fora da abordagem de combinação.

Um *Ensemble* é criado a partir de um grupo de modelos que é obtido seguindo um formalismo de geração. Os Sistemas Híbridos Residuais utilizam pelo menos dois modelos executados de forma sequencial e possui fases específicas para modelar os padrões lineares e não lineares da série temporal. Nas próximas subseções são exibidos detalhes sobre essas estratégias de combinação.

2.3.1 Ensemble

Uma das principais capacidades de um *ensemble* é a redução da variância do erro. Há uma prova teórica dessa capacidade por meio da equação de decomposição do viés-variância-covariância (*bias-variance-covariance*):

$$E[\hat{z} - z]^2 = bias^2 + \frac{1}{m}var + \left(1 - \frac{1}{m}\right)covar, \quad (2.20)$$

no qual o viés é definido por:

$$bias = \frac{1}{m} \sum_i (E[\hat{z}_i] - z), \quad (2.21)$$

a variância:

$$var = \frac{1}{m} \sum_i E[\hat{z}_i - E[\hat{z}_i]]^2, \quad (2.22)$$

e a covariância:

$$covar = \frac{1}{m(m-1)} \sum_i \sum_{i \neq j} E[\hat{z}_i - E[\hat{z}_i]](\hat{z}_i - E[\hat{z}_i]), \quad (2.23)$$

z representa o valor desejado, \hat{z}_i a saída do modelo i , m é a quantidade de modelos utilizados na combinação. Então, o erro da abordagem de combinação é composto pela soma das médias do viés, variância e covariância dos modelos. Desta forma, esta decomposição exhibe que caso a combinação seja feita por modelos que possuam baixa correlação entre si, é esperada uma melhora nos resultados (REN; ZHANG; SUGANTHAN, 2016).

Desta forma, é denominada como *ensemble* a abordagem de combinação que possua m modelos e seja composta de três fases: (1) – Geração (Subsubseção 2.3.1.1), (2) – Seleção (Subsubseção 2.3.1.2) e (3) – Agregação (Subsubseção 2.3.1.3).

2.3.1.1 Geração

A fase de geração é responsável por gerar a diversidade do *Ensemble*, existindo dois grupos de estratégia heterogênea ou homogênea (BRITTO; SABOURIN; OLIVEIRA, 2014).

Nas abordagens heterogêneas são usadas diferentes arquiteturas de modelos para criar a diversidade. Desta forma, muitas possibilidades de geração podem ser obtidas, alguns exemplos da literatura são listados a seguir. Sergio, Lima e Ludermir (2016) utilizam uma MLP com uma camada escondida, outra MLP com duas camadas escondidas, SVR e Rede profunda de crenças (*Deep Belief Network* (DBN)) como *ensemble*. Adhikari (2015) combina as previsões de um ARIMA, MLP e rede neural recorrente. Wang et al. (2018) utilizam a MLP, rede neural de arquitetura dinâmica, rede neural de Elman e rede de estado de eco (*Echo State Network* (ESN)) como componentes do ensemble. De Mattos Neto et al. (2021b) combinam as previsões do AR, ARIMA, MLP, ESN, rede neural de base radial, máquinas de aprendizado extremo e sistema de inferência *neuro-fuzzy* adaptável. Che (2015) utiliza 7 SVRs com configurações diferentes, ARIMA e análise exponencial. Como mostrado, existem muitas variações de combinações heterogêneas na literatura, sendo que na maioria dos casos, o grupo de modelos é definido pelas suas particularidades, com os autores sugerindo que a combinação aplicada traz a diversidade necessária para obter bons resultados em termos de acurácia.

As abordagens homogêneas, por sua vez, normalmente utilizam um grupo do mesmo modelo, no qual a diferença é caracterizada em termos de conjunto dados. Bagging (BREIMAN, 1996), *Random subspace* (HO, 1998) e *Random Patches* (LOUPPE; GEURTS, 2012a) são exemplos de abordagens de geração homogêneas amplamente aplicadas na literatura.

A geração por Bagging (BREIMAN, 1996) utiliza uma amostra *bootstrap* do conjunto de treinamento. Desta forma, utiliza-se uma porcentagem do conjunto de treinamento (*Bagging Size – bs*) com a amostra sendo selecionada de forma aleatória com reposição (QIU et al., 2017). A geração por *Random Subspace* foi desenvolvida por Ho (HO, 1998) e cada modelo utiliza uma porcentagem aleatória sem reposição do espaço de entradas (*Feature size – fs*). Portanto, não é permitido ter entradas repetidas no conjunto selecionado. A abordagem de geração *Random Patches* (LOUPPE; GEURTS, 2012a) utiliza a combinação das estratégias de geração *Bagging* e *Random Subspace*. Cada modelo utiliza uma amostra aleatória do treinamento e das entradas. Nesta estratégia, ambos os parâmetros *bs* and *fs* precisam ser definidos. Alguns trabalhos da literatura utilizam essa abordagem com o objetivo de maximizar a diversidade dos modelos e, conseqüentemente, obter melhores resultados em termos de acurácia (LOUPPE;

GEURTS, 2012a). Esta abordagem tem também a capacidade de reduzir os recursos de memória necessários e tempo de computação para o treinamento, sendo que cada modelo do *ensemble* utiliza um subconjunto de treinamento representado por uma porcentagem do todo, em termos de exemplos e entradas (PANOV; DŽEROSKI, 2007; LOUPPE; GEURTS, 2012b).

Abordagens heterogêneas e homogêneas são estratégias amplamente aplicadas na tarefa de geração de um *ensemble* diverso. Contudo, os métodos heterogêneos podem ter dificuldades sobre o controle da diversidade dos modelos (MENDES-MOREIRA et al., 2012), pois modelos com arquiteturas diferentes podem convergir para representações semelhantes dos dados. As abordagens homogêneas, por outro lado, conseguem um melhor controle sobre a diversidade, uma vez que os modelos explicitamente usam diferentes visões do conjunto de dados e são capazes capacidade de produzir muitas variações de modelos, aumentando a probabilidade de criar um grupo diverso e acurado (MENDES-MOREIRA et al., 2012).

2.3.1.2 Seleção

Dado que um grupo de modelos foi gerado, é possível selecionar um modelo ou um subgrupo de modelos. Este passo tem o objetivo de reduzir a complexidade do *ensemble*, ou mesmo de manter apenas os modelos com melhor desempenho seguindo algum critério. Desta forma, a seleção do *ensemble* pode ser caracterizada de forma estática ou dinâmica.

A seleção estática utiliza um modelo ou um subgrupo de modelos para prever todo o conjunto de teste. Em previsão de séries temporais é comum a utilização de estratégias de seleção estática baseadas na acurácia de treinamento ou validação, nas quais os k melhores modelos são selecionados para prever o conjunto de teste (MA; DAI, 2016; de HOLANDA; de OLIVEIRA, 2020). A seleção dinâmica, por sua vez, usa um modelo ou um subgrupo de modelos para cada padrão de teste. Essa seleção é normalmente baseada no conceito de região de competência, no qual os melhores modelos nesta região são selecionados (CRUZ; SABOURIN; CAVALCANTI, 2018; BRITTO; SABOURIN; OLIVEIRA, 2014). Esta região de competência é criada sobre a similaridade (distância euclidiana, por exemplo) entre a entrada do teste e as entradas de treino ou validação (SILVA et al., 2020).

É importante mencionar que esta fase é opcional (CRUZ; SABOURIN; CAVALCANTI, 2018) e todos os modelos podem ser usados para a próxima fase (HILLEBRAND; MEDEIROS, 2010; JIN; SU; ULLAH, 2014; QIU et al., 2017). Abordagens de seleção possuem o risco de selecionar um subgrupo de modelos que tenham bons resultados na validação ou treino e o contrário

no conjunto de teste. Há inclusive a possibilidade de retirar modelos que individualmente não tenham bons resultados, mas que poderiam ser importantes para os resultados finais do *ensemble*.

2.3.1.3 Agregação

Esta fase é responsável por juntar as previsões dos modelos e criar a previsão do final ensemble (\hat{Z}_{t+1}). Como previsão de séries temporais pode ser considerado um problema de regressão é comum agregar a previsão de m modelos por meio da média ou mediana (KOURENTZES; BARROW; CRONE, 2014; SERGIO; LIMA; LUDERMIR, 2016). Estes, por sua vez, são métodos para agregação de *ensemble* bem conhecidos (REN; ZHANG; SUGANTHAN, 2016), nos quais podem ser alcançados resultados semelhantes aos das estratégias mais complexas (KOURENTZES; BARROW; CRONE, 2014). Além disso, média e mediana são estratégias não treináveis, significando que é necessário pouco esforço computacional para agregar as previsões.

A agregação por média é definida pela Equação 2.24, em que M^i é a previsão do modelo i e m é quantidade de modelos usados no *ensemble*.

$$\hat{Z}_{t+1} = \frac{1}{m} \sum_{i=1}^m M^i. \quad (2.24)$$

A agregação por mediana é obtida por meio da ordenação das previsões dos modelos e capturando o valor do meio se m for ímpar ou a média dos dois valores do meio, caso contrário (KOURENTZES; BARROW; CRONE, 2014).

2.3.2 Sistemas Híbridos Residuais

Esta classe de combinação de preditores realiza a agregação de modelos lineares com não lineares por meio da modelagem residual (ZHANG, 2003; SANTOS JÚNIOR; de OLIVEIRA; de MATTOS NETO, 2019; PAI; LIN, 2005; PANIGRAHI; BEHERA, 2017; XU et al., 2021). A motivação desta abordagem é que séries temporais reais são compostas por ambos os padrões lineares e não lineares, sendo então interessante criar uma arquitetura que modele esses padrões em etapas específicas.

Este tipo de abordagem foi introduzido por Zhang (2003), supondo que a série temporal Z_t pode ser descrita pela soma dos padrões lineares (L_t) e não lineares (N_t), conforme a Equação 2.25.

$$Z_t = L_t + N_t. \quad (2.25)$$

Esta abordagem aplica um modelo estatístico linear (M_L) para realizar a previsão da série temporal ($M_L(Z_t)$). Após isto, a série residual é obtida pela diferença entre o valor real e a previsão linear $M_L(Z_t)$ (Equação 2.26).

$$E_t = Z_t - M_L(Z_t). \quad (2.26)$$

Se o modelo linear for bem especificado, E_t não apresentará nenhuma correlação linear estatisticamente significativa (BOX et al., 2015). Desta forma, uma abordagem de AM pode ser aplicada para modelar os padrões não lineares no resíduo. Então, um modelo de AM (M_N) é usado para prever o resíduo do modelo estatístico, obtendo a previsão não linear da série temporal ($M_N(E_t)$) (Equação 2.27).

$$M_N(E_t) = f(e_{t-1}, e_{t-2}, \dots, e_{t-n}) + \epsilon_t, \quad (2.27)$$

A previsão final \hat{Z}_{t+1} é gerada pela adição das previsões lineares e não lineares (Equação 2.28).

$$\hat{Z}_{t+1} = M_L(Z_t) + M_N(E_t). \quad (2.28)$$

Utiliza-se $M_L + M_N$ como nomenclatura para representar os sistemas híbridos que usam a adição para combinar $M_L(Z_t)$ e $M_N(E_t)$. Por exemplo, sistemas híbridos que aplicam o ARIMA para prever a série temporal, MLP para prever o resíduo e soma para combinação são referidos como ARIMA + MLP. Esta arquitetura aditiva é aplicada em muitos trabalhos da literatura, tais como: ARIMA + MLP (ZHANG, 2003; BELMAHDI; LOUZAZNI; BOUARDI, 2020), ARIMA + SVR (PAI; LIN, 2005), ARIMA + Rede de Funções de Base Radial (*Radial Basis Function network* (RBF)) (YAN; MA, 2016), Suavização Exponencial + MLP (PANIGRAHI; BEHERA, 2017), ARIMA + Rede neural recorrente com memória de longo prazo (*Long Short-Term Memory network* (LSTM)) (XU et al., 2021), e ARIMA + DBN (XU et al., 2019; QIN; LI; DU, 2017).

Além das variações de arquiteturas, as abordagens $M_L + M_N$ são amplamente aplicadas em diferentes séries temporais reais: exportações futuras (DAVE et al., 2021), radiação solar (BELMAHDI; LOUZAZNI; BOUARDI, 2020), velocidade do vento (ZHANG; WEI; TAN, 2020), qualidade da água (XU et al., 2019), índices financeiros (ZHANG, 2003; PAI; LIN, 2005; PANIGRAHI;

BEHERA, 2017), passageiros em voos internacionais (PANIGRAHI; BEHERA, 2017), uso de energia em aquecimento (SRETENOVIC et al., 2021), consumo de energia elétrica (XU et al., 2019), maré vermelha (QIN; LI; DU, 2017) e material particulado (WONGSATHAN; CHANKHAM, 2016).

A partir da abordagem aditiva, algumas variações de arquitetura são encontradas na literatura. Wang et al. (2013) propõem usar a abordagem de multiplicação em lugar da adição, na qual a série de resíduo é obtida pela divisão da previsão linear com o valor real da série temporal. A previsão final da abordagem de Wang et al. (2013) é a multiplicação da previsão da série temporal e previsão do resíduo. Este tipo de combinação sugere que podem existir casos nos quais o relacionamento entre os padrões lineares e não lineares sejam multiplicativos em vez de aditivo.

Existem também arquiteturas de sistemas híbridos em que os modelos não lineares realizam a modelagem do resíduo e a combinação (KHASHEI; BIJARI, 2011; SANTOS JÚNIOR; de OLIVEIRA; de MATTOS NETO, 2019; de MATTOS NETO et al., 2020). Neste tipo de abordagem, um modelo linear é utilizado para prever a série temporal e o resíduo é obtido de acordo com a Equação 2.26. Em seguida, existem duas formas de realizar a modelagem residual e combinação das previsões: 1 – utilizando um modelo não linear que é responsável por realizar a previsão do resíduo e combinação (em um único passo), obtendo assim a previsão final (KHASHEI; BIJARI, 2011; AGUILAR; TURIAS; COME, 2014; XU; CHAN; ZHANG, 2019). 2 – aplicando dois modelos não lineares, um para prever o resíduo e outro para combinar as previsões da série temporal e resíduo, obtendo a previsão final (SANTOS JÚNIOR; de OLIVEIRA; de MATTOS NETO, 2019; de MATTOS NETO et al., 2020; de MATTOS NETO et al., 2021a). Estes sistemas híbridos supõem que a combinação entre os padrões lineares e não lineares se apresenta de forma não linear. Então, utilizar uma abordagem de AM para esta tarefa pode ser uma boa estratégia. O ponto negativo é que essa combinação não linear adiciona complexidade nas etapas de previsão residual e combinação, podendo intensificar os problemas de sobre ajuste, sub ajuste e má especificação de parâmetros dos modelos de AM.

Recentemente de Holanda e de Oliveira (2020) e de Oliveira, Silva e de Mattos Neto (2021) propuseram métodos de *ensemble* para previsão residual.

A estratégia de Holanda e de Oliveira (2020) utiliza o ARIMA para prever a série temporal (modelagem linear) e um *ensemble* de SVRs baseado em otimização de enxame para prever os resíduos (modelagem não linear), seguindo as Equações 2.26 e 2.27. É utilizada a otimização por enxame de partículas (*Particle Swarm Optimization* (PSO)) para seleção dos parâmetros dos modelos que compõem o *ensemble*. Desta forma, o PSO é aplicado para controlar a

diversidade dos modelos por meio da manipulação do conjunto dos parâmetros (MENDES-MOREIRA et al., 2012). O trabalho supracitado avalia quatro variações do *ensemble* obtido pelo processo de otimização do PSO utilizando: a média das partículas, mediana das partículas, a média das 10 melhores partículas finais e a mediana das 10 melhores partículas finais. Os resultados são avaliados em seis séries temporais e mostram que alguma das configurações são capazes de melhorar os resultados dos modelos simples e sistemas híbridos de Zhang (2003) e Panigrahi e Behera (2017).

De Oliveria, Silva e de Mattos Neto (2021) exibem um sistema híbrido chamado de Previsor Residual Dinâmico (*Dynamic Residual Forecasting* (DReF)). Duas versões do DReF foram propostas usando ARIMA ou ARIMA Sazonal para modelagem linear da série temporal. Em ambos casos, a previs(ão) residual é feita por um dos cinco modelos disponíveis: MLP, SVR, RBF, LSTM e Rede neural Convolutacional. Neste trabalho, um modelo não linear é selecionado dinamicamente baseado em uma região de competência criada a partir dos padrões mais similares para cada novo padrão de teste. São propostas sete formas diferentes para a criação dessa similaridade. Este trabalho enfoca o estudo de uma estratégia de seleção de um modelo para cada padrão de teste. Desta forma, não é avaliada a possibilidade de utilização das previsões residuais agregadas. DReF mostra que algumas de suas arquiteturas alcançam resultados promissores quando comparadas aos modelos simples e sistemas híbridos da literatura (ZHANG, 2003; PAI; LIN, 2005; WANG et al., 2013; YAN; MA, 2016; XU et al., 2019).

O método proposto ERF se diferencia significativamente dos trabalhos de Holanda e de Oliveira (2020) e de Oliveria, Silva e de Mattos Neto (2021) em três principais aspectos:

1. O estágio de geração do método ERF é obtido por meio de duas estratégias de seleção aleatória da amostra de treino, *Bagging* e *Random Subspace*. Estas abordagens têm o principal objetivo de criar um grupo de modelos diversos e acurados. Desta forma, os modelos do *ensemble* utilizam a mesma arquitetura e parâmetros, mas possuindo visões diferentes do conjunto de treino. Esta é uma abordagem mais segura para geração de diversidade (MENDES-MOREIRA et al., 2012);
2. É realizada uma busca baseada em dados para determinar uma única arquitetura do método proposto, selecionado um conjunto de parâmetro do *ensemble* baseado na validação;
3. É desenvolvida uma arquitetura flexível que permite aplicar diferentes modelos nos passos

de previsão da série temporal e resíduo.

