



Universidade Federal de Uberlândia
Faculdade de Matemática

Bacharelado em Estatística

**ANÁLISE DE SÉRIES TEMPORAIS
APLICADA AO PROGRAMA DE
COLETA SELETIVA NO MUNICÍPIO DE
UBERLÂNDIA (MG)**

Cássio Antonio Andrade

Uberlândia-MG

2020

Cássio Antonio Andrade

**ANÁLISE DE SÉRIES TEMPORAIS
APLICADA AO PROGRAMA DE
COLETA SELETIVA NO MUNICÍPIO DE
UBERLÂNDIA (MG)**

Trabalho de conclusão de curso apresentado à Coordenação do Curso de Bacharelado em Estatística como requisito parcial para obtenção do grau de Bacharel em Estatística.

Orientador: Prof^ª Dra. Nádia Giaretta Biase

**Uberlândia-MG
2020**



**Universidade Federal de Uberlândia
Faculdade de Matemática**

Coordenação do Curso de Bacharelado em Estatística

A banca examinadora, conforme abaixo assinado, certifica a adequação deste trabalho de conclusão de curso para obtenção do grau de Bacharel em Estatística.

Uberlândia, ____ de _____ de 20____

BANCA EXAMINADORA

Prof^a Dra. Nádia Giaretta Biase

Prof. Dr. Ednaldo Carvalho Guimarães

Prof^a Dra. Maria Imaculada de Sousa Silva

**Uberlândia-MG
2020**

AGRADECIMENTOS

Agradeço primeiramente a Deus, por minha vida e por iluminar o meu caminho.

Agradeço também aos meus pais Jandir e Vera, à minha irmã Jéssica e à minha noiva Marcella, que mais uma vez estiveram ao meu lado em um segundo curso de graduação, me apoiando, incentivando e sempre acreditando em mim.

Muito obrigado também à professora e orientadora Nádia, que desde o início foi receptiva e me auxiliou desde a disciplina Séries Temporais, durante a submissão de artigos, até neste trabalho de conclusão de curso.

Aos professores Ednaldo Carvalho Guimarães e Maria Imaculada de Sousa Silva, com os quais não tive a oportunidade de cursar disciplinas, mas que aceitaram prontamente o convite de participar da banca examinadora.

Expresso também gratidão aos servidores do Departamento Municipal de Água e Esgoto (DMAE), que me receberam muito bem e contribuíram com informações a respeito do funcionamento do programa de coleta seletiva do município de Uberlândia.

Por fim, agradeço aos meus amigos e familiares, que independente de qualquer circunstância, torcem por minha felicidade e sucesso.

RESUMO

O presente estudo faz uso de técnicas de análise de séries temporais em um conjunto de dados referente à coleta seletiva no município de Uberlândia (MG) durante os anos de 2011 a 2019. A série temporal explorada é a quantidade mensal de material reciclável coletado na cidade pelo serviço de coleta seletiva, disponibilizada no site da Prefeitura Municipal de Uberlândia. As técnicas de Box e Jenkins e de suavização exponencial foram utilizadas para construção de modelos que capturam as informações contidas na série, a fim de avaliar o comportamento da variável analisada ao longo do tempo, bem como efetuar previsões dos valores futuros. Tal previsão pode contribuir na formulação das políticas públicas responsáveis por uma melhor gestão dos resíduos. Ao aplicar o modelo de Box e Jenkins, foi adotado um modelo de previsão do tipo SARIMA, que considera tanto a tendência como a sazonalidade na série, sendo selecionado aquele com menor Critério de Informação de Akaike (AIC). A fim de comparar as duas técnicas de previsão, os critérios adotados foram raiz do erro quadrático médio (RMSE), erro absoluto médio (MAE) e média absoluta percentual do erro (MAPE), sendo que o modelo de Suavização Exponencial obteve menores erros, ou melhor performance, em relação ao modelo SARIMA.

Palavras-chave: tendência, sazonalidade, modelo SARIMA, suavização exponencial.

ABSTRACT

The present study makes use of time series analysis techniques in a data set referring to the selective collection in the city of Uberlândia (MG) during the years 2011 to 2019. The time series explored is a monthly amount of recyclable material collected in the city by the selective collection service, available on the Uberlândia City Hall website. The Box e Jenkins and exponential smoothing techniques were used to build models that capture the information contained in series, in order to evaluate the behavior of the variable analyzed over time, as well as predict future values. Forecasting can contribute to the policymaking process responsible for better waste management. When applying the Box e Jenkins model, a SARIMA forecasting model was adopted, which considers both the trend and seasonality of the series, with the one with the lowest Akaike Information Criterion (AIC) being selected. For the purpose of comparing two forecasting techniques, the root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) were adopted. The Exponential Smoothing model had fewer errors, or better performance, than SARIMA model.

Keywords: trend, seasonality, SARIMA model, exponential smoothing.

SUMÁRIO

Lista de Figuras	I
Lista de Tabelas	III
1 Introdução	1
2 Fundamentação Teórica	5
2.1 Modelos para Séries Temporais	5
2.2 Decomposição Clássica da Série	7
2.3 Modelos de Suavização Exponencial e Modelos ARIMA	10
2.4 Comparação entre os modelos	18
3 Metodologia	21
4 Resultados	23
4.1 Cenário A	23
4.1.1 Descrição Inicial	23
4.1.2 Modelagem SEHW	26
4.1.3 Modelagem SARIMA	26
4.1.4 Comparação entre os modelos SEHW e SARIMA	30
4.2 Cenário B	32
4.2.1 Descrição Inicial	32
4.2.2 Modelagem SEHW	33
4.2.3 Modelagem SARIMA	34
4.2.4 Comparação entre os modelos SEHW e SARIMA	37
4.3 Cenário C	39
4.3.1 Modelagem SEHW	39
4.3.2 Modelagem SARIMA	40
4.3.3 Comparação entre os modelos SEHW e SARIMA	43
4.4 Comparação Geral	44
5 Conclusões	45
Referências Bibliográficas	47

LISTA DE FIGURAS

4.1	Análise do comportamento anual da série ao longo dos meses.	24
4.2	Análise do comportamento mensal da série ao longo dos anos.	24
4.3	Série original da quantidade de material reciclável coletado pela coleta seletiva em Uberlândia de 2011 a 2019 e suas componentes.	25
4.4	Previsões do modelo SEHW (Cenário A).	26
4.5	Gráficos de FAC (a) e FACp (b) da série diferenciada (Cenário A).	27
4.6	Gráficos de FAC (a) e FACp (b) dos resíduos (Cenário A).	28
4.7	Periodograma Acumulado (Cenário A).	28
4.8	Previsões do modelo SARIMA (Cenário A).	29
4.9	Intervalos de confiança das previsões do modelo SARIMA (Cenário A).	29
4.10	Série original e previsões dos modelos (Cenário A).	30
4.11	Previsões do modelo SEHW (Cenário B).	34
4.12	Gráficos de FAC (a) e FACp (b) da série diferenciada (Cenário B).	34
4.13	Gráficos de FAC (a) e FACp (b) dos resíduos (Cenário B).	35
4.14	Periodograma Acumulado (Cenário B).	36
4.15	Previsões do modelo SARIMA (Cenário B).	36
4.16	Intervalos de confiança das previsões do modelo SARIMA (Cenário B).	37
4.17	Série original e previsões dos modelos (Cenário B).	37
4.18	Previsões do modelo SEHW (Cenário C).	40
4.19	Gráficos de FAC (a) e FACp (b) da série diferenciada (Cenário C).	41
4.20	Gráficos de FAC (a) e FACp (b) dos resíduos (Cenário C).	41
4.21	Periodograma Acumulado (Cenário C).	42
4.22	Previsões do modelo SARIMA (Cenário C).	42
4.23	Intervalos de confiança das previsões do modelo SARIMA (Cenário C).	43
4.24	Série original e previsões dos modelos (Cenário C).	43
4.25	Série original e previsões dos modelos nos três diferentes cenários.	44

LISTA DE TABELAS

4.1	Medidas descritivas dos materiais recicláveis coletados pelo programa de coleta seletiva em Uberlândia (Cenário A).	23
4.2	Testes para verificação de tendência na série (Cenário A).	25
4.3	Testes para verificação de sazonalidade na série (Cenário A).	25
4.4	Parâmetros estimados do modelo SARIMA (3,1,0)(0,0,2) _[12] (Cenário A).	27
4.5	Previsões realizadas e erros de previsão para o ano de 2019 (Cenário A).	31
4.6	Erros RMSE, MAE e MAPE dos modelos SEHW e SARIMA (Cenário A).	31
4.7	Previsões realizadas para o ano de 2020 (Cenário A).	32
4.8	Medidas descritivas dos materiais recicláveis coletados entre 2012 e 2019 pelo programa de coleta seletiva em Uberlândia (Cenário B).	32
4.9	Testes para verificação de tendência na série (Cenário B).	33
4.10	Testes para verificação de sazonalidade na série (Cenário B).	33
4.11	Parâmetros estimados do modelo SARIMA (2,1,0)(0,0,2) _[12] (Cenário B).	35
4.12	Previsões realizadas e erros de previsão para o ano de 2019 (Cenário B).	38
4.13	Erros RMSE, MAE e MAPE dos modelos SEHW e SARIMA (Cenário B).	38
4.14	Previsões realizadas para o ano de 2020 (Cenário B).	39
4.15	Parâmetros estimados do modelo SARIMA (2,1,0)(0,0,2) _[12] (cenário C).	40
4.16	Previsões realizadas para o ano de 2020 (Cenário C).	44

1. INTRODUÇÃO

A coleta seletiva consiste na prévia separação dos resíduos conforme sua composição ou constituição, disponibilizando-os também de forma separada para coleta. Há várias formas de coleta seletiva, sendo as mais comuns no Brasil, a coleta porta a porta e a coleta por Pontos de Entrega Voluntária (PEV). Na primeira, um veículo passa em frente às residências e comércios recolhendo os resíduos separados pela população, enquanto na segunda, locais são estrategicamente designados para entrega dos resíduos segregados.

No fim do século XIX, ainda de forma incipiente, a coleta seletiva surgiu nos Estados Unidos e depois no continente europeu, mas até a segunda metade do século XX, a questão da destinação final do lixo era muito precária. No Brasil, o primeiro registro da coleta seletiva foi no bairro São Francisco em Niterói (RJ), implantada em 1985 por meio do esforço conjunto de uma associação de moradores e da Universidade Federal Fluminense. A nível municipal, Curitiba (PR) foi a primeira cidade do país a ter o sistema em 1988 [16].

Um país exemplar na gestão de resíduos sólidos é a Alemanha. Alguns fatores que explicam esse sucesso é a tradição na cobrança de taxas municipais para a coleta de lixo e a utilização de vasilhames padronizados e adequados ao acondicionamento do lixo [16]. O constante acompanhamento e a definição de metas pelos integrantes da União Europeia também concorrem para o excelente desempenho. O bloco europeu tem como objetivo chegar a uma meta máxima de 10% de resíduos sólidos dispostos em aterros sanitários até 2030, sendo que alguns países já alcançaram um índice inferior a 5% [40].

No Brasil a situação é bem diferente. Em 2017 cerca de metade das cidades brasileiras não tinham um plano integrado para o manejo do lixo [23]. E mesmo com um avanço quantitativo e qualitativo no alcance da coleta de lixo nos últimos anos, ainda existem pessoas que não são atendidas pelo serviço de coleta. Tratando-se de coleta seletiva o déficit é ainda maior. Estima-se que, em 2018, no Brasil, de todo o montante potencialmente reciclável, foi coletado de forma seletiva no máximo 13,7% [36].

É grande a dificuldade em se apontar as prováveis barreiras à coleta seletiva. Muitos municípios não souberam definir a causa da interrupção do programa de coleta seletiva em uma pesquisa realizada em 2012. Alguns dos motivos relatados foram a falta de local para triagem ou estocagem dos resíduos, a limitação das campanhas de conscientização e a má aceitação por parte da comunidade [8]. A falta de investimentos é outro empecilho. No Brasil, entre 2017 e 2018, os aportes na coleta e nos demais serviços de limpeza urbana recuaram, representando um declínio de 1,47% nos recursos aplicados na gestão dos RSU [1].

Em relação à viabilidade econômica do sistema de coleta, um estudo, no município de Palmas (TO), verificou que as disposições a pagar (DAP) estão muito abaixo dos custos estimados para a prestação dos serviços, tanto para o sistema convencional quanto para o sistema de coleta seletiva [35]. A DAP refere-se ao maior valor monetário que o indivíduo disponibiliza em prol da qualidade e preservação dos recursos ambientais e pode ser utilizada como medida de benefício [37]. Comparando então os custos estimados e os benefícios (DAP), o sistema convencional obteve uma razão de 0,182, ou seja, os benefícios equivalem apenas a 18,2% dos custos totais do sistema. Já em relação à coleta seletiva, a DAP foi superior ao valor do sistema convencional mas insuficiente para custear uma possível adoção do sistema de coleta seletiva, verificando uma relação benefício-custo de apenas 0,109. Verificou-se que há uma grande resistência por parte da população no pagamento da taxa de recolhimento do lixo, mas conclui-se que, apesar da inviabilidade econômica, cabe ao poder público a decisão política de custear o excedente para implantação da coleta seletiva vislumbrando seus benefícios em outras vertentes além da econômica [35].

A gestão adequada dos resíduos sólidos urbanos (RSU) representa um grande desafio para diversos municípios brasileiros e apesar da difícil implementação da coleta seletiva, esta é uma grande aliada no reaproveitamento adequado [9]. Haveria um grande impacto econômico caso todo o montante que é indevidamente encaminhado aos aterros e lixões fosse reciclado. Estimou-se que poderiam ter sido gerados benefícios de quase R\$ 8 bilhões em 2010, bem superiores aos valores que figuraram entre R\$ 1,4 e R\$ 3,3 bilhões [25]. Outros benefícios são listados em um artigo sobre o impacto da implantação da coleta seletiva no bairro de Santa Rosa, em Campina Grande (PB): melhoria da qualidade ambiental, melhoria da qualidade de vida, geração de emprego e renda aos catadores de materiais recicláveis, inclusão social e elevação da autoestima dos catadores de materiais recicláveis, maior facilidade na coleta de resíduos orgânicos e secos, disposição adequada dos resíduos, redução do material enviado ao lixão ou aterro sanitário, obediência à legislação e promoção da educação ambiental [27].

Uberlândia, município de Minas Gerais, é o segundo mais populoso do estado, com uma população estimada em quase 700 mil habitantes em 2019 [24]. Entre os anos de 2000 e 2010, estima-se que houve um crescimento populacional de 20,5% [38]. Esse crescimento acelerado também reflete no aumento significativo de descarte de RSU, com uma produção per capita subindo de 0,674 kg/habitante/dia em 2000 para 0,987 kg/habitante/dia em 2019, ano em que foram geradas mais de 200 mil toneladas de RSU [13].

A coleta seletiva em Uberlândia é gerenciada pelo Departamento Municipal de Água e Esgoto (DMAE) e executada pela empresa Limpebrás, que detém a concessão do serviço. O programa objetiva incentivar a formação de Associações e Cooperativas de Catadores e conscientizar a população da importância de se reduzir, reutilizar e reciclar os resíduos sólidos, tentando resolver os problemas de acumulação de lixo nos centros urbanos, reintegrando-os ao ciclo industrial, trazendo vantagens ambientais e econômicas [14].

Desde 1998, vários projetos-piloto foram tentados sem sucesso, esbarrando nas mudanças de

gestão de governo e falta de vontade política [38]. Atendendo a Política Nacional de Resíduos Sólidos, a coleta seletiva em Uberlândia, na modalidade porta a porta, foi iniciada em 2011 como um projeto piloto em dois bairros. Em 2020, atingiu-se a marca de 31 bairros do perímetro urbano de Uberlândia atendidos. Apesar disso, em 2019, apenas 46,1% da população uberlandense era abrangida, evidenciando o grande potencial para ampliação do serviço. Há também outras modalidades como a coleta seletiva em duas comunidades da zona rural, o serviço cata-treco que recolhe objetos domiciliares como sofás, televisores e móveis em geral, além de 13 Ecopontos (coleta por PEV) [14]. Santos (2019) [38] afirma que, de 2011 a 2018, não houve evolução significativa no programa de coleta seletiva, com poucos investimentos financeiros e técnicos.

A previsão da quantidade de material coletado traz grandes benefícios ao planejamento dos programas de coleta seletiva municipais, bem como da gestão dos RSU como um todo. Grande parte da matéria-prima que poderia ser reciclada é destinada aos aterros sanitários, aterros controlados e lixões. Assim, ter uma projeção dos resíduos sólidos recicláveis coletados contribui com o planejamento do descarte e tratamento dos resíduos não absorvidos pela coleta seletiva, com o planejamento das atividades de cooperativas e associações dos catadores e com a tomada de decisão dos gestores a fim de melhorar e ampliar o serviço da coleta seletiva.

Várias abordagens já foram utilizadas em projeções relacionadas aos RSU. Adamović *et al.* (2017) [2] propuseram um modelo para previsão anual da geração de RSU de 44 países utilizando redes neurais artificiais obtendo um modelo que considera a influência econômica com uma média absoluta percentual dos erros (MAPE) de apenas 4%. Já Mushtaq *et al.* (2014) [33] fizeram a previsão para a geração anual em uma cidade do Himalaia através de um modelo de regressão linear relacionando variáveis como população e número de domicílios e conseguiram boa precisão na previsão com um erro MAPE igual a 8%. Já para a previsão de diferentes tipos de resíduos produzidos, Günsel (2016) [20] ajustou modelos de Holt-Winters e modelos do tipo autorregressivo integrado de média móvel (ARIMA) utilizando dados mensais de 2013 a 2015, reservando os três últimos meses para comparar com as previsões. Dentre as várias análises, para o lixo doméstico, o modelo de suavização exponencial teve um MAPE de 5,81%, enquanto um modelo ARIMA conseguiu um MAPE de 2,9%. Merelles *et al.* (2019) [29] ajustaram um modelo ARIMA sazonal conforme influência do tempo nos dados analisados, atingindo um excelente ajuste com coeficiente de determinação R^2 de 0,97 e MAPE de 14,62%.

Não há um método universal que sempre apresentará o melhor desempenho em todas as circunstâncias. No entanto, as análises de séries temporais têm inúmeros recursos que podem ser levados em consideração com grande potencial para seleção de um método com melhor desempenho [26]. Assim, optou-se, nesse estudo, por utilizar duas técnicas de análise de séries temporais.

A suavização exponencial de Holt-Winters tem como vantagens o fácil entendimento, a aplicação não dispendiosa e a flexibilidade e como desvantagens a dificuldade em determinar o valor das constantes de suavização, a impossibilidade de estudar propriedades estatísticas como média e variância das previsões, e conseqüentemente, definir um intervalo de confiança [32].

Outro ponto fraco é a sensibilidade a eventos não usuais e outliers, que afetam as previsões [18]. A segunda abordagem é a modelagem ARIMA, com grande flexibilidade e desempenho, mas apresenta como restrições a necessidade de um número mínimo de 50 observações sequenciais, a unidade de tempo deve ser consistente entre todas as variáveis medidas e cada série temporal deve ser uniforme e ininterrupta [42].

