

Modelos de previsão de demanda são uma ferramenta importante para auxiliar na busca pela excelência no gerenciamento de recursos hídricos locais, pois auxiliam na tomada de decisões por parte das companhias de abastecimento. Neste trabalho, propõe-se a utilização de análise de séries temporais de consumo para a seleção de um modelo estatístico apropriado para estabelecer boas previsões mensais de demanda de água potável em ambiente urbano. Os procedimentos adotados para a aplicação dessa metodologia contemplam o ajuste e testes de modelos de previsão de curto prazo, por categoria de consumo, para a cidade de Joinville, Santa Catarina. Neste estudo são desenvolvidos preliminarmente dez modelos de previsão, sendo um modelo de suavização exponencial e um modelo autorregressivo integrado de médias móveis sazonal (SARIMA) para cada uma das categorias analisadas (residencial, comercial, industrial, público e total). Os resultados apontam que os modelos SARIMA tiveram melhor desempenho na previsão nas categorias comercial, industrial, público e total, enquanto o método de suavização exponencial se mostrou mais apropriado para realizar previsões na categoria residencial. Os resultados obtidos na previsão da categoria total indicam que a realização de previsões considerando um consumo urbano unificado pode ser mais assertiva.

Orientadora: Andreza Kalbusch

Coorientadora: Elisa Hening

Joinville, 2020

ANO  
2020

DANIELLE CRISTINA MANFRIN | MODELOS DE PREVISÃO DE DEMANDA DE ÁGUA  
PARA A CIDADE DE JOINVILLE UTILIZANDO ANÁLISE DE SÉRIES TEMPORAIS



**UDESC**

UNIVERSIDADE DO ESTADO DE SANTA CATARINA – UDESC  
CENTRO DE CIÊNCIAS TECNOLÓGICAS – CCT  
PROGRAMA DE PÓS GRADUAÇÃO EM ENGENHARIA CIVIL

DISSERTAÇÃO DE MESTRADO

## MODELOS DE PREVISÃO DE DEMANDA DE ÁGUA PARA A CIDADE DE JOINVILLE UTILIZANDO ANÁLISE DE SÉRIES TEMPORAIS

DANIELLE CRISTINA MANFRIN

JOINVILLE, 2020

**DANIELLE CRISTINA MANFRIN**

**MODELOS DE PREVISÃO DE DEMANDA DE ÁGUA PARA A CIDADE DE  
JOINVILLE UTILIZANDO ANÁLISE DE SÉRIES TEMPORAIS**

Dissertação de mestrado em Engenharia Civil,  
apresentado como requisito parcial para  
obtenção do título de mestre na Universidade  
do Estado de Santa Catarina.  
Orientadora: Dra. Andreza Kalbusch.  
Coorientadora: Dra. Elisa Henning.

**JOINVILLE  
2020**

Ficha catalográfica elaborada pelo programa de geração automática da  
Biblioteca Setorial do CCT/UEDESC,  
com os dados fornecidos pelo(a) autor(a)

Ristow, Danielle Cristina Manfrin  
MODELOS DE PREVISÃO DE DEMANDA DE ÁGUA  
PARA A CIDADE DE JOINVILLE UTILIZANDO ANÁLISE DE  
SÉRIES TEMPORAIS / Danielle Cristina Manfrin Ristow. –  
2020.  
146 p.

Orientadora: Andreza Kalbusch  
Coorientadora: Elisa Hening  
Dissertação (mestrado) – Universidade do Estado de  
Santa Catarina, Centro de Ciências Tecnológicas, Programa  
de Pós-Graduação em Engenharia Civil, Joinville, 2020.

1. Previsão de Demanda de Água. 2. Séries Temporais. 3.  
Suavização Exponencial. 4. ARIMA. 5. Categorias de  
Consumo de Água. I. Kalbusch, Andreza. II. Hening, Elisa. III.  
Universidade do Estado de Santa Catarina, Centro de  
Ciências Tecnológicas, Programa de Pós-Graduação em  
Engenharia Civil. IV. Título.

**Modelos de Previsão de Demanda de Água para a Cidade de Joinville  
Utilizando Análise de Séries Temporais**

por

**Danielle Cristina Manfrin**

Esta dissertação foi julgada adequada para obtenção do título de

**MESTRA EM ENGENHARIA CIVIL**

Área de concentração em "Engenharia Urbana e da Construção Civil"  
e aprovada em sua forma final pelo

**CURSO DE MESTRADO ACADÊMICO EM ENGENHARIA CIVIL  
DO CENTRO DE CIÊNCIAS TECNOLÓGICAS DA  
UNIVERSIDADE DO ESTADO DE SANTA CATARINA.**

Banca Examinadora:

---

Profa. Dra. Andreza Kalbusch  
CCT/UDESC  
(Orientadora/Presidente)

---

Prof. Dr. Custodio da Cunha Alves  
UNIVILLE

---

Prof. Dr. Leonardo Romero Monteiro  
CCT/UDESC

Joinville, SC, 21 de maio de 2020.



## **AGRADECIMENTOS**

À Deus por ter me ajudado a superar todos os obstáculos desta caminhada.

À minha filha, Olivia, por ser luz na minha vida e principal força motriz para a conclusão deste trabalho.

Ao meu grande amor, Bernardo, por acreditar em mim e por ser meu porto seguro, me apoiando das mais diversas maneiras para que eu pudesse chegar até aqui.

Aos meus pais que me ensinaram tanto ao longo da vida. Especialmente à minha mãe, minha maior incentivadora.

À minha sogra, Deisi, e à Katia, por cuidarem da minha filha com tanto amor para que eu tivesse tranquilidade para escrever.

Ao meu cunhado, Eduardo, por ser meu amigo e companheiro nos meses em que residi em Joinville.

À minha orientadora Andreza Kalbusch, e coorientadora Elisa Hening por dividirem seus conhecimentos comigo e por toda compreensão que me destinaram.

À todas as competentes e maravilhosas professoras do mestrado de Engenharia Civil da UDESC, que me mostraram a força da união feminina.

À Companhia Águas de Joinville por disponibilizar os dados para a realização deste trabalho, especialmente ao Sr. Almir de Oliveira.

À CAPES: O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.



## RESUMO

Modelos de previsão para a demanda de água potável fornecem fundamentos para a tomada de decisões operacionais, táticas e estratégicas por parte das companhias de abastecimento, além de contribuírem para a melhoria do desempenho dos sistemas de distribuição de água, antecipando os valores de consumo. Neste cenário, prever o consumo de água em áreas urbanas é fundamental para alcançar a excelência no gerenciamento dos recursos hídricos locais. A metodologia proposta neste trabalho envolve a utilização da análise de séries temporais de consumo na seleção de um modelo estatístico apropriado para estabelecer previsões mensais da demanda de água potável em ambiente urbano. Os procedimentos adotados para a aplicação dessa metodologia objetivam contemplar o ajuste e testes de modelos de previsão de curto prazo, por categoria de consumo, para previsão do consumo de água potável para o primeiro trimestre de 2018 na cidade de Joinville, Santa Catarina. Para esse estudo são desenvolvidos preliminarmente dez modelos de previsão, sendo um modelo de suavização exponencial e um modelo autorregressivo integrado de médias móveis sazonal (SARIMA) para cada uma das categorias analisadas (residencial, comercial, industrial, público e total). A avaliação da qualidade dos modelos de previsão é fundamentada na análise dos resíduos obtidos, considerando o menor erro percentual absoluto médio (MAPE) como medida estatística de desempenho e definição de modelos mais apropriados. Os resultados apontam que os modelos SARIMA tiveram melhor desempenho na previsão nas categorias comercial, industrial, público e total, enquanto o método de suavização exponencial se mostrou mais apropriado para realizar previsões na categoria residencial. Os resultados obtidos na previsão da categoria total indicam que a realização de previsões considerando o consumo urbano unificado pode ser mais assertiva.

**PALAVRAS-CHAVE:** Previsão de Demanda de Água. Séries Temporais. Suavização Exponencial. ARIMA. Categorias de Consumo de Água. Joinville.





## ABSTRACT

Water demand forecasting models provide the basis for Utilities to make operational, tactical and strategic decisions, and can help improve the performance of water distribution systems, by anticipating consumption values. Thus, predicting consumption in urban areas is important to achieve excellence in local water resources management. The proposed methodology involves the analysis of consumption time series to select of an appropriate statistical model to establish monthly water demand forecasts in an urban environment. The procedures contemplate the adjustment and testing of short-term forecasting models, by consumption category, for the first quarter of 2018 in the city of Joinville, Santa Catarina. Ten forecast models for performance evaluation are preliminarily developed for this study, which include an exponential smoothing model and a seasonal autoregressive integrated moving average model (SARIMA) for each of the residential, commercial, industrial, public and total categories. The evaluation of the quality of the forecasting models is based on the analysis of the residuals obtained, considering the lowest mean absolute percentage error (MAPE) as a statistical measure of performance and definition of more appropriate models. The results show that the SARIMA models had better performance for the commercial, industrial, public and total categories, while the exponential smoothing method was more appropriate in the residential category. The results obtained for the total category forecast indicate that predictions considering unified urban consumption may be more assertive.

**KEYWORDS:** Short-term Water Demand Forecast. Time Series. Exponential Smoothing. ARIMA. Water Consumption Categories. Joinville.



## LISTA DE TABELAS

Tabela 1 – Demandas e impactos das ações de saneamento sobre os recursos hídricos.....	26
Tabela 2 – Relação entre o horizonte e demais características das previsões de demanda. ....	32
Tabela 3 – Classificação dos modelos de Suavização.....	38
Tabela 4 – Classificação dos modelos de suavização exponencial. ....	40
Tabela 5 – Representação dos modelos de suavização exponencial sem sazonalidade por meio de modelos de espaço de estado.....	41
Tabela 6 – Identificação de modelos por meio das funções de autocorrelação e autocorrelação parcial. ....	51
Tabela 7 – Evolução do PIB de Joinville, de 2012 a 2015. ....	67
Tabela 8 – Estatísticas descritivas da série temporal de consumo da categoria residencial. ....	81
Tabela 9 – Parâmetros de suavização e erro padrão do modelo RSE.....	82
Tabela 10 – Erros e AIC obtidos no modelo RSE. ....	83
Tabela 11 – Coeficientes do modelo RA. ....	84
Tabela 12 – Erros e AIC obtidos no modelo RA.....	86
Tabela 13 – Estatísticas descritivas da série temporal de consumo da categoria comercial.....	87
Tabela 14 – Parâmetros de suavização e erro padrão do modelo CSE.....	89
Tabela 15 – Erros e AIC obtidos no modelo CSE. ....	90
Tabela 16 – Coeficientes do modelo CA. ....	91
Tabela 17 – Erros e AIC obtidos no modelo CA.....	91
Tabela 18 – Estatísticas descritivas da série temporal de consumo da categoria industrial.....	94
Tabela 19 – Parâmetros de suavização e erro padrão do modelo ISE. ....	95
Tabela 20 – Erros e AIC obtidos no modelo ISE.....	96
Tabela 21 – Coeficientes do modelo IA.....	97
Tabela 22 – Erros e AIC obtidos no modelo IA. ....	98
Tabela 23 – Estatísticas descritivas da série temporal de consumo da categoria público.....	100

Tabela 24 – Parâmetros de suavização e erro padrão do modelo PSE. ....	101
Tabela 25 – Erros e AIC obtidos no modelo PSE. ....	102
Tabela 26 – Coeficientes do modelo RA.....	103
Tabela 27 – Erros e AIC obtidos no modelo PA. ....	104
Tabela 28 – Estatísticas descritivas da série temporal de consumo da categoria total. .....	105
Tabela 29 – Parâmetros de suavização e erro padrão do modelo TSE. ....	106
Tabela 30 – Erros e AIC obtidos no modelo TSE. ....	107
Tabela 31 – Coeficientes do modelo TA.....	108
Tabela 32 – Erros e AIC obtidos no modelo TA.....	109
Tabela 33 – Resumo dos resultados obtidos na modelagem. ....	110
Tabela 34 – Consumos de água residencial previstos e observados em Joinville, para o ano de 2018.....	114
Tabela 35 – Consumos de água comercial previstos e observados em Joinville, para o ano de 2018.....	116
Tabela 36 – Consumos de água industrial previstos e observados em Joinville, para o ano de 2018.....	118
Tabela 37 – Consumos de água pública previstos e observados em Joinville, para o ano de 2018.....	120
Tabela 38 – Consumos de água total previstos e observados em Joinville, para o ano de 2018.....	121

## LISTA DE ILUSTRAÇÕES

Figura 1 – Proporção entre população rural e urbana no Brasil.....	25
Figura 2 – Características da demanda de água urbana. ....	27
Figura 3 – Caminho para realização de previsões de demanda de água. ....	31
Figura 4 – Série temporal do consumo residencial de água em Joinville de janeiro de 2007 a dezembro de 2017. ....	36
Figura 5 – Série temporal de dados de consumo de água. a) Série original. b) série estacionária obtida após a diferenciação da sazonalidade anual.....	36
Figura 6 – Representação gráfica para os modelos de suavização exponencial.....	39
Figura 7 – Performance de previsão do modelo combinado. ....	42
Figura 8 – Previsão obtida para o modelo Holt-Winters aditivo.....	43
Figura 9 – Comparação da demanda de água prevista versus observada usando o melhor modelo ARIMA para previsão de um dia à frente durante o período de testes (a) verão de 2008 e (b) verão de 2009.....	53
Figura 10 – Resultado da previsão do modelo ARMA de fevereiro de 2009 a outubro de 2009. ....	54
Figura 11 – Resultados da previsão do modelo ARIMA para a demanda real e prevista. ....	54
Figura 12 – Previsão do modelo ARIMA para demanda de água. a) Demanda real versus demanda do modelo. b) Demanda prevista no intervalo de 95% de confiança. ....	55
Figura 13 – Comparação entre os resultados de previsão diária do modelo SARIMA, demanda real e produção de água potável. ....	56
Figura 14 – Etapas metodológicas. ....	58
Figura 15 – Gráfico da função de autocorrelação de resíduos.....	62
Figura 16 – Gráfico quantil-quantil. ....	62
Figura 17 – Localização e divisão política e administrativa do município de Joinville. ....	65
Figura 18 – Evolução da ocupação urbana em Joinville de 1860 a 2011. ....	66
Figura 19 – Estimativa populacional de Joinville de 2018 a 2030. ....	67
Figura 20 – Distribuição da precipitação total anual no estado de Santa Catarina. ...	68
Figura 21 – Distribuição da temperatura média anual no estado de Santa Catarina.	69

Figura 22 – Bacias hidrográficas de Joinville.....	70
Figura 23 – Distribuição espacial da rede de abastecimento no perímetro urbano de Joinville. ....	71
Figura 24 – ETA/Cubatão. ....	72
Figura 25 – ETA/Piraí. ....	72
Figura 26 – População atendida pelo abastecimento de água da CAJ em Joinville entre 2004 e 2016.....	73
Figura 27 – Número de economias de água ativas registrado pela CAJ em Joinville entre 2004 e 2016.....	73
Figura 28 – Consumo anual total de água em Joinville, dividido por categoria de consumo, nos anos de 2007 a 2017.....	74
Figura 29 – Consumo mensal de água na categoria residencial em Joinville no período de janeiro de 2013 a dezembro de 2017 .....	75
Figura 30 – Consumo mensal de água na categoria comercial em Joinville no período de janeiro de 2013 a dezembro de 2017. ....	76
Figura 31 – Consumo mensal de água na categoria industrial em Joinville no período de janeiro de 2007 a dezembro de 2017. ....	76
Figura 32 – Consumo mensal de água na categoria público em Joinville no período de janeiro de 2007 a dezembro de 2017. ....	77
Figura 33 – Consumo anual de água por categoria de consumo analisada no período de 2013 a 2017.....	78
Figura 34 – Série temporal do consumo residencial de água em Joinville de janeiro de 2013 a dezembro de 2017.....	80
Figura 35 – Decomposição da série de dados de consumo residencial de água em Joinville. a) Observado. b) Tendência. c) Sazonalidade. d) Ruído Branco.....	81
Figura 36 – Análise dos resíduos do modelo RSE. (a) Gráfico quantil-quantil. (b) ACF. ....	83
Figura 37 – Análise dos resíduos do modelo RA. (a) Gráfico quantil-quantil. (b) ACF. ....	85
Figura 38 – Série temporal do consumo comercial de água em Joinville de janeiro de 2013 a dezembro de 2017.....	86
Figura 39 – Decomposição da série de dados de consumo comercial de água em Joinville. a) Observado. b) Tendência. c) Sazonalidade. d) Ruído Branco.....	87

Figura 40 – Análise dos resíduos do modelo CSE. (a) Gráfico quantil-quantil. (b) ACF.	89
Figura 41 – Análise dos resíduos do modelo CA. (a) Gráfico quantil-quantil. (b) ACF.	92
Figura 42 – Série temporal do consumo industrial de água em Joinville de janeiro de 2013 a dezembro de 2017.	93
Figura 43 – Decomposição da série de dados de consumo industrial de água em Joinville. a) Observado. b) Tendência. c) Sazonalidade. d) Ruído Branco.	93
Figura 44 – Análise dos resíduos do modelo ISE. (a) Gráfico quantil-quantil. (b) ACF.	95
Figura 45 – Análise dos resíduos do modelo IA. (a) Gráfico quantil-quantil. (b) ACF.	97
Figura 46 – Série temporal do consumo público de água em Joinville de janeiro de 2007 a dezembro de 2017.	99
Figura 47 – Decomposição da série de dados de consumo público de água em Joinville. a) Observado. b) Tendência. c) Sazonalidade. d) Ruído Branco.	99
Figura 48 – Análise dos resíduos do modelo PSE. (a) Gráfico quantil-quantil. (b) ACF.	101
Figura 49 – Análise dos resíduos do modelo PA. (a) Gráfico quantil-quantil. (b) ACF.	103
Figura 50 – Série temporal do consumo total de água em Joinville de janeiro de 2013 a dezembro de 2017.	105
Figura 51 – Decomposição da série de dados de consumo total de água em Joinville. a) Observados. b) Tendência. c) Sazonalidade. d) Ruído Branco.	105
Figura 52 – Análise dos resíduos do modelo TSE. (a) Gráfico quantil-quantil. (b) ACF.	107
Figura 53 – Análise dos resíduos do modelo TA. (a) Gráfico quantil-quantil. (b) ACF.	109
Figura 54 – Gráfico de previsão de consumo de água na categoria residencial em Joinville para o modelo RSE.	113
Figura 55 – Curvas de previsão e de demanda de água na categoria residencial em Joinville, em m <sup>3</sup> , pelo modelo RSE, no período de janeiro a março de 2018.	114
Figura 56 – Gráfico de previsão de consumo de água na categoria comercial em Joinville para o modelo CA.	115