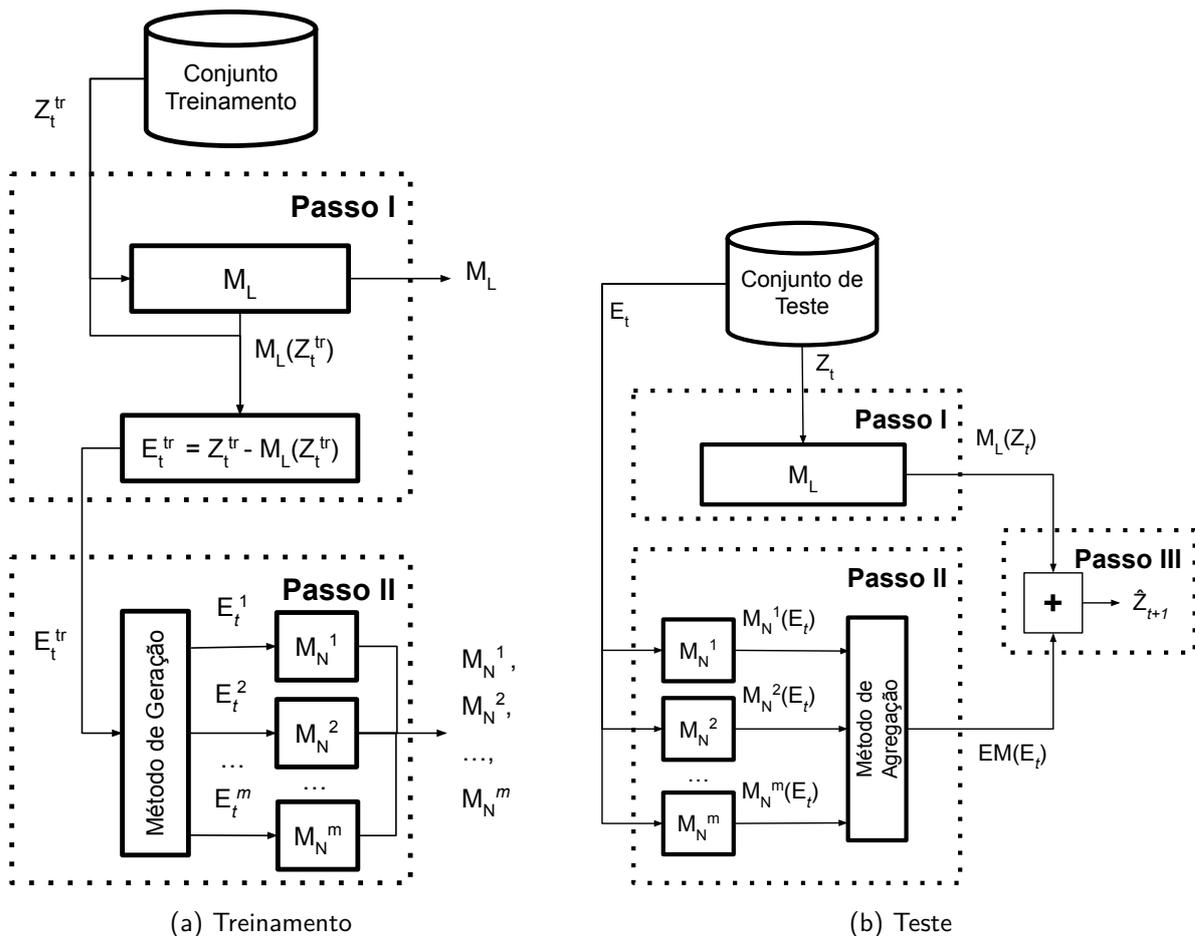
2.4 RESUMO DO CAPÍTULO

Série temporal é um conjunto de variáveis expostas no tempo que podem ser afetadas por componentes chamados de sazonalidade, tendência, ciclos e irregularidades. Na tarefa de previsão de séries temporais são utilizados modelos estatísticos e de aprendizagem de máquina que comumente são caracterizados como modelos lineares e não lineares, respectivamente. A família dos modelos ARIMA são exemplos de modelos estatísticos lineares, já MLP e SVR são exemplos de abordagem de AM não lineares. Contudo, ambos os modelos estatísticos e de AM possuem limitações. Os modelos estatísticos lineares podem encontrar dificuldade de modelar séries temporais reais, uma vez que estas normalmente apresentam padrões não lineares. Os modelos de AM, por sua vez, podem apresentar problemas, tais como má especificação de parâmetros, sobre ajuste e sub ajuste. Considerando esses problemas, nasce a motivação de combinar modelos diferentes por meio da agregação de suas previsões, alcançando um resultado final melhor que os resultados dos modelos separados. Nesse contexto, é comum utilizar métodos de combinação chamados de *ensemble* e Sistemas Híbridos residuais. *Ensemble* é caracterizado por um método que possui as fases de geração, seleção e agregação e os Sistemas Híbridos residuais criam um método sequencial que busca modelar os padrões lineares e não lineares da série temporal em etapas específicas.

3 MÉTODO PROPOSTO

O método de *ensemble* para previsão residual (*Ensemble method for Residual Forecast - ERF*) é a abordagem proposta por este trabalho. O ERF é composto por duas fases principais que são apresentadas em seções específicas: treinamento, representado no fluxograma da Figura 1a e detalhado na Seção 3.1 e teste, representado no fluxograma da Figura 1b) e descrito na Seção 3.2. Existem ao todo três passos principais na abordagem proposta: (I) modelagem da série temporal utilizando um modelo linear (M_L), (II) modelagem da série de resíduo utilizando um *ensemble* (EM) de modelos de AM e (III) geração da previsão final (\hat{Z}_{t+1}) por meio da combinação das saídas dos passos (I) e (II).

Figura 1 – Fases do método proposto



Fonte: Elaborada pelo autor (2022)

3.1 FASE DE TREINAMENTO

A Figura 1a exibe que os dados de entrada para esta fase são o respectivo conjunto da série temporal de treinamento (Z_t^{tr}). A saída desta fase, por sua vez, são os modelos lineares e não lineares treinados ($M_N^1, M_N^2, \dots, M_N^m$). Os modelos não-lineares compreendem ao *ensemble* utilizado na modelagem do resíduo.

No passo (I) da fase de treinamento, o modelo M_L é criado a partir do dados reais da série temporal de treinamento (Z_t^{tr}). Neste passo, o treinamento do M_L tem o objetivo de capturar padrões lineares existentes na série temporal. Em seguida, a série de resíduo (E_t^{tr}) é gerada por meio da diferença entre o valor real da série temporal (Z_t^{tr}) e a previsão linear ($M_L(Z_t^{tr})$). Se E_t^{tr} não apresentar correlações lineares estatisticamente relevantes seguindo algum teste (BOX et al., 2015), o modelo M_L foi então bem especificado.

No passo (II) da fase de treinamento, um *ensemble* de modelos de AM é aplicado para capturar padrões não lineares presentes na série de resíduo E_t^{tr} . O método proposto aplica um *ensemble* homogêneo (BREIMAN, 1996; HO, 1998; LOUPPE; GEURTS, 2012a) composto por m modelos que possuem a mesma arquitetura. Com o objetivo de criar um grupo de modelos (*pool*) diverso e acurado, cada modelo não linear é treinado com um conjunto de amostras diferente da série temporal (BREIMAN, 1996; HO, 1998; LOUPPE; GEURTS, 2012a).

A abordagem chamada de *Random Patches* (LOUPPE; GEURTS, 2012a) é aplicada como Método de Geração com o objetivo de criar diferentes conjuntos (visões) da série temporal.

O *Random Patches* combina as estratégias de Bagging (BREIMAN, 1996) e *Random Subspace* (HO, 1998), dois métodos de geração bem conhecidos na literatura. Estes tipos de métodos de geração selecionam aleatoriamente uma amostra de exemplos e características do treino, com o objetivo de maximizar a diversidade e reduzir os recursos de memória necessários e tempo de computação para treinar os modelos do *ensemble* (PANOV; DŽEROSKI, 2007; LOUPPE; GEURTS, 2012a; GOMES; READ; BIFET, 2019).

Para o uso do método de geração *Random Patches*, é necessário transformar a série de resíduo E_t^{tr} para uma matriz que mapeie as entradas-saídas necessárias para treinamento de um modelo de AM supervisionado. A saída representa o alvo (*Target*) do modelo de AM (previsão do resíduo) e as entradas são o conjunto de dados passados (*lags*) considerando a saída. A partir da matriz de treino, o *Random Patches* seleciona aleatoriamente uma porcentagem da amostra de treino com reposição (*Bootstrap Size* – bs) e entradas sem reposição (*Feature Size* – fs). Este processo é repetido m vezes para gerar $E_t^1, E_t^2, \dots, E_t^m$.

a agregação por média é obtida pelo somatório de $M_N^1(E_t)$, $M_N^2(E_t)$, \dots , $M_N^m(E_t)$ dividido por m . A agregação por mediana é obtida por meio da ordenação das previsões do resíduo do *ensemble* e selecionando o valor do meio se m for ímpar, caso contrário, é utilizada a média dos dois pontos centrais.

Finalmente, no passo (III) da fase de teste, a previsão da série temporal (\hat{Z}_{t+1}) é gerada por meio da soma das previsões linear ($M_L(Z_t)$) e não-linear ($EM(E_t)$).

3.3 RESUMO DO CAPÍTULO

Este capítulo apresenta detalhes sobre a arquitetura do ERF exibindo os passos necessários para treinamento e teste. Esta abordagem utiliza um modelo linear para previsão da série temporal com o objetivo de capturar padrões lineares e gerar a série residual, esta última, por sua vez, é modelada por um *ensemble*. Este *ensemble* é obtido por meio da abordagem de *Random Patches*, que gera um *ensemble* homogêneo cuja diversidade é obtida a partir de uma visão diferente do conjunto de treino que cada modelo de AM utilizado. As previsões dos modelos de AM do *ensemble* são agregadas por média ou mediana. A previsão final do método proposto é obtida pela soma da previsão do modelo linear com a previsão do resíduo realizada pelo *ensemble*.

4 METODOLOGIA

Este capítulo exhibe metodologia aplicada para a geração dos resultados expostos no próximo capítulo, mostrando as bases de dados utilizadas, assim como suas características estatísticas e divisões de treinamento, validação e teste (Seção 4.1). Por fim, na Seção 4.2 é exibida a métrica de desempenho, métodos de previsão e estratégia para seleção de parâmetros.

4.1 BASE DE DADOS

São aplicadas doze séries temporais amplamente utilizadas na literatura de Sistemas Híbridos (ZHANG, 2003; KHASHEI; BIJARI, 2011; PANIGRAHI; BEHERA, 2017; SANTOS JÚNIOR; de OLIVEIRA; de MATTOS NETO, 2019). A Tabela 1 exhibe a lista dos nomes das doze séries temporais juntamente a seus respectivos títulos adotados. Esta tabela também possui o número total de observações e o valor absoluto utilizado para treinamento, validação e teste, correspondendo à porcentagem de 60%, 20% e 20% das observações disponíveis, respectivamente. Todas as séries temporais estão disponíveis na biblioteca *Time Series Data Library* (TSDL) ¹.

Tabela 1 – Lista de séries temporais aplicadas

Série Temporal	Tamanho absoluto			
	Total	Treinamento	Validação	Teste
<i>Airlines Passengers</i>	144	86	29	29
<i>Canadian Lynx</i>	114	68	23	23
<i>Colorado River</i>	744	446	149	149
<i>IBM</i>	369	221	74	74
<i>Lake Erie</i>	600	360	120	120
<i>Milk</i>	156	94	31	31
<i>Ozon</i>	180	96	48	36
<i>Paper</i>	120	72	24	24
<i>Pollution</i>	130	78	26	26
<i>Star Bright</i>	600	360	120	120
<i>Sunspot</i>	288	172	58	58
<i>Temperature</i>	240	144	48	48

Fonte: Elaborada pelo autor (2022)

Nas próximas subseções são exibidos para cada série temporal:

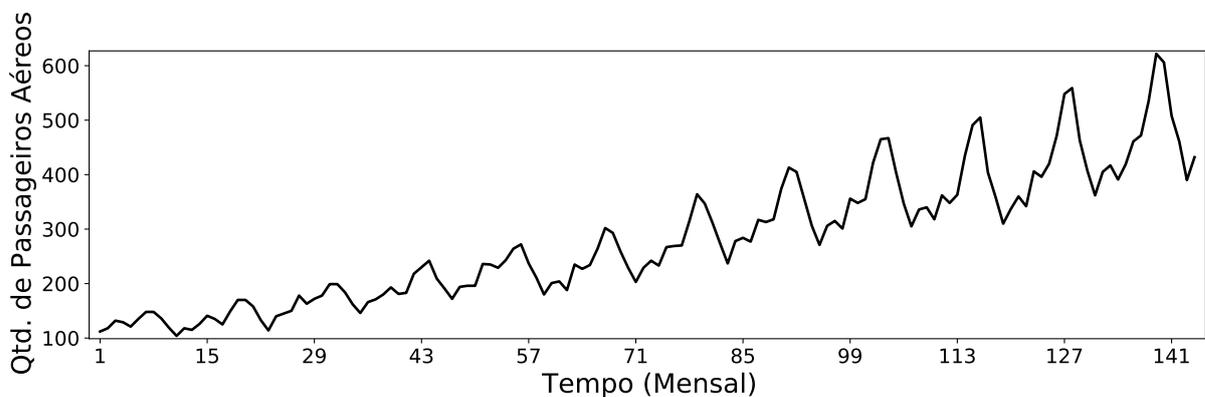
¹ <https://pkg.yangzhuoranyang.com/tsdl/>

- Período e unidade temporal utilizada;
- Estatísticas descritivas, tais como: média, desvio padrão, mínimos e máximos;
- Categorização em estacionária ou não-estacionária por meio do teste de KPSS (considerando 95% de significância). Este teste possui a hipótese nula de que uma série temporal é estacionária contra a alternativa de uma raiz unitária (SMITH et al., 2017), sendo este amplamente usado para determinar o termo de diferença do ARIMA (d) (HYNDMAN; KHANDAKAR et al., 2007);
- Análise de FAC e FACP utilizando vinte como *lag* máximo.

4.1.1 *Airlines Passengers*

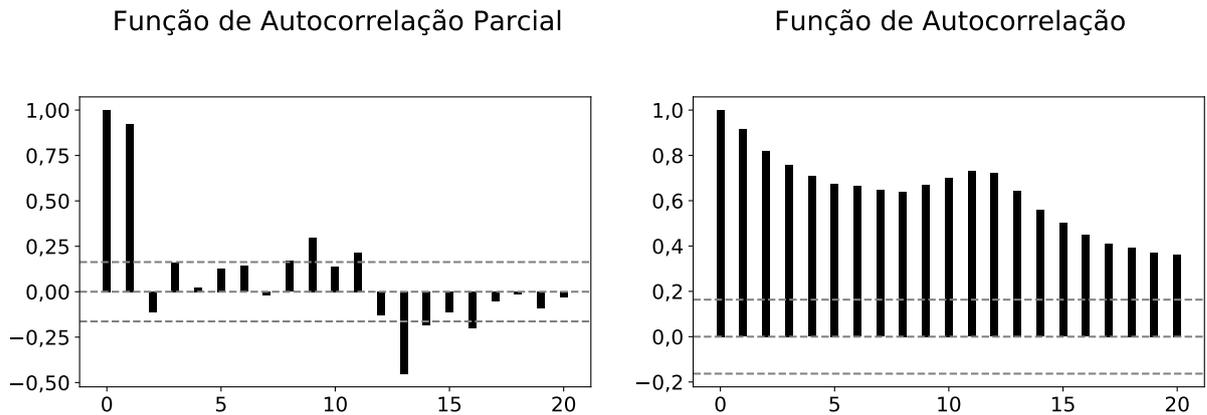
A série temporal *Airlines Passengers* (Figura 3) consiste na quantidade de passageiros em voos internacionais. É utilizada a unidade temporal mensal entre os anos de 1949 e 1960. Esta série possui 280 de valor médio, 120 de desvio padrão, 104 de valor mínimo e 622 de valor máximo.

Figura 3 – Série temporal *Airlines Passengers*



Fonte: Elaborada pelo autor (2022)

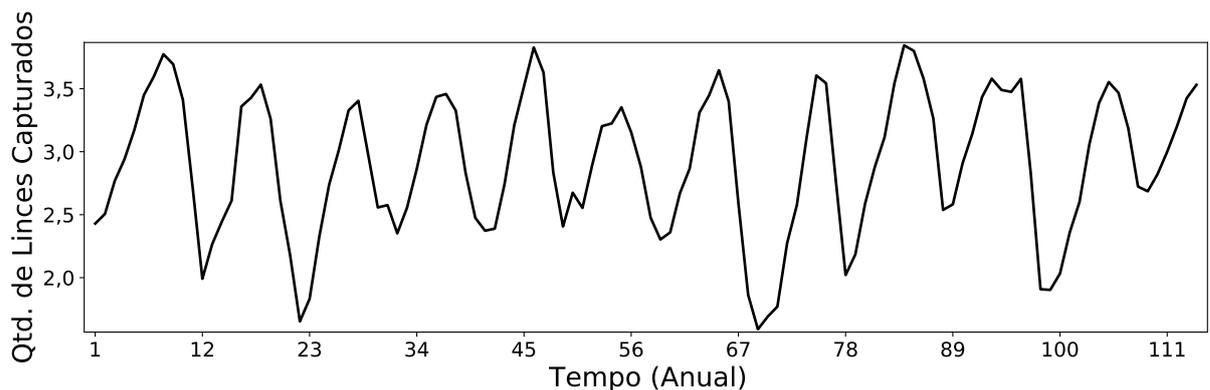
A Figura 4 exibe a FAC e FACP da série *Airlines Passengers*. No FAC, observa-se que a série em questão possui autocorrelações lineares relevantes, como também uma sazonalidade de doze meses. No FACP, por sua vez, é possível observar que existem autocorrelações relevantes até o *lag* dezesseis. Pelo teste de KPSS é possível definir esta série temporal como não estacionária, essa afirmação pode também ser visualizada na Figura 3, em que é identificada uma tendência positiva.

Figura 4 – FACP e FAC da série *Airlines Passengers*

Fonte: Elaborada pelo autor (2022)

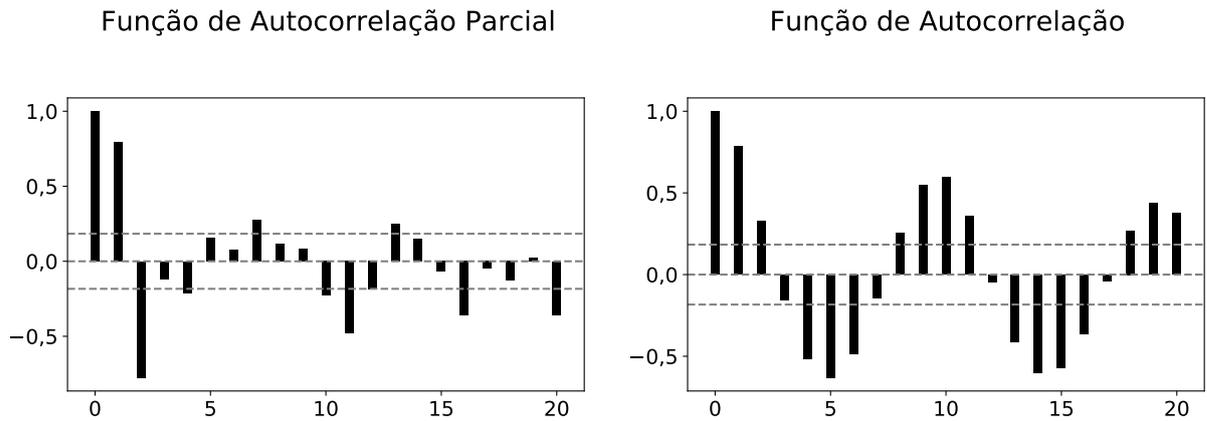
4.1.2 *Canadian Lynx*

A série temporal *Canadian Lynx* (Figura 5) consiste na quantidade de lincas capturados no norte do Canadá. É utilizada a unidade temporal anual entre os anos de 1821 e 1934, aplicando a suavização logarítmica na base 10, com o objetivo de reduzir a variância (ZHANG, 2003; KHASHEI; BIJARI, 2011). Esta série possui o valor médio de 2,9, desvio padrão de 0,6, valor mínimo de 1,6 e máximo de 3,8.