Dentro do contexto apresentado, os objetivos deste estudo foram a modelagem da quantidade de resíduos sólidos recicláveis coletados pelo programa de coleta seletiva do município de Uberlândia, em Minas Gerais, por meio de diferentes modelos de séries temporais, a comparação da eficiência dos modelos obtidos e a previsão mensal da quantidade a ser coletada no ano de 2020.

2. FUNDAMENTAÇÃO TEÓRICA

Uma série temporal é um conjunto de observações coletadas em determinado tempo [6]. Os dados têm como característica principal a dependência entre observações vizinhas, diferentemente de um modelo de regressão em que a ordem das observações é irrelevante para análise. O interesse é analisar e modelar essa dependência, que pode ser relacionada a outra variável além do tempo, como espaço, profundidade, entre outras [15].

A série temporal pode ser classificada como contínua ou discreta. O primeiro caso ocorre quando as observações são feitas continuamente no tempo. Já no segundo, as observações são feitas em tempos específicos. Séries contínuas podem ser discretizadas e é importante ressaltar que a classificação da série independe da classificação da variável observada em contínua ou discreta [15].

O estudo por meio de técnicas de séries temporais tem como propósito o entendimento do passado e a previsão do futuro, auxiliando o processo de tomada de decisão por responsáveis e autoridades. [10]. Além da previsão de valores futuros, outros objetivos são a investigação do mecanismo gerador da série, o estudo do comportamento da série e a procura por periodicidades relevantes [32].

2.1 MODELOS PARA SÉRIES TEMPORAIS

De acordo com Morettin e Toloi (2006)[32], para descrição das séries temporais, os modelos utilizados são processos controlados por leis probabilísticas denominados processos estocásticos. Sejam $(\Omega; \mathcal{F}; \mathcal{P})$ um espaço de probabilidade, $(\mathcal{S}, \mathcal{G})$ um espaço mensurável e $\{X_t\}_{t \in T}$ uma família de variáveis aleatórias onde, para cada $t \in T$, $X_t : (\Omega; \mathcal{F}; \mathcal{P}) \rightarrow (\mathcal{S}, \mathcal{G})$ é uma variável aleatória. Diz-se que $\{X_t\}_{t \in T}$ é um processo estocástico. Informalmente, um processo estocástico é um fenômeno que varia em algum grau, de forma imprevisível, a medida que o tempo passa, ou ainda, em termos gerais, é uma família de variáveis aleatórias que descreve a evolução de alguma característica sob análise ao longo do tempo.

Assim, para cada $w \in \Omega$ fixado, obtém-se uma série temporal, ou seja, uma particular realização do processo estocástico. No presente trabalho, a série temporal será denotada por $Z(t)$ e observações feitas em instantes discretos e equiespaçados no tempo por $Z_1, Z_2, \dots, Z_N$.

Algumas suposições simplificadoras são necessárias em certas situações quando se pretende utilizar modelos para descrever séries temporais, de modo a conduzir às seguintes classes de processos estocásticos:

- processos estacionários ou não estacionários, de acordo com a independência ou não relativamente à origem dos tempos;
- processos normais (Gaussianos) ou não normais, conforme as funções de distribuição de probabilidade que caracterizam os processos;
- processos Markovianos ou não-Markovianos, segundo a independência dos valores do processo, em dado instante, de seus valores em instantes precedentes.

Um processo é dito *estacionário* quando as observações ao longo do tempo flutuam em torno de um valor médio constante [19]. Morettin e Toloi (2006)[32] explicam que um processo Z é estacionário se ele se desenvolve no tempo de modo que a escolha da origem dos tempos não é importante. De forma técnica, há duas formas de estacionariedade: a fraca (ou ampla ou de segunda ordem) e a forte (ou estrita) que serão definidas a seguir.

Se todas as distribuições finito-dimensionais permanecem as mesmas sob translações no tempo, um processo estocástico $Z = \{Z(t), t \in T\}$ é *estritamente estacionário*, isto é:

$$F(z_1, \dots, z_n; t_1 + \tau, \dots, t_n + \tau) = F(z_1, \dots, z_n; t_1, \dots, t_n) \quad \forall t_1, \dots, t_n, \tau \text{ de } T. \quad (2.1)$$

Neste caso, todas as distribuições unidimensionais são invariantes sob translações no tempo, resultando em uma média $\mu(t) = \mu$ e em uma variância $V(t) = \sigma^2$ constantes para todo $t \in T$. É comum supor $\mu(t) = 0$ ou considerar o processo $\{Z(t) - \mu\}$.

Já um processo $Z = \{Z(t), t \in T\}$ é *fracamente estacionário* se, e somente se:

$$\begin{aligned} E\{Z(t)\} &= \mu(t) = \mu, \text{ constante, } \forall t \in T; \\ E\{Z^2(t) < \infty\} &\forall t \in T; \\ \gamma(t_1, t_2) &= Cov\{Z(t_1), Z(t_2)\} \text{ é uma função de } |t_1 - t_2|. \end{aligned} \quad (2.2)$$

O interesse do presente trabalho é na classe dos processos definidos nas Equações (2.1) e (2.2), denominados de *processos estacionários*. Em relação aos processos não estacionários, serão tratados apenas os *homogêneos*, nos quais o nível ou inclinação mudam com o decorrer do tempo. Estes podem se tornar estacionários por meio da aplicação de diferenças sucessivas.

Outra classificação dos modelos para séries temporais é em *modelos paramétricos*, em que o número de parâmetros é finito, e em *modelos não paramétricos*, que envolvem infinitos parâmetros. Na primeira classe de modelos, a análise é feita no domínio do tempo, sendo os modelos mais usados os modelos de erro ou regressão, os modelos autorregressivos e de médias móveis (ARMA), os modelos autorregressivos integrados e de médias móveis (ARIMA), modelos de memória longa (ARFIMA), modelos estruturais e modelos não lineares. Já na segunda, a análise é feita no domínio de frequências, sendo comumente utilizados a função de autocovariância ou autocorrelação e sua transformada de Fourier, o espectro.

2.2 DECOMPOSIÇÃO CLÁSSICA DA SÉRIE

Além da estacionariedade, os conceitos de tendência e sazonalidade são essenciais para as aplicações das séries temporais. Uma série temporal pode ser decomposta na soma, ou multiplicação, de três componentes: a tendência (T_t), a sazonalidade (S_t) e uma componente aleatória (a_t) que segue uma distribuição normal e variância zero, também chamada de erro. Essa é a decomposição clássica, descrita na Equação (2.3), quando se trabalha com um modelo aditivo, e na Equação (2.4), no caso de um modelo multiplicativo [15].

$$Z_t = T_t + S_t + a_t. \quad (2.3)$$

$$Z_t = T_t * S_t * a_t. \quad (2.4)$$

Segundo Morettin e Toloi (2006)[32], o modelo aditivo é mais adequado quando a sazonalidade não depende das outras componentes. Já quando a componente sazonal varia com a tendência, o modelo multiplicativo é indicado.

TENDÊNCIA

Para Gutiérrez (2003)[19], a tendência indica o comportamento “de longo prazo” de uma série, ou seja, se ela cresce, decresce ou permanece estável, e quão rápidas se dão essas mudanças.

Há vários métodos para estimação de T_t . Os mais comuns, de acordo com Morettin e Toloi (2006)[32], são:

- ajuste de uma função polinomial;
- suavização dos valores da série ao redor de um ponto, para estimar a tendência naquele ponto (médias móveis, medianas móveis);
- sucessivos ajustes de retas de mínimos quadrados ponderados (“lowess”).

Para verificar a existência da tendência em uma série temporal, a inspeção gráfica pode ser tomada como ponto de partida. Essa análise inicial mais subjetiva pode ser complementada com a utilização de testes de hipóteses estatísticos, que podem ser feitos depois da estimação de T_t ou, preferivelmente, antes. Caso haja outra componente na série (S_t por exemplo), esta deve ser eliminada antes de se verificar a presença da tendência.

Há testes não paramétricos que são bastante utilizados para testar se há tendência no conjunto de dados observados. Entretanto, alguns deles se baseiam em hipóteses que não podem ser verificadas para o caso de uma série temporal. Assim, os mesmos devem ser utilizados com cautela. Os testes a seguir são citados por Morettin e Toloi (2006)[32].

O teste de sequências, também chamado de teste de Wald-Wolfowitz, compara cada observação com a mediana e atribui o símbolo A caso a observação seja maior ou igual à mediana, e B caso seja menor. A estatística T_1 utilizada no teste é o número total de sequências e é

comparada com o valor tabelado w_α , que é o α -quantil da distribuição de T_1 , para rejeição ou não da hipótese nula H_0 : *Não há tendência*. Seja n_1 a quantidade de símbolos A, n_2 a quantidade de B e N a soma de A e B. Caso um desses valores seja maior que 20 é possível utilizar uma aproximação normal, isto é, $T_1 \sim \mathcal{N}(\mu, \sigma^2)$ onde:

$$\mu = \frac{2n_1n_2}{N} + 1; \quad (2.5)$$

$$\sigma^2 = \frac{2n_1n_2(2n_1n_2 - N)}{N^3 - N}. \quad (2.6)$$

Outra alternativa é o teste do sinal, ou de Cox-Stuart, no qual são agrupadas as observações em pares $(Z_1, Z_{1+c}), (Z_2, Z_{2+c}), \dots, (Z_{N-c}, Z_N)$ onde $c = \frac{N}{2}$, se N for par, e $c = \frac{N+1}{2}$, se N for ímpar. A cada par é associado o sinal + se $(Z_i < Z_{i+c})$ e o sinal - se $(Z_i > Z_{i+c})$, eliminando os empates. A estatística T_2 é o número de pares com sinal +, sendo a regra de decisão baseada na distribuição binomial quando o número de observações é menor ou igual a 20 ou na aproximação normal quando é maior que 20.

Por fim, outra opção é o teste baseado no coeficiente de correlação de Spearman. Seja R_t o posto de Z_t dentre as N observações. A estatística do teste é

$$T_3 = \sum_{t=1}^N [R_t - t]^2. \quad (2.7)$$

onde $t = 1, 2, \dots$ são os postos “naturais” dos instantes de tempo. A estatística do teste pode ser comparada com o valor tabelado da distribuição de T_3 ou então em função do coeficiente de correlação de Spearman

$$\rho = 1 - \frac{6T_3}{N^3 - N}. \quad (2.8)$$

A distribuição do coeficiente de Spearman pode ser aproximada de uma distribuição T de Student, com $n - 2$ graus de liberdade. Para verificar a hipótese nula de independência, neste caso, ausência de tendência, é possível aplicar um teste dado por

$$t = \sqrt{\frac{(n-2)\rho^2}{(1-\rho^2)}}. \quad (2.9)$$

Uma vez estimada a tendência por $\hat{T}_t$, é possível obter a série ajustada para tendência, ou livre de tendência,

$$Y_t = Z_t - \hat{T}_t. \quad (2.10)$$

Também há a possibilidade de eliminar a tendência de uma série tomando, em geral, uma ou duas diferenças sucessivas. A primeira diferença pode ser expressa como:

$$\Delta Z_t = Z_t - Z_{t-1}. \quad (2.11)$$

SAZONALIDADE

A sazonalidade corresponde às oscilações que sempre ocorrem em períodos específicos [19]. Exemplos de casos de sazonalidade são o aumento das vendas de flores no dia dos namorados, o aumento de vendas de casacos no inverno, ou seja, variações sazonais que ocorrem regularmente de um ano para outro, de um mês para outro.

No estudo das séries temporais, objetiva-se estimar S_t e obter a série sazonalmente ajustada, conforme Equação (2.12) para modelos aditivos ou Equação (2.13) para modelos multiplicativos [32].

$$Z_t^{SA} = Z_t - \hat{S}_t. \quad (2.12)$$

$$Z_t^{SA} = \frac{Z_t}{\hat{S}_t}. \quad (2.13)$$

A sazonalidade pode ser determinística, quando o padrão sazonal é regular e estável no tempo e o comportamento sazonal pode ser previsto a partir de dados anteriores. A estimação por meio do método de regressão é adequado para este tipo de sazonalidade. Já quando a componente sazonal da série varia com o tempo, trata-se da sazonalidade estocástica. Neste caso, utiliza-se o método de médias móveis para estimação de S_T .

Para verificar a existência da sazonalidade determinística, testes não paramétricos podem ser utilizados, segundo Morettin e Toloi (2006)[32]. O primeiro é o teste de Kruskal-Wallis. O conjunto de dados da série são dispostos em uma tabela em que cada coluna é tratada como uma amostra da população, ou seja, há k amostras (ou colunas) de tamanho n_j (linhas), sendo que as observações são denotadas por

$$Y_{ij}, j = 1, \dots, k, i = 1, \dots, n_j, N = \sum_{j=1}^k n_j. \quad (2.14)$$

As observações Y_{ij} são substituídas por seus postos R_{ij} , obtidos ordenando todas as N observações. Calcula-se então a soma dos postos associada à j -ésima amostra, $R_{.j}$, dado por

$$R_{.j} = \sum_{i=1}^{n_j} R_{ij}. \quad (2.15)$$

A estatística do teste, T_4 , é então calculada

$$T_4 = \frac{12}{N(N+1)} \sum_{j=1}^k \frac{R_{.j}^2}{n_j} - 3(N+1). \quad (2.16)$$

Rejeita-se a hipótese nula H_0 : *Não existe sazonalidade* se T_4 for maior ou igual ao valor crítico T_{4c} , onde T_{4c} é tal que $P_H(T_4 > T_{4c}) = \alpha$, onde α é o nível de significância do teste. Para n_j suficientemente grande, ou $k \geq 4$, a distribuição de T_4 pode ser aproximada por uma variável χ^2 com $k - 1$ graus de liberdade. Algumas suposições do teste de Kruskal-Wallis como

variáveis independentes dentro de cada amostra e amostras independentes entre si geralmente não são verificadas no caso de séries temporais.

Uma alternativa é a aplicação do teste de Friedman para amostras relacionadas. Normalmente, os meses são considerados “tratamentos” e os anos tratados como “blocos”. Diferentemente do teste de Kruskal-Wallis, a ordenação é feita dentro de cada bloco e não entre todas as observações. A estatística do teste é dada por

$$T_5 = \frac{12}{pk(k+1)} \sum_{j=1}^k R_{.j}^{*2} - 3p(k+1), \quad (2.17)$$

onde p = número de blocos (ou anos), k = número de tratamentos (ou meses) e $R_{.j}^{*2}$ denota a soma dos postos da j -ésima coluna. A distribuição de T_5 pode ser aproximada por uma variável χ^2 com $k - 1$ graus de liberdade.

Outro teste para verificação da existência de peridiocidades na série, mas que está no contexto da análise espectral, é o teste de Fisher. A ideia básica do teste é rejeitar a hipótese nula H_0 : *Não existe peridiocidade* se o periodograma contém um valor significantemente maior que o valor médio [3]. Ou seja, observam-se os períodos são significativos, comparando os valores de pico com a variabilidade total da série [28].

Um enfoque diferente que pode ser adotado é a incorporação tanto da sazonalidade, quanto da tendência, em um modelo ARIMA.

2.3 MODELOS DE SUAVIZAÇÃO EXPONENCIAL E MODELOS ARIMA

Dois modelos de previsão amplamente utilizados para analisar séries temporais são os modelos de suavização exponencial e os modelos autorregressivos integrados de médias móveis (ARIMA). Eles se destacam devido à alta generalidade, sendo capazes de ajustar séries com padrão de comportamento mais complexo, a exemplo de séries não estacionárias, com sazonalidade, tendência e correlação serial ao longo do tempo [7].

SUAVIZAÇÃO EXPONENCIAL

Os métodos de suavização exponencial assumem que os valores extremos representam a aleatoriedade. Com a suavização destes extremos, pretende-se então identificar o padrão básico da série, distinguindo o padrão de algum ruído. As vantagens dessa abordagem são a simples utilização, eficiência computacional e razoável precisão [32].

O método de suavização exponencial geral, ou Suavização Exponencial de Holt-Winters (SEHW), é indicado para séries com tendência e sazonalidade. Ele apresenta três constantes de suavização associadas ao nível (A) da série, tendência (C) e sazonalidade (D) que possuem valores no intervalo $[0,1]$. Para o modelo aditivo, utilizam-se as Equações em (2.18), chamadas de equações de suavização, para estimar as componentes do fator sazonal, do nível e da

tendência, respectivamente.

$$\begin{aligned}\hat{F}_t &= D(Z_t - \bar{Z}_t) + (1 - D)\hat{F}_{t-s}; \\ \bar{Z}_t &= A(Z_t - \hat{F}_{t-s}) + (1 - A)(\bar{Z}_{t-1} + \hat{T}_{t-1}); \\ \hat{T}_t &= C(\bar{Z}_t - \bar{Z}_{t-1}) + (1 - C)\hat{T}_{t-1}.\end{aligned}\tag{2.18}$$

Utilizando as equações de suavização é possível, então, realizar previsões por meio das Equações em (2.19), sendo h o horizonte da previsão desejada.

$$\begin{aligned}\hat{Z}_t(h) &= \bar{Z}_t + h\hat{T}_t + \hat{F}_{t+h-s}, \quad h = 1, 2, \dots, s, \\ \hat{Z}_t(h) &= \bar{Z}_t + h\hat{T}_t + \hat{F}_{t+h-2s}, \quad h = s + 1, \dots, 2s.\end{aligned}\tag{2.19}$$

ARIMA

A metodologia ARIMA, proposta por Box e Jenkins, assume que a série temporal em estudo representa uma realização de um processo estocástico que foi gerado pela passagem sucessiva de um ruído branco (processo estocástico cujas variáveis aleatórias são independentes e identicamente distribuídas) por dois filtros lineares: um estável e outro instável. Da teoria de sistemas lineares, a passagem do ruído branco pelo filtro linear estável produzirá na saída um processo estacionário. Este, quando tratado por um certo tipo de filtro instável, produzirá na saída um processo não estacionário (do tipo homogêneo) que caracteriza a maioria dos processos reais [41].

A construção desses modelos segue um ciclo iterativo. Primeiramente uma classe geral é considerada para análise (especificação). Depois, com base na análise de autocorrelações, autocorrelações parciais e outros critérios, identifica-se um modelo ou mais candidatos (identificação). Escolhidos os modelos, estimam-se seus parâmetros (estimação). E por fim, averigua-se a adequação dos modelos, por meio da análise dos resíduos (verificação). Caso o modelo não seja adequado, retorna-se à fase de identificação [32].

ESPECIFICAÇÃO

Segundo Morettin e Toloi (2006)[32], os modelos lineares estacionários são casos particulares de um modelo de filtro linear. Denota-se a entrada por a_t , sobre a qual é aplicada uma função de transferência $\psi(B)$, resultando na saída Z_t . Pode-se escrever

$$Z_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots = \mu + \psi(B)a_t.\tag{2.20}$$

Se $\tilde{Z} = Z_t - \mu$, a Equação (2.20) pode ser reescrita como

$$\tilde{Z}_t = \psi(B)a_t.\tag{2.21}$$

Outra forma é representar $\tilde{Z}_t$ como uma soma ponderada de valores passados $\tilde{Z}_{t-1}, \tilde{Z}_{t-2}, \dots$

mais um ruído a_t :

$$\tilde{Z}_t = \pi_1 \tilde{Z}_{t-1} + \pi_2 \tilde{Z}_{t-2} + \cdots + a_t = \sum_{j=1}^{\infty} \pi_j \tilde{Z}_{t-j} + a_t. \quad (2.22)$$

Reescrevendo a Equação (2.22):

$$(1 - \sum_{j=1}^{\infty} \pi_j B^j) \tilde{Z}_j = a_t, \quad (2.23)$$

em que $\pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \cdots = \psi^{-1}(B)$.