Figura 57 – Curvas de previsão e de demanda de água na categoria comercial em Joinville, em m <sup>3</sup> , pelo modelo CA, no período de janeiro a março de 2018.....	116
Figura 58 – Gráfico de previsão de consumo de água na categoria industrial em Joinville para o modelo IA.....	117
Figura 59 – Curvas de previsão e de demanda de água na categoria industrial em Joinville, em m <sup>3</sup> , pelo modelo IA, no período de janeiro a março de 2018.....	117
Figura 60 – Gráfico de previsão de consumo de água na categoria público em Joinville para o modelo PA. ....	118
Figura 61 – Curvas de previsão e de demanda de água na categoria público em Joinville, em m <sup>3</sup> , pelo modelo PA, no período de janeiro a março de 2018. ....	119
Figura 62 – Gráfico de previsão de consumo de água na categoria total em Joinville para o modelo TA. ....	120
Figura 63 – Curvas de previsão e de demanda de água na categoria total em Joinville, em m <sup>3</sup> , pelo modelo TA, no período de janeiro a março de 2018. ....	121

## SUMÁRIO

<b>1 INTRODUÇÃO .....</b>	<b>19</b>
1.1 JUSTIFICATIVA .....	21
1.2 OBJETIVOS .....	22
1.2.1 Objetivos específicos .....	22
1.3 ESTRUTURA DA DISSERTAÇÃO .....	23
<b>2 REVISÃO BIBLIOGRÁFICA .....</b>	<b>24</b>
2.1 AMBIENTE URBANO E A DEMANDA DE ÁGUA .....	24
2.2 PREVISÃO DE DEMANDA URBANA DE ÁGUA .....	28
2.3 MÉTODOS DE MODELAGEM DE PREVISÃO DE DEMANDA .....	33
2.3.1 Análise de séries temporais .....	35
2.3.1.1 <i>Suavização exponencial</i> .....	36
2.3.1.2 <i>Modelos autorregressivos integrados de médias móveis ARIMA</i> .....	43
<b>3 METODOLOGIA .....</b>	<b>57</b>
3.1 COLETA E SISTEMATIZAÇÃO DE DADOS .....	59
3.2 MODELAGEM .....	59
3.3 ACEITAÇÃO DOS MODELOS .....	61
3.4 AVALIAÇÃO DOS MODELOS .....	63
3.5 ESCOLHA DOS MODELOS .....	64
<b>4 ESTUDO DE CASO .....</b>	<b>65</b>
<b>5 RESULTADOS .....</b>	<b>79</b>
5.1 CATEGORIA RESIDENCIAL .....	80
5.1.1 <i>Suavização exponencial</i> .....	81
5.1.2 <i>ARIMA</i> .....	84
5.2 CATEGORIA COMERCIAL .....	86
5.2.1 <i>Suavização exponencial</i> .....	88
5.2.2 <i>ARIMA</i> .....	90
5.3 CATEGORIA INDUSTRIAL .....	92
5.3.1 <i>Suavização exponencial</i> .....	94
5.3.2 <i>ARIMA</i> .....	96
5.4 CATEGORIA PÚBLICO .....	98
5.4.1 <i>Suavização exponencial</i> .....	100

5.4.2 ARIMA .....	102
5.5 CATEGORIA TOTAL .....	104
5.5.1 Suavização exponencial .....	106
5.5.2 ARIMA .....	108
5.6 ESCOLHA DOS MODELOS .....	110
5.7 PREVISÃO .....	113
<b>6 CONSIDERAÇÕES FINAIS.....</b>	<b>123</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>126</b>
<b>APÊNDICE A.....</b>	<b>135</b>
<b>APÊNDICE B.....</b>	<b>137</b>
<b>APÊNDICE C.....</b>	<b>139</b>
<b>APÊNDICE D.....</b>	<b>141</b>
<b>APÊNDICE E .....</b>	<b>143</b>

## 1 INTRODUÇÃO

A Política Nacional de Recursos Hídricos define a água como um recurso natural limitado, de domínio público, essencial à existência e manutenção da vida, e necessário para garantir o bem-estar social e o desenvolvimento socioeconômico de um povo. Por isso, é imprescindível a utilização racional deste recurso para assegurar às atuais e futuras gerações, sua disponibilidade (BRASIL, 1997).

Segundo Setti *et al.* (2001) a água é necessária para a realização de diversas atividades humanas, especialmente no ambiente urbano. Dessa forma, entende-se que o maior impacto do consumo de água nas cidades esteja vinculado à questão ambiental. Ou seja, o abastecimento pode ser entendido tanto como demanda de recursos hídricos, para a produção de água potável, como impacto, quando se considera seu esgotamento.

Para gerenciamento sustentável do uso da água é preciso analisar a demanda de água de um determinado local em suas diversas perspectivas (BRASIL, 2006). Diversos episódios de escassez de água, relacionados ao grande crescimento populacional e às mudanças climáticas, têm colocado o gerenciamento de recursos hídricos na agenda global, e em várias regiões do planeta, satisfazer o crescimento da demanda de água tem sido um desafio (HOFFMANN; WORTHINGTON; HIGGS, 2006; ADAMOWSKI; KARATAPAKI, 2010; WENTZ; GOBBER, 2007; ROMANO; SALVATI; GUERRINI, 2014). Como consequência é possível perceber um interesse maior na implementação de estratégias sustentáveis para a exploração desse recurso. Para que a sustentabilidade seja possível, é necessário, além de outras medidas, servir-se de um sistema de previsão de demanda que englobe os diferentes setores de consumo, e que seja confiável a curto, médio e longo prazos (SEBRI, 2016).

De acordo com Bai *et al.* (2014) com o crescimento da urbanização e melhoria do padrão de vida, o consumo de água nas cidades tende a aumentar dramaticamente. Esse acréscimo na demanda pode sobrecarregar os sistemas distribuição de água, que devem ser dimensionados de modo a promover um abastecimento apropriado. Uma maneira bastante eficaz de garantir a adequação da operação dos sistemas é a utilização de métodos de previsão de demanda (BAI *et al.*, 2014).

Segundo Donkor *et al.* (2014) existe uma série de métodos de previsão que podem ser aplicados nos modelos de previsão de demanda, no entanto, a escolha do método depende diretamente na quantidade e qualidade dos dados, do horizonte de previsão almejado, e também da disponibilidade de tempo e recursos. Historicamente, os métodos de modelagem mais utilizados pelas empresas de abastecimento têm sido os de análise de regressão, análise de séries temporais ou uma combinação dos dois métodos. É possível encontrar na literatura diversos trabalhos de previsão de demanda a médio prazo fazendo uso dessas técnicas (SANTOS, 2011).

Odan e Reis (2012) listam dois tipos de abordagem para modelar séries temporais: linear e não-linear. A abordagem linear, empregada neste trabalho, faz uso de suavização exponencial e modelos autorregressivos de médias móveis (ARIMA). De acordo com Adamowski e Karapataki (2010) os modelos que fazem uso das séries temporais têm como premissa a dependência linear entre os valores futuros e passados. Esses modelos não consideram outras variáveis, como climáticas ou socioeconômicas durante o processo de modelagem já que exploram unicamente a relação entre o consumo atual e o passado.

A utilização dos métodos de modelagem que fazem uso exclusivo de uma determinada série temporal de dados é bastante difundida por possuir duas grandes vantagens: (i) são métodos mais simples entre os disponíveis e baseiam-se na premissa de que as tendências históricas do comportamento da série se mantêm ao longo do tempo, ou seja, assume-se uma dependência linear entre os valores futuros e passados (ADAMOWSKI; KARAPATAKI, 2010); (ii) são métodos diretos, que realizam as previsões sem considerar outras variáveis externas, que muitas vezes podem ser complexas e incertas (ZHAI *et al.*, 2012).

Apesar da relativa simplicidade dos métodos de análise de séries temporais, quando comparados aos métodos mais modernos, como as redes neurais, Billings e Jones (2008) afirmam que métodos mais complexos ou sofisticados nem sempre mostram melhor desempenho, já que aumentar a quantidade de informação disponível para o previsor, não necessariamente aumenta a qualidade da previsão, segundo os autores.

## 1.1 JUSTIFICATIVA

A previsão de demanda de água é uma questão bastante relevante para as companhias de abastecimento ao redor do mundo. De acordo com Billings e Jones (2008) ela fornece base para a tomada de decisões operacionais, táticas e estratégicas, e pode ajudar a melhorar a performance dos sistemas de distribuição de água, antecipando os valores de consumo. Entretanto, prever o consumo de água é uma tarefa desafiadora. Os padrões de demanda mostram variações diárias, semanais e sazonais, que também sofrem influência direta de indicadores socioeconômicos e fatores meteorológicos, como as características da população consumidora, grau de industrialização, temperatura do ar e nível de precipitação (ROMANO; KAPELAN, 2014).

Conforme Rinaudo (2015) as previsões realizadas podem ser de curto, médio e longo prazo, dependendo do interesse das companhias de abastecimento. Previsões de curto prazo visam antecipar a demanda de água nas próximas horas, dias, semanas ou meses, objetivando otimizar a operação de sistemas de água, estimar as receitas de vendas e planejar despesas de curto prazo. A previsão de médio prazo é usada como parte integrante do processo de planejamento econômico, com a aplicação dos resultados direcionada para a definição dos orçamentos anuais e novos investimentos (SMITH, 2003). A previsão de longo prazo fornece informações para o planejamento de construção de infraestruturas de abastecimento de água de longa duração, como definição de novas fontes de captação, armazenamentos ou transferências entre bacias de grande capacidade (RINAUDO, 2015).

A realização de modelos de previsão, de maneira geral, visa facilitar a compreensão do comportamento de consumo de água (STOKER; ROTHFEDER, 2014) e colaborar para o refinamento do planejamento econômico (DEWS, 2010), bem como, auxiliar no desenvolvimento de estratégias que visem a economia de água (BABEL; GUPTA, PRADHAN, 2007), de energia e a destinação adequada dos efluentes (HERRERA *et al.*, 2010).

Além disso, desenvolver modelos de previsão considerando apenas os dados de consumo facilita a replicação dos modelos em outras localidades, tanto no Brasil, como no mundo, porque não considera outras variáveis que influenciam o consumo,

que muitas vezes são inconsistentes, complexas (ZHAI *et al.*, 2012), e específicas de cada local (BABEL; SHINDE, 2011).

## 1.2 OBJETIVOS

O objetivo deste trabalho é elaborar uma metodologia para a aplicação e escolha de um modelo adequado para a realização de previsões mensais de demanda de água de curto prazo no ambiente urbano, por meio de análise de séries temporais de consumo.

### 1.2.1 Objetivos específicos

Os objetivos específicos deste trabalho são:

- Elaborar uma metodologia para a modelagem da série de dados de consumo de água pelo método de suavização exponencial e por modelos autorregressivos integrados de médias móveis (ARIMA);
- no estudo de caso, criar e testar modelos de previsão de demanda de água, para as diferentes categorias de consumo, por meio de suavização exponencial;
- no estudo de caso, criar e testar modelos de previsão de demanda de água, para as diferentes categorias de consumo por meio de modelos ARIMA;
- avaliar qual modelo, entre os adotados, apresenta menor erro na previsão;
- realizar previsões de demanda de água para a cidade de Joinville utilizando os modelos escolhidos para cada categoria.

### 1.3 ESTRUTURA DA DISSERTAÇÃO

O trabalho está estruturado em seis capítulos, incluindo a presente introdução. O capítulo 2 apresenta o referencial teórico que inclui a revisão de literatura sobre os modelos estatísticos propostos para estabelecer a previsão de demanda urbana de água tratada via análise de séries temporais. As previsões envolvem a utilização de modelos de suavização exponencial e modelos autorregressivos integrados de média móvel (ARIMA), além de complementos básicos fundamentais que auxiliam nas interpretações e análise inerentes a estas técnicas de previsão. Além disso, há uma abordagem sobre demandas de águas urbanas e impactos de ações de saneamento sobre os recursos hídricos relacionados à expansão urbana. No capítulo 3 é apresentada a metodologia utilizada cuja sistemática de aplicação consiste em selecionar o modelo estatístico de previsão apropriado para obter a melhor previsão de demanda. O capítulo 4 apresenta o estudo de caso realizado na empresa pública Companhia Águas de Joinville (CAJ) do ramo tratamento de água responsável pelo abastecimento de água na cidade de Joinville. No capítulo 5 estão os resultados obtidos e, na sequência, a escolha dos modelos e as previsões. Finalmente, o capítulo 6 apresenta as considerações finais.



## 2 REVISÃO BIBLIOGRÁFICA

Para entender a importância de se realizar modelos de previsão de demanda de água é necessário analisar as interações entre a ocupação urbana e a demanda de água, além dos métodos de previsão a serem aplicados.

### 2.1 AMBIENTE URBANO E A DEMANDA DE ÁGUA

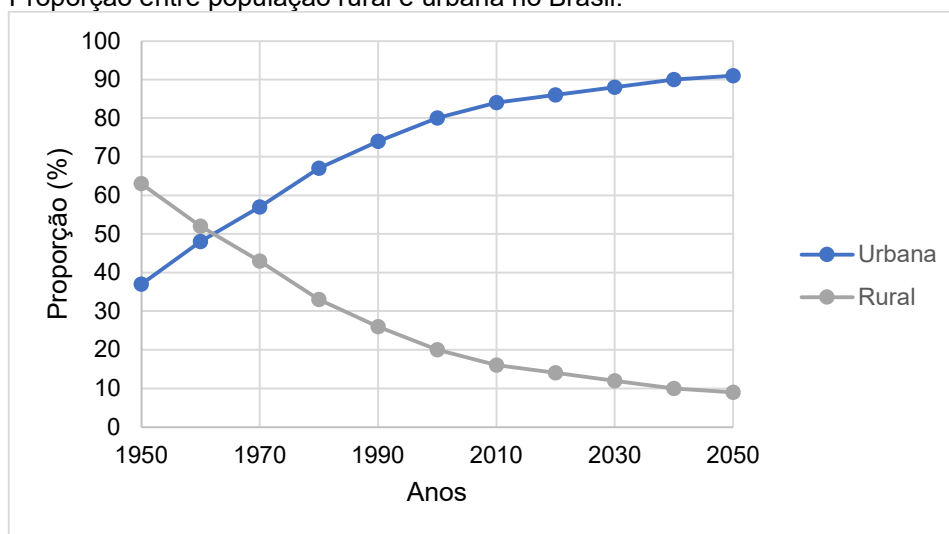
De acordo com Cabral *et al.* (2013), o ambiente urbano pode ser conceituado como um sistema complexo de interação entre indústria, habitação e pessoas, representado por variáveis e fluxos. É possível afirmar que o ambiente urbano se tornou o novo habitat da humanidade, com mais da metade da população mundial vivendo em áreas urbanas pela primeira vez na história (STIGSDOTTER *et al.*, 2017).

A expansão urbana no mundo em desenvolvimento tem sido dramática. Entre 1950 e 2014, o total da população urbana nos países em desenvolvimento aumentou dez vezes, de cerca de 300 milhões para 3 bilhões (ONU, 2014). Considerando a população mundial, passou de 17% para 53,3%. Esse processo tende a continuar. Atualmente mais de 85% da população brasileira vive na área urbana e esse valor deve chegar a 91% até 2050 (Figura 1). Nesse mesmo ano, espera-se que 66,4% da população mundial esteja vivendo em cidades (ONU, 2014).

No Brasil a migração da população rural para as áreas urbanas, em busca de melhores condições de vida, vem ocorrendo nos últimos 60 anos, e pode ser entendida como reflexo da injeção de capital – e consequente mecanização – na agricultura, do aumento de oportunidades de emprego nos setores da indústria, comércio e serviços (PERZ, 2000).

Como resultado dessa migração, observa-se a concentração da população em espaços reduzidos, produzindo grande competição pelos mesmos recursos naturais (solo e água), ocasionando a ocupação irregular das cidades e por conseguinte, problemas urbanísticos (PARKINSON; GOLDENFUM; TUCCI, 2010).

Figura 1 – Proporção entre população rural e urbana no Brasil.



Fonte: adaptada de ONU (2014).

Segundo Tucci (2008), os principais indicadores do desenvolvimento urbano são relativos à população (taxa de crescimento, migração e densificação urbana), aos aspectos econômicos (renda, produto bruto e perfil de produção) e à ocupação do solo (distribuição por tipo de uso do espaço urbano em residencial, comercial e industrial, áreas públicas). Todos esses indicadores dependem diretamente da disponibilidade de água potável, fator limitante para o desenvolvimento de uma sociedade (BRASIL, 2006).