Figura 5 – Série temporal *Canadian Lynx*

Fonte: Elaborada pelo autor (2022)

A Figura 6 exibe a FAC e FACP da série *Canadian Lynx*. No FAC, observa-se que a série em questão possui autocorrelações lineares relevantes, como também ciclos de cerca de dez anos. No FACP, por sua vez, é possível observar que a autocorrelação relevante mais distante é no *lag* vinte. Pelo teste de KPSS é possível definir esta série temporal como estacionária,

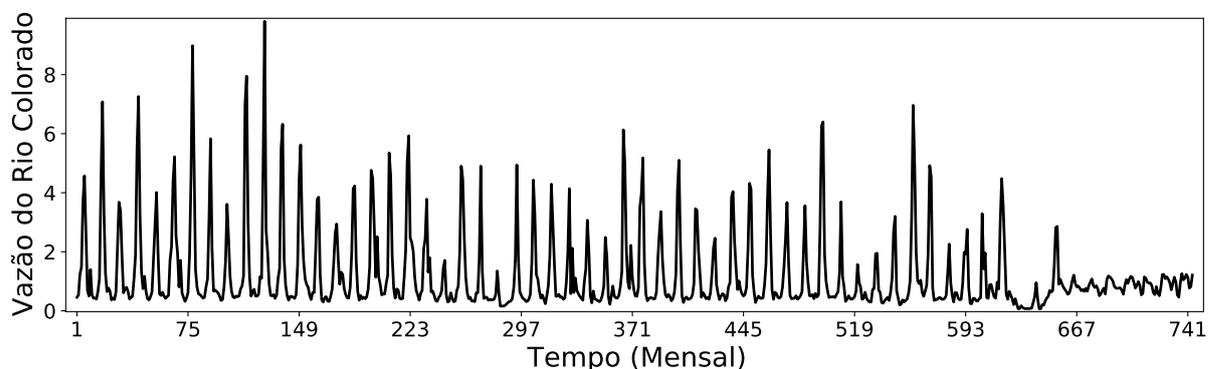
Figura 6 – FACP e FAC da série *Canadian Lynx*

Fonte: Elaborada pelo autor (2022)

essa afirmação pode também ser visualizada na Figura 5, em que é identificado que os valores da série permanecem próximos da média geral.

4.1.3 Colorado River

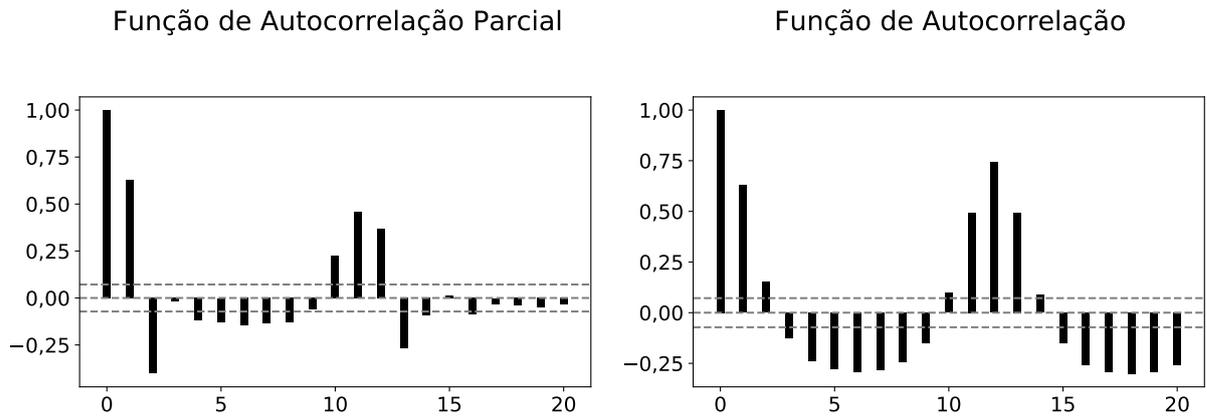
A série temporal *Colorado River* (Figura 7) consiste na vazão do afluente do rio Colorado situado em Lees Ferry (Estados Unidos da América). É utilizada a unidade temporal mensal entre os anos de 1911 e 1972. Esta série possui o valor médio, desvio padrão, mínimo e máximo de 1,2, 1,4, 0,1 e 9,8 respectivamente.

Figura 7 – Série temporal *Colorado River*

Fonte: Elaborada pelo autor (2022)

A Figura 8 exibe a FAC e FACP da série *Colorado River*. No FAC, é observado que a série em questão possui autocorrelações lineares relevantes, como também uma sazonalidade de

Figura 8 – FACP e FAC da série Colorado River



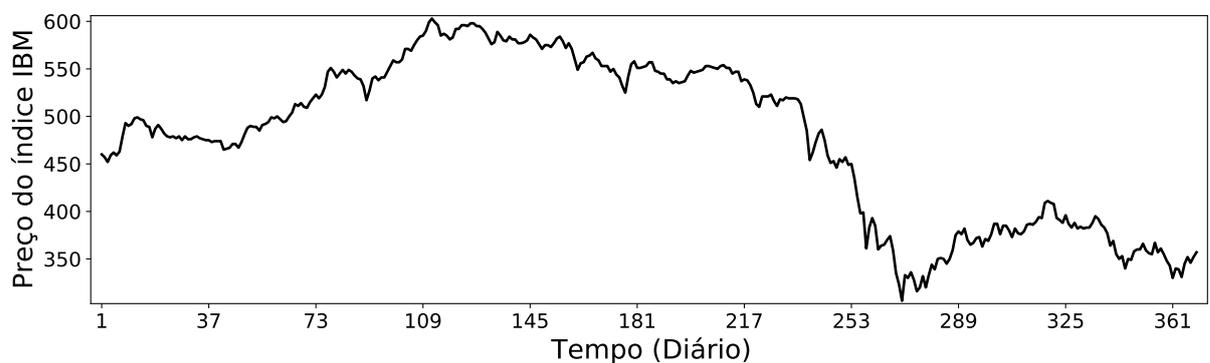
Fonte: Elaborada pelo autor (2022)

doze meses. No FACP, por sua vez, é possível observar que a autocorrelação relevante mais distante é no *lag* dezesseis. Pelo teste de KPSS é possível definir esta série temporal como não-estacionária.

4.1.4 IBM

A série temporal IBM (Figura 9) consiste nos preços de fechamento de ações da IBM. É utilizada a unidade temporal diária entre os anos de 1961 e 1962. Esta série possui o valor médio de 478,5, desvio padrão de 84,2, valor mínimo de 306 e máximo de 603.

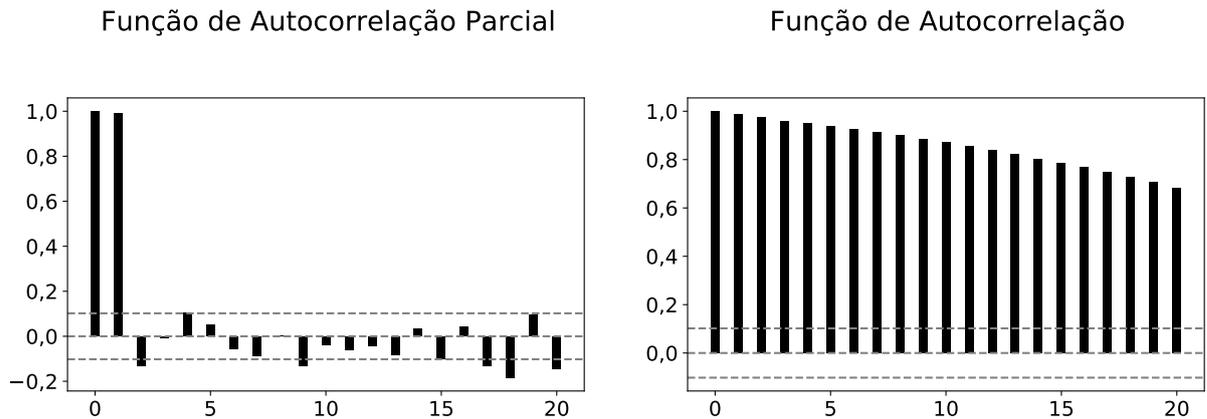
Figura 9 – Série temporal IBM



Fonte: Elaborada pelo autor (2022)

Pelo teste de KPSS é possível definir esta série temporal como não-estacionária. A Figura 10 exibe a FAC e FACP da série IBM. Pelo teste de FACP é observado que a série em

Figura 10 – FACP e FAC da série IBM



Fonte: Elaborada pelo autor (2022)

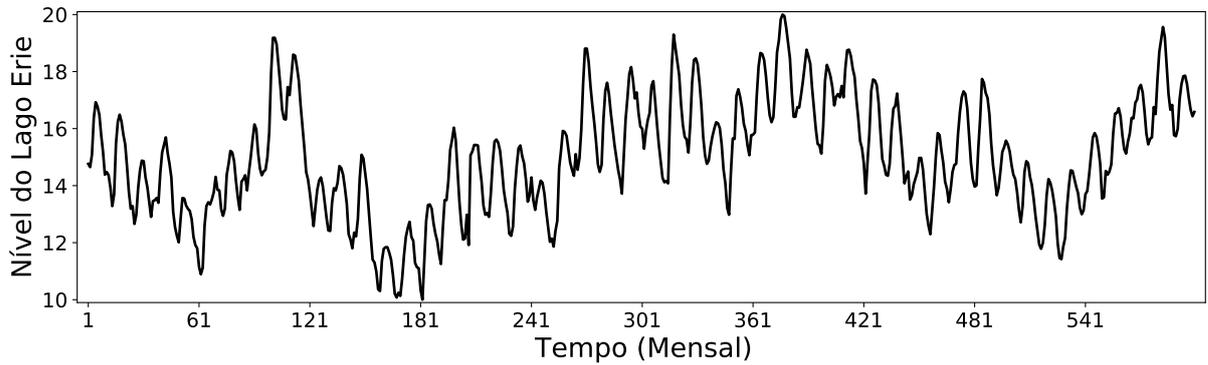
questão é altamente correlacionada apenas ao *lag* um, possuindo a autocorrelação relevante mais distante no *lag* vinte. Esse comportamento é comum em séries financeiras e é chamado de passeio aleatório (*Random Walk*). A Equação 4.1 representa o funcionamento do *Random Walk*, em que o valor da série temporal (z_t) será igual ao valor de z_{t-1} (janela 1) mais um ruído r_t , no tempo t .

$$z_t = z_{t-1} + r_t. \quad (4.1)$$

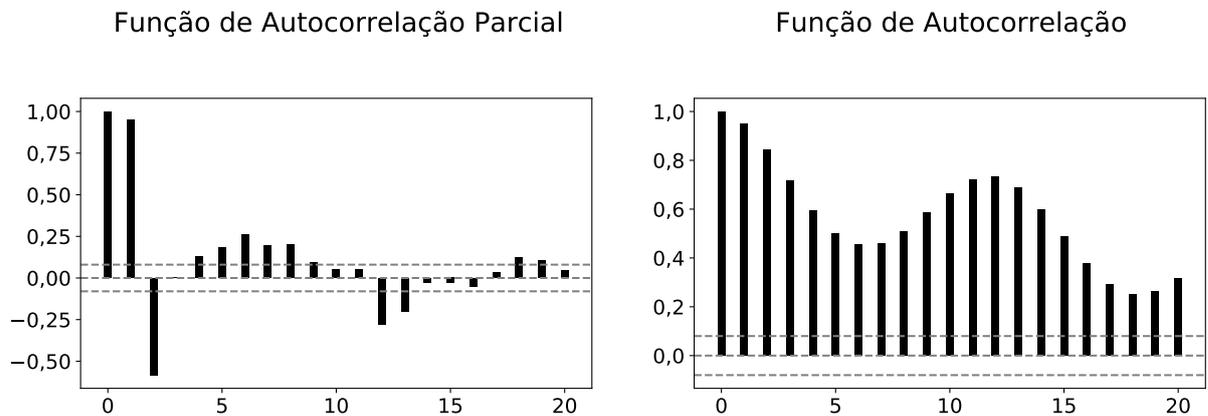
4.1.5 Lake Erie

A série temporal *Lake Erie* (Figura 11) consiste no nível do lago Erie. Este lago está situado entre o Canadá e Estados Unidos da América. É utilizada a unidade temporal mensal entre os anos de 1921 e 1970. Esta série possui o valor médio, desvio padrão, mínimo e máximo de 15, 2, 10 e 20, respectivamente.

A Figura 12 exibe a FAC e FACP da série *Lake Erie*. No FAC, observa-se que a série em questão possui autocorrelações lineares relevantes, como também uma sazonalidade de doze meses. No FACP, por sua vez, é possível observar que a autocorrelação relevante mais distante é no *lag* dezenove. Pelo teste de KPSS é possível definir esta série temporal como não-estacionária.

Figura 11 – Série temporal *Lake Erie*

Fonte: Elaborada pelo autor (2022)

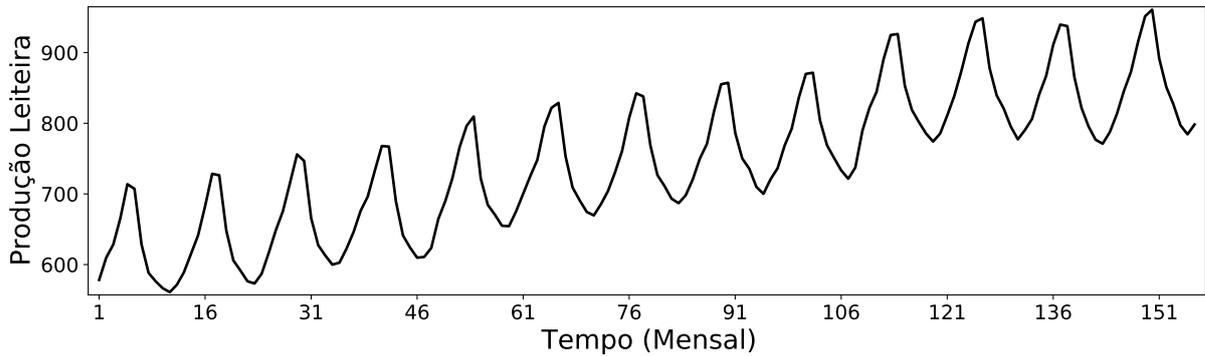
Figura 12 – FACP e FAC da série *Lake Erie*

Fonte: Elaborada pelo autor (2022)

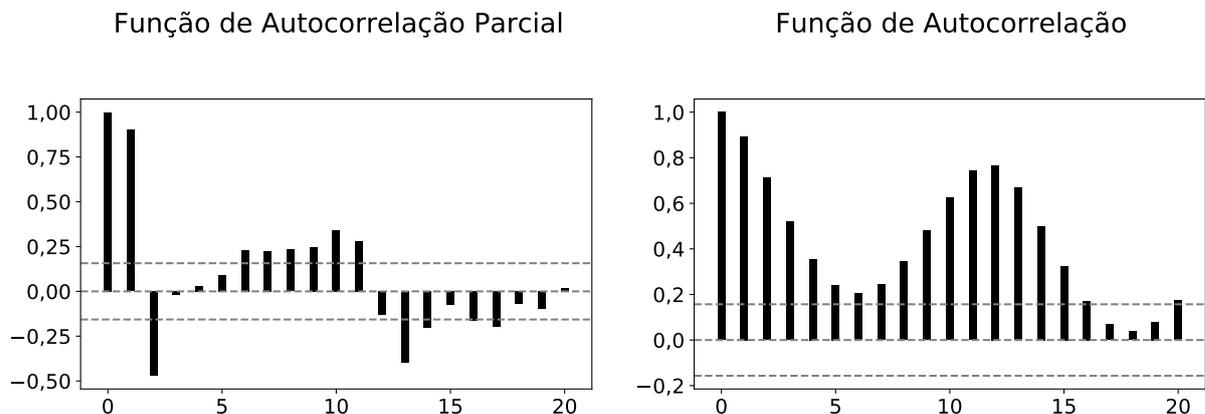
4.1.6 *Milk*

A série temporal *Milk* (Figura 13) consiste na produção de leite (libras por vaca). É utilizada a unidade temporal mensal entre os anos de 1962 e 1975. Esta série possui o valor médio, desvio padrão, mínimo e máximo de 746,5, 98,6, 561,1 e 960,8, respectivamente.

A Figura 14 exibe as Funções de Autocorrelação (FAC) e Autocorrelação Parcial (FACP) da série *Milk*. No FAC, observa-se que a série em questão possui autocorrelações lineares relevantes, como também uma sazonalidade de doze meses. No FACP, por sua vez, é possível observar que a autocorrelação relevante mais distante é no *lag* dezessete. Pelo teste de KPSS é possível definir esta série temporal como não-estacionária, essa afirmação pode também ser visualizada na Figura 13, em que é identificado um aumento da produção mensal ao longo dos anos.