Um *modelo autorregressivo de ordem p*, denotado por AR(p), é obtido quando $\pi_j = 0, j > p$ e renomeando os pesos π_j por ϕ_j , tem-se:

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \phi_2 \tilde{Z}_{t-2} + \cdots + \phi_p \tilde{Z}_{t-p} + a_t. \quad (2.24)$$

Definindo $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ como o operador autorregressivo estacionário de ordem p , outra forma de representar um processo AR(p) é:

$$\phi(B) \tilde{Z}_t = a_t. \quad (2.25)$$

Já se $\psi_j = 0, j > p$ no processo linear representado em (2.20), obtém-se um *modelo*, denotado por MA(q) e renomeando ψ_j pde médias móveis de ordem q por θ_j na Equação (2.21), tem-se:

$$\tilde{Z}_t = \theta(B) a_t = (1 - \theta_1 B - \cdots - \theta_q B^q), \quad (2.26)$$

sendo $\theta(B)$ o operador de médias móveis de ordem q.

Outra classe de modelos bastante utilizada ocorre quando há a inclusão tanto de termos autorregressivos, quanto de termos de médias móveis. Estes modelos são denotados por ARMA(p,q), cuja representação é:

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \cdots + \phi_p \tilde{Z}_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}, \quad (2.27)$$

ou ainda,

$$\phi(B) \tilde{Z}_t = \theta(B) a_t. \quad (2.28)$$

Os modelos AR(p), MA(q) e ARMA(p,q) são apropriados quando se trabalha com séries estacionárias, que se desenvolvem no tempo ao redor de uma média constante. A modelagem ARIMA é comumente aplicada a séries temporais não estacionárias. Um procedimento para eliminar a tendência de uma série é a diferenciação. Assim, aplicam-se diferenças de ordem (d) na série temporal em análise. O parâmetro (p) é relacionado ao modelo autorregressivo e o parâmetro (q) à média móvel aplicada. Estimando-se então os parâmetros, obtém-se o modelo

ARIMA (p,d,q), que é representado pela Equação (2.29).

$$\phi(B)\Delta^d Z_t = \theta(B)a_t. \quad (2.29)$$

Em que:

$\phi(B) = 1 - \phi_1 B - \phi_1 B^2 - \dots - \phi_p B^p$ é o operador regressivo de ordem p;

$\theta(B) = 1 - \theta_1 B - \theta_1 B^2 - \dots - \theta_q B^q$ é o operador de médias móveis de ordem q;

$B^m Z_t = Z_{t-m}$ é operador translação para passado;

$\Delta^d = Z_t - Z_{t-1} = Z_t - B Z_t = (1 - B) Z_t$ é o operador diferença simples, sendo d o número de diferenças.

Quando a série apresenta sazonalidade determinística, o modelo ARIMA pode ser utilizado. Já quando há sazonalidade estocástica, é necessário ajustar a série original a um modelo ARIMA Sazonal (SARIMA). Neste, há a inclusão de termos sazonais nos modelos ARIMA(p,d,q), cuja notação é SARIMA(p, d, q)(P, D, Q)_[s], em que o parâmetro (P) é relacionado ao operador autorregressivo sazonal, o parâmetro (Q) ao operador de médias móveis sazonal e o parâmetro (D) ao operador diferença sazonal, com sazonalidade de período [s], conforme Equação (2.30):

$$\phi(B)\Phi(B^s)\Delta^d \Delta_s^D Z_t = \theta(B)\Theta(B^s)a_t. \quad (2.30)$$

Em que:

$\Phi(B) = 1 - \Phi_1 B - \Phi_1 B^2 - \dots - \Phi_P B^P$ é o operador regressivo sazonal de ordem P;

$\Theta(B) = 1 - \Theta_1 B - \Theta_1 B^2 - \dots - \Theta_Q B^Q$ é o operador de médias móveis sazonal de ordem Q;

$\Delta_s^D = (1 - B^s)$ é o operador diferença sazonal, sendo D o número de diferenças sazonais.

IDENTIFICAÇÃO

Suponha $\{Z_t, t \in Z\}$ um processo estacionário real discreto e de média zero. A função de autocovariância (facv), segundo Morettin e Toloi (2006)[32], é definida por

$$\gamma_\tau = E\{Z_t Z_{t+\tau}\} \quad (2.31)$$

e atende às seguintes propriedades:

- $\gamma_0 > 0$;
- $\gamma_{-\tau} = \gamma_\tau$;
- $|\gamma_\tau| \leq \gamma_0$;
- γ_τ é não negativa definida.

De posse da facv, é possível, então, definir a função de autocorrelação (fac) como

$$\rho_\tau = \frac{\gamma_\tau}{\gamma_0}, \quad \tau \in Z, \quad (2.32)$$

A fac atende as mesmas propriedades da facv, exceto que $\rho_0 = 1$.

Coletadas as observações da série temporal $Z_1, \dots, Z_N$, a fac ρ_j é estimada por

$$r_j = \frac{c_j}{c_0}, \quad j = 0, 1, \dots, N-1, \quad (2.33)$$

onde c_j é estimativa da facv γ_τ ,

$$c_j = \frac{1}{N} \sum_{t=1}^{N-j} [(Z_t - \bar{Z})(Z_{t+j} - \bar{Z})], \quad j = 0, 1, \dots, N-1, \quad (2.34)$$

sendo $\bar{Z}$ a média amostral.

Além do uso da fac no procedimento de identificação do modelo, há também a função de autocorrelação parcial (facp). Denotando o j -ésimo coeficiente de um modelo autorregressivo AR(k) por ϕ_{kj} , é possível mostrar que:

$$\rho_j = \phi_{k1} \rho_{j-1} + \phi_{k2} \rho_{j-2} + \dots + \phi_{kk} \rho_{j-k}, \quad j = 1, \dots, k, \quad (2.35)$$

Escrevendo a Equação (2.35) na forma matricial, obtém-se as equações Yule-Walker:

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-2} \\ \vdots & & & & \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \dots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \dots \\ \rho_k \end{bmatrix} \quad (2.36)$$

Para $k = 1, 2, 3, \dots$, obtém-se de (2.36)

$$\begin{aligned} \phi_{11} &= \rho_1 \\ \phi_{22} &= \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \\ &\vdots \\ \phi_{kk} &= \frac{|\mathbf{P}_k^*|}{|\mathbf{P}_k|} \end{aligned} \quad (2.37)$$

onde $\mathbf{P}_k$ é a matriz de autocorrelações e $\mathbf{P}_k^*$ é a matriz de autocorrelações com a última coluna substituída pelo vetor de autocorrelações. Assim, a quantidade ϕ_{kk} é chamada de função de autocorrelação parcial.

Para estimar a facp de um processo, pode-se estimar, sucessivamente, modelos autorregressivos de ordem $p = 1, 2, 3, \dots$ por mínimos quadrados e tomar as estimativas do último coeficiente de cada ordem. Outra alternativa é substituir as fac ρ_j por suas estimativas r_j nas

Esquações de Yule-Walker, de modo que a Equação (2.35) será:

$$r_j = \phi_{k1} r_{j-1} + \phi_{k2} r_{j-2} + \cdots + \phi_{kk} r_{j-k}, \quad j = 1, \dots, k, \quad (2.38)$$

Definidas a fac e a facp, algumas observações podem ser sumarizadas de forma a facilitar a etapa de identificação:

- no processo AR(p), a fac decai conforme exponenciais ou senóides amortecidas em infinita extensão e a facp $\phi_{kk} \neq 0$, para $k \leq p$ e $\phi_{kk} = 0$, para $k > p$;
- no processo MA(q), a fac é finita, tendo um corte após o “lag” q , enquanto a facp se comporta similarmente à fac de um processo AR(p): com exponenciais ou senóides amortecidas, em infinita extensão;
- no processo ARMA(p,q), a fac, em infinita extensão, decai após o “lag” $q - p$ segundo exponenciais ou senóides amortecidas e a facp é similar à facp de um processo MA puro.

Analisando os padrões dos gráficos da fac, destaca-se [30]:

- Um grande pico no “lag” 1 que diminui depois de alguns “lags” indica um termo autorregressivo nos dados, de modo que a função de autocorrelação parcial determina a ordem do termo autorregressivo;
- Um grande pico no “lag” 1 seguido por uma onda decrescente que alterna entre correlações positivas e negativas sugere um termo autorregressivo de ordem superior nos dados. Use a função de autocorrelação parcial para determinar a ordem do termo autorregressivo;
- Correlações significativas no primeiro ou segundo “lag”, seguidas por correlações que não são significativas aponta para um termo de média móvel nos dados, sendo o número de correlações significativas a possível ordem do termo da média móvel.

Já para os gráficos da facp, verifica-se [31]:

- Um grande pico no “lag” 1 que diminui depois de alguns “lags” evidencia um termo de média móvel nos dados, cuja ordem pode ser determinada pela função de autocorrelação;
- Um grande pico no “lag” 1 seguido por uma onda minimizada que alterna entre correlações positivas e negativas recomenda um termo de média móvel da ordem superior nos dados, sendo que a função de autocorrelação auxilia na determinação da ordem do termo de média móvel;
- Correlações significativas no primeiro ou segundo lag, seguidas por correlações que não são significativas, sugerem um termo autorregressivo nos dados, cuja ordem é igual ao número de correlações significativas.

ESTIMAÇÃO

Com o modelo identificado, a próxima etapa é a estimação de seus parâmetros. Vários métodos são apresentados na literatura, como o método dos momentos, a estimação não linear, mas optou-se por apresentar o método da máxima verossimilhança devido sua ampla aplicação, inclusive no pacote *forecast*[21] que será utilizado mais adiante no software R[34]. Seja um modelo ARIMA(p,d,q). Quando $d > 0$, supõe-se $\mu_w = 0$ e os $p + q + 1$ parâmetros do modelo são colocados no vetor $\xi = (\phi, \theta, \sigma_a^2)$, onde $\phi = (\phi_1, \dots, \phi_p)$ e $\theta = (\theta_1, \dots, \theta_q)$. Se $d = 0$, μ é incluído como mais um parâmetro a ser estimado, tendo $p + q + 2$ parâmetros. Coletadas as N observações $Z_1, \dots, Z_N$, os estimadores de máxima verossimilhança (EMV) de ξ serão os valores que maximizam a função de verossimilhança $L(\xi|Z_1, \dots, Z_N)$.

Ao efetuar d diferenças para atingir estacionariedade, o número de observações $W = (W_1, \dots, W_n)$ se reduz a $n = N - d$, onde $W_t = \Delta^d Z_t$. O modelo ARMA(p,q) resultante é estacionário e invertível, de modo que

$$a_t = \tilde{W}_t - \phi_1 \tilde{W}_{t-1} - \dots - \phi_p \tilde{W}_{t-p} + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q}, \quad (2.39)$$

onde $\tilde{W}_t = W_t - \mu_w$.

Seja $\eta = (\phi, \theta)$. Para inicializar $[W_{-j}]$ e $[a_{-j}]$, $j = 0, 1, 2, \dots$, valores “gerados” antes do início da série, é utilizado um procedimento chamado “backforecasting”. Maximizar $L(\xi|Z_1, \dots, Z_N)$ é equivalente a maximizar $l = \log L$. Box *et al.* [5], demonstram que o logaritmo da função de verossimilhança a ser maximizado pode ser aproximado por

$$l(\xi) \simeq -n \log \sigma_a - \frac{S(\eta)}{2\sigma_a^2}, \quad (2.40)$$

onde $S(\eta)$ é a soma de quadrados não-condicional

$$S(\eta) = S(\phi, \theta) = \sum_{t=-\infty}^n [a_t(\eta, W)]^2, \quad (2.41)$$

sendo

$$[a_t(\eta, W)] = E(a_t|\eta, W). \quad (2.42)$$

VERIFICAÇÃO

Após obtenção do modelo, os resíduos devem ser analisados para checar a adequabilidade, isto é, se o modelo capturou adequadamente as informações nos dados. Resíduos não correlacionados e com média zero são características de um bom método de previsão. Outras duas propriedades desejáveis, mas não indispensáveis, são resíduos com variância constante e normalmente distribuídos [22].

O resíduo é obtido pela diferença entre a observação e o valor ajustado pelo modelo. Como não se desejam evidências de correlação serial nos resíduos, a análise dos gráficos de autocorrelação e autocorrelação parcial dos resíduos e o correlograma são maneiras para verificar a

adequação do modelo. Já o teste de Box-Pierce, em vez de olhar as autocorrelações residuais individualmente, verifica um grupo de autocorrelações. Para pequenas amostras, este teste não tem bom desempenho, sendo uma alternativa bastante utilizada o teste de Ljung-Box [15]. A seguir são descritos alguns testes de diagnóstico para o modelo ajustado à série temporal apresentados por Morettin e Toloi (2006)[32].

- **Teste de autocorrelação residual**

Estimados os parâmetros do modelo, é possível calcular os resíduos estimados:

$$\hat{a}_t = \hat{\theta}^{-1}(B)\hat{\phi}(B)W_t. \quad (2.43)$$

Em um modelo adequado, $\hat{a}_t$ devem estar próximos de a_t , ou seja, devem ser aproximadamente não-correlacionados. Seja $\hat{r}_k$ as autocorrelações dos resíduos $\hat{a}_t$, então $\hat{r}_k \sim \mathcal{N}(0, \frac{1}{n})$. Ruído branco pode ser definido como um processo estocástico onde todas as variáveis aleatórias seguem uma distribuição normal de média zero, variância constante e as covariâncias são nulas [12]. Uma boa indicação para verificar o comportamento de ruído branco em a_t é a comparação de $\hat{r}_k$ com os limites $\pm \frac{2}{\sqrt{n}}$.

- **Teste de Box-Pierce**

Box e Pierce formularam um teste cuja estatística é dada por

$$Q(K) = n(n+2) \sum_{j=1}^K \frac{\hat{r}_j^2}{(n-j)}, \quad (2.44)$$

que terá distribuição aproximada de χ^2 com $K - p - q$ graus de liberdade caso o modelo seja apropriado. Grandes valores de $Q(K)$ rejeitam a hipótese nula de ruído branco para os resíduos.

- **Teste do periodograma acumulado**

O teste do periodograma acumulado é proveniente da análise no domínio de frequência. Denotando as observações de um processo estocástico por $a_t, t = 1, \dots, n$, um estimador do espectro $p_a(f)$ do processo é dado por:

$$I_a(f_i) = \frac{1}{2\pi n} \left[\left(\sum_{t=1}^n a_t \cos\left(\frac{2\pi i}{n}t\right) \right)^2 + \left(\sum_{t=1}^n a_t \sin\left(\frac{2\pi i}{n}t\right) \right)^2 \right], 0 < (f_i) < 1/2. \quad (2.45)$$

Esse estimador do espectro é denominado periodograma. O espectro acumulado, $P_a(f)$, pode ser estimado por $\frac{1}{n} \sum_{i=1}^j I_q(f_i)$. Já a estimativa de $P_a(f)/\sigma_a^2$, denotada por C_{f_j} , é chamada de periodograma acumulado (normalizado). Prova-se que se a_t for ruído branco, então seu espectro é constante e igual a $2\sigma^2$ no intervalo $[0, 1/2]$. Analisando o gráfico de $C_{f_j} \times f_j$, se o mesmo aproxima de uma reta de inclinação 1 para $f_j \in [0, 1/2]$, então os resíduos podem ser considerados como ruído branco.

PREVISÕES

Um dos objetivos da análise de séries temporais é a realização de previsões de valores futuros. Dadas as observações $Z_1, Z_2, \dots, Z_{t-1}, Z_t$, até o instante t , chamado de origem das previsões, há um interesse em prever o valor de Z_{t+h} , $h \geq 1$. A previsão de origem t e horizonte h é denotada por $\hat{Z}_t(h)$.

Supondo que a previsão $\hat{Z}_t(h)$ é uma função das observações até o instante t , e, consequentemente, função das componentes aleatórias $a_t, a_{t-1}, \dots$, tal que

$$\hat{Z}_t(h) = \psi_h^* a_t + \psi_{h+1}^* a_{t-1} + \psi_{h+2}^* a_{t-2} + \dots, \quad (2.46)$$

deseja-se determinar os pesos ψ_j^* que minimizem o erro quadrático médio (EQM) de previsão:

$$E[e_t(h)]^2 = E[Z_{t+h} - \hat{Z}_t(h)]^2 = E\left[\sum_{j=0}^{\infty} \psi_j a_{t+h-j} - \sum_{j=0}^{\infty} \psi_{h+j}^* a_{t-j}\right]^2. \quad (2.47)$$

Demonstra-se que EQM é mínimo se $\psi_{h+j}^* = \psi_{h+j}$, $j = 0, 1, 2, \dots, h$ fixo. Assim, a previsão de EQM mínimo é

$$\hat{Z}_t(h) = \psi_h a_t + \psi_{h+1} a_{t-1} + \dots = \sum_{j=0}^{\infty} \psi_{h+j} a_{t-j}, \quad (2.48)$$

e o erro de previsão

$$e_t(h) = a_{t+h} + \psi_1 a_{t+h-1} + \dots + \psi_{h-1} a_{t-1} = \sum_{j=1}^{\infty} \psi_j a_{t+h-j}. \quad (2.49)$$

Pode-se concluir que:

- $\hat{Z}_t(h) = E(Z_{t+h} | Z_t, Z_{t-1}, \dots)$, ou seja, a previsão de EQM mínimo é a esperança condicional de Z_{t+h} , dadas as observações passadas da série;
- A esperança condicional de $e_t(h)$, dadas as observações passadas da série, é nula e a variância do erro de previsão é $V(h) = (1 + \psi_1^2 + \psi_2^2 + \dots + \psi_{h-1}^2) \sigma_a^2$;
- Os erros de previsão a um passo são não-correlacionados, o que não acontece para horizontes maiores que um.

Para realização de uma previsão $\hat{Z}_n(k)$ nos modelos ARIMA, utiliza-se a equação do modelo obtido substituindo os valores futuros de erros a_t por zero, valores futuros $(Z_{n+1}, Z_{n+2}, \dots)$ pelas esperanças condicionais e os valores passados de Z pelos valores observados [15].