Conforme a Agência Nacional de Águas (ANA, 2017), considerando que a demanda de água no Brasil deve ter um aumento de 30% até 2030, se faz necessária a aplicação de métodos e sistemas alternativos para garantir o desenvolvimento sustentável, equilibrando a oferta e demanda de água. A taxa de estresse hídrico experimentada por 31 países fica entre 25% e 70%, o que mostra a necessidade de se buscar soluções para resolver esse problema. As soluções sugeridas para redução do consumo vão desde a utilização de equipamentos economizadores até a busca de fontes alternativas de abastecimento (BERTOLAZZI & CUSTÓDIO, 2020).

As águas urbanas são as demandas resultantes da ocupação das cidades, com impacto direto nos recursos hídricos locais, englobam o sistema de abastecimento de água e esgotos sanitários, a limpeza pública e drenagem pluvial, como mostra a Tabela 1 (BRASIL, 2006).

Dentre as ações de saneamento citadas, encontra-se o objeto de estudo deste trabalho: o abastecimento urbano de água. No Brasil, a categorização do consumo urbano de água é realizada conforme o art. 47 do Decreto nº 7.217 (BRASIL, 2010),

baseada na estrutura de remuneração e de cobrança dos serviços de água, distribuídas por faixas ou quantidades crescentes de utilização ou de consumo (BEZ, 2012).

De maneira geral consideram-se quatro categorias de consumo: residencial, comercial, industrial e público (TSUTIYA, 2013). A determinação dos usuários dentro de cada categoria de consumo fica a cargo das empresas de abastecimento (BRASIL, 2010).

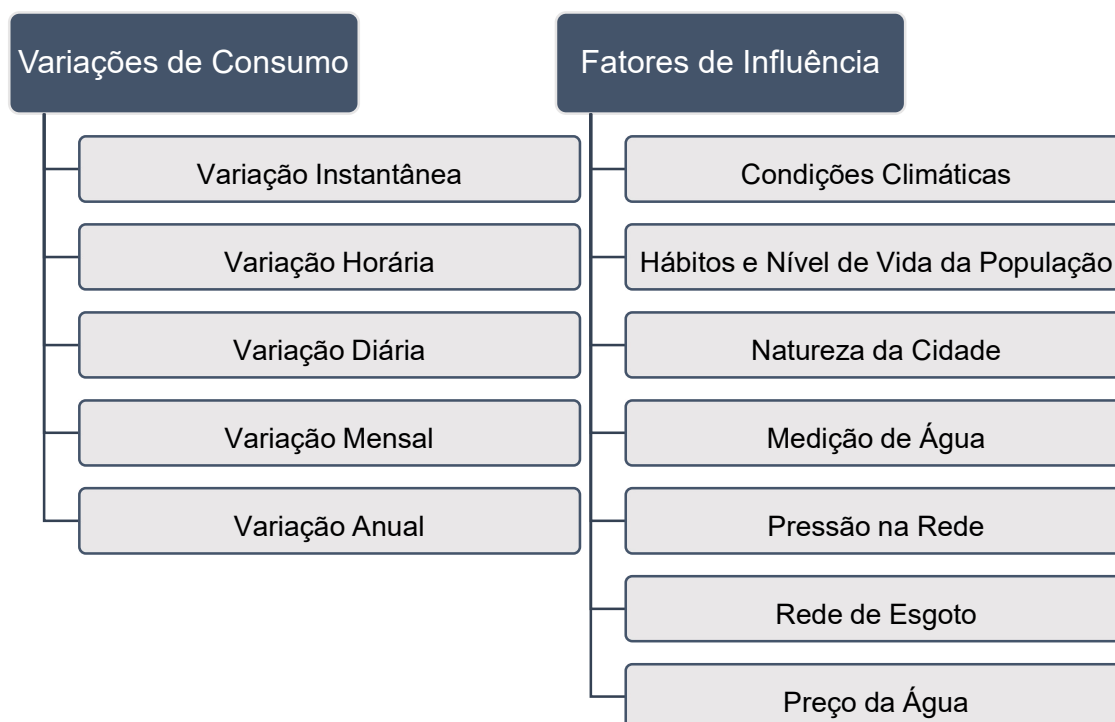
Tabela 1 – Demandas e impactos das ações de saneamento sobre os recursos hídricos.

<b>Ação de Saneamento</b>	<b>Demanda sobre os recursos hídricos</b>	<b>Impacto sobre os recursos hídricos</b>
Abastecimento de água	Grande demanda para abastecimento urbano.	Impacto devido às atividades desenvolvidas no processo (retirada, resíduos de tratamento, etc.).
Esgotamento sanitário	Pequenas demandas para operação e manutenção dos sistemas, porém a mesma demanda do abastecimento de água, ao se considerar o esgotamento sanitário como o “esgotamento do abastecimento de água, após sua utilização”.	Impactos potencialmente elevados, sendo função do sistema de esgotamento sanitário minimizá-los por meio da disposição adequadas dos efluentes.
Limpeza pública	Pequenas demandas para operação e manutenção dos sistemas.	Impactos potencialmente elevados, sendo função do manejo dos resíduos sólidos minimizá-los por meio da disposição adequada dos efluentes.
Drenagem pluvial	Pequenas demandas para operação e manutenção dos sistemas.	Impactos pela elevação da concentração das águas pluviais e fluviais.

Fonte: Brasil (2006).

De acordo com Tsutiya (2013), em uma observação mais ampla, levando em conta o consumo agrupado dessas categorias, duas considerações precisam ser feitas a respeito da demanda de água urbana. São elas: os tipos de variação de consumo e os fatores de influência (Figura 2).

Figura 2 – Características da demanda de água urbana.



Fonte: adaptado de Tsutiya (2013).

Billings e Jones (2008) também nomeiam os principais componentes da demanda dos sistemas de abastecimento. São eles: a demanda residencial de água, que engloba o consumo interno e externo nas residências e leva em consideração as horas de maior consumo no dia; as demandas comerciais e industriais; e as perdas na distribuição.

Em todo mundo sabe-se que entender e gerenciar a água no contexto urbano é um fator crítico para a conquista da sustentabilidade, portanto, um problema de vital importância (MARLOW *et al.*, 2013). Segundo Daniell *et al.* (2015) considera-se a gestão da água nas cidades, juntamente com seus sistemas de energia, alimentação, materiais, sistemas ambientais e sistemas socioeconômicos, um componente importante dentro dos desafios globais de sustentabilidade. Para se obter uma boa qualidade de vida nos ambientes urbanos, é necessário que se tenha água em quantidade e qualidade adequadas para suprir uma série de necessidades, como o bem-estar social, desenvolvimento econômico e saúde ambiental (DANIELL *et al.*, 2015).

De acordo com Bach *et al.* (2014), é possível observar um aumento do reconhecimento da complexidade inerente ao ambiente urbano. Daniell *et al.* (2015) afirmam que é preciso ponderar que o gerenciamento dos sistemas de abastecimento urbano é multifacetado. Além disso, os autores apontam que é preciso caminhar para um gerenciamento combinado das questões de armazenamento, demanda, uso, qualidade, sistemas de saneamento, gerenciamento de chuvas, cheias e os ecossistemas dependentes.

Objeto de análise deste trabalho, a avaliação da demanda de água é uma área de crescente interesse entre planejadores urbanos e gestores de água. A modelagem desses elementos é uma ferramenta útil para compreender a interação entre os fatores humanos e naturais do ambiente urbano que impulsionam o consumo de água (SHARVELLE *et al.*, 2017).

Segundo Zhai *et al.* (2012), recentemente, a previsão de demanda de água tornou-se um componente essencial no planejamento e gerenciamento eficaz dos recursos hídricos porque fornece uma visão simulada do futuro e contribui para identificar a alternativa de gestão adequada ao equilíbrio entre a oferta e a demanda de água.

## 2.2 PREVISÃO DE DEMANDA URBANA DE ÁGUA

Uma das primeiras pesquisas relacionadas à previsão de demanda de água foi realizada para a cidade de Birmingham, no Reino Unido, pelo engenheiro James Mansergh em 1892. A justificativa do estudo foi que as oportunidades percebidas na época para o desenvolvimento de novas fontes locais eram limitadas, talvez atendendo não mais do que outros 10 ou 20 anos e as fontes existentes estavam experimentando um risco crescente de deterioração da qualidade da água (BRADFORD; BRIDGEMAN; GATERELL, 2010).

Segundo Bradford, Bridgeman e Gaterell (2010), enquanto a demanda anual média de abastecimento público de água, na época, era de 73.000.000 litros/dia, com uma demanda de pico no verão de 20% acima da média, aumentando em mais 14.000.000 litros/dia, a produção máxima das fontes locais de águas superficiais e

subterrâneas era de 87.000.000 litros/dia. A demanda crescia em 3% ao ano e o limite de oferta estava próximo de zero.

Apresentados ao parlamento britânico os resultados da análise de Mansergh, explicando os indicadores de crescimento da demanda, foram considerados razoáveis, e assim, os reservatórios de captação Elan Valley e um esquema de aquedutos, localizados na região central do País de Gales foi aprovado. Projetado e construído pela a Birmingham Corporation, desde então, o esquema fornece um suprimento de água abundante, confiável e barato (BRADFORD; BRIDGEMAN; GATERELL, 2010).

Historicamente, no entanto, a prioridade dos governos ao redor do planeta era desenvolver o abastecimento e aumentar a porcentagem de lares conectados à rede, ignorando a evolução da demanda de maneira geral, especialmente nos anos que se seguiram à Segunda Guerra Mundial (BILLINGS; JONES, 2008).

Conforme Rinaudo (2015), como a água era um insumo de grande disponibilidade e baixo custo, as infraestruturas de saneamento eram superdimensionadas de maneira a evitar qualquer crise de abastecimento.

A necessidade de se otimizar o planejamento dessas infraestruturas surgiu nas décadas de 1970 e 1980, com o aumento do custo marginal para desenvolver novas fontes de captação e declínios não antecipados de consumo. O objetivo dessa otimização foi evitar sistemas de abastecimento superestimados e barrar o custo crescente dos erros de previsão (RINAUDO, 2015).

De acordo com Beecher e Chestnut (2012), a construção de instalações de grandes dimensões com base em uma estimativa de demanda de "limite superior" levaria a encargos de custo extra significativos que teriam que ser transferidos para os clientes por meio de aumentos tarifários. Por outro lado, a subestimação da demanda futura de água resultaria em escassez (BILLINGS; JONES, 2008).

Ao longo do tempo, outros trabalhos relacionados à análise de demanda de água foram realizados. Howe e Linaweaver (1967), Maidment e Parzen (1984), e Billings e Agthe (1998) estimaram a demanda de água por métodos de regressão. Smith (1988), Jowitt e Xu (1992), e Zhou *et al.* (2000) criaram modelos de previsão de demanda analisando séries temporais de consumo. Na década seguinte Jain, Vershney e Joshi (2001) e Ghiassi, Zimbra e Saidane (2008) fizeram uso de inteligência artificial para determinar as demandas futuras, aplicando modelos de previsão de redes neurais artificiais (RNA).

Segundo Merrett (2004), são quatro as perspectivas quando se fala de previsão de demanda de água: uso, consumo, necessidade e demanda econômica.

- a) O uso de água é o primeiro conceito e o de mais fácil compreensão. De uma maneira genérica, tabulam-se os usuários regionais de água, dividindo-os em classes de consumo: residencial, comercial, industrial, rural, serviços públicos e outros usos. Em cada caso observa-se o volume de água por unidade de tempo que chega à propriedade de um usuário (uma habitação, uma fazenda, uma mina, uma fábrica, um hospital, um escritório, etc.);
- b) O segundo conceito é sobre o consumo. Durante o curso do uso da água, parte dela retorna ao ciclo hidrológico por intermédio da evapotranspiração. O que resta após a conclusão do processo de uso são as águas residuais e os retornos de irrigação. Esses fluxos são reutilizados internamente ou externamente por grupos de usuários, ou reciclados como escorrência para águas superficiais ou recarga das águas subterrâneas. O consumo de água é de particular interesse porque a proporção de consumo para águas residuais mais os retornos variam significativamente entre os tipos de uso;
- c) O terceiro dos conceitos é a necessidade da água. Para famílias, a necessidade de água é um conceito social, cultural e relacionado à saúde; refere-se a níveis desejáveis ou recomendados de quantidade e qualidade. No setor agrícola, a necessidade de água refere-se aos requisitos de irrigação líquida. Para os setores de mineração, fabricação e serviços, a necessidade de água é determinada tecnicamente e institucionalmente. Portanto, a diferença central entre o uso da água e a necessidade de água é que o primeiro representa um fluxo real para os usuários, enquanto o último representa um fluxo desejado ou recomendado;
- d) O quarto conceito é a demanda econômica da água. Para os economistas, o conceito de "demanda" refere-se ao relacionamento, em um determinado momento e dentro de um mercado definido, entre o preço por unidade de qualquer produto ou serviço e a quantidade, em cada período, que os usuários estão dispostos a comprar e pagar. O contexto cultural e econômico de qualquer função de demanda é referido como as condições de demanda. A primeira dessas condições é o hábito do usuário. O segundo é o preço, a qualidade e a disponibilidade de *commodities* que os usuários consideram como substitutos do produto. Estes dois juntos devem explicar a vontade dos

usuários de comprar. A terceira condição é a renda, os ativos e o acesso ao crédito dos usuários, que contabilizam a capacidade de compra.

Considerando essas perspectivas, o conceito aplicado neste trabalho para a demanda de água foi o de necessidade (MERRET, 2004). Ou seja, o interesse foi estudar volume de água necessário para atender às demandas dos diferentes tipos de consumidores por um certo período (BILLINGS; JONES, 2008), um dos principais objetivos de se realizar modelos de previsão de demanda de água (ZHAI *et al.*, 2012).

Conforme Billings e Jones (2008), o caminho para se realizar previsões de demanda de água (Figura 3) deve incluir: a definição do horizonte de projeto; a escolha das variáveis influentes no consumo de água; a determinação do método de modelagem de acordo com a disponibilidade de dados; e a realização da previsão propriamente dita.

Figura 3 – Caminho para realização de previsões de demanda de água.



Fonte: adaptado de Billings e Jones (2008).

Segundo Billings e Jones (2008), a escolha da metodologia (incluindo o horizonte de previsão) está diretamente ligada à finalidade pretendida para os resultados da previsão. Donkor *et al.* (2014) afirmam que a adequação dos modelos aos objetivos por trás da realização de previsões de demanda é função do horizonte



de previsão, que deve ser definido pelo nível de planejamento almejado. Todavia, não há uma nomenclatura conhecidamente estabelecida para esses horizontes. Billings e Jones (2008) e Smith (2003) apresentam diferentes horizontes temporais sobre o que constitui previsões de longo, médio e curto prazo. Este trabalho adota a classificação apresentada por Donkor *et al.* (2014), mostrada na Tabela 2, que categoriza os horizontes temporais, relacionando a periodicidade da previsão, o nível de planejamento e aplicação.

Tabela 2 – Relação entre o horizonte e demais características das previsões de demanda.

<b>Horizonte de Previsão</b>	<b>Período de Previsão</b>	<b>Nível de Planejamento</b>	<b>Aplicação</b>
Curto prazo (menos de 1 ano)	Horário, diário, semanal e mensal	Operacional	Gerenciamento e otimização dos sistemas operacionais
Médio prazo (1 a 10 anos)	Mensal e anual	Tático	Previsão de receita, planejamento de investimentos, melhorias no sistema de tratamento
Longo prazo (mais de 10 anos)	Anual	Estratégico	Expansão da capacidade

Fonte: adaptada de Donkor *et al.* (2014).

Considerando a aplicação dos resultados, Herrera *et al.* (2010) afirmam que as previsões de curto prazo permitem a otimização das regulagens de bombeamento, aumentando a eficiência energética e auxiliam na escolha das fontes de captação mais adequadas, para que se atinjam os padrões de qualidade necessários. Na comparação entre as medições de vazão previstas e as reais, as previsões ajudam a identificar as possíveis falhas de rede como vazamentos e rompimento de tubulações (BRENTAN *et al* 2017), possibilitando o estabelecimento de um gerenciamento de alerta antecipado (TIWARI; ADAMOWSKI, 2015). As previsões de demanda de água de curto prazo também ajudam na tomada de decisões precisas, como quando implementar restrições regulatórias de uso de água em épocas de estresse hídrico ou seca (TIWARI; ADAMOWSKI, 2015).

As previsões de médio prazo concentram-se na variabilidade do consumo de água por uma base de clientes consolidada ou que aumenta lentamente. Considera as mudanças impulsionadas pelos ciclos climáticos, mudanças na composição ou características da base de clientes, ou ciclos econômicos (RINAUDO, 2015). São

comumente desenvolvidas para planejar melhorias no sistema de distribuição e tratamento de água, definição de tarifas (BILLINGS; JONES, 2008), para estimação dos orçamentos anuais e planejar os investimentos de capital (SMITH, 2003).

Previsões de longo prazo têm aplicações diretas nas programações de expansão da capacidade das empresas de abastecimento. São úteis para o planejamento e dimensionamento de novos sistemas de distribuição de água (BRENTAN *et al.*, 2017) – como grandes reservatórios, plantas de dessalinização, transferências entre bacias (RINAUDO, 2015) – e sistemas de esgoto (WU; ZHOU, 2010).

## 2.3 MÉTODOS DE MODELAGEM DE PREVISÃO DE DEMANDA

Segundo Donkor *et al.* (2014), muitos são os métodos de previsão disponíveis hoje, entretanto, a escolha de um método específico está intimamente ligada à disponibilidade de dados, tempo e recursos, bem como ao horizonte de previsão. Conforme Billings e Jones (2008), esses métodos podem variar de simples previsões informais, nas quais os tomadores de decisão julgam que o futuro será como suas lembranças do passado, a modelos formais complexos que exigem muitas variáveis, grandes quantidades de dados e um comprometimento significativo de recursos.