Figura 13 – Série temporal *Milk*

Fonte: Elaborada pelo autor (2022)

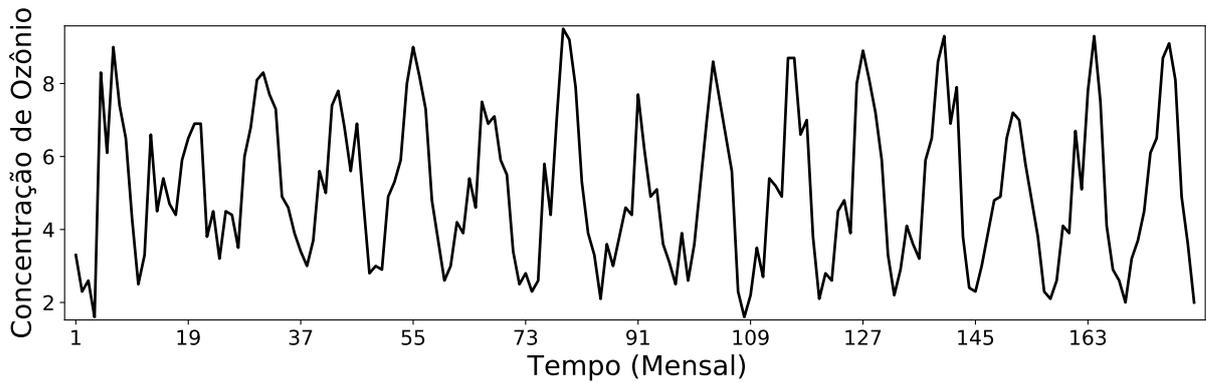
Figura 14 – FACP e FAC da série *Milk*

Fonte: Elaborada pelo autor (2022)

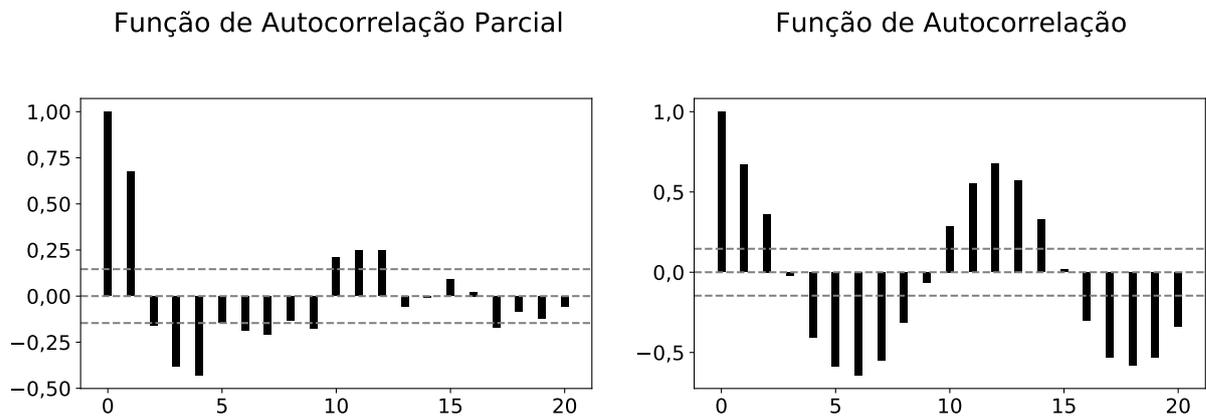
4.1.7 Ozon

A série temporal *Ozon* (Figura 15) representa a concentração de ozônio na cidade de Azusa (Estados Unidos da América). É utilizada a unidade temporal mensal entre os anos de 1956 e 1970. Esta série possui o valor médio, desvio padrão, mínimo e máximo de 5,1, 2,1, 1,6 e 9,5, respectivamente.

A Figura 16 exibe a FAC e FACP da série *Ozon*. No FAC, observa-se que a série em questão possui autocorrelações lineares relevantes, como também uma sazonalidade de doze meses. No FACP, por sua vez, é possível observar que a autocorrelação relevante mais distante é no *lag* dezessete. Pelo teste de KPSS é possível definir esta série temporal como estacionária.

Figura 15 – Série temporal *Ozon*

Fonte: Elaborada pelo autor (2022)

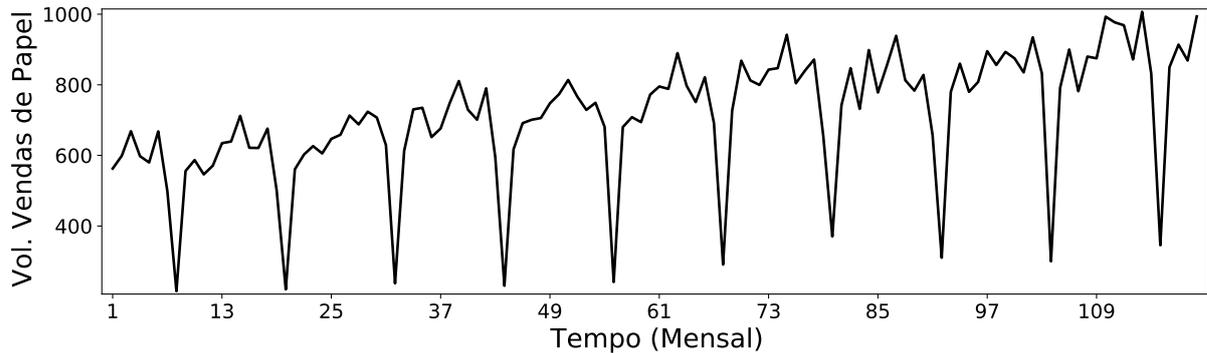
Figura 16 – FACP e FAC da série *Ozon*

Fonte: Elaborada pelo autor (2022)

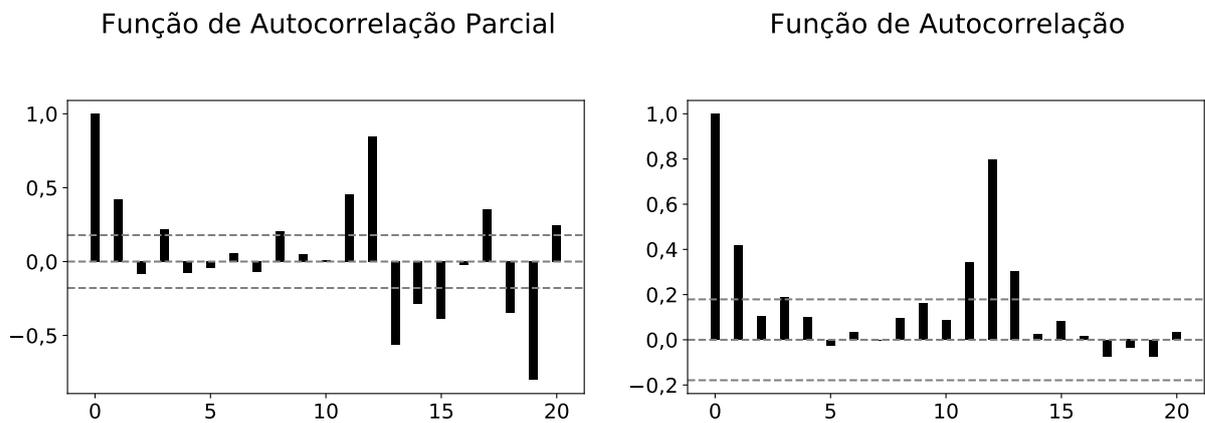
4.1.8 *Paper*

A série temporal *Paper* (Figura 17) consiste no volume de venda de papéis na França. É utilizado a unidade temporal mensal entre os anos de 1963 e 1972. Esta série possui o valor médio, desvio padrão, mínimo e máximo de 713,7, 174,1, 215,2 e 1006,9, respectivamente.

A Figura 18 exibe a FAC e FACP da série *Paper*. No FAC, observa-se que a série em questão possui autocorrelações lineares relevantes, como também uma sazonalidade de doze meses. No FACP, por sua vez, é possível observar que a autocorrelação relevante mais distante é no *lag* vinte. Pelo teste de KPSS é possível considerar esta série temporal como estacionária.

Figura 17 – Série temporal *Paper*

Fonte: Elaborada pelo autor (2022)

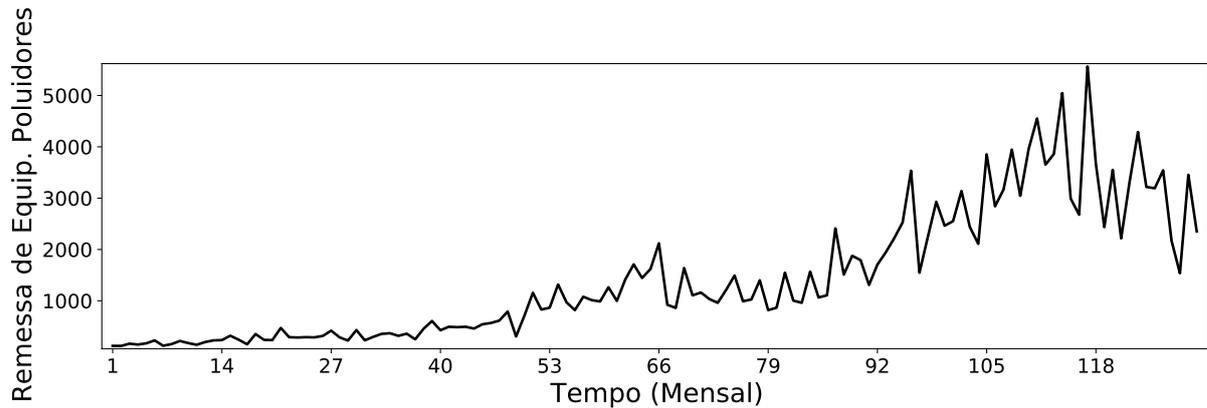
Figura 18 – FACP e FAC da série *Paper*

Fonte: Elaborada pelo autor (2022)

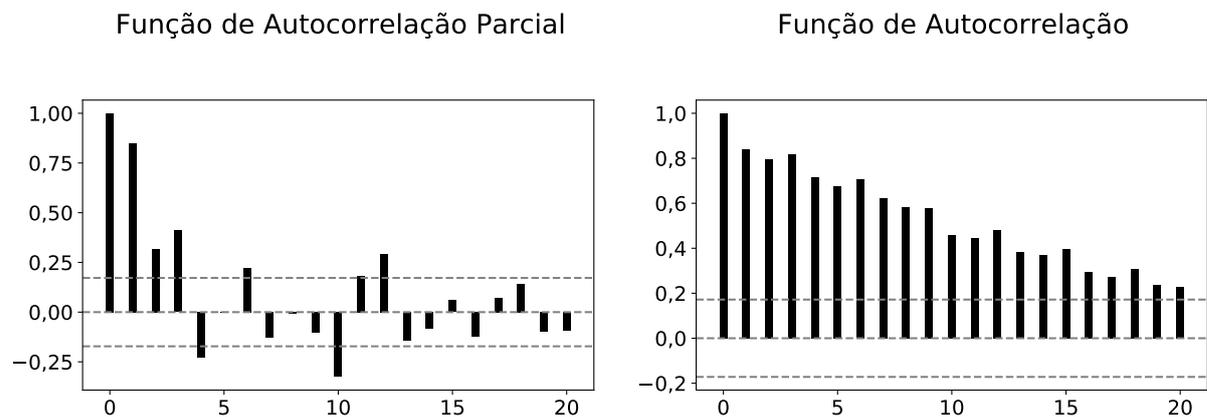
4.1.9 *Pollution*

A série temporal *Pollution* (Figura 19) consiste na remessa mensal de equipamentos de poluição (em milhares de francos franceses) entre os anos 1986 e 1997. Esta série possui o valor médio, desvio padrão, mínimo e máximo de 1439,3, 1261,2, 120,9 e 5566,1 respectivamente.

A Figura 20 exibe a FAC e FACP da série *Pollution*. No FAC é observado que a série em questão possui autocorrelações lineares relevantes. No FACP por sua vez, é possível observar que a autocorrelação relevante mais distante é no *lag* doze. Pelo teste de KPSS é possível definir esta série temporal como não-estacionária, essa afirmação pode também ser visualizada na Figura 7, em que é identificada uma alta variação na remessa ao longo dos anos.

Figura 19 – Série temporal *Pollution*

Fonte: Elaborada pelo autor (2022)

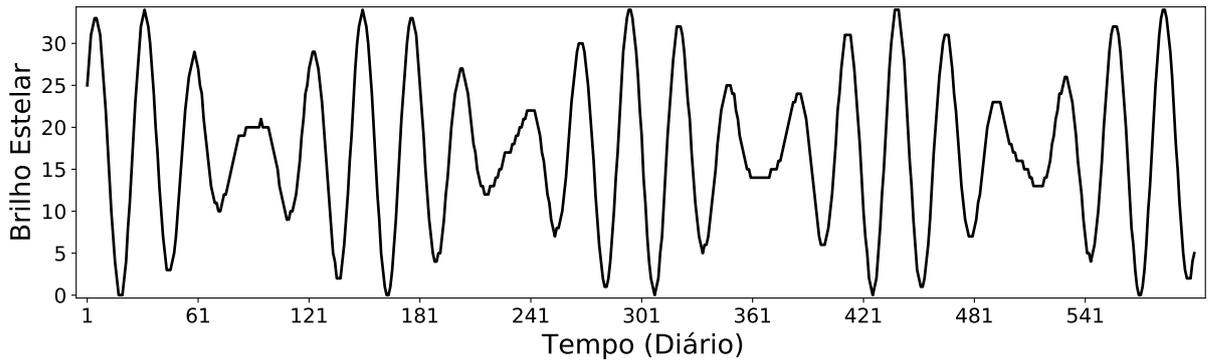
Figura 20 – FACP e FAC da série *Pollution*

Fonte: Elaborada pelo autor (2022)

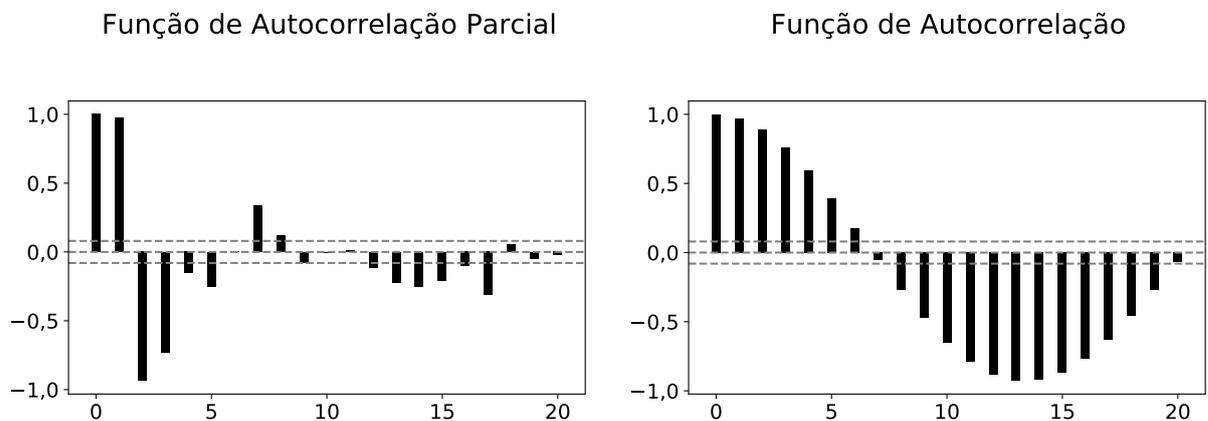
4.1.10 *Star Bright*

A série temporal *Star Bright* (Figura 21) compreende as observações diárias do brilho da estrela (mesma hora e local). Esta série possui o valor médio, desvio padrão, mínimo e máximo de 17, 9, 0 e 34, respectivamente.

A Figura 22 exibe a FAC e FACP da série *Star Bright*. No FAC, observa-se que a série em questão possui autocorrelações lineares relevantes, como também uma sazonalidade de treze dias. No FACP, por sua vez, é possível observar que a autocorrelação relevante mais distante é no *lag* dezessete. Pelo teste de KPSS é possível definir esta série temporal como estacionária.

Figura 21 – Série temporal *Star Bright*

Fonte: Elaborada pelo autor (2022)

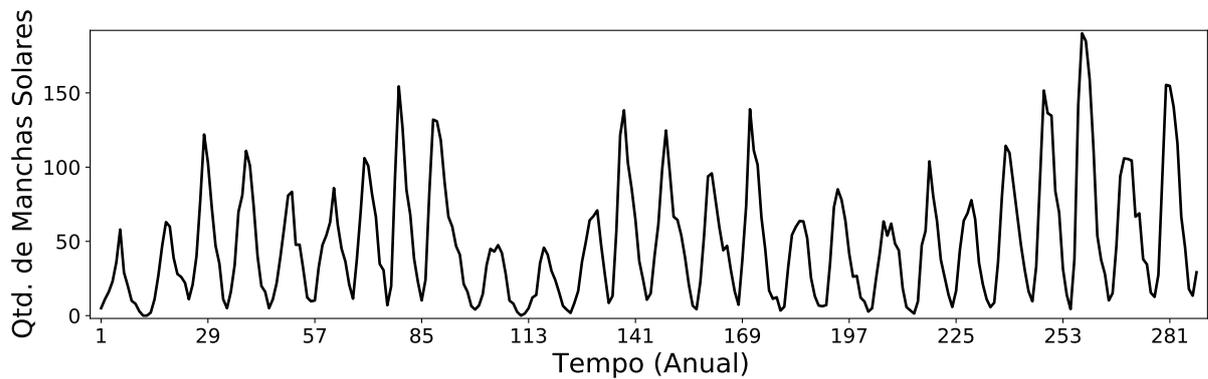
Figura 22 – FACP e FAC da série *Star Bright*

Fonte: Elaborada pelo autor (2022)

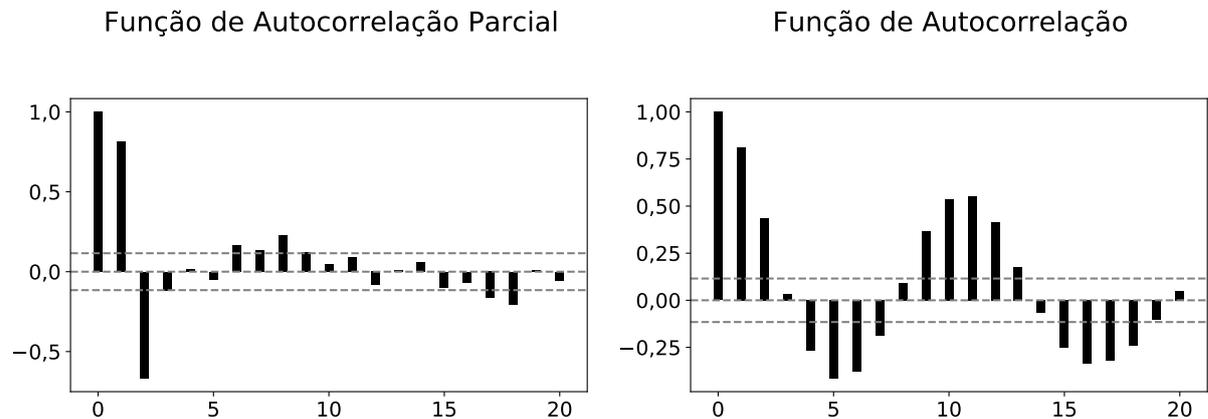
4.1.11 *Sunspot*

A série temporal *Sunspot* (Figura 23) consiste na quantidade de manchas na superfície do sol. É utilizada a unidade temporal anual entre 1700 e 1987. Esta série possui o valor médio 48,4, desvio padrão de 39,4, valor mínimo de 0 e máximo de 190,2.

A Figura 24 exibe a FAC e FACP da série *Sunspot*. No FAC, observa-se que a série em questão possui autocorrelações lineares relevantes, como também são observados ciclos de onze anos. No FACP, por sua vez, é possível observar que a autocorrelação relevante mais distante é no *lag* dezoito. Pelo teste de KPSS é possível definir esta série temporal como estacionária.