2.4 COMPARAÇÃO ENTRE OS MODELOS

Além de checar a adequabilidade do modelo, a análise residual permite a comparação entre modelos por meio da performance preditiva, com a construção de medidas baseadas nos erros

de previsão. Bertolo (2013)[4] apresenta estatísticas bastante utilizadas como o erro quadrático médio (EQM), a raiz do erro quadrático médio (RMSE), o erro absoluto médio (MAE), a média absoluta percentual dos erros (MAPE), apresentados nas Equações (2.50), (2.51), (2.52), (2.53). O EQM destaca os grandes erros, comparados aos erros de menor magnitude e o RMSE é sua raiz quadrada. O MAE consiste na diferença absoluta entre valores reais e preditos. Já o MAPE não depende de escala, podendo ser utilizado para comparar modelos aplicados a diferentes conjuntos de dados, mas não é indicado quando a série temporal apresenta valores nulos.

$$EQM = \frac{1}{n} \sum_{i=1}^n (z_i - \hat{z}_i)^2, \quad (2.50)$$

$$RMSE = \sqrt{EQM} = \sqrt{\frac{1}{n} \sum_{i=1}^n (z_i - \hat{z}_i)^2}, \quad (2.51)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |z_i - \hat{z}_i|, \quad (2.52)$$

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{z_i - \hat{z}_i}{z_i} \right|, \quad (2.53)$$

em que z_i é o valor observado, $\hat{z}_i$ é o valor previsto e n é o tamanho da amostra.

Outra forma de selecionar os modelos são os critérios de informação, que consideram a qualidade do ajuste mas penalizam a inclusão de parâmetros extras. Os critérios mais utilizados são o critério de informação de Akaike (AIC) e o critério de informação bayesiano (BIC), descritos nas Equações (2.54) e (2.55). Quanto menor o valor do critério utilizado, melhor o modelo [15].

$$AIC = -2 \log (L) + 2m, \quad (2.54)$$

$$BIC = -2 \log (L) - m \log(n), \quad (2.55)$$

em que L é a verossimilhança maximizada dos dados, m é o número de parâmetros do modelo e n é tamanho da amostra.

3. METODOLOGIA

A série temporal em estudo é disponibilizada no site da prefeitura de Uberlândia [14]. Os dados referem-se ao quantitativo mensal de material reciclável coletado por meio da coleta seletiva, em quilogramas, não contendo informações adicionais sobre o tipo de material ou porcentagem do material reciclável coletado que é aproveitado. Outras informações adicionais são a população abrangida pelo programa de coleta seletiva e a produção per capita em quilogramas por habitante por dia.

Foram analisados os registros referente ao período de janeiro de 2011 a dezembro de 2019 com o intuito de construir modelos de previsão por meio das técnicas de suavização exponencial e de Box e Jenkins para prever o comportamento da quantidade de material reciclável a ser coletada em 2020, bem como verificar o comportamento da série ao longo desses nove anos.

Com o auxílio do software R[34], foram realizados os cálculos e a construção dos modelos. Optou-se pela utilização de um software livre e com recursos computacionais que apresentam funcionalidades específicas para séries temporais univariadas, especialmente o pacote *forecast* [21], que auxilia principalmente nas etapas de identificação e estimação da modelagem ARIMA.

Inicialmente, uma análise descritiva da série temporal completa foi efetuada para verificação do comportamento geral dos dados. Em seguida, por meio da função *decompose()* do pacote *forecast* [21], as componentes da tendência, sazonalidade e resíduos foram plotados separadamente, bem como a série original. Assim, foi possível ter indícios sobre a existência ou ausência de tendência e sazonalidade nos dados observados.

Para confirmar ou rejeitar esses indícios, foram realizados testes abordados anteriormente. Para a tendência, aplicaram-se o teste das sequências, o teste do sinal e o teste baseado no coeficiente de correlação por postos de Spearman. Já para sazonalidade, foram utilizados o teste de Kruskal-Wallis, o teste de Friedman e o teste de Fisher. Este último foi executado com apoio do software R por meio da função *fisher.g.test()* do pacote *GeneCycle* [3].

O modelo de Suavização Exponencial de Holt-Winters foi construído com auxílio do software de planilhas LibreOffice Calc [17] para realização dos cálculos e determinação dos valores das constantes de suavização A, C e D. Segundo Bertolo (2013)[4], a determinação desses valores é arbitrária e o melhor valor pode ser obtido por meio de um método iterativo que minimiza alguma medida de qualidade da previsão. Assim, optou-se por estipular as constantes que minimizam o EQM.

As doze primeiras observações são utilizadas para iniciar o modelo. Assim, o nível estimado referente ao instante $t = 12$ é a média das 12 primeiras observações. A tendência é nula e a

componente sazonal, nos instantes de 1 a 12, é definida pela diferença entre o valor observado em cada instante e essa média calculada [11]. Iniciado o processo, as estimativas a partir do instante $t = 13$ são calculadas de acordo com as Equações em (2.18).

Depois de calculadas todas as estimativas até o instante referente à última observação disponível para modelagem, é possível então calcular as previsões com base nesse instante para o horizonte desejado conforme Equações em (2.19). De posse das previsões e dos valores reais, calcula-se o erro de previsão e verifica-se a performance preditiva do modelo. Há também a opção de atualizar as previsões à medida que novas observações são coletadas.

Para a modelagem proposta por Box e Jenkins, foi utilizado o procedimento iterativo já descrito. Primeiramente, especificou-se uma classe geral de modelos. Em seguida, vários modelos candidatos foram identificados e partiu-se para estimação dos parâmetros. Depois, é selecionado o modelo que apresenta o menor valor para o critério de informação Akaike (AIC) e por fim, investiga-se a adequação deste. Caso o modelo não seja adequado, seleciona-se outro, com o menor AIC dentre o conjunto de modelos restantes, até a obtenção de um que seja capaz de capturar as informações contidas na série. Para auxiliar na identificação e na estimação dos diferentes modelos, foi utilizada a função *auto.arima()* do pacote *forecast* [21].

Três cenários foram considerados para a construção dos modelos por meio da metodologia ARIMA e SEHW:

- **Cenário A:** considera as 108 observações, separando-as em um conjunto de treinamento com dados de 2011 a 2018 e em um conjunto de teste com os valores de 2019;
- **Cenário B:** descarta os valores referente ao ano de 2011, em que o programa ainda era um projeto-piloto, de forma que o conjunto de treinamento é formado pelas quantidades referente aos anos 2012 a 2018 e o conjunto de teste com as observações de 2019;
- **Cenário C:** mantém o primeiro ano fora da análise, similarmente ao Cenário B, mas desconsidera a divisão dos dados em dois conjuntos, incluindo na modelagem os dados de 2019 em vez de separá-lo em um conjunto de teste.

4. RESULTADOS

4.1 CENÁRIO A

Neste cenário foram consideradas as observações desde 2011, ano em que foi iniciado o programa de coleta seletiva em Uberlândia, até 2019. As observações referentes a este último ano foram separadas a fim de comparar com as previsões realizadas pelo modelo com base nas observações de 2011 a 2018.

4.1.1 DESCRIÇÃO INICIAL

As estatísticas descritivas das 108 observações analisadas no estudo estão apresentadas na Tabela 4.1. Verifica-se uma grande dispersão dos dados, pois há um alto coeficiente de variação e uma grande diferença entre o valor mínimo, que é a primeira observação referente a janeiro de 2011, e o valor máximo, atingido em outubro de 2019. Os baixos valores das observações iniciais podem ser explicados pela fase inicial de implantação do serviço de coleta no ano de 2011, ainda em fase de projeto-piloto.

Tabela 4.1: Medidas descritivas dos materiais recicláveis coletados pelo programa de coleta seletiva em Uberlândia (Cenário A).

Medidas Descritivas	Valor	Medidas Descritivas	Valor
Média (ton)	159.876,04	Mínimo (ton)	13.602,00
Desvio-padrão (ton)	39.056,84	1º Quartil (ton)	151.667,50
Coeficiente de variação (%)	24,43	Mediana (ton)	168.400,00
Assimetria	-1,66	3º Quartil (ton)	180.629,00
Curtose	3,09	Máximo (ton)	222.508,00

Esse desempenho inicial é evidente na Figura 4.1, que facilita a comparação do comportamento dos dados ao longo dos anos. Já na Figura 4.2, o comportamento em cada mês ao longo dos anos é destacado nas linhas pretas e os traços horizontais representam as médias mensais. Assim, verifica-se que geralmente há uma queda do material coletado nos meses de fevereiro e junho, que foram os que apresentaram menores médias mensais. É importante salientar que fevereiro é um mês de curta duração e junho apresenta baixas temperaturas, fatores que podem explicar a baixa quantidade.

Na Figura 4.3 é apresentada a série original e sua decomposição nas componentes aleatória,

sazonal e tendência. Nota-se que a parte aleatória varia em torno do nível zero, enquanto a componente sazonal apresenta picos nas mudanças de ano. Já em relação à tendência, percebe-se um crescimento ao longo dos anos, exceto durante o ano de 2016, em que há uma queda.

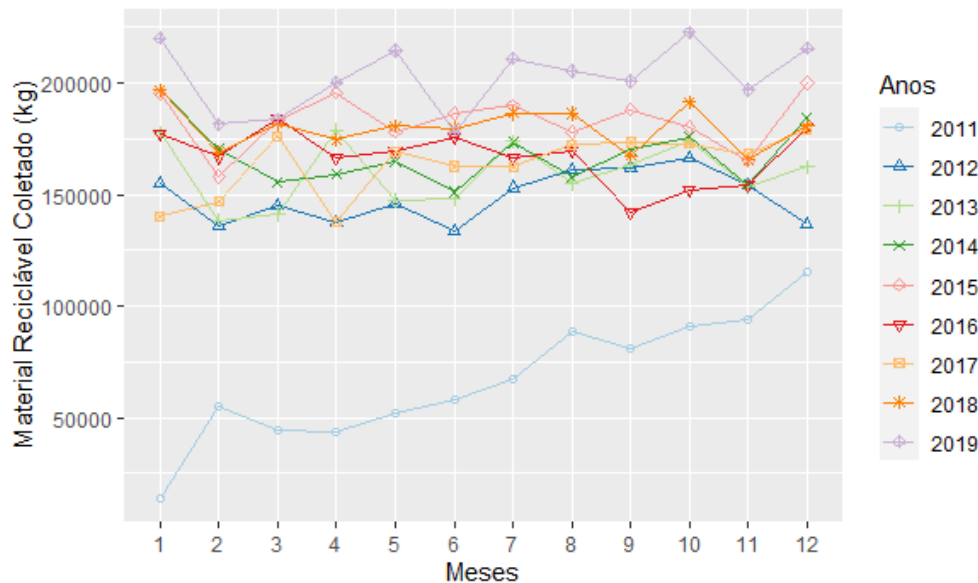


Figura 4.1: Análise do comportamento anual da série ao longo dos meses.

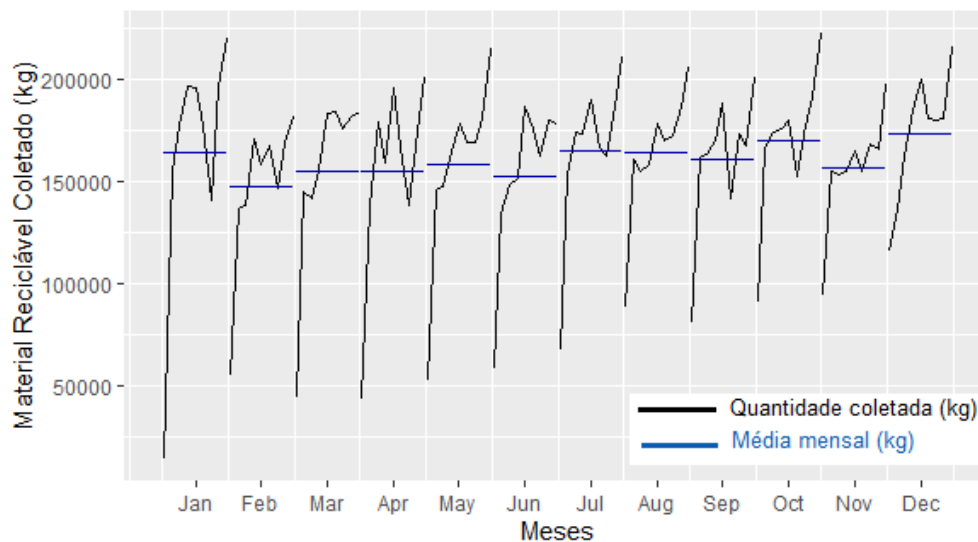


Figura 4.2: Análise do comportamento mensal da série ao longo dos anos.

As hipóteses levantadas sobre tendência e sazonalidade são analisadas por meio da execução de testes. Para a tendência, no teste das sequências e no teste dos sinais, houve evidências para rejeição da hipótese nula H_0 : “Não há tendência”, considerando um nível de significância de 5%, conforme resultados apresentados na Tabela 4.2. Na mesma direção, por meio do teste de correlação por postos de Spearman rejeitou-se a hipótese nula de independência, obtendo um coeficiente de correlação $\rho = 0,73$, valor que indica uma tendência de crescimento quanto mais próximo de 1.

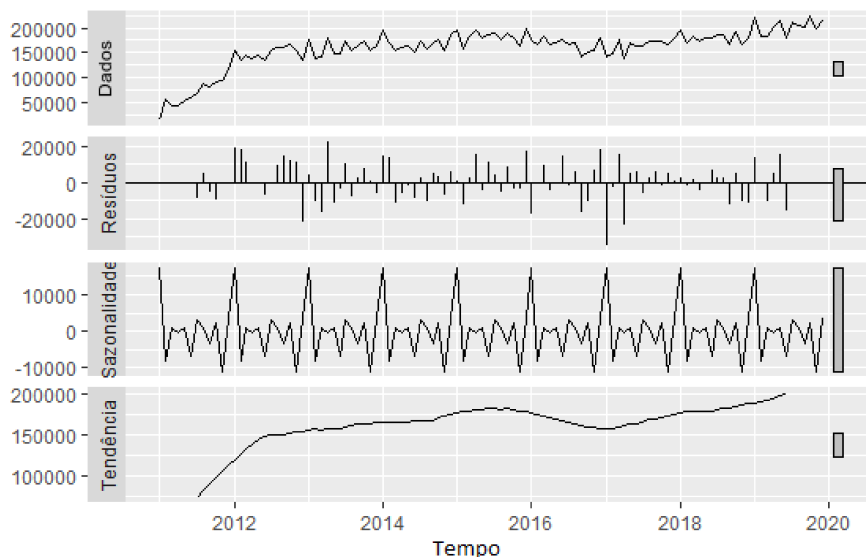


Figura 4.3: Série original da quantidade de material reciclável coletado pela coleta seletiva em Uberlândia de 2011 a 2019 e suas componentes.

Tabela 4.2: Testes para verificação de tendência na série (Cenário A).

Teste	Estatística do teste	Decisão	Conclusão
Wald-Wolfowitz	$T_1 = 40$	$z_{cal} = -3,00 < z_{tab} = -1,96$	Rejeita-se H_0
Cox-Stuart	$T_2 = 46$	$z_{cal} = 5,17 > z_{tab} = 1,96$	Rejeita-se H_0
Correlação de Spearman	$T_3 = 57056$	$t_{cal} = 10,94 > t_{tab} = 1,98$	Rejeita-se H_0

Ao checar a existência da sazonalidade por meio do teste de Kruskal-Wallis, não se rejeitou a hipótese nula H_0 : “Não há sazonalidade”, enquanto que para o teste de Friedman há evidências para rejeição. Morettin e Tolói (2006) [32] alertam que uma das suposições para aplicação do teste de Kruskal-Wallis é a independência das variáveis dentro de cada amostra e a independência das amostras entre si, o que pode explicar a conclusão no sentido oposto do teste de Friedman, que é recomendado para amostras relacionadas. Diante das conclusões divergentes dos testes mencionados, executou-se o teste de Fisher que teve como resposta um pequeno p-valor, de $2,08e^{-3}$, indicando a existência de periodicidade considerando um nível de significância de 5%. Assim, optou-se por ajustar um modelo que considera a existência de sazonalidade. Na Tabela 4.3 são apresentados os resultados encontrados na execução dos testes para verificação da existência de sazonalidade na série.

Tabela 4.3: Testes para verificação de sazonalidade na série (Cenário A).

Teste	Estatística do teste	Decisão	Conclusão
Kruskal-Wallis	$T_4 = 7,75$	$T_4 < \chi^2_{tab} = 19,67$	Não se rejeita H_0
Friedman	$T_5 = 27,22$	$T_5 > \chi^2_{tab} = 19,67$	Rejeita-se H_0
Fisher	$Valor-p = 2,08e^{-3}$	$Valor-p < \alpha$	Rejeita-se H_0

4.1.2 MODELAGEM SEHW

Após a investigação inicial da série, verificou-se que a mesma não é estacionária. Por apresentar tendência e sazonalidade, o SEHW é o mais adequado dentre os métodos de suavização exponencial. Assim, foram usadas as equações de suavização do modelo SEHW para o modelo aditivo, já que a magnitude do padrão sazonal não muda à medida que a série aumenta ou diminui [39].

De modo a minimizar o EQM, determinou-se o valor das constantes de suavização, sendo $A = 0,0714$, $C = 0,3385$ e $D = 0,3159$. Com esses valores, é possível então calcular as estimativas das três componentes do modelo e realizar as previsões do modelo SEHW, ilustradas na Figura 4.4. As previsões para o ano de 2019 foram bem próximas aos valores observados. Analisando as previsões para 2020, espera-se um crescimento em relação aos valores observados em 2019.

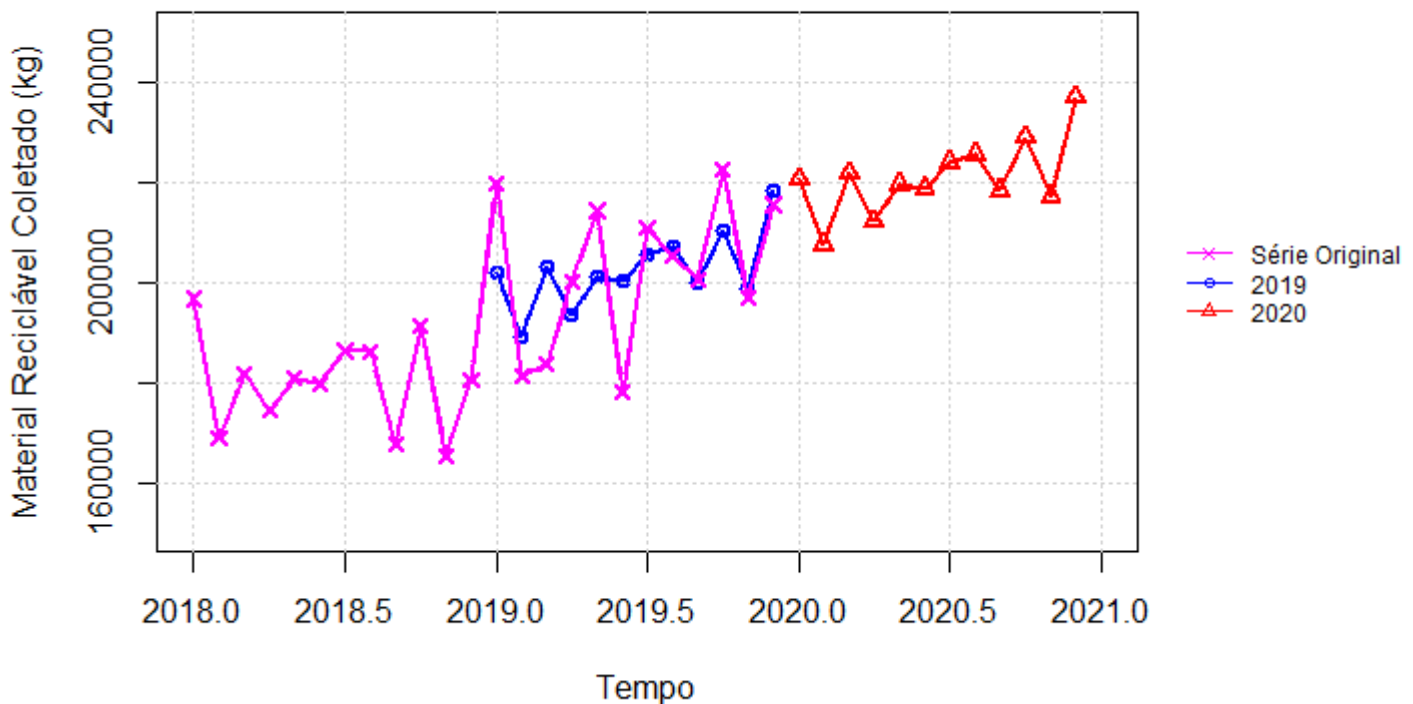


Figura 4.4: Previsões do modelo SEHW (Cenário A).