Gardiner e Herrington (2003) e Billings e Jones (2008) classificam os métodos de previsão como: métodos baseados em julgamento subjetivo; análise de séries temporais; e métodos regressivos. Billings e Jones (2008) adicionam ainda à essa classificação, os métodos não paramétricos.

Os métodos baseados em julgamentos subjetivos, de acordo com Gardiner e Herrington (2003), dependem da experiência de um indivíduo ou grupo, e podem ser de natureza totalmente subjetiva ou uma modificação de resultados mais objetivos, derivados de outras abordagens. Esses métodos incluem: acatar a opinião de especialistas, aplicação do método Delphi, construção de cenários alternativos e pesquisas. Podem ser úteis para pequenas empresas de abastecimento, que enfrentam uma mudança lenta nos padrões de consumo (BILLINGS; JONES, 2008).

Os métodos de previsão de séries temporais para demanda de água envolvem a consideração apenas dos dados de consumo e a previsão de valores futuros pela

projeção de tendências passadas (GARDINER; HERRINGTON, 2003). Por meio da análise de séries temporais, pode-se obter desde modelos que calculam apenas médias simples de dados históricos a modelos relativamente complexos de Box-Jenkins e ARIMA que fazem uso de variações passadas na série de dados para projetar seus valores futuros. A vantagem dos modelos de análise de séries temporais é que os únicos dados necessários são os dados históricos sobre a variável que está sendo prevista (BILLINGS; JONES, 2008).

O método regressivo ou causal é aquele em que é feita uma tentativa de prever a variável dependente em função de outras variáveis que, supõe-se, a controlam ou influenciam (GARDINER; HERRINGTON, 2003). Esse método reconhece que a mudança na demanda decorre de muitos fatores, incluindo tarifas, renda familiar, clima, atividade econômica, programas de conservação de água, nível de atividade econômica (emprego ou rotatividade), e características habitacionais (GARDINER; HERRINGTON, 2003; BILLINGS; JONES, 2008; RINAUDO, 2015). O método consiste em estimar a relação estatística entre consumo (a variável dependente) e um conjunto de variáveis explicativas (RINAUDO, 2015).

Os métodos não paramétricos, como redes neurais artificiais (RNA) e lógica *fuzzy*, fazem uso de inteligência artificial e se apresentam como alternativas interessantes aos modelos de regressão usuais. Nesses métodos o programa de estimativa “aprende” sobre as relações entre as variáveis no modelo e usa esse conhecimento para produzir previsões (BILLINGS; JONES, 2008; QI; CHANG, 2011). Pode-se ainda realizar previsões com a aplicação combinada dos métodos descritos anteriormente, buscando maior sinergia entre os resultados (MOHAMED; AL-MUALLA, 2010; QI; CHANG, 2011; RINAUDO, 2015).

Rinaudo (2015) ainda destaca outros métodos de previsão: a análise de demanda unitária, onde o coeficiente de demanda unitária de água é multiplicado pelo número de usuários em cada categoria; a modelagem de microcomponentes, onde é realizada uma simulação dos usos finais de consumidores residenciais; e modelos baseados no uso e ocupação do solo, onde a demanda é avaliada espacialmente usando o coeficiente de demanda unitária por categoria de consumo.

A aceitação dos modelos pode ser feita por meio da análise de resíduos e observação das funções de autocorrelação, do coeficiente de determinação ( $R^2$ ) (MAKRIDAKIS; WHEELWRIGHT; HYNDMAN, 1997). Considera-se como resíduo de um modelo a demanda atual subtraída dos valores estimados pelo modelo potencial

(MAPHYNDMAN; ATHANASOPOULOS, 2013). A avaliação pode ser realizada pela avaliação dos erros de previsão, como o erro quadrático médio (RMSE), o desvio médio absoluto (MAD) e o erro percentual absoluto médio (MAPE) (BILLINGS; JONES, 2008).

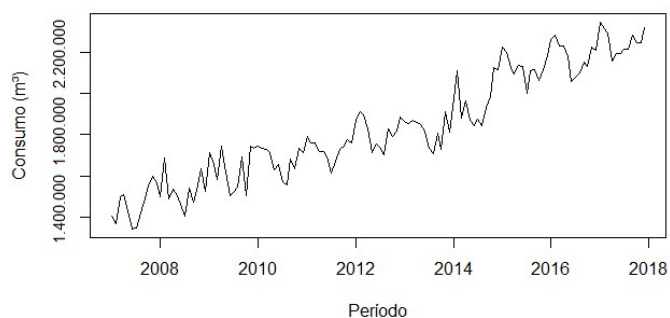
Considerando os métodos disponíveis, pode-se afirmar que talvez um dos maiores desafios da previsão seja combinar com sucesso as técnicas de previsão estatística com os ajustes de julgamento necessários, porque as tendências futuras podem ser diferentes dos padrões passados (GARDINER; HERRINGTON, 2003). Não se deve presumir, porém, que aumentar a quantidade de informação disponível para o previsor necessariamente aumenta a qualidade da previsão. Ou seja, métodos mais complexos ou sofisticados nem sempre apresentam um melhor desempenho, quando comparados aos mais simples (BILLINGS; JONES, 2008).

### 2.3.1 Análise de séries temporais

De acordo com Box, Jenkins e Reinsel (2008), uma série temporal é uma sequência de observações tomadas ao longo do tempo (Figura 4). Apresenta a característica intrínseca de que, tipicamente, as observações adjacentes são dependentes. Podem ser compostas de quatro elementos: tendência, onde se observa o sentido de deslocamento da série; sazonalidade, que descreve um movimento ondulatório ao longo de vários anos e tende a ser periódico; e ruído aleatório, que compreende a variabilidade intrínseca aos dados e que não pode ser modelado (MAKRIDAKIS; WHEELWRIGHT; HYNDMAN, 1997).

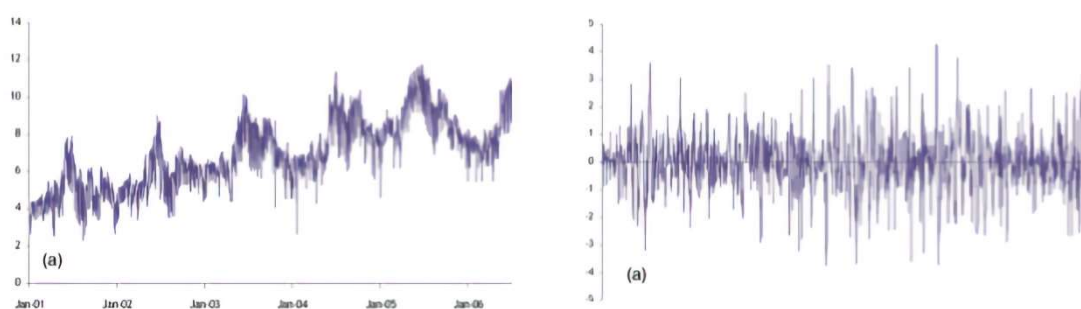
Uma série temporal estacionária é aquela cujas propriedades não dependem do tempo em que a série é observada, ou seja, quando ela se desenvolve no tempo aleatoriamente ao redor de uma média constante. Portanto, as séries temporais com tendências, ou com sazonalidade não são estacionárias – a tendência e a sazonalidade afetarão o valor das séries temporais em diferentes momentos, em uma análise de curto prazo. Para diminuir essa influência, pode-se eliminar a sazonalidade e tendência das séries temporais por meio de diferenciações e transformações matemáticas, como a aplicação de logaritmos (Figura 5) (MAKRIDAKIS; WHEELWRIGHT; HYNDMAN, 1997).

Figura 4 – Série temporal do consumo residencial de água em Joinville de janeiro de 2007 a dezembro de 2017.



Fonte: a autora (2020).

Figura 5 – Série temporal de dados de consumo de água. a) Série original. b) série estacionária obtida após a diferenciação da sazonalidade anual.



Fonte: Caiado (2010).

Conforme Billings e Jones (2008), as séries temporais de consumo de água são utilizadas nos métodos de análise, que incluem suavização exponencial e os modelos autorregressivos de médias móveis (ARIMA), que serão abordados neste trabalho.

### 2.3.1.1 Suavização exponencial

Modelos de previsão realizados por meio de suavização exponencial são bastante utilizados por possuírem duas grandes vantagens. É o método mais simples entre os disponíveis e baseia-se na premissa de que as tendências históricas do comportamento da série se mantêm ao longo do tempo. É um método direto, que

realiza as previsões sem considerar outras variáveis externas, que muitas vezes podem ser complexas e incertas (ZHAI *et al.*, 2012).

Segundo Billah *et al.* (2006), as aplicações de suavização exponencial para modelos de previsão que fazem uso de séries temporais, geralmente, baseiam-se em três métodos básicos: suavização exponencial simples, suavização exponencial dupla, também conhecida como Método de Holt, e o Método de Holt-Winters para séries que possuem variação sazonal.

Conforme Hyndman e Athanasopoulos (2013), as previsões produzidas usando métodos de suavização exponencial são médias ponderadas de observações passadas, com seus pesos decaindo exponencialmente quanto mais antigas forem as observações. Em outras palavras, quanto mais recente a observação, maior o peso associado. Esta estrutura gera previsões confiáveis para um amplo espectro de séries temporais, segundo os autores.

Segundo Miranda (2009), o funcionamento característico dos modelos de suavização exponencial inclui a separação das componentes da série temporal, a suavização de seus valores passados, ou seja, a consideração de pesos diferentes para cada dado, que decaem exponencialmente com o passar do tempo, e posteriormente a recomposição dos elementos para realização das previsões.

As três componentes na suavização exponencial são: nível, tendência e sazonalidade. O nível é o valor médio da observação no período (valor observado excluindo, quando há, a sazonalidade e o erro aleatório), a tendência é diferença sequencial considerando dois níveis consecutivos e a sazonalidade é um episódio que se repete com uma periodicidade constante. Considerando essas componentes, o modelo de suavização exponencial simples, por exemplo, apresenta apenas a componente de nível. O modelo de Holt possui nível e tendência, e o modelo de Holt-Winters contempla as três componentes, apresentando nível, tendência e sazonalidade, que pode ser aditiva ou multiplicativa (MIRANDA, 2009; MAKRIDAKIS; WHEELWRIGHT; HYNDMAN, 1997).

A origem do método de suavização exponencial de séries temporais remonta ao trabalho desenvolvido por Brown durante a Segunda Guerra Mundial, e publicado em 1959. Paralelamente, Holt (1957) desenvolveu um modelo de suavização exponencial similar, que levava em consideração a tendência aditiva das séries. O trabalho de Holt ganhou destaque quando Winters testou o método em um modelo ampliado que considerava, além da tendência, a sazonalidade (MIRANDA, 2009).

Ao longo do tempo o método de suavização exponencial foi ampliado (HYNDMAN; ATHANASOPOULOS, 2013). Pegels (1969) introduziu um modelo com tendência multiplicativa, Gardner Jr. e McKenzie (1985) propuseram um modelo com tendência aditiva amortecida, e Taylor (2003) desenvolveu um modelo com tendência multiplicativa amortecida (MIRANDA, 2009).

Gardner Jr. (2006) revisou os trabalhos de suavização exponencial realizados até então, e propôs equações para produção de modelos espaço de estado que consideram as mudanças de nível, tendência e sazonalidade ao longo do tempo – com uma única fonte de erros, e cujas previsões são pontuais. Esta classificação dos modelos segue o padrão (YZ), onde (i) Y indica o tipo de tendência: N, nenhuma; A, aditiva, Ad, aditiva amortecida; M, multiplicativa, Md, multiplicativa amortecida; e (ii) Z indica o tipo de sazonalidade: N, nenhuma; A, aditiva; M, multiplicativa. A classificação dos modelos, inicialmente sugerida por Taylor (2003), pode ser observada na Tabela 3.

Tabela 3 – Classificação dos modelos de Suavização.

<b>Tendência</b>	<b>Sazonalidade</b>		
	Nenhuma (N)	Aditiva (A)	Multiplicativa (M)
Nenhuma (N)	NN	NA	NM
Aditiva (A)	AN	AA	AM
Aditiva Amortecida (Ad)	AdN	AdA	AdM
Multiplicativa (M)	MN	MA	MM
Multiplicativa Amortecida (Md)	MdN	MdA	MdM

Fonte: adaptado de Taylor (2003).

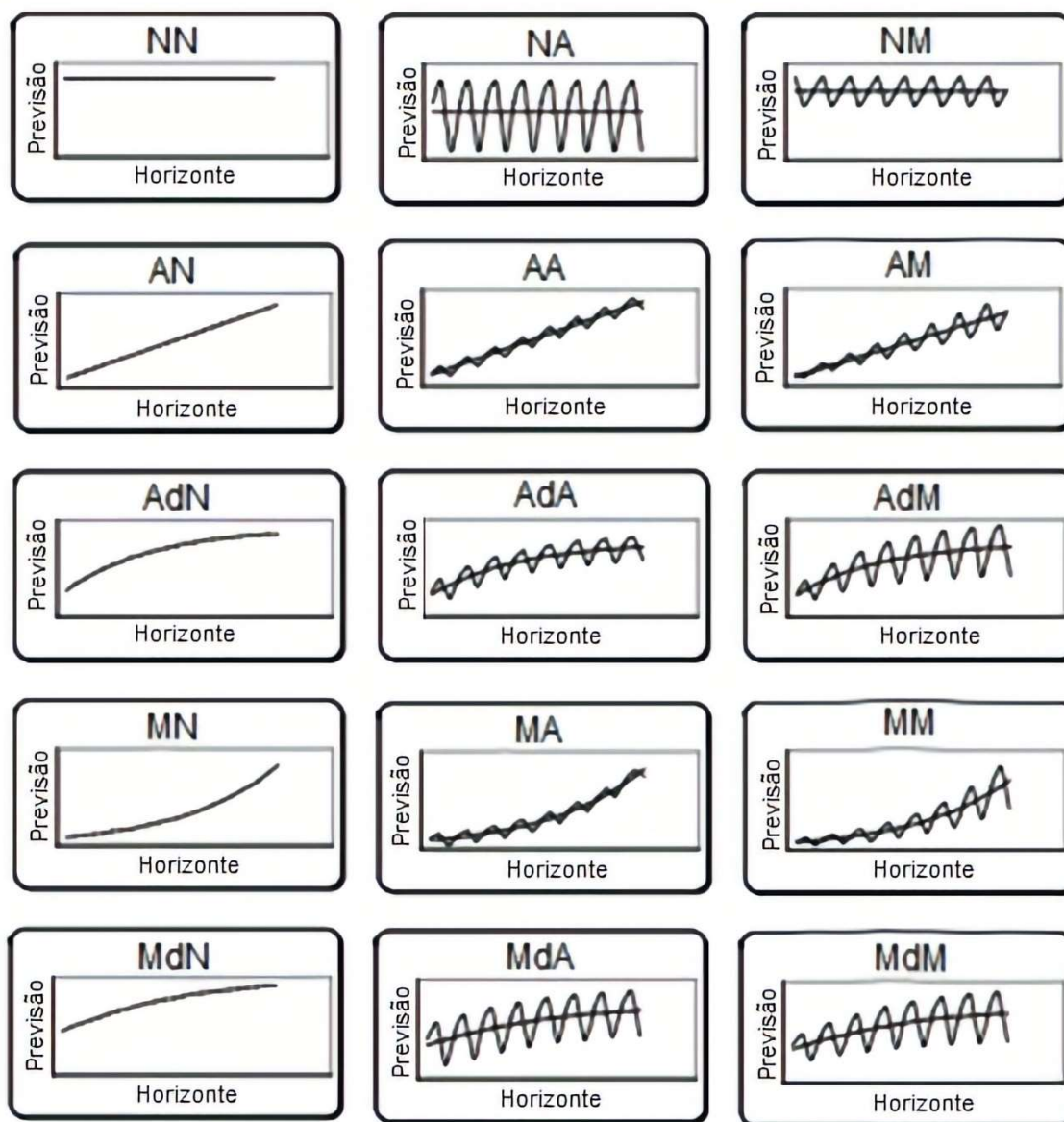
A representação gráfica de todos os modelos descritos pode ser observada na Figura 6, e serve de base para a identificação do modelo mais adequado para a análise de uma determinada série histórica dos dados (MIRANDA, 2009).

Em seu trabalho, Hyndman *et al.* (2005) apresentam modelos estatísticos subjacentes aos métodos exponenciais de suavização. Esses modelos geram previsões de pontos idênticas aos métodos atualizados por Gardner Jr. (2006) mas também geram intervalos de previsão.

Cada modelo consiste em uma equação de medição que descreve os dados observados e algumas equações de transição que descrevem como os componentes ou estados não observados (nível, tendência, sazonalidade) mudam ao longo do tempo. Para cada método existem dois modelos: um com erros aditivos e um com

erros multiplicativos. As previsões pontuais produzidas pelos modelos são idênticas se eles usam os mesmos valores de parâmetros de suavização. Entretanto, gerarão diferentes intervalos de previsão (HYNDMAN; ATHANASOPOULOS, 2013).

Figura 6 – Representação gráfica para os modelos de suavização exponencial.



Fonte: adaptado de Miranda (2009).

Para distinguir entre um modelo com erros aditivos e um com erros multiplicativos (e também distinguir os modelos dos métodos), Hyndman *et al.* (2005) mostram a adição de uma terceira letra à classificação sugerida anteriormente, passando para (XYZ). Onde X agora representa o tipo de erro: sendo A para aditivo; e M para multiplicativo, conforme a Tabela 4.



Tabela 4 – Classificação dos modelos de suavização exponencial.