Figura 23 – Série temporal *Sunspot*

Fonte: Elaborada pelo autor (2022)

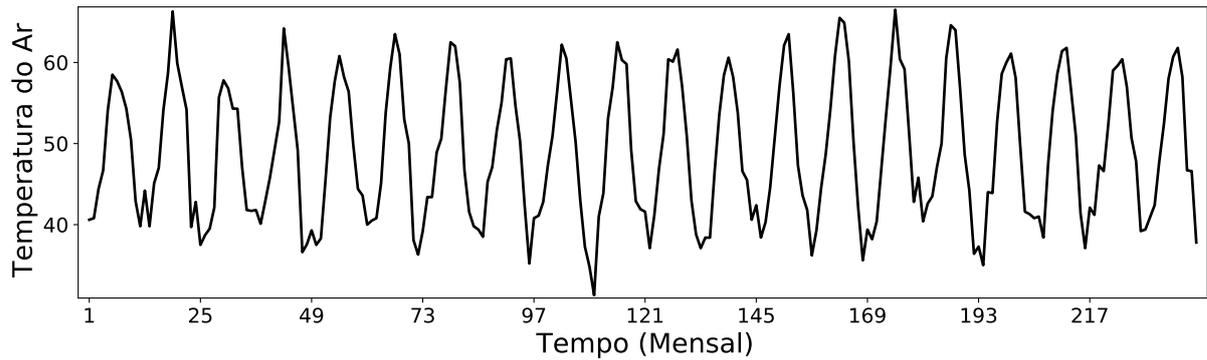
Figura 24 – FACP e FAC da série *Sunspot*

Fonte: Elaborada pelo autor (2022)

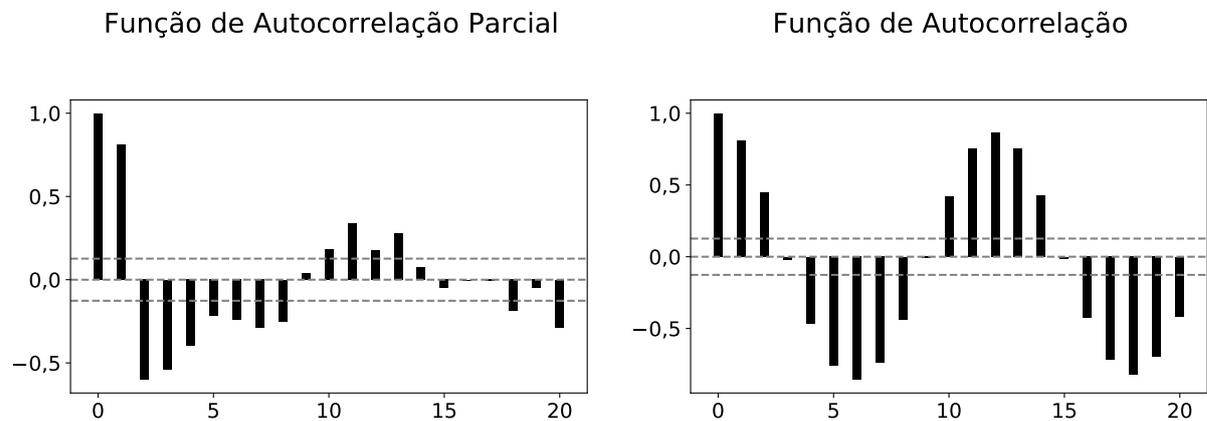
4.1.12 *Temperature*

A série temporal *Temperature* (Figura 7) consiste na temperatura mensal do ar na cidade de Nottingham (Inglaterra) entre os anos de 1920 e 1939. Esta série possui o valor médio, desvio padrão, mínimo e máximo de 49,0, 8,6, 31,3 e 66,5, respectivamente.

A Figura 26 exibe a FAC e FACP da série *Temperature*. No FAC, é observa-se que a série em questão possui autocorrelações lineares relevantes, como também uma sazonalidade de doze meses. No FACP, por sua vez, é possível observar que a autocorrelação relevante mais distante é no *lag* vinte. Pelo teste de KPSS é possível definir esta série temporal como estacionária, essa afirmação pode também ser visualizada na Figura 25, em que é identificado ao redor do tempo uma média constante.

Figura 25 – Série temporal *Temperature*

Fonte: Elaborada pelo autor (2022)

Figura 26 – FACP e FAC da série *Temperature*

Fonte: Elaborada pelo autor (2022)

4.2 MÉTODOS APLICADOS

Para avaliação dos métodos descritos a seguir é utilizada a métrica disponível na Equação 4.2, a qual descreve a métrica RMSE. Nesta equação, o K representa o tamanho da base de dados, z_t o valor atual e \hat{z}_t o valor previsto no tempo t . O RMSE possui a característica de evidenciar (punir) erros maiores, devido à sua sensibilidade a *outliers*. Quanto mais curta for o valor dessa métrica, melhor a acurácia. O RMSE tem sido amplamente empregado em trabalhos recentes da literatura de Sistemas Híbridos como métrica de avaliação (PANIGRAHI; BEHERA, 2017; PUROHIT et al., 2021; XU et al., 2021; SRETENOVIC et al., 2021).

$$RMSE = \sqrt{\frac{1}{K} \sum_{t=1}^K (z_t - \hat{z}_t)^2}. \quad (4.2)$$

A Tabela 2 exibe o nome e referência dos modelos aplicados nas comparações existentes no próximo capítulo. Nesta tabela existem informações sobre as duas versões da abordagem proposta (ERF_{MLP} e ERF_{SVR}). Ainda na Tabela 2, são exibidos os três modelos simples (ARIMA, MLP e SVR) e oito Sistemas Híbridos da literatura (ZHANG, 2003; PAI; LIN, 2005; KHASHEI; BIJARI, 2011; WANG et al., 2013; XU et al., 2019; XU et al., 2021; de HOLANDA; de OLIVEIRA, 2020). Estes sistemas híbridos são selecionados pois foram amplamente aplicados em diferentes séries temporais da literatura.

O ARIMA é utilizado como previsor da série temporal em ambas as abordagens propostas. ERF_{MLP} e ERF_{SVR} realizam a modelagem da série residual usando um *ensemble* de MLPs e um *ensemble* SVRs, respectivamente.

Tabela 2 – Nomenclatura dos métodos propostos e da literatura aplicados

Abordagem	Referência	Modelos
Modelo Simples	Hyndman, Khandakar et al. (2007)	ARIMA
	Pedregosa et al. (2011)	MLP
	Pedregosa et al. (2011)	SVR
Sistema Híbrido	Zhang (2003)	ARIMA e MLP
	Pai e Lin (2005)	ARIMA e SVR
	Xu et al. (2019)	ARIMA e DBN
	Xu et al. (2021)	ARIMA e LSTM
	Yan e Ma (2016)	ARIMA e RBF
	Wang et al. (2013)	ARIMA e MLP
	Khashei e Bijari (2011)	ARIMA e MLP
	de Holanda e de Oliveira (2020)	ARIMA e <i>ensemble</i> de SVRs por PSO
Método Proposto	ERF_{MLP}	ARIMA e <i>ensemble</i> de MLPs
	ERF_{SVR}	ARIMA e <i>ensemble</i> de SVRs

Fonte: Elaborada pelo autor (2022)

O ARIMA é desenvolvido utilizando a biblioteca *Forecast* disponível em R (HYNDMAN; KHANDAKAR et al., 2007). Para os modelos MLP e SVR é utilizada a biblioteca *python Scikit-*

learn (PEDREGOSA et al., 2011). O modelo RBF é desenvolvido por meio da classe *BaseEstimator* do *Scikit-learn* (PEDREGOSA et al., 2011). A LSTM e DBN são desenvolvidos utilizando a biblioteca TensorFlow para *python* (ABADI et al., 2015; ALBERTBUP, 2017). A classe *BaggingRegressor* da biblioteca *Scikit-learn* (PEDREGOSA et al., 2011) é utilizada para implementar o método de geração *Random Patches* presente no Passo II da fase de treinamento do método proposto (Figura 1a). Para este método de geração três parâmetros são necessários $n_estimators$, $max_samples$ e $max_features$ denominados como m , bs , e fs , respectivamente.

A Tabela 3 mostra a seleção de parâmetros das abordagens aplicadas neste trabalho. Para selecionar os parâmetros do ARIMA, p , d e q , é utilizada a abordagem automática de Hyndman, Khandakar et al. (2007). Para os métodos que utilizam DBN, LSTM, MLP, RBF e SVR, é feita a abordagem de Busca em Grade (*Grid Search*) para a seleção dos parâmetros. Este processo consiste em treinar e validar todas as possíveis combinações utilizando um subespaço de valores dos parâmetros. Cada combinação é executada 30 vezes e a combinação com menor RMSE médio na validação é selecionada. Utilizando esses parâmetros selecionados, os modelos são treinados e testados 30 vezes para análises subsequentes. Os parâmetros do ensemble m , bs , fs e Método de Agregação também são selecionados por meio do método *Grid Search* sobre os parâmetros disponíveis na Tabela 3.

Os resultados do método de de Holanda e de Oliveira (2020) são gerados segundo o código fonte disponível no GitHub do autor². Esse método usa um PSO para criar o ensemble de SVRs. O intervalo de parâmetros utilizado é: $C = (1e-3, \dots, 1e4)$, $\epsilon = (1e-3, \dots, 1e4)$, e $\gamma = (1e-3, \dots, 1e3)$. Os parâmetros específicos do PSO aplicados são: número de partículas = 100, $c_1 = 2$, $(c_2) = 2$, e inércia=1. A população final do processo de otimização é selecionado como ensemble, e suas previsões são agregadas por média ou mediana (baseada no conjunto de validação).

² <https://github.com/rbbh/ARIMA-SVR-PSO-ensemble>

Tabela 3 – Valores dos parâmetros para os modelos lineares e não lineares usados pelo ERF e métodos da literatura.

Modelo	Parâmetros	Valor
ARIMA	p, d, q	Hyndman, Khandakar et al. (2007)
DBN	Entradas	Análise de FACP
	Nós Escondidos	10; 20; 50
	Tx. aprendizado da máquina Boltzmann restrita - RBM	0,1; 0,01; 0,001
	Tx. aprendizado	0,1; 0,01; 0,001
LSTM	Entradas	Análise de FACP
	Nós Escondidos	20; 50; 100
	Algoritmo	Adam
	Função de Ativação	ReLU
	Interações	500
MLP	Nós Escondidos	20; 50; 100
	Entradas	Análise de FACP
	Algoritmo	lbfgs
	Função de Ativação	Sigmoide
	Interações	1000
	Tx. aprendizado	0,1; 0,9; 0,001; 0,0001
RBF	Nós Escondidos	2; 5; 10; 15; 20; 50
	Entradas	Análise de FACP
	Tx. aprendizado	0,1; 0,9; 0,001; 0,0001
	Interações	200
SVR	Nº of K-means <i>clusters</i>	2; 5; 10; 15; 20
	Entradas	Análise de FACP
	<i>Kernel</i>	RBF
	C	10; 100; 1000
	ϵ	0,9; 0,1; 0,01
Ensemble	γ	0,9; 0,1; 0,01
	m	10; 20; 50; 100
	bs (%)	40; 60; 80
	fs (%)	40; 60; 80
	Método de Agregação	média; mediana

Fonte: Elaborada pelo autor (2022)

4.3 RESUMO DO CAPÍTULO

Este capítulo exhibe as doze séries temporais estudadas com o objetivo de avaliar a abordagem proposta juntamente aos métodos da literatura. Essas séries possuem características diferentes em termos de quantidade de pontos, tendência, sazonalidade e estacionariedade. Além disso, estas são amplamente aplicadas na literatura de previsão de séries temporais. Para mensurar a qualidade dos métodos aplicados, é utilizada a métrica RMSE. Os métodos propostos e os da literatura utilizam modelos lineares e não lineares em suas arquiteturas, nos quais os parâmetros dos modelos não lineares (abordagens de AM) são selecionados utilizando a estratégia de *Grid Search* baseados no conjunto de validação, já os parâmetros do modelo linear (abordagem estatística ARIMA) utilizam a abordagem automática de Hyndman, Khandakar et al. (2007).

5 ANÁLISE EXPERIMENTAL

Este capítulo exhibe os resultados obtidos nos experimentos e apresenta os detalhes sobre a seleção de parâmetros do método proposto ERF (5.1 Análise de Sensibilidade), a comparação entre o método ERF e abordagens da literatura (5.2 Resultados Experimentais) e a avaliação entre todos os métodos em todas as séries temporais, identificando as abordagens com melhores classificações (*ranks*) e se são diferentes (5.3 Análise Final).

5.1 ANÁLISE DE SENSIBILIDADE

A Tabela 4 exhibe os parâmetros selecionados do *ensemble* no ERF_{MLP} e ERF_{SVR} para todas as séries temporais estudadas. Cada configuração de parâmetros do *ensemble* é selecionada com base no RMSE médio de validação. Na Tabela 4 é possível observar que os parâmetros mais comumente selecionados são o $w = 100$, $bs = 80\%$, $fs = 80\%$ e mediana como método de agregação. Para ambos os métodos propostos, somente o $m = 10$ não é selecionado para nenhuma série temporal.

A Figura 27 exhibe alguns exemplos de como os parâmetros do ERF (m , bs , fs e método de agregação) podem influenciar os resultados do RMSE no teste e validação para as séries *Milk*, *Paper*, *Pollution* e *River*. Estas séries são usadas como exemplo porque a variação dos parâmetros influenciou os seus resultados em diferentes formas. As figuras da análise de sensibilidade das outras oito séries temporais podem ser encontradas no Anexo A. Nessas figuras cada parâmetro é analisado independentemente, sendo que um parâmetro é variado e os demais são fixados como MLP, 100, 80%, 80% e mediana como modelo base, m , bs , fs e método de agregação, respectivamente.

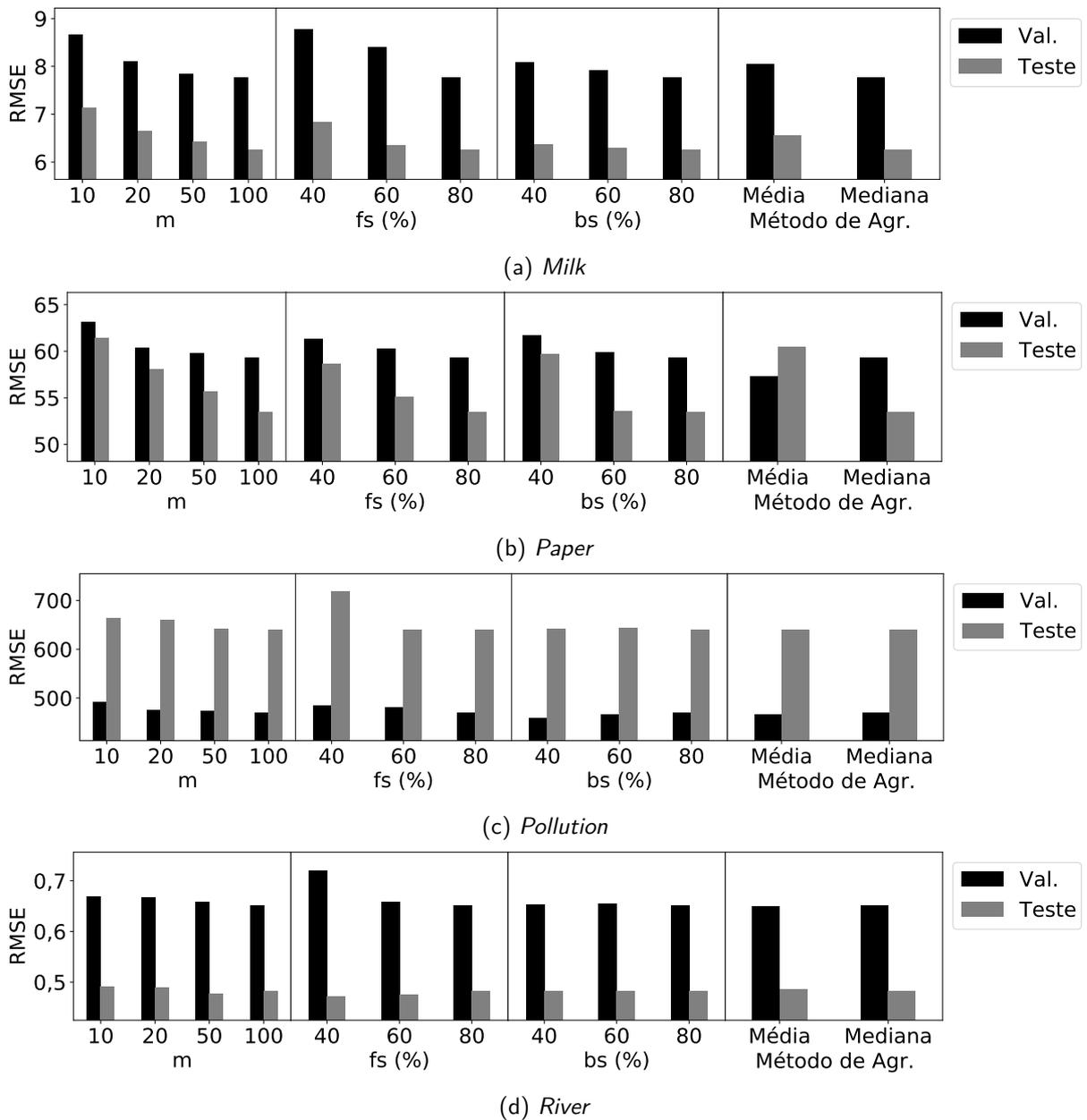
A Figure 27 mostra a capacidade de generalização do método ERF em que na comparação do grupo de parâmetros, a mesma configuração obteve bons resultados em ambos os conjuntos de validação e teste. Valores maiores para os parâmetros m e fs tendem a encontrar melhores resultados. Este padrão não é observado para o parâmetro bs , no qual valores menores podem também encontrar bons resultados (exemplo: Figura 27c). Considerando o método de agregação, observa-se que a mediana é geralmente a melhor opção em ambos os conjuntos de validação e teste.