4.1.3 MODELAGEM SARIMA

Conforme as fases de construção dos modelos de Box e Jenkins, na primeira especificou-se a busca por modelos do tipo SARIMA que contam com a presença dos termos sazonais no modelo, já que foi detectada a sazonalidade. Há também a necessidade da diferenciação para eliminar os efeitos da tendência na série. Na fase de identificação do modelo, examinou-se

os gráficos de FAC e FACp da série diferenciada, apresentados na Figura 4.5. As defasagens significativas, na FAC, seguem um padrão, reforçando a escolha pelo ajuste de um modelo do tipo autorregressivo integrado de médias móveis sazonal e o primeiro “lag” significativo na FACp dá sinais da presença do parâmetro regressivo p no modelo.

Dentre os vários modelos candidatos identificados e estimados, o modelo com menor AIC com valor de 2070,9 foi o SARIMA (2,2,1)(0,0,2)_[12]. Verifica-se que o modelo apresenta duas diferenças, entretanto, analisando a série com apenas uma diferenciação, esta operação já é suficiente para eliminar a tendência. Assim, optou-se por escolher, dentre os modelos com operador de diferença de ordem 1, aquele com menor AIC. Obteve-se então o modelo SARIMA (3,1,0)(0,0,2)_[12] com AIC de 2093,9 e BIC de 2109,22. Os parâmetros estimados são apresentados na Tabela 4.4.

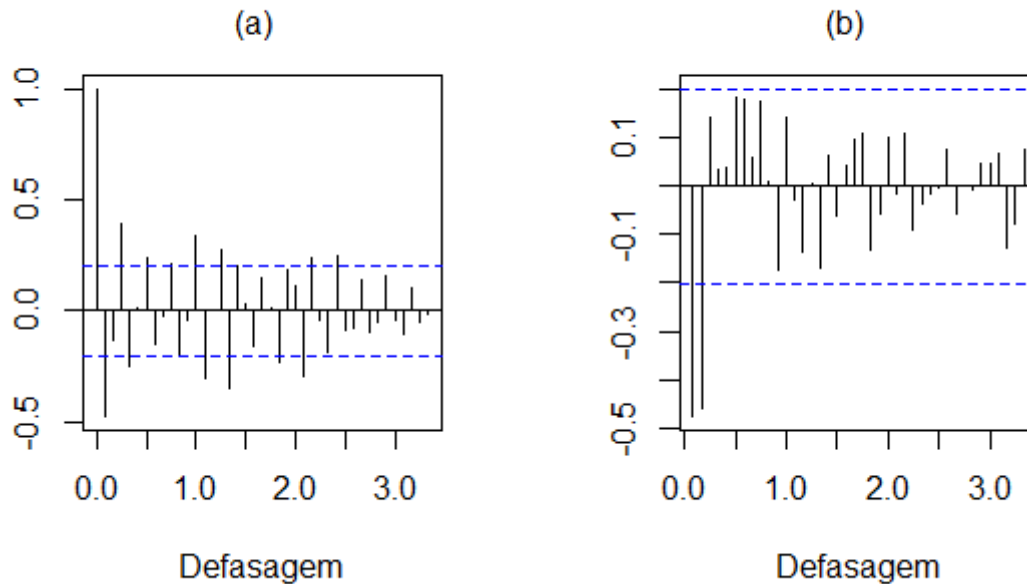


Figura 4.5: Gráficos de FAC (a) e FACp (b) da série diferenciada (Cenário A).

Tabela 4.4: Parâmetros estimados do modelo SARIMA (3,1,0)(0,0,2)_[12] (Cenário A).

SARIMA	Parâmetro	Estimativa	Desvio-Padrão
(3,1,0)(0,0,2) _[12]	ϕ_1	-0,6020	0,1060
	ϕ_2	-0,3351	0,1197
	ϕ_3	0,2254	0,1055
	Θ_1	0,1410	0,5156
	Θ_2	0,0972	0,2196

Para verificar a adequabilidade do modelo, foram avaliadas as funções de autocorrelação (FAC) e autocorrelação parcial (FACp) dos resíduos, ilustradas na Figura 4.6. Não são observadas defasagens significativas, indicando que os resíduos são não correlacionados. O teste de

Box-Pierce permite avaliar a hipótese nula de ruído branco para os resíduos. Com um valor-p de 0,735, conclui-se que os resíduos são um processo de ruído branco, mesma conclusão obtida pelo teste do periodograma acumulado, apresentado na Figura 4.7.

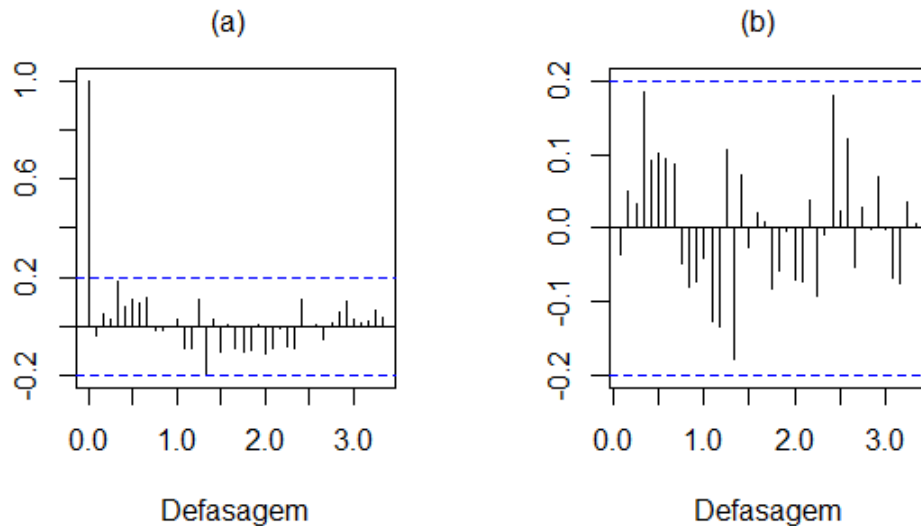


Figura 4.6: Gráficos de FAC (a) e FACp (b) dos resíduos (Cenário A).

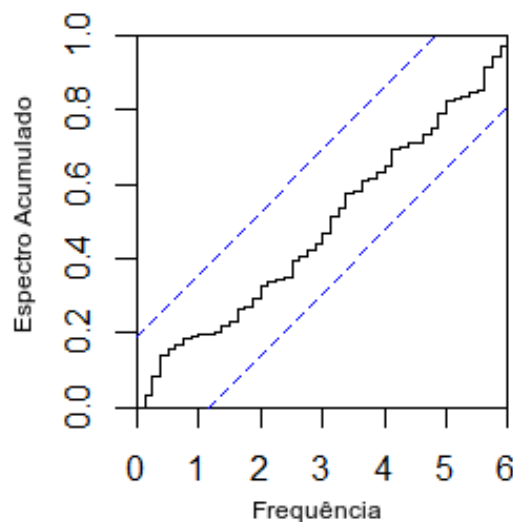


Figura 4.7: Periodograma Acumulado (Cenário A).

Concluída a fase de verificação, é possível realizar as previsões para o ano de 2019, confrontando-as com os valores observados, e para o ano de 2020. Na Figura 4.8, verifica-se que as previsões para o ano de 2019 foram inferiores aos valores reais. E para o ano de 2020, as previsões projetam um leve crescimento em relação às previsões de 2019, mas em relação aos valores reais do ano anterior, em geral, espera-se uma redução na quantidade de material reciclável a ser coletado. A modelagem de Box e Jenkins permite calcular um intervalo de confiança, estabelecendo assim limites de confiança para a previsão realizada. Na Figura 4.9,

podem ser visualizadas a série original e as previsões efetuadas com limites de confiança de 95%. Observa-se que quanto maior o horizonte da previsão, maior a incerteza acerca desta.

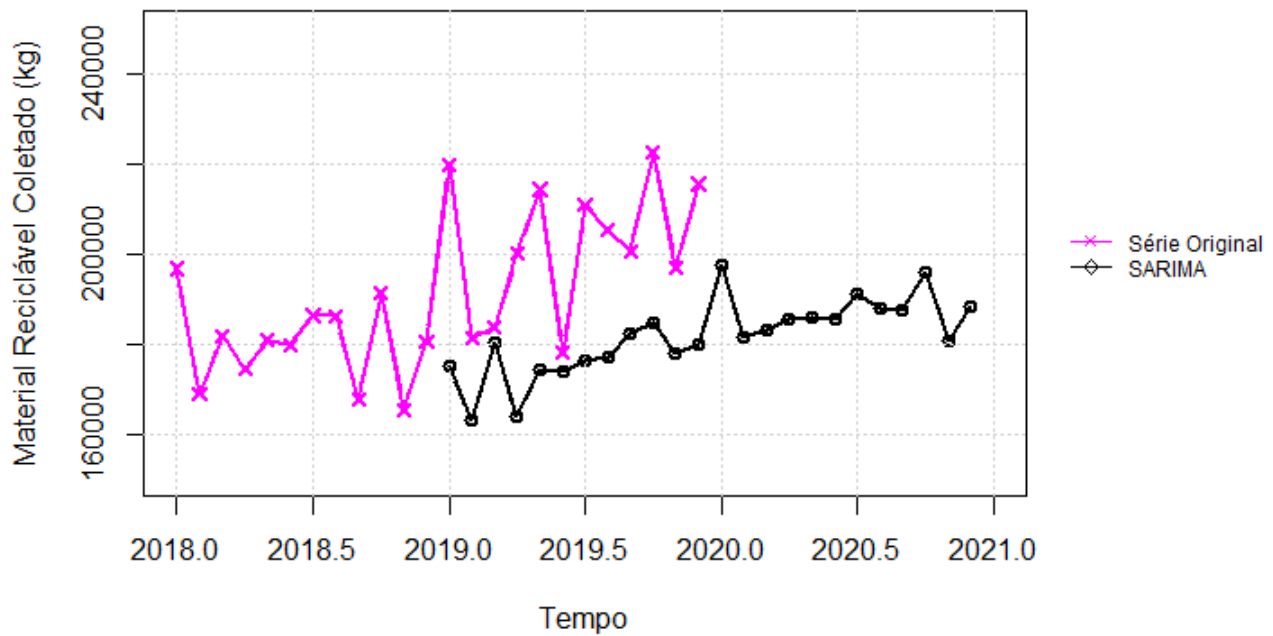


Figura 4.8: Previsões do modelo SARIMA (Cenário A).

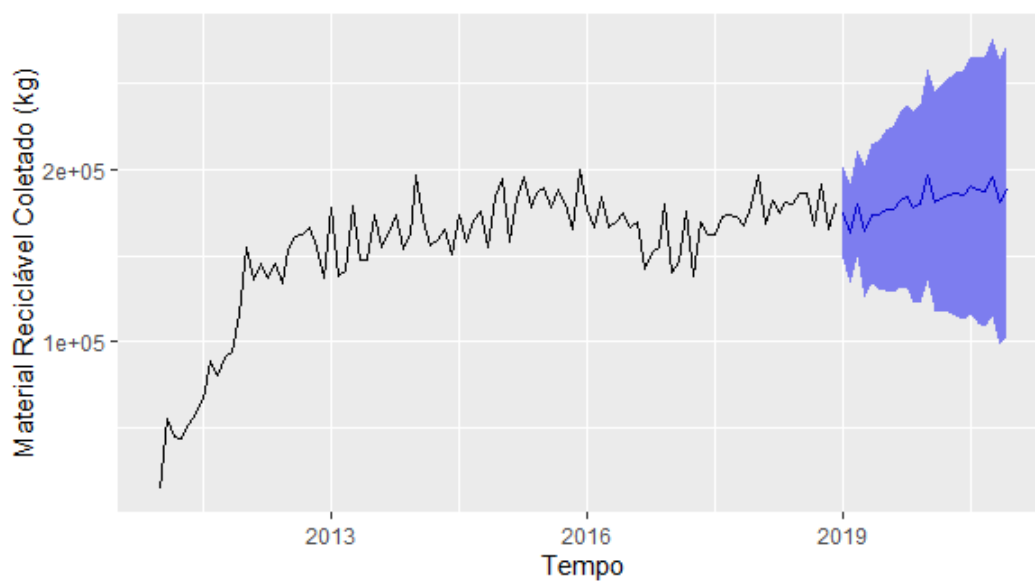


Figura 4.9: Intervalos de confiança das previsões do modelo SARIMA (Cenário A).

4.1.4 COMPARAÇÃO ENTRE OS MODELOS SEHW E SARIMA

O método de previsão de SEHW aplica pesos desiguais às observações passadas da série temporal, com maior peso para as observações mais recentes [32]. Assim, o crescimento observado nos anos de 2017 e 2018 é refletido nas previsões realizadas em 2019 e 2020. Já nos modelos de Box e Jenkins, parte-se da ideia de que os valores de uma série temporal são altamente dependentes, ou seja, cada valor pode ser explicado por valores prévios da série. O modelo pode ser visto como um filtro que separa o sinal do ruído e então projeta esse sinal para o futuro para fazer as previsões [5]. Assim, as variações nas previsões do modelo SARIMA foram mais suaves. Os valores reais e previsões do ano 2019 são mostradas na Figura 4.10.

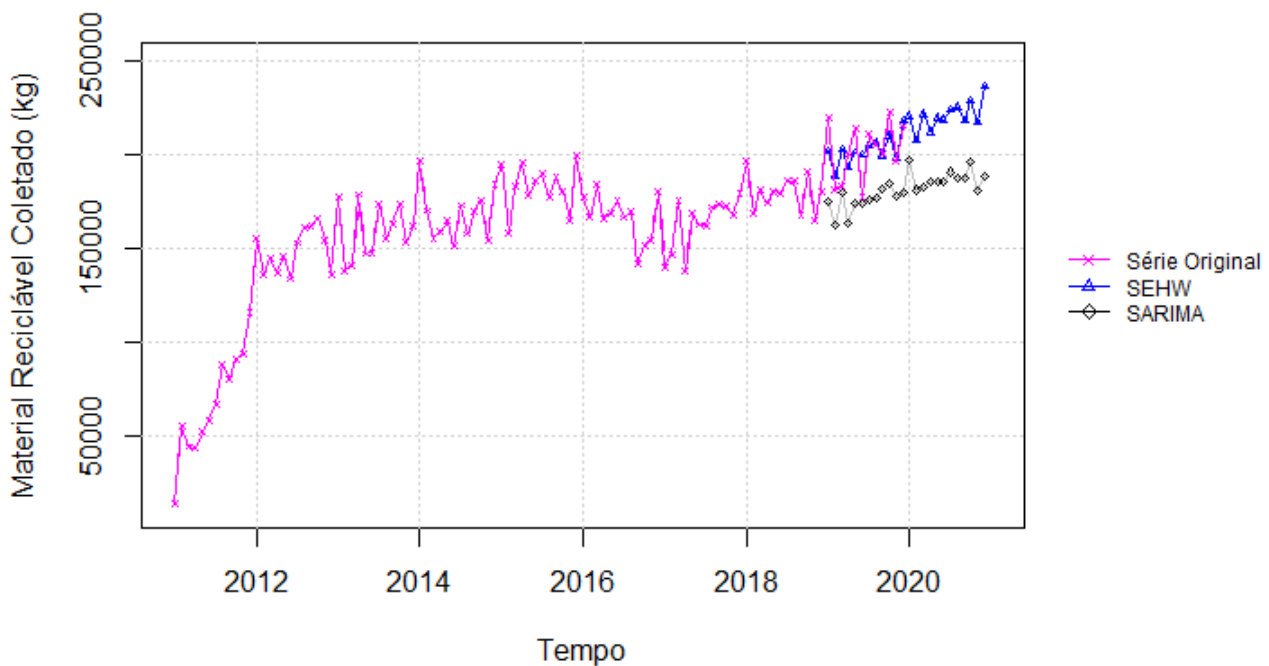


Figura 4.10: Série original e previsões dos modelos (Cenário A).

A utilização do modelo SARIMA resultou em maiores erros quando comparados com o modelo SEHW, ou seja, pior desempenho. Os valores previstos para o ano de 2019 foram mais próximos dos valores reais utilizando o modelo de suavização exponencial, conforme Tabela 4.5. Além disso, o modelo SEHW prevê um aumento na quantidade de material reciclável a ser coletado em 2020, o que é mais condizente com a realidade, enquanto o modelo SARIMA subestima esse montante. O modelo SARIMA permite o cálculo do intervalo de confiança das previsões e verifica-se na Figura 4.9 que quanto maior o horizonte da previsão, maior é a amplitude desse intervalo, ou seja, maior a incerteza.

Na Tabela 4.6, os erros RMSE, MAE e MAPE são apresentados, destacando o melhor desempenho do modelo SEHW, diferentemente do que foi observado no estudo de Günsel (2016) [20]. A alta magnitude dos dois primeiros critérios dá a impressão de serem erros significativos,

mas o erro percentual MAPE, independente da escala dos dados, ilustra a boa performance preditiva obtida.

Tabela 4.5: Previsões realizadas e erros de previsão para o ano de 2019 (Cenário A).

Mês-Ano	Valor Real	SEHW		SARIMA	
		Valor Previsto	Erro Previsão	Valor Previsto	Erro Previsão
Jan-19	219979,00	202055,87	17923,13	197408,50	22570,50
Fev-19	181329,00	188933,89	7604,89	181399,50	70,50
Mar-19	183661,00	203266,76	19605,76	183264,70	396,30
Abr-19	200218,00	193662,83	6555,17	185530,40	14687,60
Mai-19	214441,00	201069,27	13371,73	185922,80	28518,20
Jun-19	177961,00	200207,95	22246,95	185347,70	7386,70
Jul-19	210856,00	205473,90	5382,10	191024,00	19832,00
Ago-19	205287,00	206981,19	1694,19	187982,90	17304,10
Set-19	200693,00	199810,34	882,66	187475,20	13217,80
Out-19	222508,00	210327,68	12180,32	196042,80	26465,20
Nov-19	196883,00	198658,75	1775,75	180879,70	16003,30
Dez-19	215604,00	218337,87	2733,87	188468,20	27135,80

Tabela 4.6: Erros RMSE, MAE e MAPE dos modelos SEHW e SARIMA (Cenário A).