Sazonalidade						
Tendência	Nenhuma (N)		Aditiva (A)		Multiplicativa (M)	
	Erro					
	Adit. (A)	Mult. (M)	Adit. (A)	Adit. (A)	Mult. (M)	Adit. (A)
Nenhuma (N)	ANN	MNN	ANA	MNA	AMN	MMN
Aditiva (A)	ANA	MNA	AAA	MAA	AAM	MAM
Aditiva Amortecida (Ad)	AAdN	MAdN	AAdA	MAdA	AAdM	MAdM
Multiplicativa (M)	AMN	MMN	AMA	MMA	AMM	MMM
Multiplicativa Amortecida (Md)	AMdN	MMdN	AMdA	MMdA	AMdM	MMdM

Fonte: adaptado de Hyndman *et al.* (2005).

As equações para os modelos de suavização exponencial podem ser na forma padrão, na forma de correção do erro, e na forma de modelos espaço de estado (MIRANDA, 2009).

Um modelo espaço de estado é formado por uma equação mensurável ou observável – que descreve a relação entre o vetor de estado (não observado) e o valor observado – e uma equação chamada de transição (ou estado), que representa como o vetor de estado se modifica ao longo do tempo (MIRANDA, 2009). As equações para os modelos de suavização exponencial de espaço de estado podem ser observadas na Tabela 5.

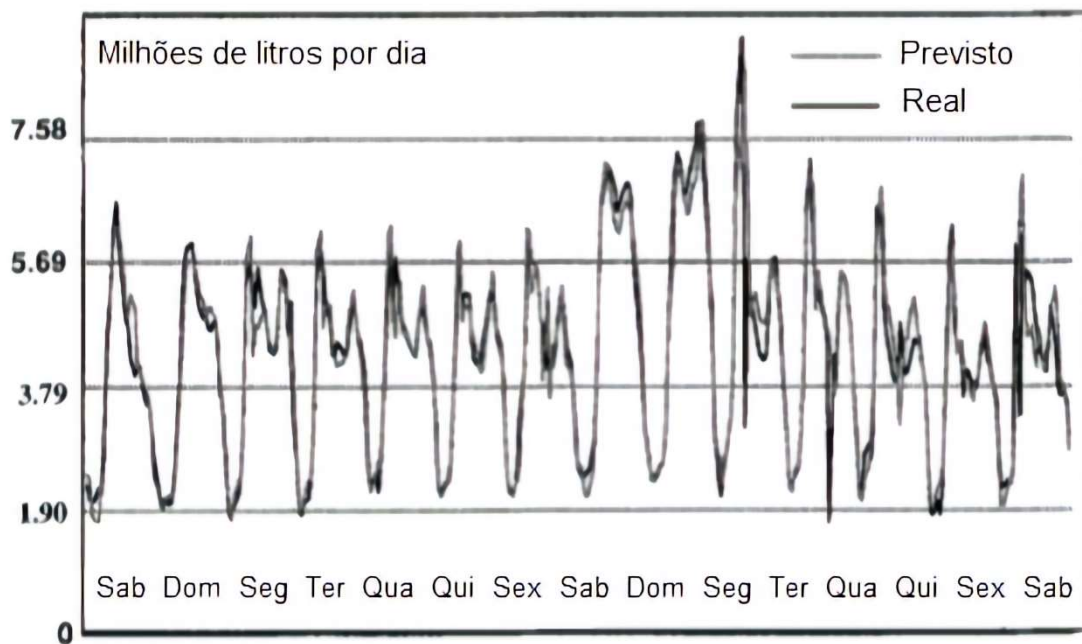
É possível encontrar na literatura alguns modelos de previsão de demanda de água baseados em suavização exponencial. Homwongs, Sastri e Foster III (1994) utilizaram a suavização exponencial de Winters combinada com o método autorregressivo integrado de médias móveis (ARIMA) sazonal para prever a demanda horária de água urbana na cidade americana de Arlington, no Texas. Os resultados (Figura 7) mostraram que a combinação dos métodos produziu um modelo com pequenos erros de previsão, com MAPE abaixo de 10% para as previsões para a próxima hora.

Tabela 5 – Representação dos modelos de suavização exponencial por meio de modelos de espaço de estado.

Tendência	Sazonalidade					
	A			M		
	Erro			Erro		
	A	M	A	M	A	M
N	$e_t = y_t - \mu_t$	$e_t = (y_t - \mu_t)/\mu_t$	$e_t = y_t - \mu_t$	$e_t = (y_t - \mu_t)/\mu_t$	$e_t = y_t - \mu_t$	$e_t = (y_t - \mu_t)/\mu_t$
	$\mu_t = l_{t-1}$	$\mu_t = l_{t-1}$	$\mu_t = l_{t-1} + S_{t-m}$	$\mu_t = l_{t-1} + S_{t-m}$	$\mu_t = l_{t-1} S_{t-m}$	$\mu_t = l_{t-1} S_{t-m}$
	$l_t = \mu_t + \alpha e_t$	$l_t = \mu_t(1 + \alpha e_t)$	$l_t = (\mu_t + \alpha e_t) - S_{t-m}$	$l_t = \mu_t(1 + \alpha e_t) - S_{t-m}$	$l_t = (\mu_t + \alpha e_t)/S_{t-m}$	$l_t = \mu_t(1 + \alpha e_t)/S_{t-m}$
		$S_t = (\mu_t + \gamma e_t) - l_{t-1}$	$S_t = (\mu_t + \alpha e_t) - l_{t-1}$	$S_t = \mu_t(1 + \alpha e_t) - l_{t-1}$	$S_t = (\mu_t + \gamma e_t)/l_{t-1}$	$S_t = \mu_t(1 + \alpha e_t)/l_{t-1}$
A	$\mu_t = l_{t-1} + b_{t-1}$	$\mu_t = l_{t-1} + b_{t-1}$	$\mu_t = l_{t-1} + b_{t-1} + S_{t-m}$	$\mu_t = l_{t-1} + b_{t-1} + S_{t-m}$	$\mu_t = (l_{t-1} + b_{t-1})S_{t-m}$	$\mu_t = (l_{t-1} + b_{t-1})S_{t-m}$
	$l_t = \mu_t + \alpha e_t$	$l_t = \mu_t(1 + \alpha e_t)$	$l_t = (\mu_t + \alpha e_t) - S_{t-m}$	$l_t = \mu_t(1 + \alpha e_t) - S_{t-m}$	$l_t = (\mu_t + \alpha e_t)/S_{t-m}$	$l_t = \mu_t(1 + \alpha e_t)/S_{t-m}$
	$b_t = (\mu_t + \beta^* e_t) - l_{t-1}$	$b_t = \mu_t(1 + \beta^* e_t) - l_{t-1}$	$b_t = (\mu_t + \beta^* e_t) - (l_{t-1} + S_{t-m})$	$b_t = \mu_t(1 + \beta^* e_t) - (l_{t-1} + S_{t-m})$	$b_t = (\mu_t + \beta^* e_t)/S_{t-m} - l_{t-1}$	$b_t = \mu_t(1 + \beta^* e_t)/S_{t-m} - l_{t-1}$
		$S_t = (\mu_t + \gamma e_t) - (l_{t-1} + b_{t-1})$	$S_t = \mu_t(1 + \gamma e_t) - (l_{t-1} + b_{t-1})$	$S_t = (\mu_t + \gamma e_t)/(l_{t-1} + b_{t-1})$	$S_t = (\mu_t + \gamma e_t)/(l_{t-1} + b_{t-1})$	$S_t = \mu_t(1 + \gamma e_t)/(l_{t-1} + b_{t-1})$
Ad	$\mu_t = l_{t-1} + \phi b_{t-1}$	$\mu_t = l_{t-1} + \phi b_{t-1}$	$\mu_t = l_{t-1} + \phi b_{t-1} + S_{t-m}$	$\mu_t = l_{t-1} + \phi b_{t-1} + S_{t-m}$	$\mu_t = (l_{t-1} + \phi b_{t-1})S_{t-m}$	$\mu_t = (l_{t-1} + \phi b_{t-1})S_{t-m}$
	$l_t = \mu_t + \alpha e_t$	$l_t = \mu_t(1 + \alpha e_t)$	$l_t = (\mu_t + \alpha e_t) - S_{t-m}$	$l_t = \mu_t(1 + \alpha e_t) - S_{t-m}$	$l_t = (\mu_t + \alpha e_t)/S_{t-m}$	$l_t = \mu_t(1 + \alpha e_t)/S_{t-m}$
	$b_t = (\mu_t + \beta^* e_t) - l_{t-1}$	$b_t = (\mu_t + \beta^* e_t) - l_{t-1}$	$b_t = (\mu_t + \beta^* e_t) - (l_{t-1} + S_{t-m})$	$b_t = \mu_t(1 + \beta^* e_t) - (l_{t-1} + S_{t-m})$	$b_t = (\mu_t + \beta^* e_t)/S_{t-m} - l_{t-1}$	$b_t = \mu_t(1 + \beta^* e_t)/S_{t-m} - l_{t-1}$
		$S_t = (\mu_t + \gamma e_t) - (l_{t-1} + \phi b_{t-1})$	$S_t = \mu_t(1 + \gamma e_t) - (l_{t-1} + \phi b_{t-1})$	$S_t = (\mu_t + \gamma e_t)/(l_{t-1} + \phi b_{t-1})$	$S_t = \mu_t(1 + \gamma e_t)/(l_{t-1} + \phi b_{t-1})$	$S_t = \mu_t(1 + \gamma e_t)/(l_{t-1} + \phi b_{t-1})$
M	$\mu_t = l_{t-1} b_{t-1}$	$\mu_t = l_{t-1} b_{t-1}$	$\mu_t = l_{t-1} b_{t-1} + S_{t-m}$	$\mu_t = l_{t-1} b_{t-1} + S_{t-m}$	$\mu_t = l_{t-1} b_{t-1} S_{t-m}$	$\mu_t = l_{t-1} b_{t-1} S_{t-m}$
	$l_t = \mu_t + \alpha e_t$	$l_t = \mu_t(1 + \alpha e_t)$	$l_t = (\mu_t + \alpha e_t) - S_{t-m}$	$l_t = \mu_t(1 + \alpha e_t) - S_{t-m}$	$l_t = (\mu_t + \alpha e_t)/S_{t-m}$	$l_t = \mu_t(1 + \alpha e_t)/S_{t-m}$
	$b_t = (\mu_t + \beta^* e_t)/l_{t-1}$	$b_t = \mu_t(1 + \beta^* e_t)/l_{t-1}$	$b_t = [(\mu_t + \beta^* e_t) - S_{t-m}]/l_{t-1}$	$b_t = [\mu_t(1 + \beta^* e_t) - S_{t-m}]/l_{t-1}$	$b_t = (\mu_t + \beta^* e_t)/(l_{t-1} S_{t-m})$	$b_t = \mu_t(1 + \beta^* e_t)/(l_{t-1} S_{t-m})$
		$S_t = (\mu_t + \gamma e_t) - l_{t-1} b_{t-1}$	$S_t = \mu_t(1 + \gamma e_t) - l_{t-1} b_{t-1}$	$S_t = (\mu_t + \gamma e_t)/(l_{t-1} b_{t-1})$	$S_t = \mu_t(1 + \gamma e_t)/(l_{t-1} b_{t-1})$	$S_t = \mu_t(1 + \gamma e_t)/(l_{t-1} b_{t-1})$
Md	$\mu_t = l_{t-1} b_{t-1}^\phi$	$\mu_t = l_{t-1} b_{t-1}^\phi$	$\mu_t = l_{t-1} b_{t-1}^\phi + S_{t-m}$	$\mu_t = l_{t-1} b_{t-1}^\phi + S_{t-m}$	$\mu_t = l_{t-1} b_{t-1}^\phi S_{t-m}$	$\mu_t = l_{t-1} b_{t-1}^\phi S_{t-m}$
	$l_t = \mu_t + \alpha e_t$	$l_t = \mu_t(1 + \alpha e_t)$	$l_t = (\mu_t + \alpha e_t) - S_{t-m}$	$l_t = \mu_t(1 + \alpha e_t) - S_{t-m}$	$l_t = (\mu_t + \alpha e_t)/S_{t-m}$	$l_t = \mu_t(1 + \alpha e_t)/S_{t-m}$
	$b_t = (\mu_t + \beta^* e_t)/l_{t-1}$	$b_t = \mu_t(1 + \beta^* e_t)/l_{t-1}$	$b_t = [(\mu_t + \beta^* e_t) - S_{t-m}]/l_{t-1}$	$b_t = [\mu_t(1 + \beta^* e_t) - S_{t-m}]/l_{t-1}$	$b_t = (\mu_t + \beta^* e_t)/(l_{t-1} S_{t-m})$	$b_t = \mu_t(1 + \beta^* e_t)/(l_{t-1} S_{t-m})$
		$S_t = (\mu_t + \gamma e_t) - l_{t-1} b_{t-1}^\phi$	$S_t = \mu_t(1 + \gamma e_t) - l_{t-1} b_{t-1}^\phi$	$S_t = (\mu_t + \gamma e_t)/(l_{t-1} b_{t-1}^\phi)$	$S_t = \mu_t(1 + \gamma e_t)/(l_{t-1} b_{t-1}^\phi)$	$S_t = \mu_t(1 + \gamma e_t)/(l_{t-1} b_{t-1}^\phi)$

Onde:  $e_t$  é o erro de previsão um passo à frente;  $y_t$  é o valor observado no período  $t$ ;  $\mu_t$  é o nível da série no período  $t$ ;  $S_t$  é a sazonalidade da série no período  $t$ ;  $m$  é o tamanho do ciclo sazonal;  $\alpha$  é o parâmetro de suavização para o nível;  $\gamma$  é o parâmetro de suavização para a sazonalidade;  $b_t$  é o crescimento da série no período  $t$ ;  $\beta^* = \alpha\beta$ ;  $\beta$  é o parâmetro de suavização para o crescimento;  $\phi$  é o parâmetro de amortecimento do crescimento. Fonte: Miranda (2009).

Figura 7 – Performance de previsão do modelo combinado.

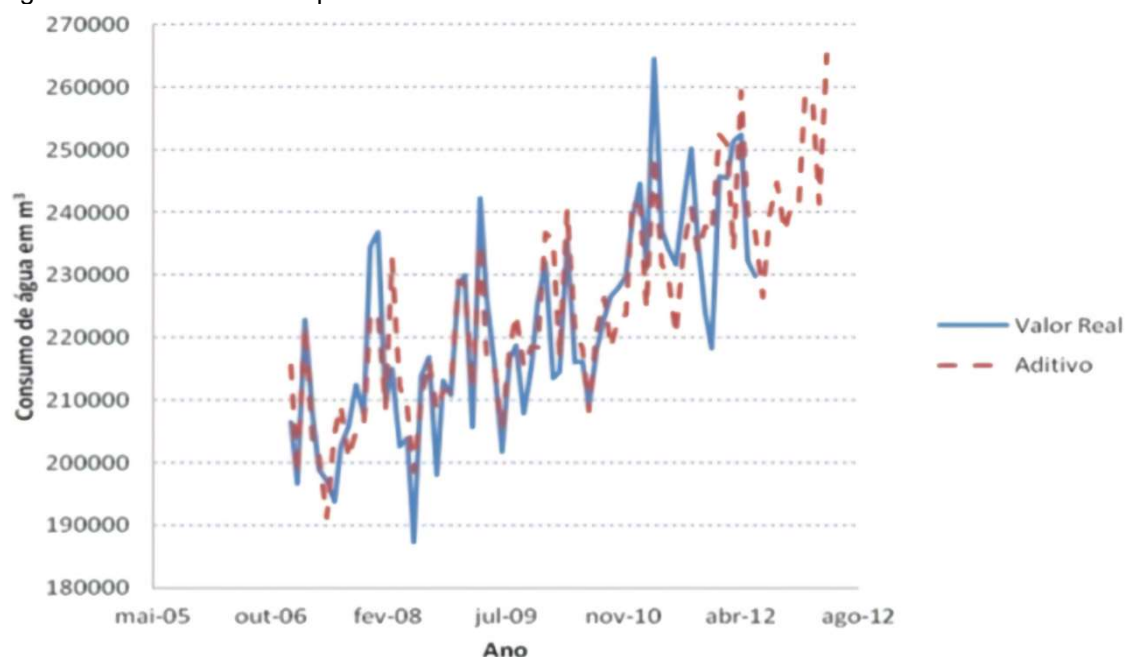


Fonte: adaptado de Homwongs, Sastri e Foster III (1994).

Caiado (2010) analisou o desempenho dos modelos de séries temporais univariadas: Holt-Winters; GARCH (Generalized Autoregressive Conditional Heterocedasticity); SARIMA de sazonalidade dupla (semanal e anual); e de uma combinação dos três métodos, para prever a demanda de água diária na Espanha. O autor investigou se a combinação de previsão de diferentes métodos poderia melhorar a precisão. Os resultados sugerem que em dias específicos da semana, o modelo Holt-Winters é o mais preciso e mais adequado para previsões com horizontes maiores.

Comunello, Daminelli e Bortolotti (2013) analisaram a demanda mensal de água urbana, no município de Medianeira, no Paraná, aplicando modelos Holt-Winters Aditivo e Holt-Winters Multiplicativo. Os resultados mostram que o modelo aditivo foi o que melhor se adequou à série de dados, com MAPE de 2,86% (Figura 8).

Figura 8 – Previsão obtida para o modelo Holt-Winters aditivo.



Fonte: Comunello, Daminelli e Bortolotti (2013).