Tabela 4 – Parâmetros do *ensemble* selecionados baseado no RMSE de validação

Série temporal	Modelo Base	m	bs (%)	fs (%)	Método de Agregação
<i>Airlines</i>	ERF _{MLP}	100	40	80	Mediana
	ERF _{SVR}	100	80	80	Mediana
IBM	ERF _{MLP}	50	80	40	Média
	ERF _{SVR}	100	80	40	Média
<i>Lake</i>	ERF _{MLP}	50	40	60	Mediana
	ERF _{SVR}	100	80	40	Média
<i>Lynx</i>	ERF _{MLP}	100	80	40	Mediana
	ERF _{SVR}	100	80	80	Média
<i>Milk</i>	ERF _{MLP}	100	80	80	Mediana
	ERF _{SVR}	100	60	80	Mediana
<i>Ozon</i>	ERF _{MLP}	100	60	60	Mediana
	ERF _{SVR}	100	80	80	Média
<i>Paper</i>	ERF _{MLP}	100	80	80	Média
	ERF _{SVR}	50	40	80	Mediana
<i>Pollution</i>	ERF _{MLP}	100	40	80	Média
	ERF _{SVR}	50	40	80	Média
<i>River</i>	ERF _{MLP}	100	40	80	Média
	ERF _{SVR}	100	40	60	Mediana
<i>Star</i>	ERF _{MLP}	100	80	80	Mediana
	ERF _{SVR}	100	80	80	Mediana
<i>Sunspot</i>	ERF _{MLP}	100	60	40	Mediana
	ERF _{SVR}	100	40	40	Média
<i>Temperature</i>	ERF _{MLP}	100	80	60	Média
	ERF _{SVR}	100	80	80	Mediana

Fonte: Elaborada pelo autor (2022)

Figura 27 – Análise de sensibilidade dos parâmetros do método proposto ERF_{MLP} para as séries temporais (a) *Milk*, (b) *Paper*, (c) *Pollution* e (d) *River*



Fonte: Elaborada pelo autor (2022)

5.2 RESULTADOS EXPERIMENTAIS

Tabela 5 – Resultados do RMSE para as séries temporais utilizando os métodos simples e propostos. O melhor e segundo melhor estão destacados com negrito e sublinhado, respectivamente

Série Temporal	ARIMA	MLP	SVR	ERF_{MLP}	ERF_{SVR}
Airlines	43,624	18,519	45,837	16,677	<u>17,663</u>
River	0,505	0,525	0,505	0,487	<u>0,493</u>
IBM	7,558	8,875	8,611	<u>7,665</u>	7,697
Lake	0,385	0,370	<u>0,375</u>	0,378	0,377
Lynx	0,201	0,398	0,296	0,212	<u>0,210</u>
Milk	23,009	9,352	11,554	6,266	<u>6,514</u>
Ozon	1,018	1,561	0,930	0,997	<u>0,994</u>
Paper	176,171	85,418	104,707	<u>60,511</u>	59,696
Pollu.	795,927	5696,494	946,711	655,797	<u>673,68</u>
Star	0,717	0,377	0,321	0,293	<u>0,294</u>
Sunspot	21,573	29,325	22,669	20,506	<u>20,898</u>
Temp.	2,654	2,887	2,560	2,642	<u>2,628</u>

Fonte: Elaborada pelo autor (2022)

A Tabela 5 exibe o RMSE médio das 30 execuções no conjunto de teste utilizando os métodos propostos (ERF_{MLP} e ERF_{SVR}) e modelos simples (ARIMA, MLP e SVR). Esta tabela mostra que o ERF_{MLP} obteve os melhores resultados em seis séries temporais e o segundo melhor resultado em duas séries temporais. O ERF_{SVR} obteve o melhor resultado em uma série temporal, por outro lado, alcançou o segundo melhor resultado em nove das séries temporais. Considerando agora apenas os modelos simples, o SVR e ARIMA ficaram empatados, uma vez que cada um conseguiu obter a melhor métrica em duas séries temporais. Ainda entre os modelos simples, o SVR alcançou o segundo melhor resultado em uma das séries temporais estudadas.

Na comparação do RMSE médio das 30 execuções no conjunto de teste utilizando os métodos propostos e sistemas híbridos da literatura (Tabela 6), o ERF_{MLP} obteve a melhor métrica em duas séries temporais e o segundo melhor resultado em sete séries. O ERF_{SVR} alcançou o melhor resultado em três séries e o segundo melhor em duas séries temporais. A Tabela 6 mostra ainda que Pai e Lin (2005) é o melhor sistema híbrido entre as abordagens da literatura, no qual alcançou o melhor resultados em três séries temporais e o segundo melhor em uma série. O Anexo B exibe os resultados médios considerando outras métricas.

Tabela 6 – Resultados do RMSE para as séries temporais utilizando os Sistemas Híbridos da literatura e propostos. O melhor e segundo melhor estão destacados em negrito e sublinhado, respectivamente

Série Temporal	Zhang (2003)	Pai e Lin (2005)	Xu <i>et al.</i> (2019)	Yan e Ma (2016)	Xu <i>et al.</i> (2021)	Khashei e Bijari (2011)	Wang <i>et al.</i> (2013)	de Holanda e de Oliveira (2020)	ERF _{MLP}	ERF _{SVR}
Airlines	32,380	16,993	30,226	43,481	54,087	46,902	26,445	17,176	16,677	17,663
River	0,522	0,514	<u>0,484</u>	0,517	0,662	0,504	0,501	0,474	0,487	0,493
IBM	7,872	8,026	7,67	7,715	7,722	17,074	7,808	7,556	<u>7,665</u>	7,697
Lake	0,374	<u>0,377</u>	0,379	0,409	0,384	0,385	0,373	0,378	0,378	<u>0,377</u>
Lynx	0,370	0,220	0,208	<u>0,201</u>	0,232	0,200	0,360	0,209	0,212	0,210
Milk	8,020	6,252	11,749	11,622	16,557	22,669	9,375	7,416	<u>6,266</u>	6,514
Ozon	1,511	1,005	1,015	1,032	1,045	1,016	1,191	1,039	<u>0,997</u>	0,994
Paper	102,703	67,512	113,508	157,666	149,275	176,172	99,917	62,684	<u>60,511</u>	59,696
Pollu.	656,57	642,522	782,125	799,766	809,117	858,316	2157,745	674,435	<u>655,797</u>	673,68
Star	0,324	0,289	0,447	0,592	0,432	0,713	8,95	0,346	<u>0,293</u>	0,294
Sunspot	35,676	24,138	20,595	22,181	20,723	21,592	37,147	21,816	20,506	<u>20,898</u>
Temp.	3,718	2,731	2,702	2,812	2,628	2,654	3,496	2,718	2,642	2,628

Fonte: Elaborada pelo autor (2022)

Tabela 7 – Teste de hipótese Wilcoxon com 95% de confiança. Os símbolos (+), (-) ou (=) representam que o método proposto ERF_{MLP} é melhor, pior ou equivalente ao outro método, respectivamente.

Model	Airlines	River	IBM	Lake	Lynx	Milk	Ozon	Paper	Pollu.	Star	Sunspot	Temp.	Total + =-
ARIMA	+	+	-	+	-	+	+	+	+	+	+	=	10
MLP	=	+	+	-	+	+	+	+	+	+	+	+	11
SVR	+	+	+	-	+	+	-	+	+	+	+	-	9
Zhang (2003)	+	+	+	-	+	+	+	+	-	+	+	+	10
Pai e Lin (2005)	+	+	+	=	+	=	+	+	-	-	+	+	10
Xu <i>et al.</i> (2019)	+	-	=	=	-	+	+	+	+	+	=	+	10
Yan e Ma (2016)	+	+	+	+	-	+	+	+	+	+	+	+	11
Xu <i>et al.</i> (2021)	+	+	+	+	+	+	+	+	+	+	=	=	12
Khashei e Bijari (2011)	+	+	+	+	-	+	+	+	+	+	+	=	11
Wang <i>et al.</i> (2013)	+	+	+	-	+	+	+	+	+	+	+	+	11
de Holanda e de Oliveira (2020)	+	-	-	=	=	+	+	+	+	+	+	+	10

Fonte: Elaborada pelo autor (2022)

Tabela 8 – Teste de hipótese Wilcoxon com 95% de confiança. Os símbolos (+), (-) ou (=) representam que o método proposto ERF_{SVR} é melhor, pior ou equivalente ao outro método, respectivamente.

Model	Airlines	River	IBM	Lake	Lynx	Milk	Ozon	Paper	Pollu.	Star	Sunspot	Temp.	Total + =
ARIMA	+	+	-	+	-	+	+	+	+	+	+	+	10
MLP	=	+	+	-	+	+	+	+	+	+	+	+	11
SVR	+	+	+	-	+	+	-	+	+	+	+	-	9
Zhang (2003)	+	+	+	-	+	+	+	+	-	+	+	+	10
Pai e Lin (2005)	-	+	+	=	+	-	+	+	-	-	+	+	8
Xu et al. (2019)	+	-	=	+	-	+	+	+	+	+	-	+	9
Yan e Ma (2016)	+	+	=	+	-	+	+	+	+	+	+	+	11
Xu et al. (2021)	+	+	=	+	+	+	+	+	+	+	=	=	12
Khashei e Bijari (2011)	+	+	+	+	-	+	+	+	+	+	+	+	11
Wang et al. (2013)	+	+	+	-	+	+	+	+	+	+	+	+	11
de Holanda e de Oliveira (2020)	-	-	-	+	=	+	+	+	=	+	+	+	9

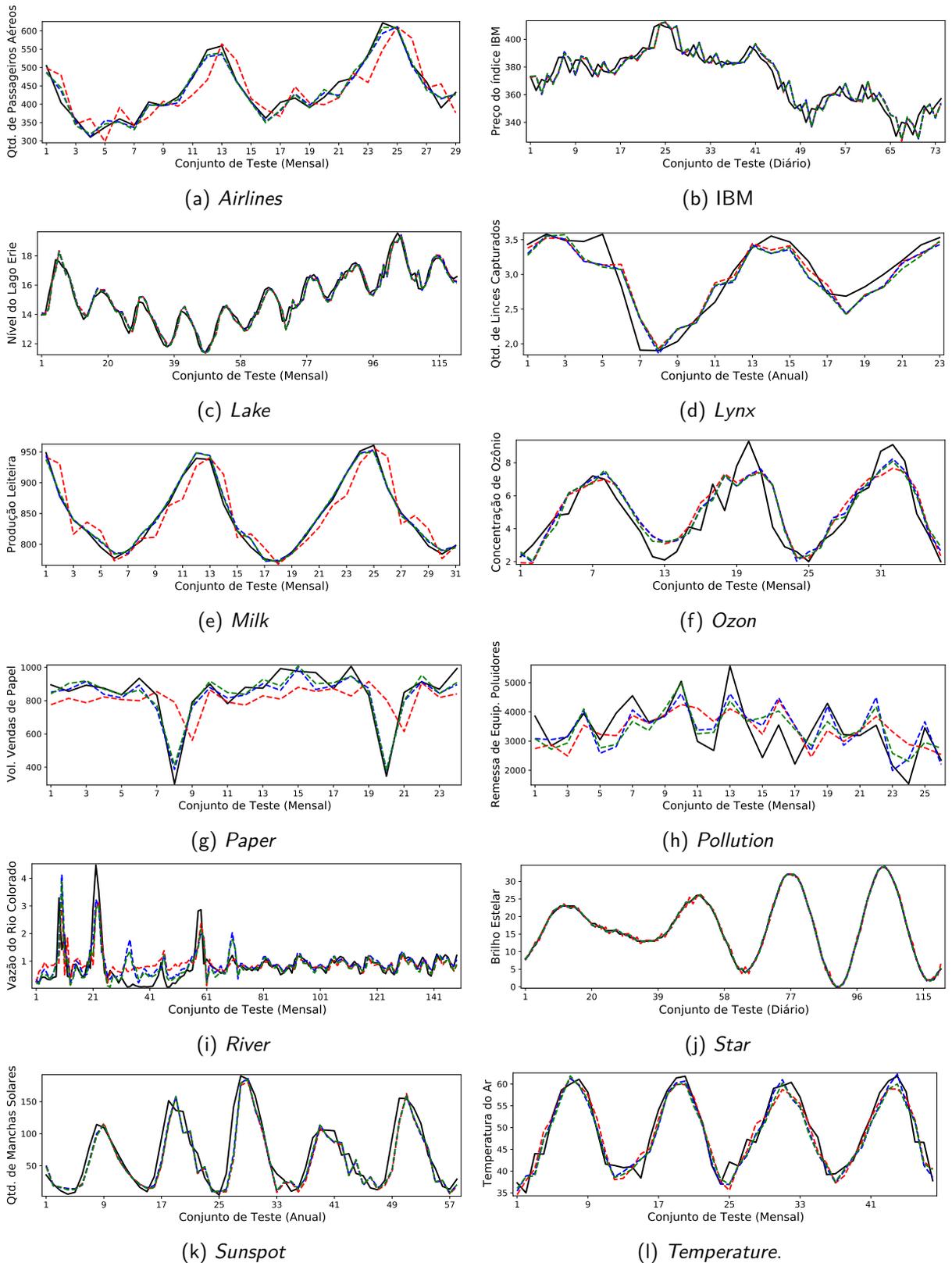
Fonte: Elaborada pelo autor (2022)

As Tabelas 7 e 8 exibem o teste de hipótese de Wilcoxon com 95% de confiança sobre as 30 execuções do RMSE de teste para as abordagens propostas ERF_{MLP} e ERF_{SVR} , respectivamente. Essas tabelas seguem a representação que o método proposto pode ser melhor (+), pior (-) ou equivalente (=) comparativamente às abordagens da literatura. Desta forma, é possível afirmar que o ERF_{MLP} é estatisticamente melhor ou igual ao de Xu et al. (2021) em todos os casos; MLP, Yan e Ma (2016), Khashei e Bijari (2011) e Wang et al. (2013) em onze casos; ARIMA, Zhang (2003), Pai e Lin (2005), Xu et al. (2019) e de Holanda e de Oliveira (2022) em dez casos e SVR em nove casos. Considerando o ERF_{SVR} (Tabela 8), este método proposto é estatisticamente melhor ou igual ao de Xu et al. (2021) em todos os casos; MLP, Yan e Ma (2016), Khashei e Bijari (2011) e Wang et al. (2013) em onze casos; ARIMA e Zhang (2003) em 10 casos; SVR, Xu et al. (2019) e de Holanda e de Oliveira (2022) em nove casos e Pai e Lin (2005) em oito casos.

A Figura 28 exibe a previsão no conjunto de teste para todas as séries temporais estudadas, aplicando o ARIMA e métodos propostos. Nesta figura, a linha preta apresenta o valor real, linha azul é a previsão do ERF_{MLP} , a linha verde é a previsão do ERF_{SVR} e a linha vermelha a

previsão do ARIMA. É possível observar que as versões do método proposto podem obter previsões mais próximas ao real quando comparadas ao ARIMA. Alguns exemplos dessa melhoria podem ser observados nas séries *Airlines*, *River*, *Milk*, *Ozon*, *Paper* e *Temperature*.

Figura 28 – Valor real da série temporal (linha preta) com as respectivas previsões, aplicando o ARIMA (linha vermelha), ERF_{MLP} (linha azul) e ERF_{SVR} (linha verde)



Fonte: Elaborada pelo autor (2022)

5.3 ANÁLISE FINAL

Os testes de hipóteses de Friedman e Nemenyi são utilizados para classificar (*rank*) todos os modelos sobre todas as execuções das séries temporais (HOLLANDER; WOLFE; CHICKEN, 2013). O teste de hipótese de Friedman compara a classificação média dos resultados dos modelos, considerando a hipótese nula de que as classificações dos modelos são equivalentes. Se a hipótese nula do teste de Friedman for rejeitada, conclui-se que as classificações dos modelos são significativamente diferentes e é possível utilizar o teste de Nemenyi (DEMŠAR, 2006), que por sua vez, aplica uma comparação pareada dos modelos sobre todas as séries temporais (PANIGRAHI; BEHERA, 2017; de OLIVEIRA; LUDERMIR, 2016). Esse teste considera um grupo de w modelos, c execuções nas séries temporais e o valor crítico (q_α) baseado no o intervalo estatístico estudentizado (DEMŠAR, 2006). Se a diferença entre as classificações for maior que CD (Equação 5.1), os modelos são diferentes.

$$CD = q_\alpha \sqrt{\frac{w(w+1)}{6c}}. \quad (5.1)$$

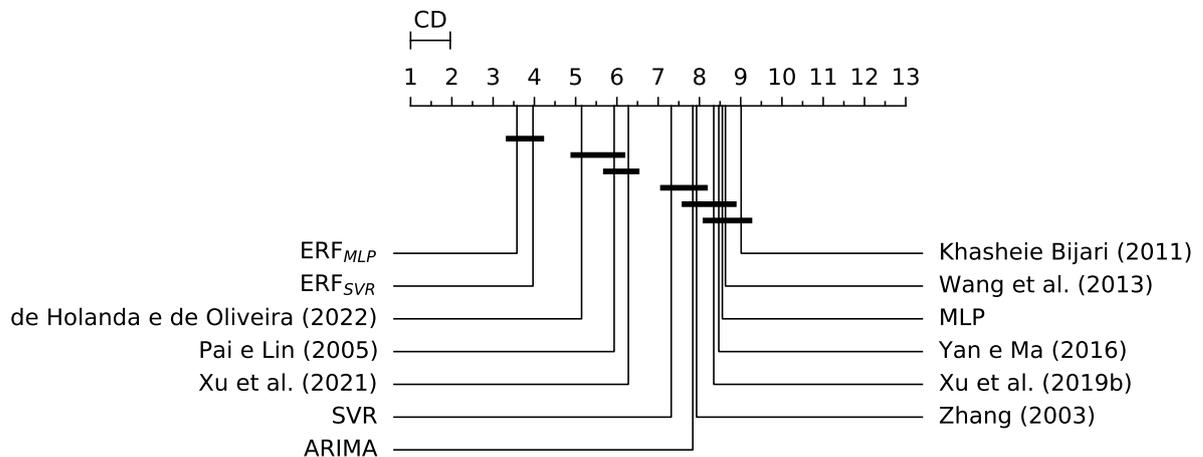
As bibliotecas *python* Scipy (VIRTANEN et al., 2020) e *Orange* (DEMŠAR et al., 2013) são usadas para desenvolver os testes de Friedman e Nemenyi, respectivamente.

Considerando a classificação (baseada no RMSE) dos treze métodos ($w = 13$) sobre as 30 execuções em doze séries temporais ($c = 30 \times 12$), a hipótese nula do teste de Friedman com 95% de confiança ($p\text{-value} = 3.36e-176$) pode ser rejeitada. Desta forma, é possível concluir que as classificações dos métodos são significativamente diferentes e pode-se prosseguir para o teste de Nemenyi. O teste de Nemenyi aplica o $q_\alpha = 3,312$ ($\alpha = 0,05$ ou 95% de confiança), obtendo $CD = 0,96$.

A Figura 29 exibe o teste de hipótese de Nemenyi, o qual indica que ambas as versões do método proposto são significativamente melhores que os modelos simples e sistemas híbridos da literatura. O ERF_{MLP} obteve a melhor classificação geral, seguido pelo ERF_{SVR} . Outro aspecto interessante nesta figura é que os métodos propostos ERF_{MLP} e ERF_{SVR} possuem classificações significativamente melhores que Zhang (2003) e Pai e Lin (2005), que representam os sistemas híbridos que utilizam uma MLP e uma SVR para prever o resíduo, respectivamente. Mostrando assim que a utilização do *ensemble* proposto para modelagem residual é uma estratégia mais segura comparativamente ao respectivo modelo simples. A Figura 29 demonstra também que as abordagens propostas atingiram classificações significan-

temente melhores em comparação aos modelos simples ARIMA, SVR e MLP. Aspecto positivo que muitos sistemas híbridos da literatura não foram capazes de alcançar.