Critério	Método	
	SARIMA (3,1,0)(0,0,2) _[12]	SEHW
RMSE	29878,08	11803,12
MAE	26762,90	9329,71
MAPE	12,85	4,71

Para o ano de 2020, o modelo SEHW prevê, em geral, um crescimento em relação aos valores observados em 2019, com valores superiores a 200 toneladas mensais. Já para o modelo SARIMA, essa marca não é alcançada pelas previsões de 2020, as quais possivelmente subestimarão a quantidade coletada. Essas previsões são mostradas na Tabela 4.7.

Tabela 4.7: Previsões realizadas para o ano de 2020 (Cenário A).

Mês-Ano	SARIMA	SEHW
Jan-20	197408,50	220713,16
Fev-20	181399,50	207591,19
Mar-20	183264,70	221924,06
Abr-20	185530,40	212320,13
Mai-20	185922,80	219726,57
Jun-20	185347,70	218865,24
Jul-20	191024,00	224131,20
Ago-20	187982,90	225638,49
Set-20	187475,20	218467,64
Out-20	196042,80	228984,98
Nov-20	180879,70	217316,05
Dez-20	188468,20	236995,16

4.2 CENÁRIO B

Analizando a Figura 4.1, verifica-se que os dados referentes ao ano 2011 são bem inferiores que os demais. Assim, no Cenário B, as observações deste ano, no qual o projeto de coleta seletiva ainda estava em sua fase inicial, são desconsideradas para fins de modelagem.

4.2.1 DESCRIÇÃO INICIAL

Inicialmente são apresentadas as medidas descritivas do conjunto de observações sem o ano de 2011 na Tabela 4.8, permitindo verificar média superior e variabilidade inferior àquelas apresentadas na Tabela 4.1 que considera as doze observações iniciais. De modo similar ao Cenário A, considerou-se um conjunto de treinamento, mas agora com as observações de 2012 a 2018, e um conjunto de testes com as observações de 2019.

Tabela 4.8: Medidas descritivas dos materiais recicláveis coletados entre 2012 e 2019 pelo programa de coleta seletiva em Uberlândia (Cenário B).

Medidas Descritivas	Valor	Medidas Descritivas	Valor
Média (ton)	171.496,40	Mínimo (ton)	13.376,00
Desvio-padrão (ton)	19.971,37	1º Quartil (ton)	157.282,50
Coeficiente de variação (%)	11,65	Mediana (ton)	171.305,00
Assimetria	0,29	3º Quartil (ton)	182.123,80
Curtose	-0,11	Máximo (ton)	222.508,00

Na Tabela 4.9 são apresentados os testes de tendência aplicados na série observada a partir

de 2012. Em todos eles há evidências para trabalhar com modelos que consideram a presença de tendência.

Tabela 4.9: Testes para verificação de tendência na série (Cenário B).

Teste	Estatística do teste	Decisão	Conclusão
Wald-Wolfowitz	$T_1 = 38$	$z_{cal} = -2,27 < z_{tab} = -1,96$	Rejeita-se H_0
Cox-Stuart	$T_2 = 38$	$z_{cal} = 4,04 > z_{tab} = 1,96$	Rejeita-se H_0
Correlação de Spearman	$T_3 = 57040$	$t_{cal} = 7,99 > t_{tab} = 1,98$	Rejeita-se H_0

Verificando a existência de sazonalidade na série, pelo teste de Kruskal-Wallis conclui-se que não há sazonalidade, enquanto que para o teste de Friedman há evidências para presença da componente sazonal. Já pelo teste de Fisher, considerando um nível de significância de 5%, há evidências de peridiocidade na série. Deste modo, opta-se por ajustar um modelo que considera a existência de sazonalidade. Os resultados dos testes são apresentados na Tabela 4.10.

Tabela 4.10: Testes para verificação de sazonalidade na série (Cenário B).

Teste	Estatística do teste	Decisão	Conclusão
Kruskal-Wallis	$T_4 = 10,95$	$T_4 < \chi^2_{tab} = 19,67$	Não se rejeita H_0
Friedman	$T_5 = 28,31$	$T_5 > \chi^2_{tab} = 19,67$	Rejeita-se H_0
Fisher	$Valor-p = 0,01$	$Valor-p < \alpha$	Rejeita-se H_0

4.2.2 MODELAGEM SEHW

Novamente com a presença de tendência e sazonalidade na série, utilizou-se o modelo de suavização exponencial de Holt-Winters. As constantes de suavização, determinadas pelo método iterativo que minimiza o EQM foram $A = 0,0774$, $C = 0,3775$ e $D = 0,1347$. Na Figura 4.11 são ilustradas as previsões para 2019 e 2020. Foi verificado um comportamento semelhante ao modelo SEHW do cenário A mostrado na Figura 4.4, com previsões de 2019 próximas aos valores reais e projeções de crescimento para 2020.

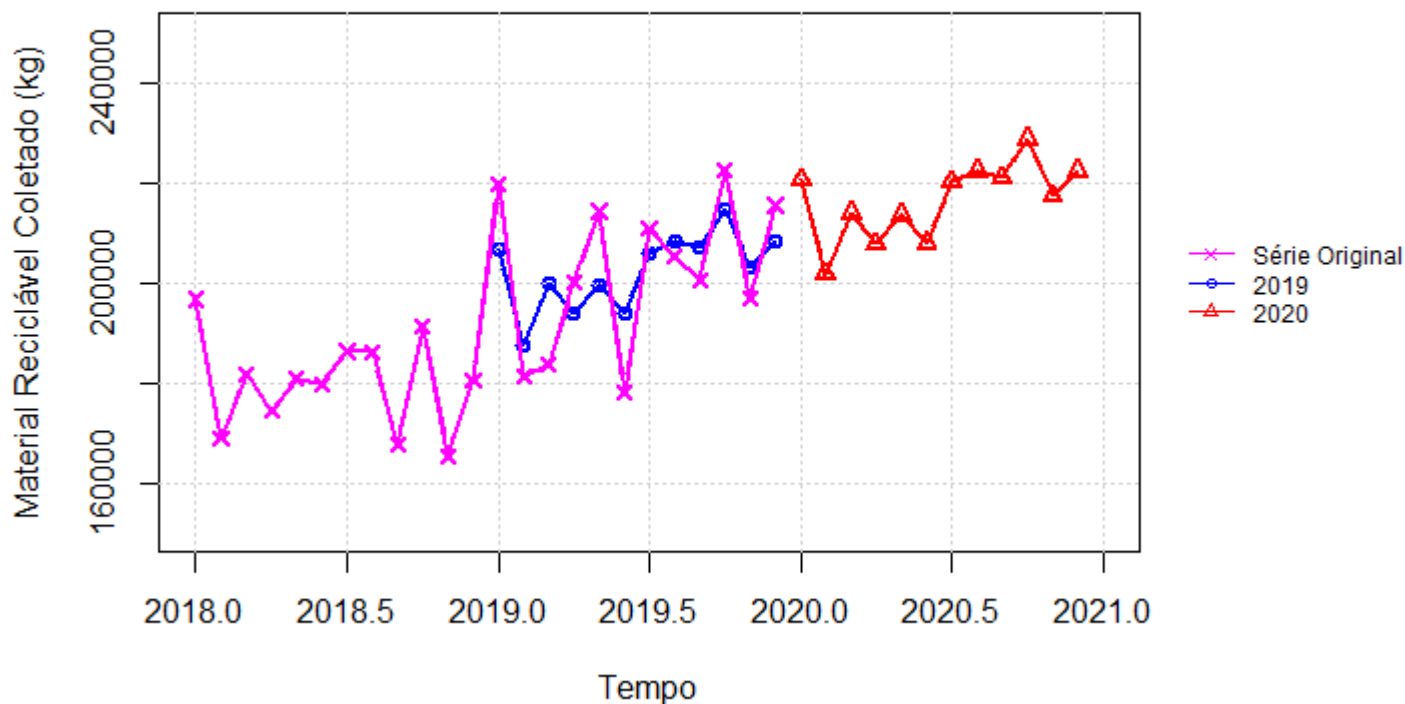


Figura 4.11: Previsões do modelo SEHW (Cenário B).

4.2.3 MODELAGEM SARIMA

Na busca por um modelo SARIMA, foram examinados os gráficos de FAC e FACp da série diferenciada apresentados na Figura 4.12. As defasagens significativas, na FAC, seguem um padrão, reforçando a escolha pelo ajuste de um modelo do tipo autorregressivo integrado de médias móveis sazonal e o primeiro “lag” significativo na FACp dá sinais da presença do parâmetro regressivo p no modelo.

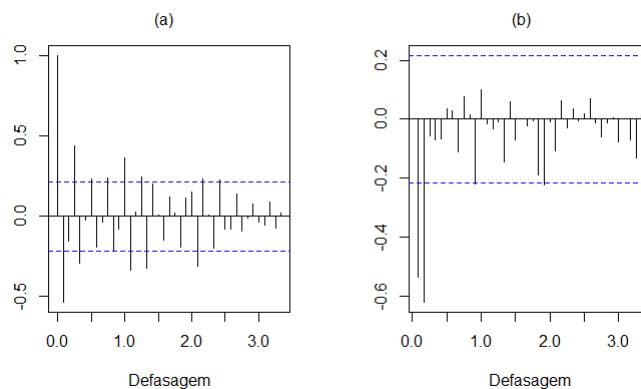


Figura 4.12: Gráficos de FAC (a) e FACp (b) da série diferenciada (Cenário B).

Com AIC de valor 1802,73 e BIC de 1814,82, o modelo $(2,1,0)(0,0,2)_{[12]}$, cujos parâmetros são listados na Tabela 4.11, foi o escolhido dentre vários.

Tabela 4.11: Parâmetros estimados do modelo SARIMA $(2,1,0)(0,0,2)_{[12]}$ (Cenário B).

SARIMA	Parâmetro	Estimativa	Desvio-Padrão
$(2,1,0)(0,0,2)_{[12]}$	ϕ_1	-0,8394	0,0892
	ϕ_2	-0,5994	0,0890
	Θ_1	0,1964	0,1141
	Θ_2	0,3204	0,1430

A FAC e FACp dos resíduos do modelo encontrado, bem como o periodograma acumulado, respectivamente nas Figuras 4.13 e 4.14, indicam a adequabilidade do modelo. O valor-p de 0,8794 do teste Box-Pierce reitera a hipótese de ruído branco.

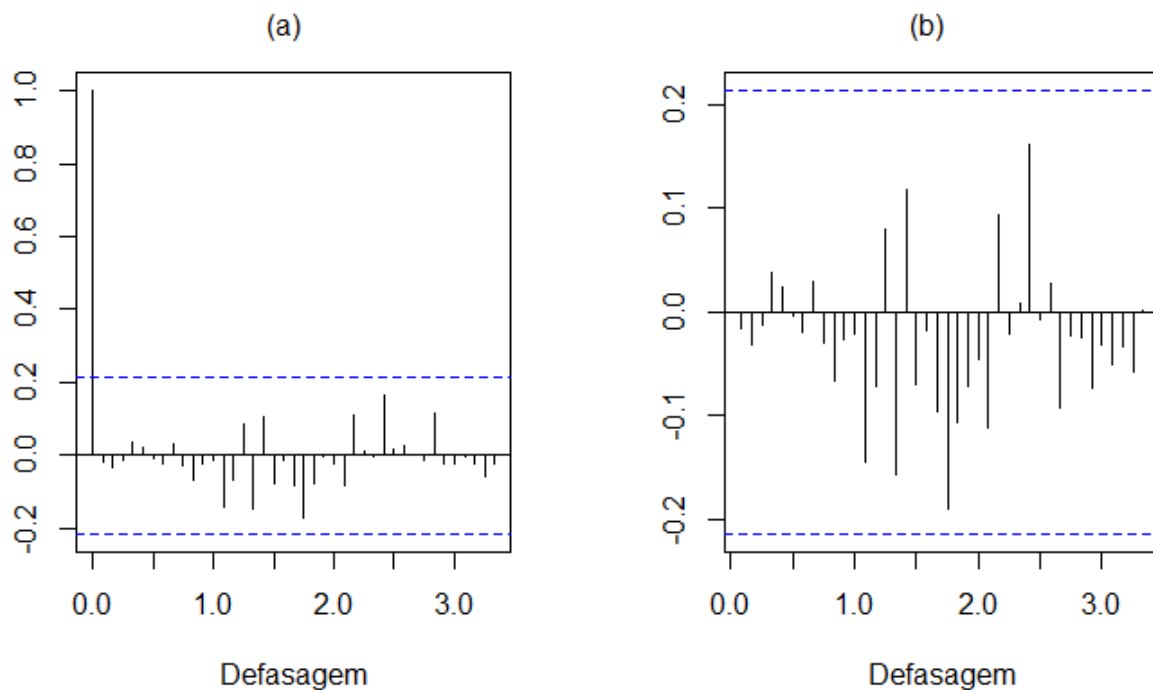


Figura 4.13: Gráficos de FAC (a) e FACp (b) dos resíduos (Cenário B).

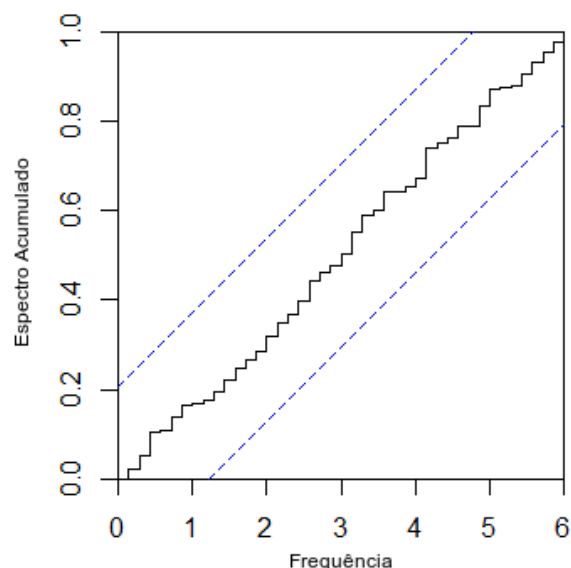


Figura 4.14: Periodogramma Acumulado (Cenário B).

As previsões do modelo obtido, para os anos 2019 e 2020, são apresentadas na Figura 4.15. Nota-se que os valores obtidos são inferiores às previsões feitas pelo SARIMA encontrado no cenário A, mostradas na 4.4, subestimando a quantidade coletada em 2019 e possivelmente a quantidade a ser coletada em 2020. Na Figura 4.16 é ilustrada a variabilidade das previsões realizadas.

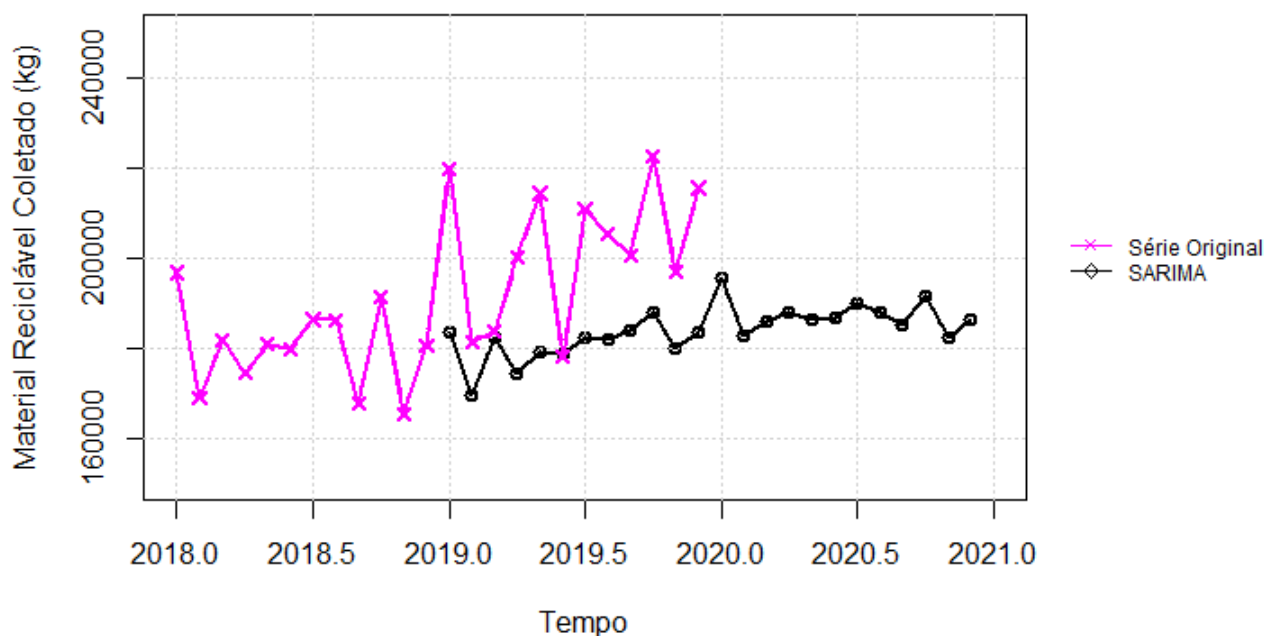


Figura 4.15: Previsões do modelo SARIMA (Cenário B).

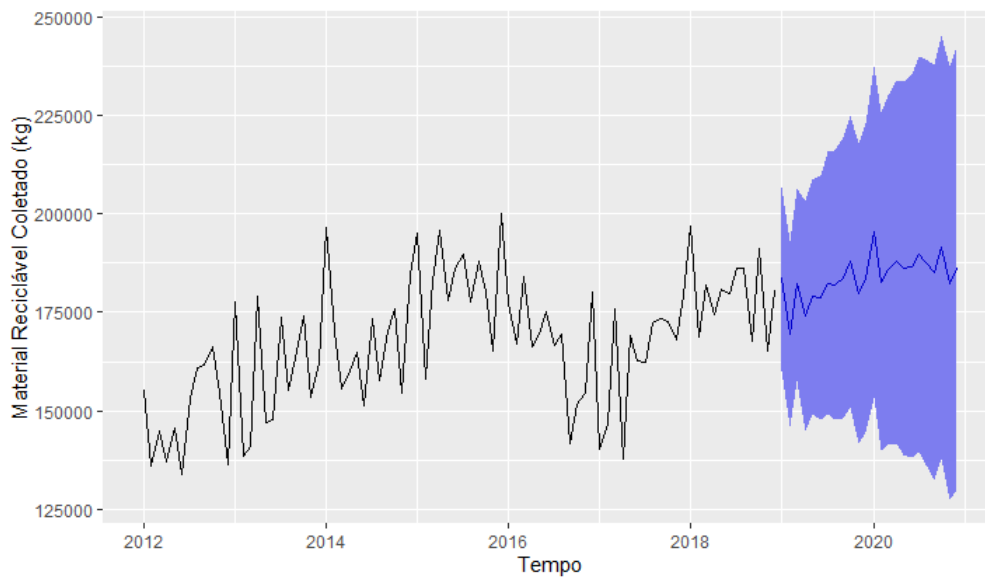


Figura 4.16: Intervalos de confiança das previsões do modelo SARIMA (Cenário B).

4.2.4 COMPARAÇÃO ENTRE OS MODELOS SEHW E SARIMA

Na Figura 4.17 é apresentada a série original e as previsões dos anos 2019 e 2020 feitas pelos modelos SARIMA e SEHW. Assim como no cenário A, verifica-se que o método da suavização exponencial obteve melhor performance preditiva, além de não subestimar as previsões.