#### 2.3.1.2 Modelos autorregressivos integrados de médias móveis ARIMA

Os modelos ARIMA (modelos autorregressivos integrados de médias móveis) são altamente estruturados, porém, pouco aplicados isoladamente em estudos de demanda de água, exceto em um contexto de curto ou curtíssimo prazo. Eles são apropriados para modelar o padrão sazonal no uso mensal de água, bem como padrões fixos que podem ocorrer para o uso diário da água (BILLINGS; JONES, 2008).

Os modelos ARIMA provêm da combinação de modelos de autorregressão (AR), onde projeta-se a variável de interesse usando uma combinação linear de valores passados da variável, e modelos de médias móveis (MA), onde os valores dos erros são uma média móvel ponderada dos últimos erros de previsão. O I (para "integrado") indica que os valores de dados foram substituídos pela diferença entre seus valores e os valores anteriores (e esse processo de diferenciação pode ter sido executado mais de uma vez) (HYNDMAN; ATHANASOPOULOS, 2013).

Para os modelos AR, MA, ARMA e ARIMA são requeridas condições de estacionariedade, média igual a zero e variância constante (BOX; JENKINS; REINSEL, 2008). No caso de séries não estacionárias, pode-se eliminar a influência

de sazonalidade e tendência das séries temporais por meio de diferenciações e transformações matemáticas, como a aplicação de logaritmos. Posteriormente pode-se inverter as transformações e obter as previsões para a série original pela aplicação de um modelo ARIMA sazonal (SARIMA), que adiciona um componente sazonal à equação (HYNDMAN; ATHANASOPOULOS, 2013).

Conforme Oliveira (2012), o consumo de água em um determinado local pode ser compreendido como uma série temporal: estocástica (mesmo sofrendo influência de inúmeros fatores, a demanda não pode ser perfeitamente explicada por intermédio destes); e não estacionária, com características específicas, variando de acordo com o local de estudo (BABEL; SHINDE, 2011). No caso da série temporal de consumo de água, pode-se assumir que os valores registrados são relacionados aos valores passados. Assim, entende-se que cada dado  $Z_t$  é função de  $p$  observações de demandas anteriores (OLIVEIRA, 2012).

Considerando o espaçamento temporal entre as observações constante, é possível descrever os valores anteriores como  $Z_{t-1}, Z_{t-2}, \dots, Z_{t-p}$ . Isto é, a presente observação da série  $Z_t$  pode ser explicada como uma função das  $p$  observações passadas, onde  $p$  determina a ordem da estrutura ou o número de passos entre as observações passadas e a previsão da próxima observação. Como essa relação não é determinística, admite-se que para o instante  $t$ , existe uma diferença entre o dado previsto e o dado verdadeiro, representada pelo erro, ou ruído branco,  $a_t$  (OLIVEIRA, 2012). Essa suposição caracteriza o conceito central de um Modelo Autorregressivo AR ( $p$ ), expresso pela equação<sup>1</sup> 1 (HYNDMAN; ATHANASOPOULOS, 2013).

$$Z_t = \psi_1 Z_{t-1} + \psi_2 Z_{t-2} + \dots + \psi_p Z_{t-p} + a_t \quad (1)$$

Onde:

$Z_t$  é a observação presente da série;

$\psi$  é o coeficiente autorregressivo;

$p$  é a ordem da estrutura autorregressiva, varia de 1 a  $p$ ;

$a_t$  é o erro ou ruído branco.

---

<sup>1</sup> A simbologia utilizada para representar os parâmetros foi modificada de forma a evitar o uso de um mesmo símbolo para diferentes parâmetros e modelos.

Nesses modelos, para a realização de previsões, extrapola-se a série para momentos futuros  $t + 1$ ,  $t + 2$ ... e assim por diante. Entretanto, na previsão, perde-se a parcela de erro. Com o objetivo de evitar a perda de dados na modelagem, surge o modelo de previsão de erro, ou médias móveis MA (OLIVEIRA, 2012). Os modelos de médias móveis MA ( $q$ ), são formados por uma combinação linear dos erros,  $a_t$ , ocorridos no período corrente e nos períodos passados, conforme a equação 2 (HYNDMAN; ATHANOSOPOULOS, 2013).

$$Z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (2)$$

Onde:

$Z_t$  é a observação presente da série;

$a_t$  é o erro ou ruído branco;

$\theta$  é o coeficiente de médias móveis;

$q$  é a ordem da estrutura de médias móveis, varia de 1 a  $q$ .

Os modelos autorregressivos de médias móveis ARMA ( $p, q$ ) (equação 3) são uma combinação desses dois métodos, e como os modelos AR e MA, só podem ser aplicados em séries estacionárias (OLIVEIRA, 2012).

$$Z_t = \psi_1 Z_{t-1} + \dots + \psi_p Z_{t-p} + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q} + a_t \quad (3)$$

Onde:

$Z_t$  é a observação presente da série;

$\psi$  é o coeficiente autorregressivo;

$p$  é a ordem da estrutura autorregressiva, varia de 1 a  $p$ ;

$a_t$  é o erro ou ruído branco;

$\theta$  é o coeficiente de médias móveis;

$q$  é a ordem da estrutura de médias móveis, varia de 1 a  $q$ .

Assim que se começa a combinar os componentes dessa maneira os modelos ficam cada vez mais complexos (HYNDMAN; ATHANOSOPOULOS, 2013). Uma maneira de facilitar a compreensão é escrever a equação 3 em função de um operador

de defasagem ( $B$ ), que é uma ferramenta de notação útil para se trabalhar com *lags* em séries temporais.

A aplicação do operador de defasagem ( $B$ ) em uma série tem o efeito de voltar o dado a um período anterior. Duas aplicações voltam dois períodos, e assim sucessivamente, como mostram as equações 4 e 5 (HYNDMAN; ATHANOSOPOULOS, 2013).

$$BZ_t = Z_{t-1} \quad (4)$$

$$B(BZ_t) = B^2Z_t = Z_{t-2} \quad (5)$$

Onde:

$B$  é o operador de defasagem;

$Z_t$  é a observação presente da série.

Neste caso, um modelo ARMA (equação 3) poderia ser definido da seguinte forma (equação 6) (OLIVEIRA, 2012):

$$(1 - \psi_1 B - \psi_2 B^2 - \dots - \psi_p B^p)Z_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)a_t \quad (6)$$

Onde:

$\psi$  é o coeficiente autorregressivo;

$B$  é o operador de defasagem;

$Z_t$  é a observação presente da série;

$\theta$  é o coeficiente de médias móveis;

$a_t$  é o erro ou ruído branco;

$p$  é a ordem da estrutura autorregressiva, varia de 1 a  $p$ ;

$q$  é a ordem da estrutura de médias móveis, varia de 1 a  $q$ .

Os modelos ARMA podem ser generalizados também para séries não estacionárias, por meio de sucessivas diferenciações da série original. O número necessário de auto diferenças é denominado ordem de integração  $d$ , resultando em

um modelo autorregressivo integrado de médias móveis ARIMA  $(p, d, q)$  (HYNDMAN; ATHANOSOPOULOS, 2013).

Uma das funções da diferenciação de uma série de dados é retirar os efeitos de tendência e sazonalidade, ao torná-la estacionária (MAKRIDAKIS; WHEELWRIGHT; HYNDMAN, 1997). Conforme o teorema de Beveridge-Nelson, qualquer série que possa ser modelada por meio de um ARIMA  $(p, d, q)$  possui uma tendência estocástica do tipo passeio casual com *drift* (CRIBARI NETO, 1990). Quando há a inclusão do *drift*, a tendência linear existente na série, que pode ser positiva ou negativa, é representada por uma constante  $c$  (HYNDMAN, 2014). Um intercepto também pode ser incluído para representar a média, quando o modelo apresenta média diferente de zero (HYNDMAN, 2014).

Assim um modelo ARIMA  $(p, d, q)$  pode ser escrito conforme a equação 7 (HYNDMAN; ATHANOSOPOULOS, 2013).

$$Z_t^d = c + \psi_1 Z_{t-1}^d + \dots + \psi_p Z_{t-p}^d + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q} + a_t \quad (7)$$

Onde:

$Z_t^d$  é a observação presente da série diferenciada  $d$  vezes;

$c$  é uma constante;

$\psi$  é o coeficiente autorregressivo;

$p$  é a ordem da estrutura autorregressiva, varia de 1 a  $p$ ;

$a_t$  é o erro ou ruído branco;

$\theta$  é o coeficiente de médias móveis;

$q$  é a ordem da estrutura de médias móveis, varia de 1 a  $q$ .

O operador de defasagem também é útil para descrever o processo de diferenciação. As equações 8 e 9 mostram como podem ser escritas as duas primeiras diferenciações dos dados (HYNDMAN; ATHANOSOPOULOS, 2013).

$$Z'_t = Z_t - Z_{t-1} = Z_t - BZ_t = (1 - B)Z_t \quad (8)$$

$$Z''_t = Z_t - 2Z_{t-1} + Z_{t-2} = (1 - 2B + B^2)Z_t = (1 - B)^2 Z_t \quad (9)$$



Onde:

$Z'_t$  é a observação presente da série diferenciada uma vez;

$B$  é o operador de defasagem;

$Z_t$  é a observação presente da série;

$Z''_t$  é a observação presente da série diferenciada duas vezes.

Generalizando a equação 9 para  $d$  diferenciações, obtêm-se a equação 10 (HYNDMAN; ATHANOSOPOULOS, 2013).

$$Z^d_t = (1 - B)^d Z_t \quad (10)$$

Onde:

$Z^d_t$  é a observação presente da série diferenciada  $d$  vezes;

$B$  é o operador de defasagem;

$d$  é o número de diferenças ou ordem de integração;

$Z_t$  é a observação presente da série.

Dessa maneira, o modelo ARIMA representado na equação 7 pode ser escrito em função do operador de defasagem ( $B$ ), como mostra a equação 11 (HYNDMAN; ATHANOSOPOULOS, 2013).

$$(1 - \psi_1 B - \dots - \psi_p B^p)(1 - B)^d Z_t = c + (1 - \theta_1 B - \dots - \theta_q B^q) a_t \quad (11)$$

Onde:

$\psi$  é o coeficiente autorregressivo;

$B$  é o operador de defasagem;

$p$  é a ordem da estrutura autorregressiva, varia de 1 a  $p$ ;

$d$  é o número de diferenças ou ordem de integração;

$Z_t$  é a observação presente da série;

$c$  é uma constante;

$\theta$  é o coeficiente de médias móveis;

$a_t$  é o erro ou ruído branco;

$q$  é a ordem da estrutura de médias móveis, varia de 1 a  $q$ .

O *software* R (R CORE TEAM, 2019) utiliza uma parametrização diferente para a equação 11, utilizando a equação 12 para representar a constante  $c$  (HYNDMAN; ATHANOSOPOULOS, 2013).

$$c = \mu(1 - \psi_1 - \dots - \psi_p) \quad (12)$$

Onde:

$\mu$  é a média dos valores observados da série diferenciada ( $Z_t^d$ );

$\psi$  é o coeficiente autorregressivo.

A equação 13 representa a parametrização utilizada pelo *software* R (R CORE TEAM, 2019) para um modelo ARIMA (HYNDMAN; ATHANOSOPOULOS, 2013).

$$(1 - \psi_1 B - \dots - \psi_p B^p)(Z_t^d - \mu) = (1 - \theta_1 B - \dots - \theta_q B^q)a_t \quad (13)$$

Onde:

$\psi$  é o coeficiente autorregressivo;

$B$  é o operador de defasagem;

$p$  é a ordem da estrutura autorregressiva, varia de 1 a  $p$ ;

$Z_t^d$  é a observação presente da série diferenciada;

$\mu$  é a média dos valores observados da série diferenciada;

$\theta$  é o coeficiente de médias móveis;

$q$  é a ordem da estrutura de médias móveis, varia de 1 a  $q$ ;

$a_t$  é o erro ou ruído branco.

Sabe-se que uma série temporal do consumo de água é um processo não estacionário que possui sazonalidade, ou seja, nota-se um efeito cíclico nas observações, separado por um período  $s$  (OLIVEIRA, 2012). Ao aplicar um modelo ARIMA  $(p, d, q)$  na série, esse efeito cíclico é retirado, e um novo modelo ARIMA  $(P, D, Q)$  pode ser empregado nos resíduos do primeiro modelo. Esses dois modelos conjugados incorporam um componente sazonal ao modelo, resultando em um modelo ARIMA sazonal  $(p, d, q) (P, D, Q)$ , ou SARIMA (OLIVEIRA, 2012).

A equação 14 descreve a relação entre os parâmetros autorregressivos e de médias móveis para o desenvolvimento de um modelo SARIMA, conforme Box e Jenkins (1976), em função do operador de defasagem ( $B$ ).

$$\psi(B)\Psi(B^s)\nabla^D\nabla^dZ_t = c + \theta(B)\Theta(B^s)a_t \quad (14)$$

Onde:

$\psi$  é o coeficiente autorregressivo  $(1 - \psi_1 B - \dots - \psi_p B^p)$ ;

$B$  é o operador não sazonal de defasagem;

$\Psi$  é o coeficiente autorregressivo sazonal  $(1 - \Psi_1 B - \dots - \Psi_p B^p)$ ;

$B^s$  é o operador sazonal de defasagem;

$s$  é o período da sazonalidade;

$\nabla^D$  é o coeficiente de integração sazonal, definido como  $\nabla^D = (1 - B)^D$ ;

$\nabla^d$  é o coeficiente de integração, definido como  $\nabla^d = (1 - B)^d$ ;

$Z_t$  é a observação presente da série;

$c$  é uma constante;

$\theta$  é o coeficiente de médias móveis  $(1 - \theta_1 B - \dots - \theta_q B^q)$ ;

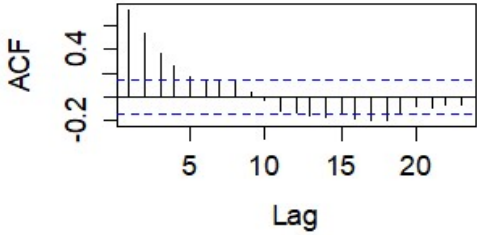
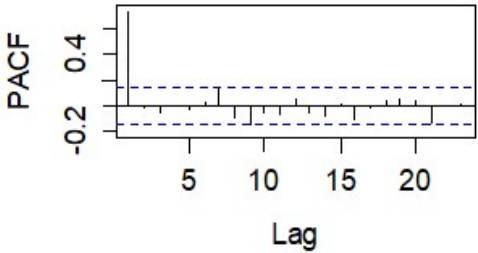
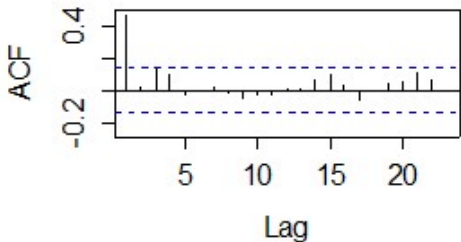
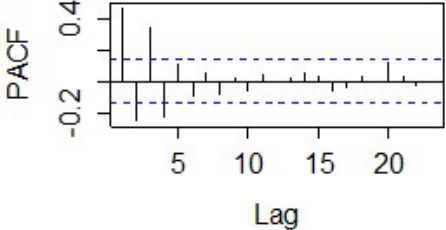
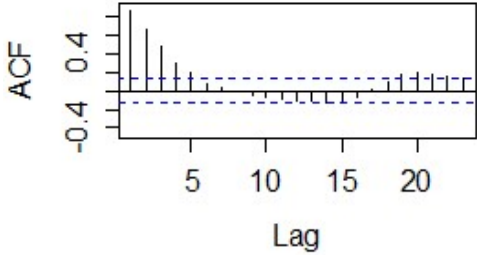
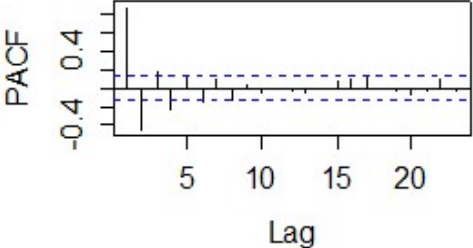
$\Theta$  é o coeficiente de médias móveis sazonal  $(1 - \Theta_1 B - \dots - \Theta_q B^q)$ ;

$a_t$  é o erro ou ruído branco.

Conforme Box e Jenkins (1976) o processo de desenvolvimento de um modelo ARIMA passa por um ciclo iterativo de quatro etapas. No estágio de identificação são definidos os valores de  $p$ ,  $d$ ,  $q$ . Primeiramente faz-se uma análise para decidir os números de diferenças  $d$  e diferenças sazonais ( $D$ ) buscando atingir a estacionaridade da série original. Em seguida, segundo os autores, calcula-se a função de autocorrelação (ACF) e função de autocorrelação parcial (PAFC) da série já diferenciada buscando-se identificar as ordens dos termos autorregressivos e de médias móveis  $p$ ,  $q$ . O número de diferenças realizadas ( $d$ ) é definido de acordo com a mudança do comportamento da ACF. O número suficiente de diferenças é alcançado, quando a ACF decai rapidamente para zero. A função de autocorrelação ACF serve de base para indicar a ordem de médias móveis ( $q$ ), enquanto a função de

autocorrelação PACF fornece dados para a ordem autorregressiva ( $p$ ), conforme a Tabela 6.

Tabela 5 – Identificação de modelos por meio das funções de autocorrelação e autocorrelação parcial.

Modelo	AFC	PAFC
	AFC decai de acordo com exponenciais ou senoides amortecidas, infinita em extensão	PACF finita, apresenta um corte após o lag $p$
AR		
	AFC finita, apresenta um corte após o lag $q$	PAFC decai de acordo com exponenciais ou senoides amortecidas, infinita em extensão
MA		
	AFC decai de acordo com exponenciais ou senoides amortecidas, infinita em extensão	PAFC decai de acordo com exponenciais ou senoides amortecidas, infinita em extensão
ARMA		

Fonte: Adaptado de Oliveira (2012) e Henning (2017).