Figura 29 – Classificação do teste de Nemenyi considerando o RMSE



Fonte: Elaborada pelo autor (2022)

5.4 RESUMO DO CAPÍTULO

Neste capítulo foram exibidos os resultados experimentais das abordagens propostas nomeadas como ERF_{MLP} e ERF_{SVR} . Inicialmente foi analisada a sensibilidade dos parâmetros dos métodos propostos. Nessa análise foi observado que um *ensemble* maior tende a atingir melhores resultados, assim como a utilização do método de agregação mediana. Foi observado também que a utilização das métricas de validação para seleção da configuração do ensemble é uma estratégia segura para obter bons resultados no conjunto de teste. As abordagens propostas foram comparadas a três modelos simples e oito sistemas híbridos da literatura. Na avaliação em doze séries temporais reais, as abordagens ERF_{MLP} e ERF_{SVR} encontram bons resultados em termos de RMSE médio, sempre mantendo suas métricas entre as melhores. Foi realizada também a avaliação estatística de Wilcoxon, um teste pareado que mostrou que as abordagens ERF_{MLP} e ERF_{SVR} são estatisticamente superiores na grande maioria das análises. Por fim, para se ter uma classificação geral levando em conta todos os modelos e séries temporais, foram aplicados os testes de hipótese de Friedman e Nemenyi. Por meio desses testes foi possível afirmar que entre as treze abordagens estudadas (dois métodos propostos, três modelos simples e oito SHs da literatura) o ERF_{MLP} e ERF_{SVR} possuem a melhor e

segunda melhor classificação geral, respectivamente.

6 CONCLUSÃO

Este capítulo apresenta as conclusões desta tese (6.1 Considerações Finais) e sugestões para próximas pesquisas (6.2 Trabalhos Futuros).

6.1 CONSIDERAÇÕES FINAIS

A proposta desta tese é um método de *Ensemble* para previsão residual (*Ensemble method for Residual Forecast* – ERF), consistindo em um sistema híbrido que combina um modelo estatístico linear com um *ensemble* de modelos de AM. Inicialmente é usado o modelo linear, obtendo a previsão da série temporal e o respectivo resíduo. Em seguida, um *ensemble* prevê esse resíduo, consistindo na previsão não linear de m modelos de AM agregados pela média ou mediana. Cada modelo de AM do *ensemble* é treinado com uma porcentagem aleatória do treino (*Bagging Size* – bs) e das entradas (*Feature Size* – fs). A previsão final da abordagem proposta consiste na soma entre a previsão linear da série temporal e a previsão do resíduo do *ensemble*.

O modelo estatístico ARIMA é aplicado como modelo linear, já como modelo não linear é usado a MLP ou SVR, obtendo assim duas versões da abordagem proposta chamadas de ERF_{MLP} e ERF_{SVR} , respectivamente.

A análise experimental é conduzida por meio de doze séries temporais da literatura, sendo que treze métodos são considerados: duas abordagens propostas, ARIMA, MLP, SVR e oito sistemas híbridos da literatura (ZHANG, 2003; PAI; LIN, 2005; KHASHEI; BIJARI, 2011; WANG et al., 2013; XU et al., 2019; XU et al., 2021; de HOLANDA; de OLIVEIRA, 2020)

Seguindo o teorema *no free lunch* (SILVA et al., 2020), nenhum método é o melhor em todas as séries temporais. Contudo, em termos de RMSE médio (30 execuções), os resultados de teste das abordagens propostas estão entre os melhores resultados considerando as séries temporais estudadas. Como também são estatisticamente melhores que as abordagens da literatura na maioria dos casos. Adicionalmente, os métodos propostos alcançaram as melhores classificações nos testes de hipótese de Friedman e Nemenyi, em que ERF_{MLP} e ERF_{SVR} obtiveram a melhor e segunda melhor classificação, respectivamente.

Observa-se também que existe um grupo de parâmetros do *ensemble* que são comumente selecionados considerando as doze séries temporais aplicadas: mediana como método de agre-

gação, $m = 100$, $bs = 80\%$ e $fs = 80\%$. Esta observação é uma importante característica do método proposto, uma vez que com pouco esforço na etapa de seleção de parâmetros é possível encontrar bons resultados nos conjuntos de validação e teste.

Desta forma, é possível listar as contribuições desta tese como a seguir:

- i. Arquitetura de Sistema Híbrido que considera ambos os conceitos de modelagem residual e *ensemble*;
- ii. Estratégia genérica que pode ser utilizada com qualquer modelo linear e não linear, podendo também ser aplicada a qualquer contexto de previsão de séries temporais;
- iii. Análise sobre como os parâmetros do *ensemble* podem afetar o resultado final do sistema híbrido proposto;
- iv. Sistema que corrige as previsões do ARIMA em diferentes séries temporais;
- v. Extensiva análise estatística entre métodos propostos, modelos simples e sistemas híbridos da literatura.

6.2 TRABALHOS FUTUROS

Esta tese propõe uma abordagem capaz de gerar um *ensemble* para previsão residual, exibindo resultados competitivos quando comparados aos métodos da literatura. Desta forma, é estabelecida uma forma segura e eficiente para se gerar um *ensemble* nesse contexto, abrindo assim oportunidades para pesquisas futuras, algumas delas são:

- Aplicação de diferentes modelos lineares e não lineares na arquitetura ERF, tais como: Suavização Exponencial (PANIGRAHI; BEHERA, 2017), ARIMA Sazonal (XU et al., 2019), Máquinas de Aprendizado Extremo (MA; DAI, 2016) e Rede Neural profunda de Crenças (XU et al., 2019).
- Aplicação da abordagem proposta em outras séries temporais, tais como: velocidade do vento (de MATTOS NETO et al., 2020), energia solar (ALKHAYAT; MEHMOOD, 2021) e unidade de manutenção de estoque (*Stock Keeping Unit (SKU)*) (MAKRIDAKIS; SPILIOTIS; ASSIMAKOPOULOS, 2021).

- Exploração em contextos de previsão que possuam variáveis exógenas (de MATTOS NETO et al., 2020). Variáveis exógenas podem adicionar ao problema de previsão novas dimensões. Nesse contexto, a abordagem proposta pode reduzir o custo do treinamento e criar especialistas locais, uma vez que cada modelo do ensemble utiliza uma visão aleatória do conjunto de treino.
- Exploração de outras estratégias da área de *ensemble*, tais como: seleção dinâmica (SILVA et al., 2020; de MATTOS NETO et al., 2021a), geração por correlação negativa (LIU; YAO, 1999) e agregação por densidade de *kernel* (KOURENTZES; BARROW; CRONE, 2014).

REFERÊNCIAS

- ABADI, M.; AGARWAL, A.; BARHAM, P.; BREVDO, E.; CHEN, Z.; CITRO, C.; CORRADO, G. S.; DAVIS, A.; DEAN, J.; DEVIN, M.; GHEMAWAT, S.; GOODFELLOW, I.; HARP, A.; IRVING, G.; ISARD, M.; JIA, Y.; JOZEFOWICZ, R.; KAISER, L.; KUDLUR, M.; LEVENBERG, J.; MANÉ, D.; MONGA, R.; MOORE, S.; MURRAY, D.; OLAH, C.; SCHUSTER, M.; SHLENS, J.; STEINER, B.; SUTSKEVER, I.; TALWAR, K.; TUCKER, P.; VANHOUCHE, V.; VASUDEVAN, V.; VIÉGAS, F.; VINYALS, O.; WARDEN, P.; WATTENBERG, M.; WICKE, M.; YU, Y.; ZHENG, X. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. 2015. Software available from tensorflow.org.
- ADHIKARI, R. A neural network based linear ensemble framework for time series forecasting. *Neurocomputing*, v. 157, p. 231 – 242, 2015. ISSN 0925-2312.
- ADHIKARI, R.; AGRAWAL, R. K. An introductory study on time series modeling and forecasting. *CoRR*, abs/1302.6613, 2013.
- AGUILAR, R. J.; TURIAS, I.; COME, J. M. Hybrid approaches based on sarima and artificial neural networks for inspection time series forecasting. *Transportation Research Part E: Logistics and Transportation Review*, Elsevier, v. 67, p. 1–13, 2014.
- ALBERTBUP. *A Python implementation of Deep Belief Networks built upon NumPy and TensorFlow with scikit-learn compatibility*. 2017. Disponível em: <<https://github.com/albertbup/deep-belief-network>>.
- ALKHAYAT, G.; MEHMOOD, R. A review and taxonomy of wind and solar energy forecasting methods based on deep learning. *Energy and AI*, v. 4, p. 100060, 2021. ISSN 2666-5468.
- BELMAHDI, B.; LOUZAZNI, M.; BOUARDI, A. E. A hybrid arima–ann method to forecast daily global solar radiation in three different cities in morocco. *The European Physical Journal Plus*, Springer, v. 135, n. 11, p. 1–23, 2020.
- BOX, G. E.; JENKINS, G. M.; REINSEL, G. C.; LJUNG, G. M. *Time series analysis: forecasting and control*. [S.l.]: John Wiley & Sons, 2015.
- BOZDOGAN, H. Akaike's information criterion and recent developments in information complexity. *Journal of mathematical psychology*, Elsevier, v. 44, n. 1, p. 62–91, 2000.
- BREIMAN, L. Bagging predictors. *Machine learning*, Springer, v. 24, n. 2, p. 123–140, 1996.
- BRITTO, A. S.; SABOURIN, R.; OLIVEIRA, L. E. Dynamic selection of classifiers—a comprehensive review. *Pattern Recognition*, v. 47, n. 11, p. 3665 – 3680, 2014. ISSN 0031-3203.
- BROCKWELL, P. J.; BROCKWELL, P. J.; DAVIS, R. A.; DAVIS, R. A. *Introduction to time series and forecasting*. [S.l.]: Springer, 2016.
- BROWN, G.; WYATT, J. L.; TINO, P.; BENGIO, Y. Managing diversity in regression ensembles. *Journal of machine learning research*, v. 6, n. 9, 2005.
- CAO, L.-J.; TAY, F. E. H. Support vector machine with adaptive parameters in financial time series forecasting. *IEEE Transactions on neural networks*, IEEE, v. 14, n. 6, p. 1506–1518, 2003.

- CHANG, C.-C.; LIN, C.-J. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, ACM, v. 2, n. 3, p. 27, 2011.
- CHATFIELD, C. *The analysis of time series: an introduction*. [S.l.]: Chapman and Hall/CRC, 2003.
- CHE, J. Optimal sub-models selection algorithm for combination forecasting model. *Neurocomputing*, Elsevier, v. 151, p. 364–375, 2015.
- COCHRANE, J. H. Time series for macroeconomics and finance. *Unpublished book manuscript*, 1997.
- COWPERTWAIT, P. S.; METCALFE, A. V. *Introductory time series with R*. [S.l.]: Springer Science & Business Media, 2009.
- CRUZ, R. M.; SABOURIN, R.; CAVALCANTI, G. D. Dynamic classifier selection: Recent advances and perspectives. *Information Fusion*, v. 41, p. 195 – 216, 2018. ISSN 1566-2535.
- DAVE, E.; LEONARDO, A.; JEANICE, M.; HANAFIAH, N. Forecasting indonesia exports using a hybrid model arima-lstm. *Procedia Computer Science*, v. 179, p. 480–487, 2021. ISSN 1877-0509. 5th International Conference on Computer Science and Computational Intelligence 2020.
- de HOLANDA, R. B. B.; de OLIVEIRA, J. F. L. Swarm based ensembles for time series residual forecasting. In: *2020 IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI)*. [S.l.: s.n.], 2020. p. 595–602.
- de MATTOS NETO, P. S.; de OLIVEIRA, J. F.; SANTOS JÚNIOR, D. S. de O.; SIQUEIRA, H. V.; MARINHO, M. H.; MADEIRO, F. An adaptive hybrid system using deep learning for wind speed forecasting. *Information Sciences*, v. 581, p. 495–514, 2021. ISSN 0020-0255.
- de MATTOS NETO, P. S.; FIRMINO, P. R. A.; SIQUEIRA, H.; TADANO, Y. D. S.; ALVES, T. A.; de OLIVEIRA, J. F. L.; MARINHO, M. H. D. N.; MADEIRO, F. Neural-based ensembles for particulate matter forecasting. *IEEE Access*, IEEE, v. 9, p. 14470–14490, 2021.
- de MATTOS NETO, P. S. G.; de OLIVEIRA, J. F. L.; SANTOS JÚNIOR, D. S. de O.; SIQUEIRA, H. V.; da NÓBREGA MARINHO, M. H.; MADEIRO, F. A hybrid nonlinear combination system for monthly wind speed forecasting. *IEEE Access*, v. 8, p. 191365–191377, 2020.
- de OLIVEIRA, J. F.; LUDERMIR, T. B. A hybrid evolutionary decomposition system for time series forecasting. *Neurocomputing*, v. 180, p. 27–34, 2016. ISSN 0925-2312. Progress in Intelligent Systems Design.
- de OLIVEIRA, J. F. L.; SILVA, E. G.; DE MATTOS NETO, P. S. G. A hybrid system based on dynamic selection for time series forecasting. *IEEE Transactions on Neural Networks and Learning Systems*, p. 1–13, 2021.
- DEMŠAR, J. Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine Learning Research*, JMLR. org, v. 7, p. 1–30, 2006.
- DEMŠAR, J.; CURK, T.; ERJAVEC, A.; GORUP, Č.; HOČEVAR, T.; MILUTINOVIČ, M.; MOŽINA, M.; POLAJNAR, M.; TOPLAK, M.; STARIČ, A. et al. Orange: Data mining toolbox in python. *Journal of Machine Learning Research*, v. 14, p. 2349–2353, 2013.

- ENGELBRECHT, A. P. *Computational intelligence: an introduction*. [S.l.]: John Wiley & Sons, 2007.
- GOMES, H. M.; READ, J.; BIFET, A. Streaming random patches for evolving data stream classification. In: IEEE. *2019 IEEE International Conference on Data Mining (ICDM)*. [S.l.], 2019. p. 240–249.
- HAYKIN, S. *Kalman filtering and neural networks*. [S.l.]: John Wiley & Sons, 2004. v. 47.
- HILLEBRAND, E.; MEDEIROS, M. C. The benefits of bagging for forecast models of realized volatility. *Econometric Reviews*, Taylor & Francis, v. 29, n. 5-6, p. 571–593, 2010.
- HO, T. K. The random subspace method for constructing decision forests. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 20, n. 8, p. 832–844, 1998.
- HOLLANDER, M.; WOLFE, D. A.; CHICKEN, E. *Nonparametric statistical methods*. [S.l.]: John Wiley & Sons, 2013. v. 751.
- HYNDMAN, R. J.; KHANDAKAR, Y. et al. *Automatic time series for forecasting: the forecast package for R*. [S.l.], 2007.
- JIN, S.; SU, L.; ULLAH, A. Robustify financial time series forecasting with bagging. *Econometric Reviews*, Taylor & Francis, v. 33, n. 5-6, p. 575–605, 2014.
- KEIJZER, M.; BABOVIC, V. Genetic programming, ensemble methods and the bias/variance tradeoff—introductory investigations. In: SPRINGER. *European Conference on Genetic Programming*. [S.l.], 2000. p. 76–90.
- KHASHEI, M.; BIJARI, M. A novel hybridization of artificial neural networks and arima models for time series forecasting. *Applied Soft Computing*, Elsevier, v. 11, n. 2, p. 2664–2675, 2011.
- KOURENTZES, N.; BARROW, D. K.; CRONE, S. F. Neural network ensemble operators for time series forecasting. *Expert Systems with Applications*, v. 41, n. 9, p. 4235 – 4244, 2014. ISSN 0957-4174.
- LIU, L.-M.; HUDAK, G. B.; BOX, G. E.; MULLER, M. E.; TIAO, G. C. *Forecasting and time series analysis using the SCA statistical system*. [S.l.]: Scientific Computing Associates DeKalb, IL, 1992. v. 1.
- LIU, Y.; YAO, X. Ensemble learning via negative correlation. *Neural Networks*, v. 12, n. 10, p. 1399–1404, 1999. ISSN 0893-6080.
- LOUPPE, G.; GEURTS, P. Ensembles on random patches. In: SPRINGER. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. [S.l.], 2012. p. 346–361.
- LOUPPE, G.; GEURTS, P. Ensembles on random patches. In: FLACH, P. A.; BIE, T. D.; CRISTIANINI, N. (Ed.). *Machine Learning and Knowledge Discovery in Databases*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. p. 346–361.
- MA, Z.; DAI, Q. Selected an stacking elms for time series prediction. *Neural Processing Letters*, Springer, v. 44, n. 3, p. 831–856, 2016.