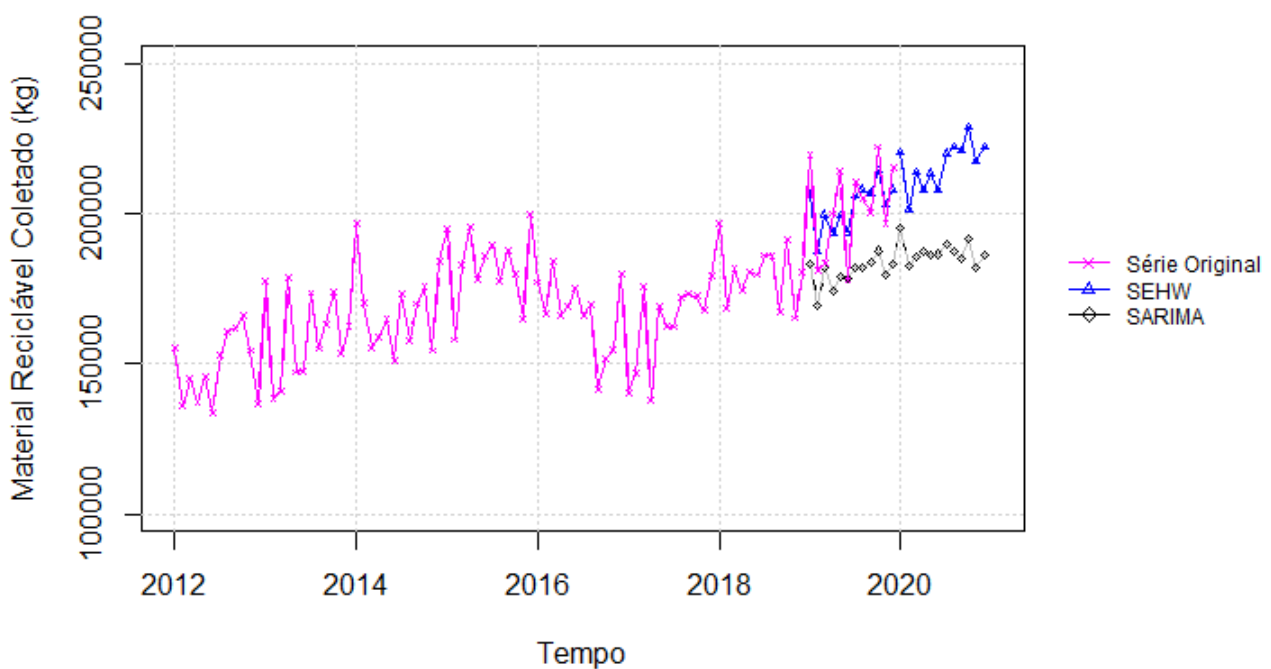


Figura 4.17: Série original e previsões dos modelos (Cenário B).

A previsões de 2019 de ambos os modelos são apresentadas na Tabela 4.12. Os erros de previsão do modelo SEHW são menores que o modelo SARIMA, este que apresentou valores previstos que subestimaram a coleta de 2019.

Tabela 4.12: Previsões realizadas e erros de previsão para o ano de 2019 (Cenário B)

Mês-Ano	Valor Real	SEHW		SARIMA	
		Valor Previsto	Erro Previsão	Valor Previsto	Erro Previsão
Jan-19	219979,00	206596,37	13382,63	183613,00	36366,00
Fev-19	181329,00	187538,14	6209,14	169388,50	11940,50
Mar-19	183661,00	199887,43	16226,43	182165,90	1495,10
Abr-19	200218,00	193759,30	6458,70	174088,80	26129,20
Mai-19	214441,00	199464,58	14976,42	179068,60	35372,40
Jun-19	177961,00	193933,43	15972,43	178817,50	856,50
Jul-19	210856,00	206118,67	4737,33	182396,20	28459,80
Ago-19	205287,00	208253,69	2966,69	181947,30	23339,70
Set-19	200693,00	207011,16	6318,16	183719,00	16974,00
Out-19	222508,00	214764,13	7743,87	187865,60	34642,40
Nov-19	196883,00	203363,65	6480,65	179916,90	16966,10
Dez-19	215604,00	208279,53	7324,47	183615,40	31988,60

Os erros de cada modelo são apresentados na Tabela 4.13, ratificando o melhor desempenho do modelo SEHW. Em comparação com o Cenário A, os erros de previsão foram ainda menores.

Tabela 4.13: Erros RMSE, MAE e MAPE dos modelos SEHW e SARIMA (Cenário B).

Critério	Método	
	SARIMA (2,1,0)(0,0,2) _[12]	SEHW
RMSE	25093,16	10119,49
MAE	22044,18	9066,41
MAPE	10,52	4,54

Já as previsões de 2020 são apresentadas na Tabela 4.14. Espera-se um crescimento, em relação à 2019, na quantidade a ser coletada pelo método SEHW, enquanto para o modelo SARIMA a previsão mensal é abaixo das 200 toneladas mensais.

Tabela 4.14: Previsões realizadas para o ano de 2020 (Cenário B).

Mês-Ano	SARIMA	SEHW
Jan-20	195427,10	220744,43
Fev-20	182743,20	201686,19
Mar-20	185820,00	214035,49
Abr-20	187797,60	207907,36
Mai-20	186159,20	213612,63
Jun-20	186668,20	208081,49
Jul-20	189779,30	220266,72
Ago-20	187727,70	222401,75
Set-20	185190,40	221159,22
Out-20	191524,50	228912,19
Nov-20	182339,70	217511,71
Dez-20	186293,30	222427,60

4.3 CENÁRIO C

Conforme pode-se observar nas Figuras 4.9 e 4.16, quanto maior o horizonte, maior a variabilidade ou incerteza das previsões realizadas. Desta maneira, com o intuito de verificar a influência na previsão do material a ser coletado em 2020, neste cenário optou-se não separar os dados de 2019 em um conjunto de teste. Isto é, as previsões para o ano de 2020 foram realizadas com base na observação referente a dezembro de 2019, em vez de tomar como ponto de partida o valor referente a dezembro de 2018 como nos cenários anteriores. Optou-se também por não utilizar os dados de 2011, de forma similar ao Cenário B.

A série é semelhante àquela apresentada no Cenário B, exceto pelo fato da utilização das observações do ano de 2019 durante a modelagem, não sendo reservadas para uma análise da performance preditiva do modelo. Assim, a descrição inicial da série é semelhante ao apresentado na Seção 4.2.1 do cenário anterior.

4.3.1 MODELAGEM SEHW

Na modelagem da suavização exponencial é possível atualizar as previsões à medida que novas observações são coletadas. Assim, com base no modelo encontrado no Cenário B, com as constantes $A = 0,0774$, $C = 0,3775$ e $D = 0,1347$, as observações de 2019 foram utilizadas para atualizar as previsões de 2020. O procedimento consiste em calcular as equações de suavização para os dados de janeiro à dezembro de 2019, conforme apresentado em (2.18) e então realizar as previsões para 2020, de acordo com (2.19), com base na última observação disponível, referente à dezembro de 2019. As previsões atualizadas são apresentadas na Figura 4.18.

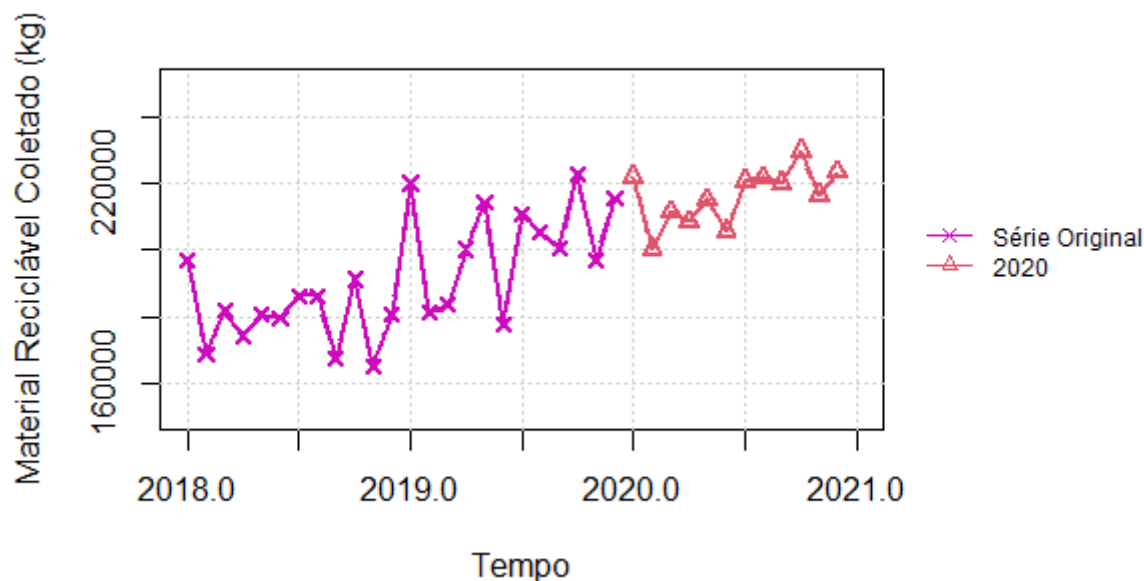


Figura 4.18: Previsões do modelo SEHW (Cenário C).

4.3.2 MODELAGEM SARIMA

Considerando as 96 observações entre janeiro de 2012 e dezembro de 2019, novamente objetivou-se estipular um modelo SARIMA que descreva bem a série analisada. Examinando os gráficos de FAC e FACp da série diferenciada dispostos na Figura 4.19, verifica-se semelhança com a análise feita para o Cenário B, com indícios de um modelo do tipo autorregressivo integrado de médias móveis sazonal e presença do parâmetro regressivo p no modelo.

Com AIC de valor 2065,60 e BIC de 2078,37, o modelo encontrado teve as mesmas ordens do modelo obtido no cenário B, com notação $(2,1,0)(0,0,2)_{[12]}$, mas com estimativas diferentes, as quais são apresentadas na Tabela 4.15.

Tabela 4.15: Parâmetros estimados do modelo SARIMA $(2,1,0)(0,0,2)_{[12]}$ (cenário C).

SARIMA	Parâmetro	Estimativa	Desvio-Padrão
$(2,1,0)(0,0,2)_{[12]}$	ϕ_1	-0,8026	0,0823
	ϕ_2	-0,6180	0,0801
	Θ_1	0,3455	0,1080
	Θ_2	0,4264	0,1587

Analisando os resíduos, observa-se por meio da análise dos gráficos da FAC e FACp residual, apresentados na Figura 4.20, e do periodograma acumulado, na Figura 4.21, que o modelo é adequado. O teste Box-Pierce, com valor-p de 0,6222, corrobora com essa conclusão.

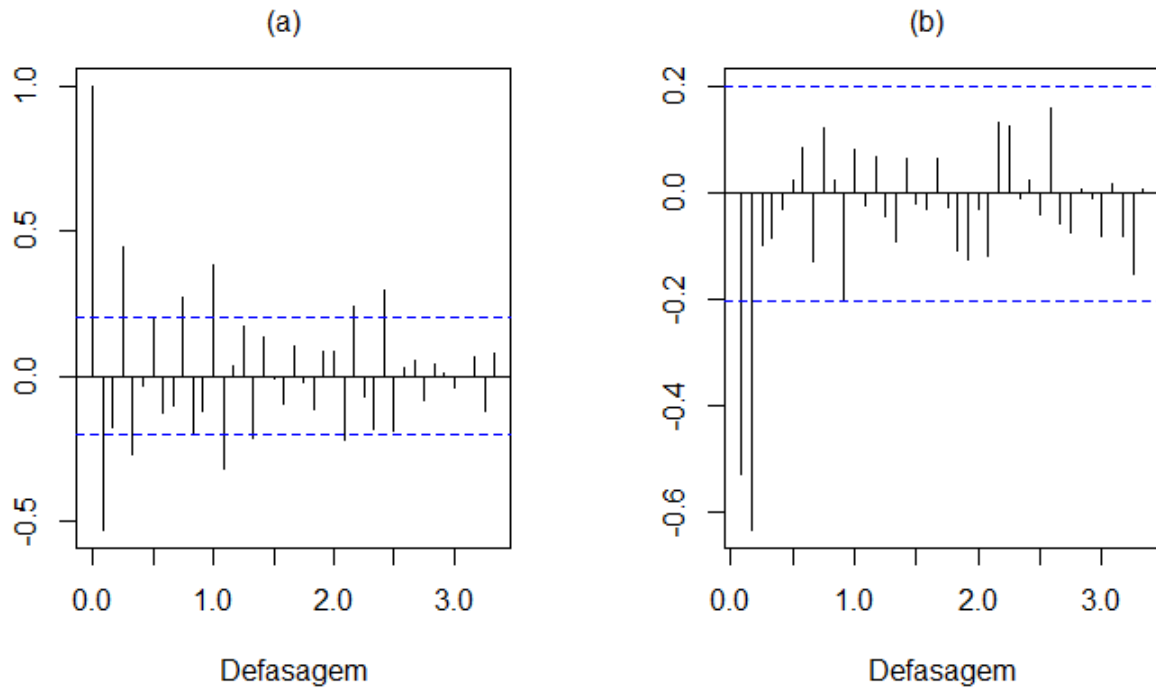


Figura 4.19: Gráficos de FAC (a) e FACp (b) da série diferenciada (Cenário C).

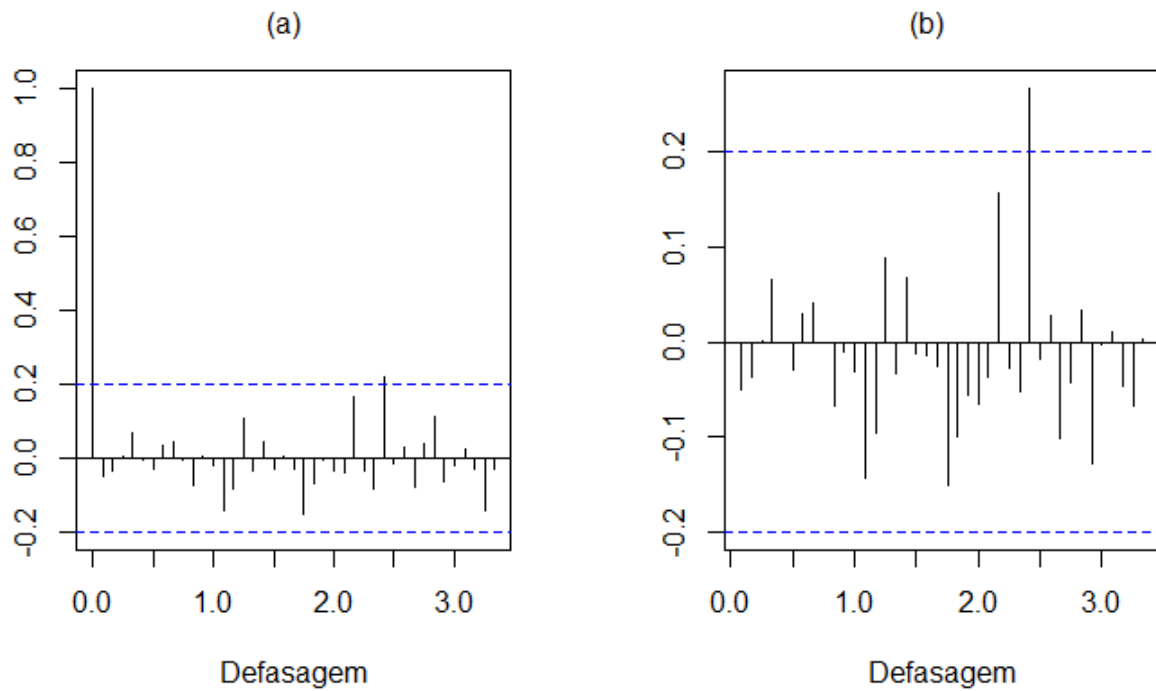


Figura 4.20: Gráficos de FAC (a) e FACp (b) dos resíduos (Cenário C).

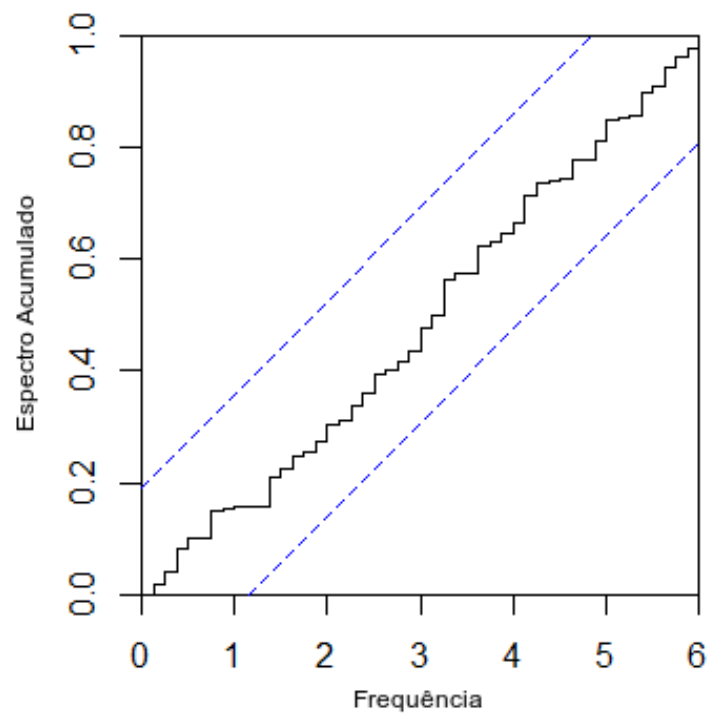


Figura 4.21: Periodograma Acumulado (Cenário C).

As previsões para 2020 do modelo obtido são apresentadas na Figura 4.22. Diferentemente dos outros cenários, espera-se um crescimento na quantidade prevista a ser coletada. Na Figura 4.23 são ilustradas as previsões e seus intervalos de confiança.