Na fase de estimação, os coeficientes do polinômio autorregressivo e de médias móveis, sazonal e não sazonal, são calculados. O processo de estimação de um modelo é realizado por meio do método de Estimativa de Máxima Verossimilhança (MLE) com a minimização dos valores quadrados dos ruídos aleatórios previstos (BOX; JENKINS, 1976).

Na etapa de diagnóstico busca-se confirmar a hipótese assumida de que os resíduos geradores do modelo são do tipo ruído branco, ou seja, distribuídos

aleatoriamente. Por meio dos testes de resíduos, faz-se o cálculo da correlação entre os próprios resíduos e entre os resíduos e a série original. Caso os resultados apontem autocorrelação, novos termos autorregressivos devem ser adicionados. Neste momento toma-se a decisão sobre encerrar o processo ou buscar um outro modelo possível na etapa de identificação. A última etapa é a de previsão, quando é realizada a previsão propriamente dita (BOX; JENKINS, 1976).

A modelagem pode também ser realizada automaticamente por um *software*, como no caso deste trabalho. No *software* R (R CORE TEAM, 2019) utiliza-se a função `auto.arima`, do pacote *forecast* (HYNDMAN; KHANDAKAR, 2008; HYNDMAN *et al.*, 2018) baseada em uma variação do algoritmo Hyndman-Khandakar (HYNDMAN; KHANDAKAR, 2008). Esse algoritmo combina testes de raiz unitária, minimização do Critério de Informação de Akaike (AIC) e da Estimativa de Máxima Verossimilhança (MLE) para a obtenção de um modelo ARIMA (HYNDMAN; ATHANASOPOULOS, 2013).

Conforme Hyndman de Athanasopoulos (2013), neste tipo de modelagem o número de diferenças ( $d$ ) entre 0 e 2 ( $0 \leq d \leq 2$ ) é determinado utilizando repetidos testes de Kwiatkowski, Phillips, Schmidt, e Shin (KPSS) – cujo objetivo é descobrir se uma série temporal é estacionária em torno de uma média ou tendência linear, ou não-estacionária devido a uma raiz unitária (KWIATKOWSKI *et al.*, 1992).

Além de testar todas as combinações, o algoritmo usa uma pesquisa gradual para percorrer o espaço do modelo e determinar os valores de  $p$  e  $q$ , por meio da minimização do AIC depois de  $d$  diferenciações (HYNDMAN; ATHANOSOPOULOS, 2013).

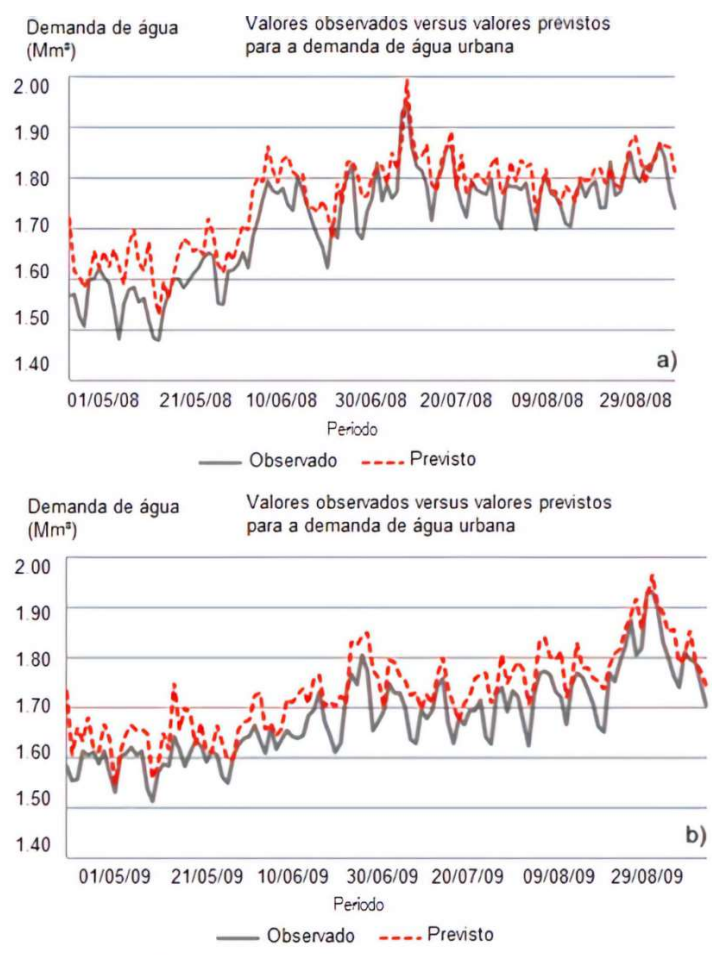
Inicialmente são ajustados quatro modelos iniciais: ARIMA (0,  $d$ , 0); ARIMA (2,  $d$ , 2); ARIMA (1,  $d$ , 0); ARIMA (0,  $d$ , 1). Nos modelos onde  $d \neq 2$ , uma constante  $c$  é incluída. Quando  $d \leq 1$ , um modelo adicional ARIMA (0,  $d$ , 0) sem constante é ajustado. São testados então, modelos com variações de  $\pm 1$  nos termos  $p$  e  $q$ , incluindo e excluindo a constante  $c$ . Realiza-se novamente a pesquisa gradual até que nenhum valor menor de AIC seja encontrado (HYNDMAN; ATHANOSOPOULOS, 2013).

Na literatura é possível encontrar diversos estudos que modelam o consumo de água por meio de modelos ARIMA e ARIMA sazonais. Na análise de desempenho de modelos de séries temporais univariadas (Holt-Winters, GARCH e SARIMA) de sazonalidade dupla (semanal e anual), Caiado (2010) afirma que para previsões de

um ou dois dias o modelo ARIMA tem o mesmo desempenho do modelo GARCH, e se adequa melhor às previsões de um a quatro dias que o modelo Holt-Winters.

Adamowski *et al.* (2012) compararam o desempenho de modelos de previsão de transformações ondulatórias discretas (WA), regressão linear múltipla (RLM), regressão não linear múltipla (RNLM), ARIMA, RNA, e WA-RNA para prever o consumo urbano de água diário entre maio e agosto nos anos de 2008 e 2009, em Montreal, no Canadá. Os resultados do modelo ARIMA são apresentados na Figura 9, com um coeficiente de determinação de 0,782.

Figura 9 – Comparação da demanda de água prevista versus observada usando o melhor modelo ARIMA para previsão de um dia à frente durante o período de testes (a) verão de 2008 e (b) verão de 2009.

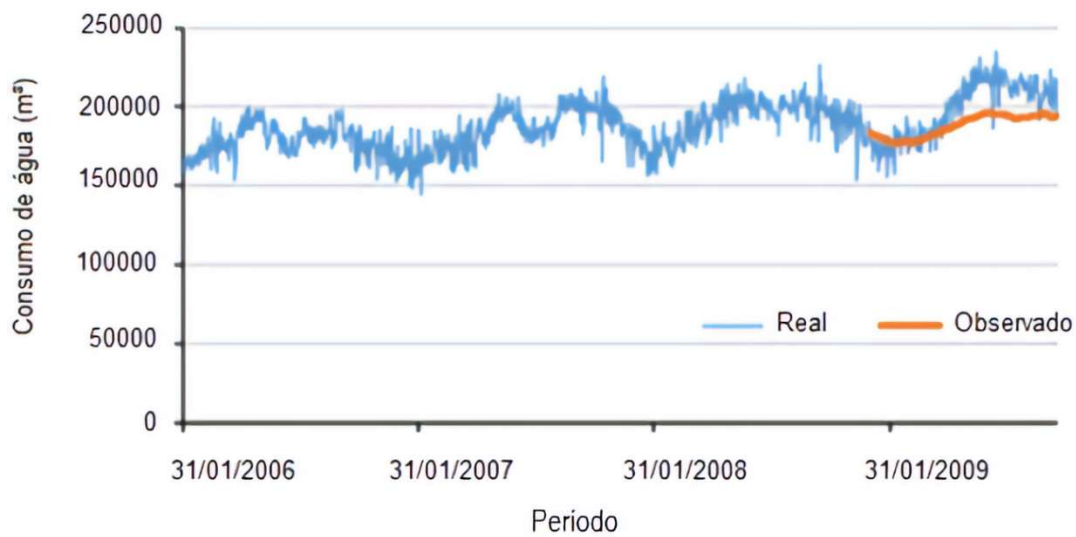


Fonte: adaptado de Adamowski *et al.* (2012).

Al-Zahrani e Abo-Monasar (2015) utilizaram regressão linear múltipla, ARMA e RNA para prever a demanda de água diária na cidade de Al-Khobar, na Árabia

Saudita, de fevereiro a outubro de 2009, e obtiveram um MAPE de 6,15% para o modelo ARMA (Figura 10).

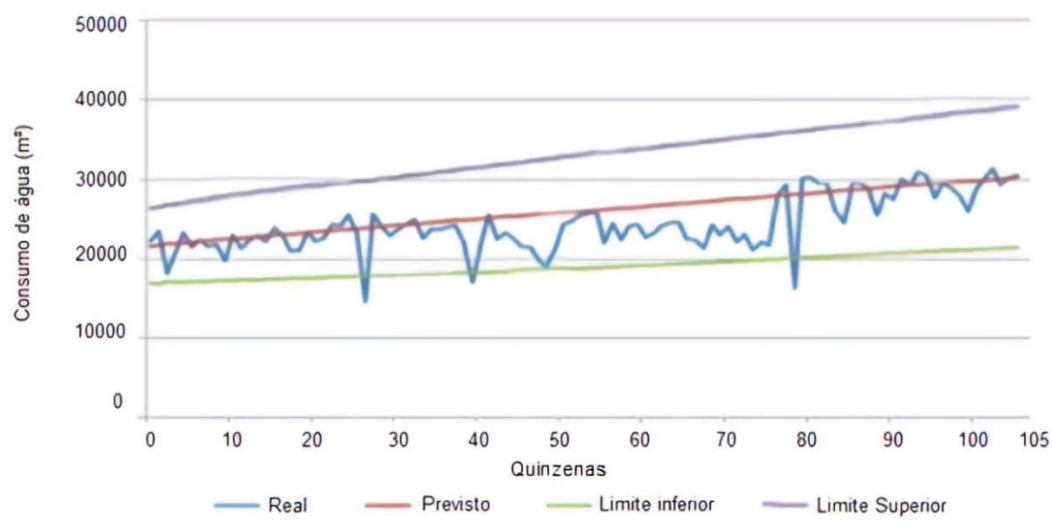
Figura 10 – Resultado da previsão do modelo ARMA de fevereiro de 2009 a outubro de 2009.



Fonte: adaptado de Al-Zahrani e Abo-Monasar (2015).

Amponsah, Otoo e Todoko (2015) fizeram uso de modelos ARIMA e SARIMA para prever o consumo anual de água na região de Volta, em Gana para 105 quinzenas, entre os anos de 2013 e 2016. Os resultados apontam que o modelo ARIMA é o que melhor se adequa à série de dados (Figura 11).

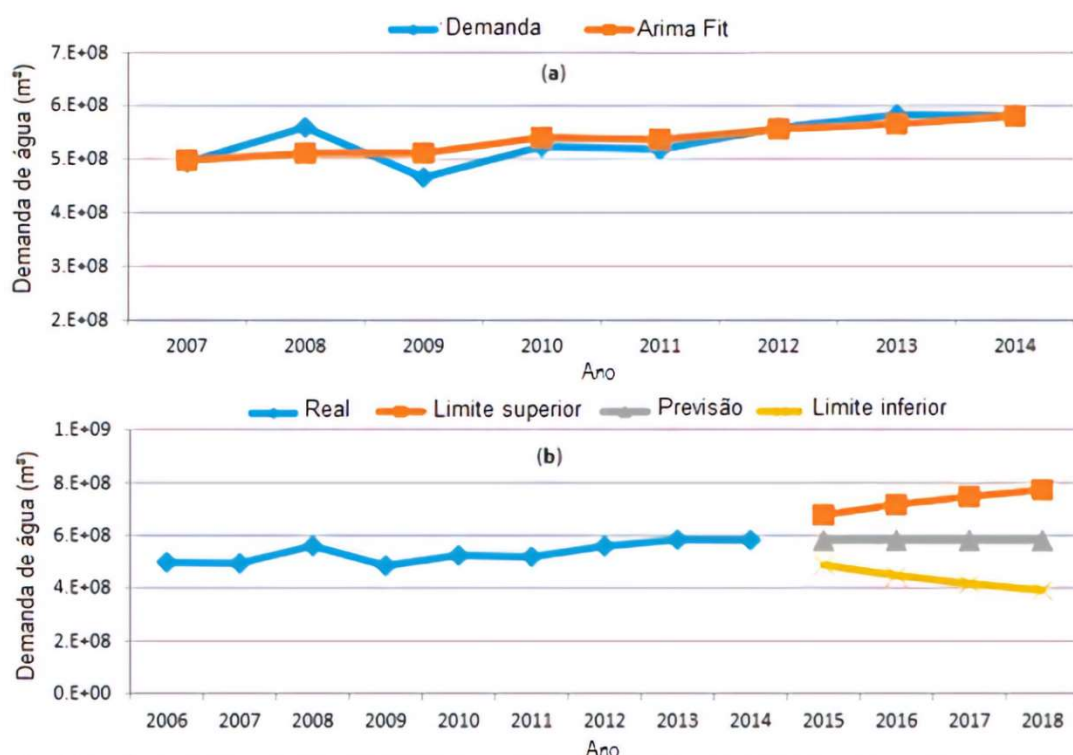
Figura 11 – Resultados da previsão do modelo ARIMA para a demanda real e prevista.



Fonte: adaptado de Amponsah, Otoo e Todoko (2015).

Yalçintaş *et al.* (2015) previram a demanda anual de água urbana em Istambul, na Turquia, por intermédio de um modelo ARIMA entre os anos de 2015 e 2018. Os resultados apontam que o modelo se adequa à série de dados com um MAPE de 6,3% (Figura 12).

Figura 12 – Previsão do modelo ARIMA para demanda de água. a) Demanda real versus demanda do modelo. b) Demanda prevista no intervalo de 95% de confiança.

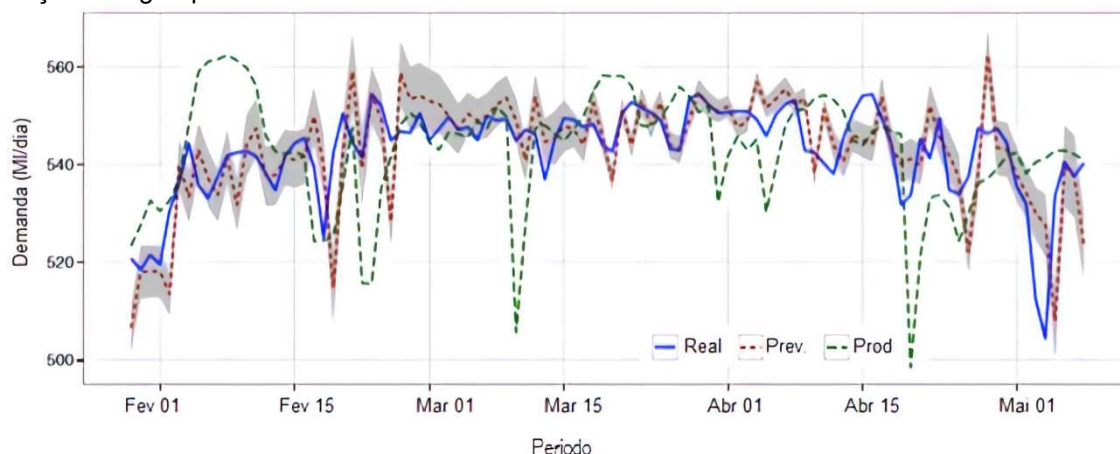


Fonte: adaptado de Yalçintaş *et al.* (2015).

Arandia *et al.* (2016) estimaram a demanda de água de curto prazo (subhorária, horária e diária) para a cidade de Dublin, na Irlanda, com aplicação de modelos SARIMA com coleta de dados *offline* e *online* (aplicando filtro Kalman). Os resultados de previsão para um dos modelos foram comparados à demanda real e aos dados de produção de água potável e podem ser observados na Figura 13.



Figura 13 – Comparação entre os resultados de previsão diária do modelo SARIMA, demanda real e produção de água potável.



Fonte: adaptado de Arandia *et al.* (2016).

Pelo poder de captar características comumente associadas a séries de consumo de água: tendência e sazonalidade, o método ARIMA têm sido uma técnica bastante utilizada na previsão do consumo de água nos últimos anos, por ser de fácil desenvolvimento e implantação (OLIVEIRA, 2012).