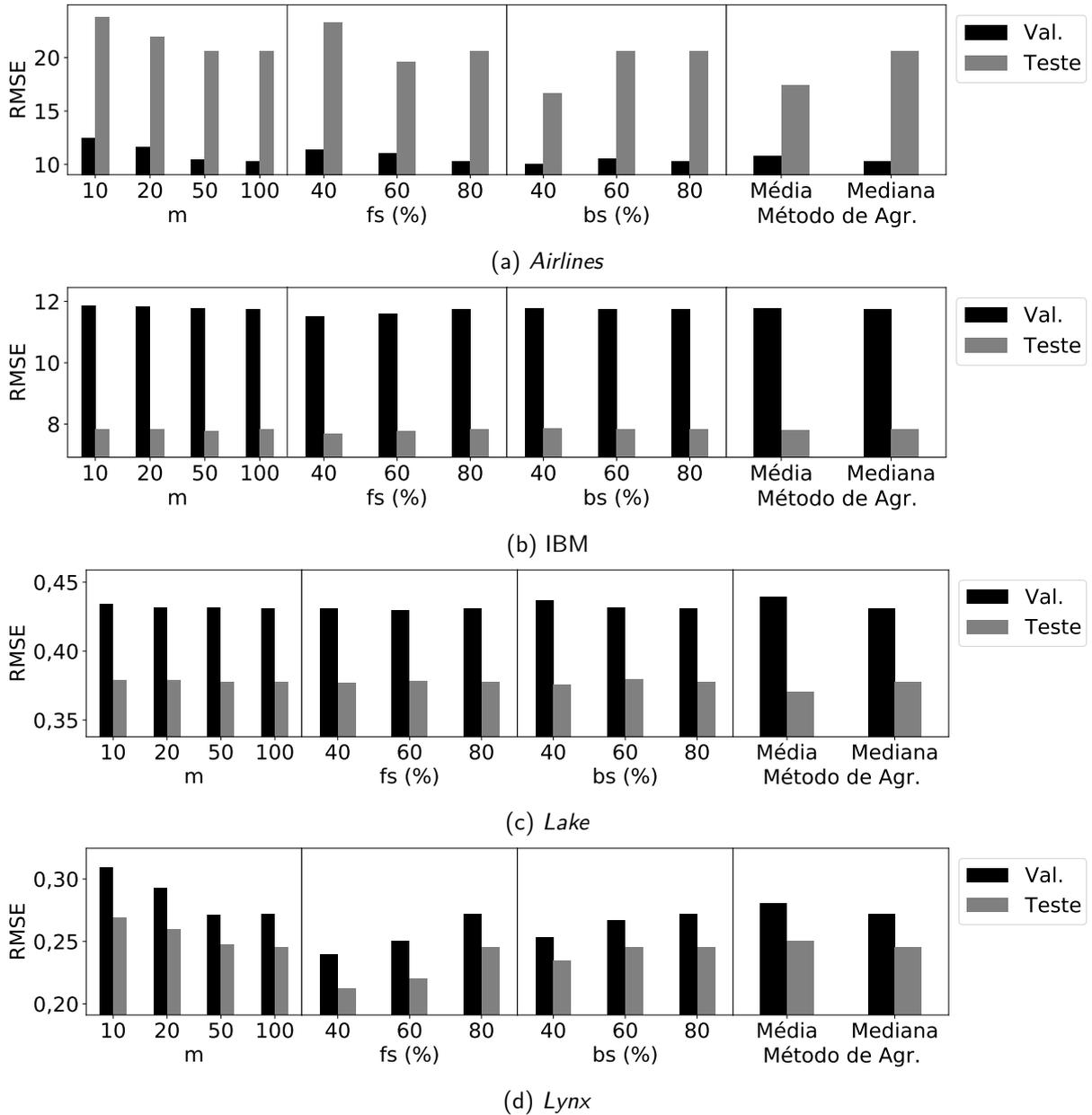
- MAKRIDAKIS, S.; SPILIOTIS, E.; ASSIMAKOPOULOS, V. The m5 competition: Background, organization, and implementation. *International Journal of Forecasting*, 2021. ISSN 0169-2070.
- MENDES-MOREIRA, J.; SOARES, C.; JORGE, A. M.; SOUSA, J. F. D. Ensemble approaches for regression: A survey. *Acm computing surveys (csur)*, ACM New York, NY, USA, v. 45, n. 1, p. 1–40, 2012.
- MITCHELL, T. M. Machine learning. *Mc G raw ill.(forthcoming)*, 1997.
- MORETTIN, P. A.; TOLOI, C. M. de C. *Modelos para previsão de séries temporais*. [S.l.]: Instituto de matematica pura e aplicada, 1981. v. 1.
- PAI, P.-F.; LIN, C.-S. A hybrid arima and support vector machines model in stock price forecasting. *Omega*, Elsevier, v. 33, n. 6, p. 497–505, 2005.
- PANIGRAHI, S.; BEHERA, H. A hybrid ets–ann model for time series forecasting. *Engineering Applications of Artificial Intelligence*, Elsevier, v. 66, p. 49–59, 2017.
- PANOV, P.; DŽEROSKI, S. Combining bagging and random subspaces to create better ensembles. In: BERTHOLD, M. R.; SHAW-TAYLOR, J.; LAVRAČ, N. (Ed.). *Advances in Intelligent Data Analysis VII*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007. p. 118–129.
- PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011.
- PUROHIT, S. K.; PANIGRAHI, S.; SETHY, P. K.; BEHERA, S. K. Time series forecasting of price of agricultural products using hybrid methods. *Applied Artificial Intelligence*, Taylor & Francis, v. 0, n. 0, p. 1–19, 2021.
- QIN, M.; LI, Z.; DU, Z. Red tide time series forecasting by combining arima and deep belief network. *Knowledge-Based Systems*, v. 125, p. 39–52, 2017. ISSN 0950-7051.
- QIU, X.; REN, Y.; SUGANTHAN, P. N.; AMARATUNGA, G. A. Empirical mode decomposition based ensemble deep learning for load demand time series forecasting. *Applied Soft Computing*, v. 54, p. 246 – 255, 2017. ISSN 1568-4946.
- REN, Y.; ZHANG, L.; SUGANTHAN, P. Ensemble classification and regression-recent developments, applications and future directions [review article]. *IEEE Computational Intelligence Magazine*, v. 11, n. 1, p. 41–53, 2016.
- SANTOS JÚNIOR, D. S. d. O.; de OLIVEIRA, J. F.; de MATTOS NETO, P. S. An intelligent hybridization of arima with machine learning models for time series forecasting. *Knowledge-Based Systems*, v. 175, p. 72–86, 2019.
- SERGIO, A. T.; LIMA, T. P. de; LUDERMIR, T. B. Dynamic selection of forecast combiners. *Neurocomputing*, Elsevier, v. 218, p. 37–50, 2016.
- SHEN, X.-J.; NI, C.; WANG, L.; ZHA, Z.-J. Sliker: Sparse loss induced kernel ensemble regression. *Pattern Recognition*, v. 109, p. 107587, 2021. ISSN 0031-3203.

- SHUMWAY, R. H.; STOFFER, D. S.; STOFFER, D. S. *Time series analysis and its applications*. [S.l.]: Springer, 2000. v. 3.
- SILVA, E. G.; CAVALCANTI, G. D. C.; de OLIVEIRA, J. F. L.; de MATTOS NETO, P. S. G. On the evaluation of dynamic selection parameters for time series forecasting. In: *2020 International Joint Conference on Neural Networks (IJCNN)*. [S.l.: s.n.], 2020. p. 1–7.
- SILVA, E. G.; SANTOS JÚNIOR, D. S. d. O.; CAVALCANTI, G. D.; de MATTOS NETO, P. S. Improving the accuracy of intelligent forecasting models using the perturbation theory. In: *2018 International Joint Conference on Neural Networks (IJCNN)*. [S.l.: s.n.], 2018. p. 1–7.
- SMITH, T. G. et al. *pmdarima: ARIMA estimators for Python*. 2017. [Online; accessed <today>]. Disponível em: <<http://www.alkaline-ml.com/pmdarima>>.
- SRETENOVIĆ, A. A.; JOVANOVIĆ, R. Ž.; NOVAKOVIĆ, V. M.; NORD, N. M.; ŽIVKOVIĆ, B. D. Hybrid artificial intelligence model for prediction of heating energy use. *Thermal Science*, n. 00, p. 152–152, 2021.
- TASKAYA-TEMIZEL, T.; CASEY, M. C. A comparative study of autoregressive neural network hybrids. *Neural Networks*, v. 18, n. 5, p. 781 – 789, 2005. ISSN 0893-6080.
- VAPNIK, V.; GOLOWICH, S. E.; SMOLA, A. et al. Support vector method for function approximation, regression estimation, and signal processing. *Advances in neural information processing systems*, MORGAN KAUFMANN PUBLISHERS, p. 281–287, 1997.
- VIRTANEN, P.; GOMMERS, R.; OLIPHANT, T. E.; HABERLAND, M.; REDDY, T.; COURNAPEAU, D.; BUROVSKI, E.; PETERSON, P.; WECKESSER, W.; BRIGHT, J.; van der Walt, S. J.; BRETT, M.; WILSON, J.; MILLMAN, K. J.; MAYOROV, N.; NELSON, A. R. J.; JONES, E.; KERN, R.; LARSON, E.; CAREY, C. J.; POLAT, Í.; FENG, Y.; MOORE, E. W.; VanderPlas, J.; LAXALDE, D.; PERKTOLD, J.; CIMRMAN, R.; HENRIKSEN, I.; QUINTERO, E. A.; HARRIS, C. R.; ARCHIBALD, A. M.; RIBEIRO, A. H.; PEDREGOSA, F.; van Mulbregt, P.; SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, v. 17, p. 261–272, 2020.
- WANG, L.; WANG, Z.; QU, H.; LIU, S. Optimal forecast combination based on neural networks for time series forecasting. *Applied Soft Computing*, v. 66, p. 1–17, 2018. ISSN 1568-4946.
- WANG, L.; ZOU, H.; SU, J.; LI, L.; CHAUDHRY, S. An arima-ann hybrid model for time series forecasting. *Systems Research and Behavioral Science*, v. 30, n. 3, p. 244–259, 2013.
- WINKLER, R. L.; MAKRIDAKIS, S. The combination of forecasts. *Journal of the Royal Statistical Society: Series A (General)*, Wiley Online Library, v. 146, n. 2, p. 150–157, 1983.
- WONGSATHAN, R.; CHANKHAM, S. Improvement on pm-10 forecast by using hybrid arimax and neural networks model for the summer season in chiang mai. *Procedia Computer Science*, Elsevier, v. 86, p. 277–280, 2016.
- XU, D.; ZHANG, Q.; DING, Y.; ZHANG, D. *Application of a Hybrid ARIMA-LSTM Model Based on The SPEI For Drought Forecasting*. [S.l.]: Research Square, 2021. 1–17 p.

-
- XU, R.; XIONG, Q.; YI, H.; WU, C.; YE, J. Research on water quality prediction based on sarima-lstm: A case study of beilun estuary. In: *2019 IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*. [S.l.: s.n.], 2019. p. 2183–2188.
- XU, S.; CHAN, H. K.; ZHANG, T. Forecasting the demand of the aviation industry using hybrid time series sarima-svr approach. *Transportation Research Part E: Logistics and Transportation Review*, Elsevier, v. 122, p. 169–180, 2019.
- XU, W.; PENG, H.; ZENG, X.; ZHOU, F.; TIAN, X.; PENG, X. A hybrid modelling method for time series forecasting based on a linear regression model and deep learning. *Applied Intelligence*, Springer, v. 49, n. 8, p. 3002–3015, 2019.
- YAN, Q.; MA, C. Application of integrated arima and rbf network for groundwater level forecasting. *Environmental Earth Sciences*, Springer, v. 75, n. 5, p. 396, 2016.
- ZHANG, G.; PATUWO, B. E.; HU, M. Y. Forecasting with artificial neural networks: The state of the art. *International journal of forecasting*, Elsevier, v. 14, n. 1, p. 35–62, 1998.
- ZHANG, G. P. Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, Elsevier, v. 50, p. 159–175, 2003.
- ZHANG, J.; WEI, Y.; TAN, Z. An adaptive hybrid model for short term wind speed forecasting. *Energy*, v. 190, p. 115615, 2020. ISSN 0360-5442.

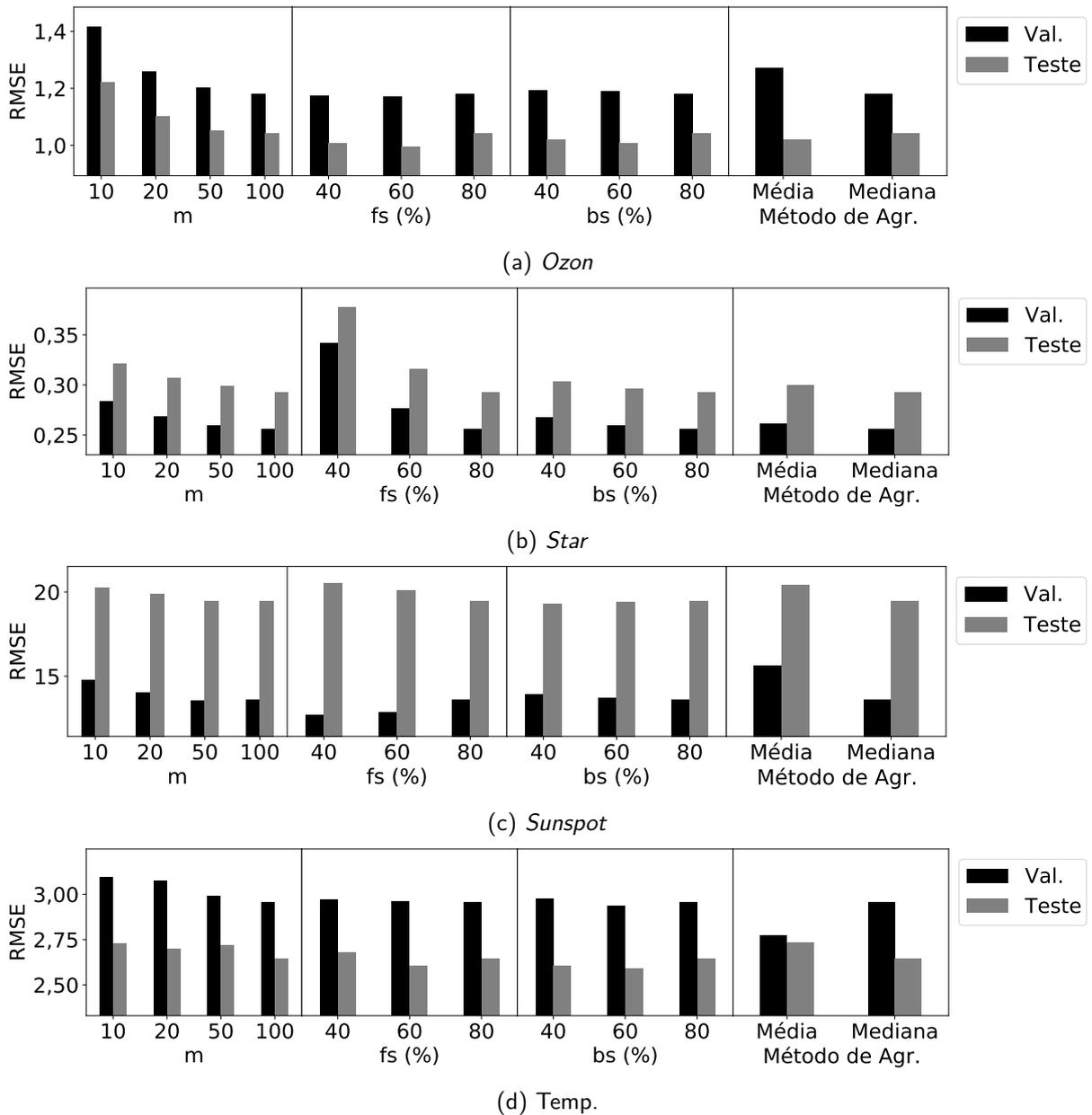
ANEXO A – ANÁLISE DE SENSIBILIDADE

Figura 30 – Análise de sensibilidade dos parâmetros do método proposto ERF_{MLP} para as séries temporais (a) *Airlines*, (b) *IBM*, (c) *Lake* e (d) *Lynx*.



Fonte: Elaborada pelo autor (2022)

Figura 31 – Análise de sensibilidade dos parâmetros do método proposto ERF_{MLP} para as séries temporais (a) Ozon, (b) Star, (c) Sunspot e (d) Temp.



Fonte: Elaborada pelo autor (2022)

ANEXO B – RESULTADOS DE MÉTRICAS ADICIONAIS

Tabela 9 – Resultados do *Mean Absolute Percentage Error* - MAPE (média das 30 execuções) para as séries temporais utilizando os métodos literatura e propostos no conjunto de teste. O melhor e segundo melhor estão destacados em negrito e sublinhado, respectivamente. Os valores reais e preditos são normalizados entre 0,1 e 0,9, devido que existem séries que possuem valores iguais a 0

Model	Airlines	River	IBM	Lake	Lynx	Milk	Ozon	Papper	Pollu.	Star	Sunspot	Temp.
ARIMA	26,0	29,7	16,6	7,1	<u>21,7</u>	21,0	28,3	49,8	39,6	3,6	20,7	18,5
MLP	11,0	28,3	17,3	6,5	35,9	11,4	41,3	15,6	255,9	2,0	30,3	18,5
SVR	20,4	24,8	19,2	6,6	27,7	15,4	24,4	26,5	32,0	1,7	17,7	16,8
Zhang (2003)	19,8	27,9	16,9	6,4	33,5	8,9	50,5	17,0	31,1	1,6	27,4	22,2
Pai e Lin (2005)	<u>9,5</u>	23,1	17,3	6,4	23,2	6,4	28,5	15,8	28,5	1,5	23,1	18,6
Xu et al. (2019)	17,8	27,0	16,5	6,8	22,9	10,4	28,7	32,3	38,6	2,4	20,8	18,8
Yan e Ma (2016)	26,6	31,0	<u>16,0</u>	7,6	21,8	12,2	28,8	43,4	41,2	3,1	22,5	19,8
Xu et al. (2021)	25,7	38,8	16,7	7,1	23,9	16,1	28,3	42,8	38,9	2,2	21,2	<u>17,9</u>
Khashei e Bijari (2011)	25,8	29,7	36,4	7,1	<u>21,7</u>	21,3	28,2	49,8	37,0	3,6	20,7	18,4
Wang et al. (2013)	13,6	26,1	16,8	6,5	41,1	10,7	32,2	18,5	87,8	34,6	30,5	20,6
de Holanda e de Oliveira (2020)	10,6	24,0	15,2	7,0	15,9	<u>7,2</u>	<u>27,0</u>	<u>12,4</u>	35,4	1,8	28,4	22,5
ERF _{SVR}	10,3	<u>23,2</u>	16,8	6,6	21,9	7,5	29,6	13,1	31,4	1,5	<u>20,1</u>	18,1
ERF _{MLP}	9,4	26,0	16,6	6,6	22,0	7,6	30,2	12,3	<u>30,0</u>	1,5	<u>20,1</u>	18,2

Fonte: Elaborada pelo autor (2022)

Tabela 10 – Resultados do *Mean Absolute Error* - MAE (média das 30 execuções) para as séries temporais utilizando os métodos literatura e propostos no conjunto de teste. O melhor e segundo melhor estão destacados em negrito e sublinhado, respectivamente

Model	Airlines	River	IBM	Lake	Lynx	Milk	Ozon	Papper	Pollu.	Star	Sunspot	Temp.
ARIMA	35,848	0,320	<u>6,166</u>	0,294	0,156	17,731	0,796	124,221	664,286	0,572	16,138	2,162
MLP	14,903	0,312	7,064	0,273	0,306	7,842	1,201	72,454	5412,845	0,310	21,687	2,274
SVR	36,053	0,292	6,752	0,283	0,227	9,500	0,740	83,570	764,764	0,275	15,650	2,118
Zhang (2003)	25,286	0,316	6,344	<u>0,272</u>	0,281	6,298	1,243	79,612	512,781	0,250	23,129	2,975
Pai e Lin (2005)	12,443	<u>0,284</u>	6,416	0,270	0,176	4,824	0,811	58,609	537,990	0,232	18,023	2,339
Xu et al. (2019)	24,570	0,302	6,247	0,284	0,167	8,981	0,794	82,396	657,006	0,369	<u>15,468</u>	2,175
Yan e Ma (2016)	35,978	0,335	6,178	0,313	0,155	8,738	0,807	119,972	673,377	0,479	16,859	2,166
Xu et al. (2021)	40,506	0,446	6,294	0,295	0,191	12,903	0,826	97,209	677,649	0,347	15,720	2,248
Khashei e Bijari (2011)	38,268	0,320	14,585	0,294	0,156	18,409	0,794	126,331	711,451	0,562	16,205	<u>2,160</u>
Wang et al. (2013)	20,647	0,327	6,308	0,274	0,294	7,583	0,940	81,151	1112,541	7,909	25,085	2,866
de Holanda e de Oliveira (2020)	<u>12,902</u>	0,285	6,161	0,282	0,166	5,955	0,865	52,652	527,117	0,276	16,306	2,383
ERF _{SVR}	14,319	0,283	6,248	0,277	0,167	5,123	0,810	<u>51,536</u>	546,712	<u>0,231</u>	15,632	2,257
ERF _{MLP}	13,089	0,292	6,183	0,280	0,170	<u>5,015</u>	0,796	48,142	<u>521,735</u>	0,228	15,464	2,188

Fonte: Elaborada pelo autor (2022)