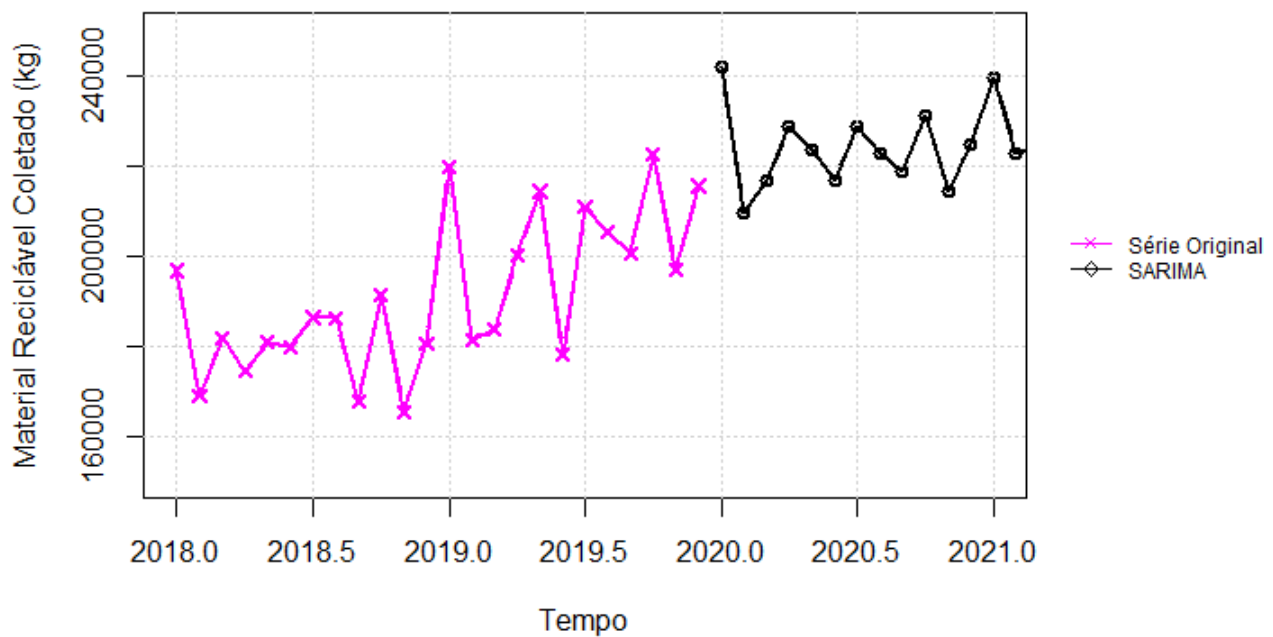


Figura 4.22: Previsões do modelo SARIMA (Cenário C).

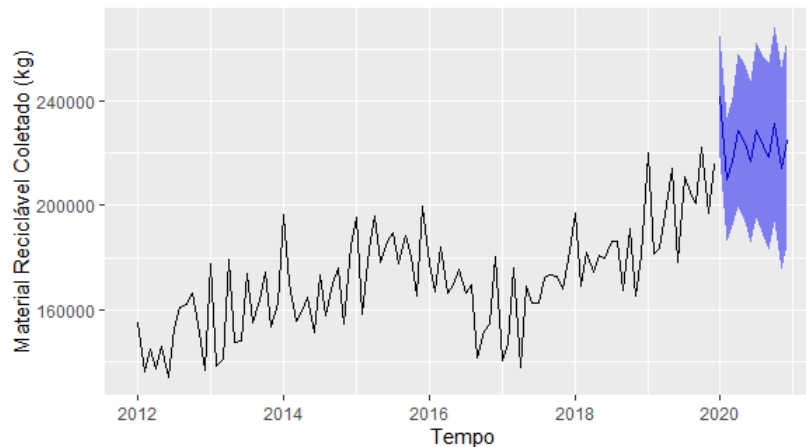


Figura 4.23: Intervalos de confiança das previsões do modelo SARIMA (Cenário C).

4.3.3 COMPARAÇÃO ENTRE OS MODELOS SEHW E SARIMA

A série original e as previsões para 2020 provenientes dos dois modelos são apresentadas na Figura 4.24. Diferentemente dos outros cenários, o modelo ARIMA apresentou previsões superiores ao modelo SEHW.

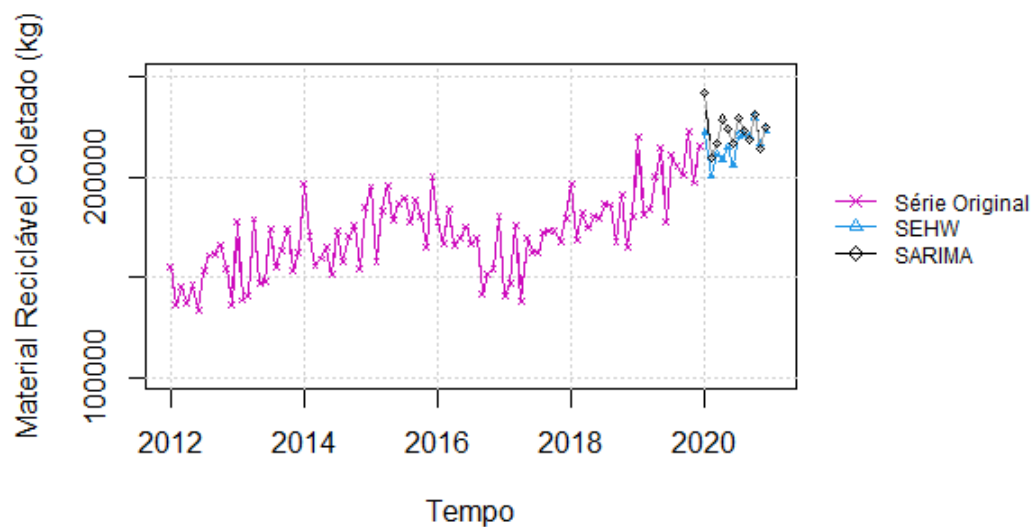


Figura 4.24: Série original e previsões dos modelos (Cenário C).

Como não foi reservada uma parte das observações para comparação da performance preditiva, não é possível calcular os erros do modelo. As previsões de 2020 são listadas na Tabela 4.16. Espera-se um crescimento, em relação à 2019, na quantidade a ser coletada tanto pelo método de Box e Jenkins, quanto pelo método da suavização exponencial.

Tabela 4.16: Previsões realizadas para o ano de 2020 (Cenário C).

Mês-Ano	SARIMA	SEHW
Jan-20	241892,40	221965,56
Fev-20	209781,10	200316,89
Mar-20	216739,30	211495,42
Abr-20	228839,60	208415,16
Mai-20	223816,50	215147,66
Jun-20	216680,00	205606,29
Jul-20	228963,00	220575,31
Ago-20	222803,30	221731,28
Set-20	218704,60	220139,08
Out-20	231345,90	229756,92
Nov-20	214223,80	216531,63
Dez-20	224924,30	223270,72

4.4 COMPARAÇÃO GERAL

A fim de comparar todos os modelos obtidos, na Figura 4.25 são plotadas as previsões feitas das duas técnicas utilizadas, ARIMA e SEHW, referente aos três cenários abordados. Observa-se que a modelagem ARIMA apresentou previsões bem distintas em relação ao três cenários, sendo que nos cenários A e B foi subestimada a coleta referente a 2019 e possivelmente subestimar a coleta para o ano de 2020. Já em relação ao modelo de suavização exponencial, a atualização das previsões não resultou em grandes diferenças, já que as previsões para 2020 nos dois cenários foram bem próximas e o cenário considerando o primeiro ano do projeto (cenário A) apresentou os maiores valores previstos para 2020 com o modelo SEHW.

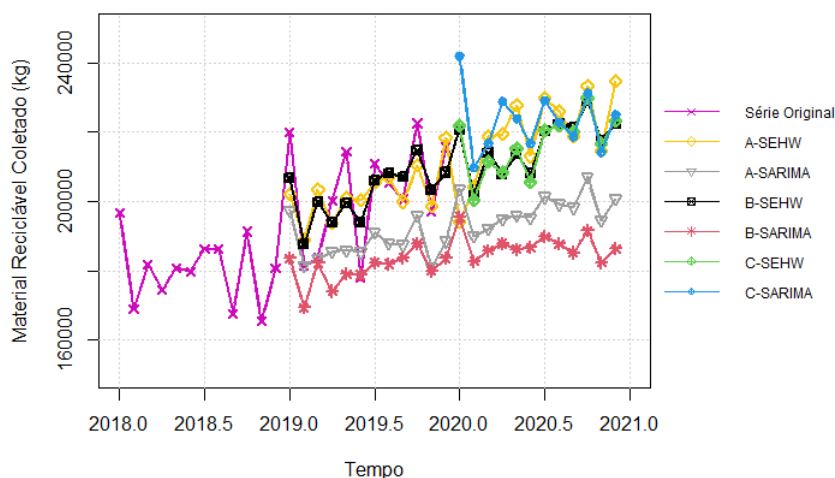


Figura 4.25: Série original e previsões dos modelos nos três diferentes cenários.

5. CONCLUSÕES

Os modelos de suavização exponencial e de Box e Jenkins foram úteis para descrever o comportamento do quantitativo mensal de material coletado pelo serviço de coleta seletiva em Uberlândia. Foram ajustados dois modelos de Suavização Exponencial de Holt-Winters para os cenários A e B, enquanto no cenário C foram atualizadas as previsões do cenário B com os valores observados de 2019. Em relação à modelagem de Box e Jenkins, no cenário A foi ajustado um modelo SARIMA $(3,1,0)(0,0,2)_{[12]}$, ao passo que nos cenários B e C foram ajustados um modelo SARIMA $(2,1,0)(0,0,2)_{[12]}$.

Os dados referentes ao ano de 2011 são bem inferiores aos valores observados em anos seguintes. Desta forma, recomenda-se tratá-los como “outliers” e não considerá-los na modelagem, conforme realizado nos cenários B e C. Além disso, caso não seja necessário trabalhar com métricas relacionadas à performance preditiva, o cenário C, que considera as observações de 2019 durante a modelagem e não reserva parte dos dados para comparação com as previsões realizadas, apresentou previsões para o ano 2020 compatíveis com o que se pode esperar da realidade, com aumento da quantidade de material reciclável a ser coletado, inclusive na modelagem ARIMA, diferente dos outros cenários em que a quantidade a ser coletada foi subestimada quando se utilizou essa técnica.

Dentre as duas técnicas utilizadas, recomenda-se a utilização da modelagem de suavização exponencial, que além de ser mais simples, apresentou melhor performance preditiva com menores erros RMSE, MAE e MAPE e nos três distintos cenários fez previsões que consideram um aumento na quantidade de material reciclável a ser coletado pelo programa de coleta seletiva em Uberlândia em 2020.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] ABRELPE: *Panorama dos resíduos sólidos no Brasil 2018 2019*, 2019. <http://abrelpe.org.br/download-panorama-2018-2019/>, acessado em 19/06/2020.
- [2] Adamović, V. M., Antanasijević, D. Z., Ristić, M., Perić-Grujić, A. A. e Pocajt, V. V.: *Prediction of municipal solid waste generation using artificial neural network approach enhanced by structural break analysis*. *Environ Sci Pollut Research*, 24:299–311, 2017. <https://doi.org/10.1007/s11356-016-7767-x>.
- [3] Ahdesmaki, M., Fokianos, K. e Strimmer, K.: *GeneCycle: Identification of Periodically Expressed Genes*, 2019. <https://CRAN.R-project.org/package=GeneCycle>, R package version 1.1.4.
- [4] Bertolo, L. A.: *Métodos básicos de previsão no Excel*, 2013. <http://www.bertolo.pro.br/MetodosQuantitativos/Simulacao/MetodosBasicosDePrevisaoDeSeriesTemporaisNoExcel.pdf>, acessado em 19/06/2020.
- [5] Box, G. E., Jenkins, G. M., Reinsel, G. C. e Ljung, G. M.: *Time Series Analysis: Forecasting and Control*. John Wiley Sons, 2015.
- [6] Brookwell, P. J. e Davis, R. A.: *Introduction to Time Series and Forecasting*. Springer International Publishing, 2016.
- [7] Chechi, L. e Bayer, F. M.: *Modelos univariados de séries temporais para previsão das temperaturas médias mensais de Erechim, RS*. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 16(12):1321–1329, 2012.
- [8] Conke, L. S.: *Barreiras ao Desenvolvimento da Coleta Seletiva no Brasil*. Tese de Doutorado, 2015.
- [9] Conke, L. S. e Nascimento, E. P.: *A coleta seletiva nas pesquisas brasileiras: uma avaliação metodológica*. *Revista Brasileira de Gestão Urbana*, 10(1):199–212, 2018. <https://doi.org/10.1590/2175-3369.010.001.ao14>.
- [10] Cowpertwait, P. S. P. e Metcalfe, A. W.: *Introductory Time Series with R*. Springer International Publishing, 2009.

- [11] Da Luz, C., Mandler, M. L., Tokoro, P. M., Lima, R. e Moro, G.: *Métodos estatísticos de suavização exponencial holt-winters para previsão de demanda em uma empresa do setor metal mecânico*. Revista Gestão Industrial, 8(4), 2012.
- [12] Davila, V. H. L.: *Introdução às Séries Temporais*, 2015. <https://www.ime.unicamp.br/~hlaachos/MaterialSeries.pdf>, acessado em 19/06/2020.
- [13] DMAE: *Coleta, transporte e disposição de resíduos sólidos urbanos no aterro sanitário*, 2020. <https://www.uberlandia.mg.gov.br/prefeitura/orgaos-municipais/dmae/servicos-dmae/coleta-de-residuos/>, acessado em 19/06/2020.
- [14] DMAE: *Programa de coleta seletiva: Coleta, transporte e destinação de resíduos sólidos recicláveis para associações e cooperativas de catadores*, 2020. <https://www.uberlandia.mg.gov.br/prefeitura/orgaos-municipais/dmae/servicos-dmae/coleta-de-residuos/>, acessado em 19/06/2020.
- [15] Ehlers, R. S.: *Análise de Séries Temporais*, 2009. <http://www.icmc.usp.br/ehlers/stemp/stemp.pdf>, acessado em 19/06/2020.
- [16] Eigenheer, E. M.: *A limpeza urbana através dos tempos*. Gráfica Pallotti, 2ª ed., 2009.
- [17] Foundation, T. D.: *LibreOffice Calc*. <https://www.libreoffice.org/discover/calc/>.
- [18] Gelper, S., Fried, R. e Croux, C.: *Robust forecasting with exponential and Holt–Winters smoothing*. Journal of Forecasting, 29(1):285–300, 2010. <https://doi:10.1002/for.1125>.
- [19] Gutiérrez, J. L. C.: *Monitoramento da instrumentação da Barragem de Corumbá-I por redes neurais e modelos de Box Jenkins*. Dissertação de Mestrado, 2003.
- [20] Günsel, C. B.: *Environmental Problems in Northern Cyprus: a Descriptive Statistical Analysis and Forecasting of Air Pollution and Solid Waste*. Master's Thesis, 2016.
- [21] Hyndman, R., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., Petropoulos, F., Razbash, S., Wang, E. e Yasmeen, F.: *forecast: Forecasting functions for time series and linear models*, 2020. <https://pkg.robjhyndman.com/forecast/>, R package version 8.13.
- [22] Hyndman, R. J. e Athanasopoulos, G.: *Forecasting: principles and practice*. OTexts, 2ª ed., 2018.
- [23] IBGE: *Perfil dos municípios brasileiros - 2017*. IBGE, 2018.
- [24] IBGE: *Cidades*, 2019. <https://www.ibge.gov.br/cidades-e-estados/mg/uberlandia.html>, acessado em 19/06/2020.
- [25] IPEA: *Pesquisa sobre Pagamento por Serviços Ambientais Urbanos para a Gestão dos Resíduos Sólidos*. IPEA, 2010.

- [26] Lemke, C. e Gabrys, B.: *On the benefit of using time series features for choosing a forecasting method*. Em *Proceedings of the European Symposium on Time Series Prediction in Porvoo*, 2008.
- [27] Maia, H. J. L., Silva, P. A., Cavalcante, L. P. S., Souza, M. A. e Silva, M. M. P.: *Coleta seletiva: benefícios da sua implantação no bairro de Santa Rosa, Campina Grande - PB*. POLÊMICA, 12(2):352–368, 2013. <https://doi.org/10.12957/polemica.2013.6437>.
- [28] Marques, R. W. C.: *Avaliação da sazonalidade do mercado de flores e plantas ornamentais no Estado de São Paulo*. Dissertação de Mestrado, 2002.
- [29] Merelles, L. R. O., Silva, C. O., Luz, M. P., Menezes, J. E. e Dias, V. S.: *Previsão de geração de resíduos sólidos para o aterro de Aparecida de Goiânia (GO) por séries temporais*. Engenharia Sanitaria e Ambiental, 24(3):537–546, 2019. <https://doi.org/10.1590/s1413-41522019183284>.
- [30] Minitab: *Função de autocorrelação (ACF)*, 2019. <https://support.minitab.com/pt-br/minitab/18/help-and-how-to/modeling-statistics/time-series/how-to/autocorrelation/interpret-the-results/autocorrelation-function-acf/>, acessado em 19/06/2020.
- [31] Minitab: *Interprete a função de autocorrelação parcial (PACF)*, 2019. <https://support.minitab.com/pt-br/minitab/18/help-and-how-to/modeling-statistics/time-series/how-to/partial-autocorrelation/interpret-the-results/partial-autocorrelation-function-pacf/>, acessado em 19/06/2020.
- [32] Morettin, P. A. e Toloi, C. M. C.: *Análise de séries temporais*. Edgard Blucher, 2ª ed., 2006.
- [33] Mushtaq, J., Dar, A. Q. e Ahsan, N.: *Spatial-temporal variations and forecasting analysis of municipal solid waste in the mountainous city of north-western Himalayas*. SN Applied Sciences, 2, 2020. <https://doi.org/10.1007/s42452-020-2975-x>.
- [34] R Core Team: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019. <https://www.R-project.org/>.
- [35] Rodrigues, W. e Santana, W. C.: *Análise econômica de sistemas de gestão de resíduos sólidos urbanos: o caso da coleta de lixo seletiva em Palmas, TO*. Revista Brasileira de Gestão Urbana, 4(2):299–312, 2012. <https://doi.org/10.7213/urbe.7402>.
- [36] Saneamento – SNS, M. do Desenvolvimento Regional Secretaria Nacional de: *Sistema Nacional de Informações sobre Saneamento: Diagnóstico do Manejo de Resíduos Sólidos Urbanos – 2018*, 2019. <http://www.snis.gov.br/diagnostico-anual-residuos-solidos/diagnostico-do-manejo-de-residuos-solidos-urbanos-2018>, acessado em 19/06/2020.

- [37] Santos, R. S. dos, Almeida, C. R. de, Silva, L. F. da, ABUD, G. e Matos, C. A. C. de: *A Disposição a pagar (DAP) e a valoração de ativos ambientais: uma aplicação do método na praia do Atalaia em Salinópolis-PA*. XIV Encontro Latino Americano de Iniciação Científica e X Encontro Latino Americano de Pós-Graduação–Universidade do Vale do Paraíba, 2010.
- [38] Santos, T. B. F.: *Coleta seletiva de resíduos sólidos urbanos em Uberlândia (MG): desafios e possibilidades de boas práticas para uma cidade sustentável*. Dissertação de Mestrado, 2019.
- [39] Silva, D., Santos, M. e Costa, D.: *A utilização do modelo Holt-Winters na elaboração de um orçamento de resultado de uma cooperativa de crédito rural*. Revista de Contabilidade do Mestrado em Ciências Contábeis da UERJ, 21(1):39–56, 2016.
- [40] Silva, V. P. M. e Capanema, L. X. D. L.: *Políticas públicas na gestão de resíduos sólidos: experiências comparadas e desafios para o Brasil*. BNDES, 25(50):153–200, 2019.
- [41] Souza, R. C.: *Metodologias para a análise e previsão de séries temporais univariadas e multivariadas*. Brazilian Review of Econometrics, 1(2):77–105, 1981.
- [42] Yanovitzky, I. e Vanlaer, A.: *Time series analysis: traditional and contemporary approaches*. The SAGE Sourcebook of Advanced Data Analysis Methods for Communication Research, pp. 125–158, 2007.