### 3 METODOLOGIA

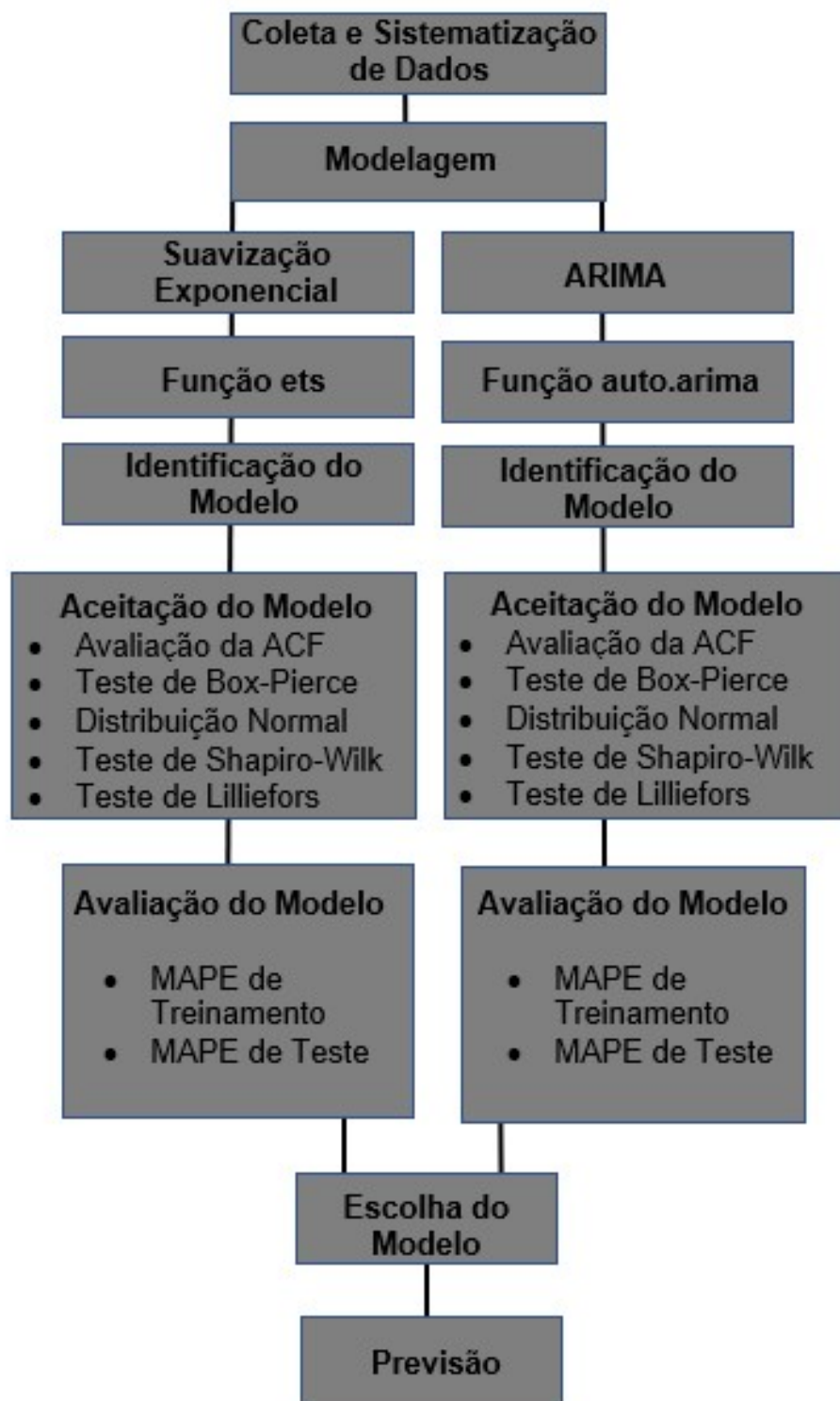
A metodologia proposta neste trabalho tem como objetivo definir qual modelo de previsão de séries temporais, entre os ajustados, é mais adequado para prever o consumo urbano mensal de água, a curto prazo, em uma determinada série de dados. Para tanto, são aplicados em séries temporais de consumo de água os modelos de Suavização Exponencial e Autorregressivo Integrado de Médias Móveis (ARIMA), visando a realização de previsões de demanda. Os modelos são avaliados e a escolha do método mais apropriado está fundamentado na comparação dos erros obtidos nos modelos aceitos em etapa precedente. Isso, implica na seleção adequada de um modelo capaz de prever acontecimentos futuros, tendo como propósito a redução de riscos na tomada de decisão.

Este capítulo é organizado em cinco subitens que descrevem o processo de realização da modelagem. O primeiro subitem diz respeito à coleta e sistematização de dados. O segundo subitem apresenta os passos para o desenvolvimento dos modelos. Após a modelagem, no terceiro subitem, são definidos os critérios de aceitação dos modelos, que abordam a análise de resíduos. No quarto subitem são avaliados os desempenhos dos modelos, onde são observados principalmente os erros de previsão. Logo após a avaliação, no quinto subitem, escolhe-se o melhor modelo e a previsão é realizada com um horizonte de 3 meses.

A sistemática de modelagem aplicada no desenvolvimento dos modelos estatísticos, desde a organização dos dados até a previsão de demanda propriamente dita deste trabalho é realizada no ambiente do software R para computação estatística (R CORE TEAM, 2019), com a utilização dos pacotes *tseries* (TRAPLETTI; HORNIK; LEBARON, 2019) e *forecast* (HYNDMAN; KHANDAKAR, 2008; HYNDMAN *et al.*, 2018).

A metodologia de aplicação proposta nesse trabalho envolve a utilização de um fluxograma que inclui basicamente um conjunto de etapas que combinam todo o referencial teórico até aqui apresentado para estabelecer a previsão de demanda via ambiente R. O resumo dessas etapas metodológicas é ilustrado conforme Figura 14.

Figura 14 – Etapas metodológicas.



Fonte: a autora (2020).

### 3.1 COLETA E SISTEMATIZAÇÃO DE DADOS

Para o estabelecimento de modelos de previsão de consumo mensal de água que fazem uso de séries temporais é necessária a criação de uma base de dados. Como será utilizada na modelagem em diferentes métodos, para o modelo ARIMA a base deve conter o consumo mensal de água com um mínimo de 50 observações (CHATFIELD, 2005).

Na sistematização de dados, os valores de consumo devem ser organizados em ordem cronológica separados por categoria de consumo. Após a organização, deve-se verificar a existência de dados faltantes e de dados extremos, também chamados de *outliers*. Caso a série apresente dados faltantes e/ou *outliers*, é possível realizar a correção desses valores fazendo uso de interpolação linear e/ou imputação de dados faltantes (NASCIMENTO *et al.*, 2012). Pode-se fazer uso das funções *tsclean* para que o R substitua automaticamente os *outliers* e dados faltantes, ou *tsoutliers* que apenas identifica os dados que apresentam problemas (HYNDMAN E KHANDAKAR, 2008; HYNDMAN *et al.*, 2018).

Neste trabalho propõe-se realizar testes de acurácia com dados mensais de pelo menos três meses de consumo, conforme o horizonte de previsão, como indicado por Hyndman e Athanasopoulos (2013). Tais dados não devem fazer parte da amostra, devendo ser dados de meses subsequentes aos meses que compõem a amostra.

### 3.2 MODELAGEM

A análise estatística pode ser desenvolvida com os pacotes *tseries* (TRAPLETTI; HORNIK; LEBARON, 2019), *forecast* (HYNDMAN; KHANDAKAR, 2008; HYNDMAN *et al.*, 2018) e *nortest* (GROSS; LIGGES, 2015) que fornecem as funções para a transformação dos dados em série temporal, para a realização da modelagem e previsão, e testes de normalidade, respectivamente.

Após a ativação dos pacotes, os dados são inseridos no programa e transformados em séries temporais com frequência mensal. Conforme Hyndman e

Athanasopoulos (2013) antes da modelagem se faz necessário plotar a série, observar suas características, verificar a existência de padrões e a presença de dados incomuns. Dessa maneira, devem ser observadas: a distribuição normal dos dados de cada série; a existência de tendência e sazonalidade, típicas de séries temporais de consumo de água; e as estatísticas descritivas.

Em casos em que os dados não seguem uma distribuição normal pode-se fazer uso de transformações matemáticas, com o propósito de simplificar os padrões nos dados históricos, removendo fontes conhecidas de variação ou tornando o padrão mais consistente em todo o conjunto de dados (HYNDMAN; ATHANASOPOULOS, 2013).

Posteriormente, inicia-se a modelagem propriamente dita. Serão desenvolvidos dois modelos de previsão, para cada categoria de consumo, sendo um por meio de suavização exponencial e outro por meio do método ARIMA.

Na aplicação do modelo de Suavização Exponencial, faz-se uso do comando *ets()*, que define automaticamente o melhor modelo para série temporal estudada, na forma (X, Y, Z), onde X representa o tipo de erro; Y o tipo de tendência; e Z o tipo de sazonalidade.

O *software* aplica em cada série todos os 30 modelos de espaço de estado possíveis, otimizando os parâmetros de suavização e os estados iniciais das variáveis (nível, tendência e sazonalidade) em cada caso. Seleciona então o modelo com menor valor do Critério de Informação de Akaike (AIC) e produz previsões pontuais, para o número de passos à frente determinado. Os intervalos de previsão são calculados utilizando os resultados analíticos de Hyndman *et al.* (2005), ou simulando caminhos de amostra para os valores preditos ( $\mu_t$ ) com t variando de  $n + 1$  a  $n + h$  passos à frente ( $\mu_{n+1}, \dots, \mu_{n+h}$ ), e encontrando os percentis relacionados ao parâmetro de suavização de nível ( $\alpha$ )  $\frac{\alpha}{2}$  e  $1 - \frac{\alpha}{2}$ , dos dados simulados para cada horizonte de previsão (HYNDMAN *et al.*, 2008).

Como resultado do modelo indicado, são obtidos os valores dos parâmetros de suavização ( $\alpha, \beta, \gamma$ ), de amortecimento ( $\phi$ ) e os valores iniciais de nível, tendência e sazonalidade ( $l, b, s$ ), além dos erros do modelo (HYNDMAN; KHANDAKAR, 2008).

Para a modelagem pelo método ARIMA, utiliza-se a função *auto.arima()* que indica automaticamente o modelo que melhor se adequa à série de dados. A modelagem automática define o número de diferenças (entre 0 e 2) por intermédio de

repetidos testes de KPSS, definindo  $p$  e  $q$  pela minimização do AIC após  $d$  diferenças. São ajustados alguns modelos iniciais, que combinam diferentes valores de  $p$  e  $q$  e incluem e excluem uma constante, e o modelo com menor valor de AIC é escolhido (HYNDMAN; ATHANASOPOULOS, 2013).

Com o modelo definido, é possível visualizar os valores de  $p$ ,  $q$  e  $d$ , e  $P$ ,  $Q$  e  $D$  (no caso de séries sazonais) que são as ordens da estrutura autorregressiva, de médias móveis e de integração (não sazonais e sazonais) do modelo mais adequado, além de seus coeficientes, erros e o valor do AIC (HYNDMAN; KHANDAKAR, 2008).

A aceitação dos modelos é realizada por meio de observações gráficas e testes estatísticos dos resíduos, verificando a existência de autocorrelação e distribuição normal entre eles (HYNDMAN; ATHANASOPOULOS, 2013). Os modelos que não apresentarem distribuição normal e/ou apresentarem autocorrelação nos resíduos devem ser desconsiderados (HYNDMAN; ATHANASOPOULOS, 2013).

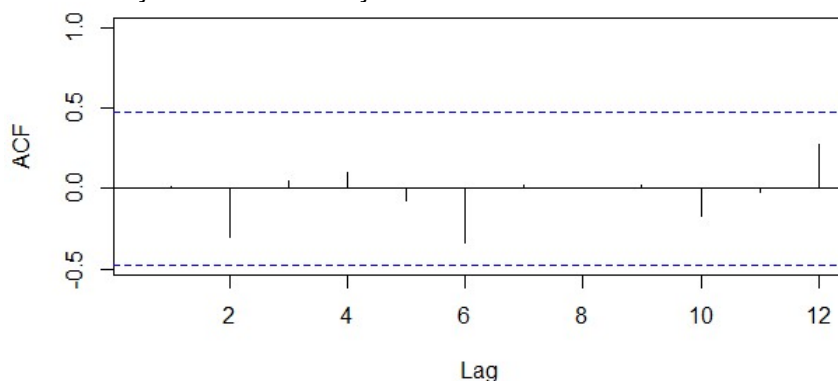
As decisões de probabilidade deste método serão baseadas na probabilidade de significância (p-valor). O p-valor é uma medida de inferência estatística utilizada para teste de hipóteses, que mede a probabilidade de se obter uma estatística de teste igual ou maior que a calculada se a hipótese nula for verdadeira (GALE; ZHANG, 2016). Pode também ser entendido como o menor nível de significância com que se rejeita a hipótese nula (HYNDMAN; ATHANASOPOULOS, 2013). Em todos os modelos, o p-valor será ser usado como medida comparativa nos testes realizados, sendo confrontado um nível de significância ( $\alpha$ ) de 5%. Nos modelos aprovados, efetua-se a previsão com um horizonte de três meses e o teste de acurácia é realizado.

### 3.3 ACEITAÇÃO DOS MODELOS

Um modelo de previsão só é considerado adequado se os resíduos forem normalmente distribuídos e não apresentarem autocorrelação (HYNDMAN; ATHANASOPOULOS, 2013). Para a verificação da existência de autocorrelação é utilizada a função de autocorrelação (ACF), que avalia as relações existentes entre os valores de consumo e seus valores adjacentes, e que pode ser definida como a razão entre a autocovariância e a variância para um conjunto de dados (OLIVEIRA, 2012).

A autocorrelação deve ser analisada visualmente, com observação gráfica (Figura 15), onde os pregos ou picos obtidos nos *lags* não devem ultrapassar a os limites de significância demarcados pelas linhas pontilhadas, e também por meio do teste de Box-Pierce (HYNDMAN; ATHANASOPOULOS, 2013).

Figura 15 – Gráfico da função de autocorrelação de resíduos.

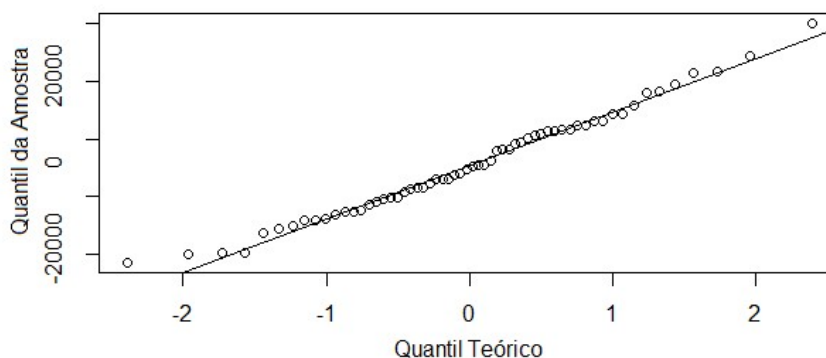


Fonte: a autora (2020).

No teste de Box-Pierce a hipótese nula é de que os resíduos não são correlacionados, e a hipótese alternativa indica que os resíduos se autocorrelacionam. A avaliação das hipóteses é feita em função do p-valor calculado pelo teste e do nível de significância ( $\alpha$ ) adotado. Mais informações sobre o teste de Box-Pierce podem ser encontradas em Hyndman e Athanasopoulos (2013).

A normalidade da distribuição dos resíduos, em todos os modelos desenvolvidos será avaliada pela observação do gráfico quantil-quantil (Figura 16), onde os pontos devem acompanhar a reta diagonal ajustada (WILK; GNANADESIKAN, 1968), e pelos testes de Shapiro-Wilk e Lilliefors (RAZALI; WAH, 2011).

Figura 16 – Gráfico quantil-quantil.



Fonte: a autora (2020).

Nos testes de Shapiro-Wilk e Lilliefors são consideradas duas hipóteses: a hipótese nula é de que os resíduos são normalmente distribuídos; na hipótese alternativa, os resíduos não têm distribuição normal (PEREIRA FILHO, 2013). A avaliação das hipóteses é realizada comparando-se o p-valor calculado pelo teste e o nível de significância ( $\alpha$ ) adotado. Maiores detalhes sobre os testes de distribuição normal podem ser vistos em Pereira Filho (2013) e Razali e Wah (2011).

### 3.4 AVALIAÇÃO DOS MODELOS

Uma das medidas de avaliação consideradas para os modelos, na escolha realizada pelo *software*, é o Critério de Informação de Akaike (AIC). Na comparação dos modelos, valores menores de AIC indicam modelos mais próximos à realidade, com menor perda de informação (GUAJARATI, 2006). O AIC é definido conforme a equação 15 (EHLERS, 2009).

$$AIC_{SE} = -2 \log(L) + 2m \quad (15)$$

Onde:

$L$  é a função de máxima verossimilhança do modelo ajustado;

$m$  é o número de parâmetros do modelo.

A avaliação dos erros de previsão dos modelos foi realizada com base no menor erro percentual absoluto médio (MAPE). O MAPE foi adotado como principal critério de escolha dos modelos neste estudo, por ser considerado relevante na comparação de modelos de previsão (ARANDIA et al., 2016), de fácil interpretação e largamente utilizado em estudos de modelagem de demanda de água (BILLAH et al., 2006; YALÇINTAŞ et al., 2015; AL-ZHARANI; ABO-MONASAR, 2015; ARANDIA et al., 2016). O cálculo do erro percentual absoluto médio é apresentado na equação 16 (AL-ZHARANI; ABO-MONASAR, 2015).

$$MAPE = \frac{\sum_{t=1}^n \frac{(y_t - \hat{y}_t)}{\mu_t}}{n} 100 \quad (16)$$



Onde:

$y_t$  é o valor observado;

$\hat{y}$  é o valor previsto;

$n$  é o tamanho da amostra.

### 3.5 ESCOLHA DOS MODELOS

Neste trabalho, a escolha do melhor modelo para cada categoria de consumo foi guiada pela comparação dos valores do erro percentual absoluto médio (MAPE) de treinamento (obtido na modelagem) e de teste (obtido na verificação da acurácia). O MAPE de treinamento indica como o modelo se ajusta aos valores passados, enquanto o MAPE de teste aponta quão preciso é o modelo na previsão dos valores futuros. Essa comparação se faz necessária pois um modelo que se ajusta muito bem a uma série de dados não necessariamente realizará boas previsões (HYNDMAN; ATHANOSOPOULOS, 2013).

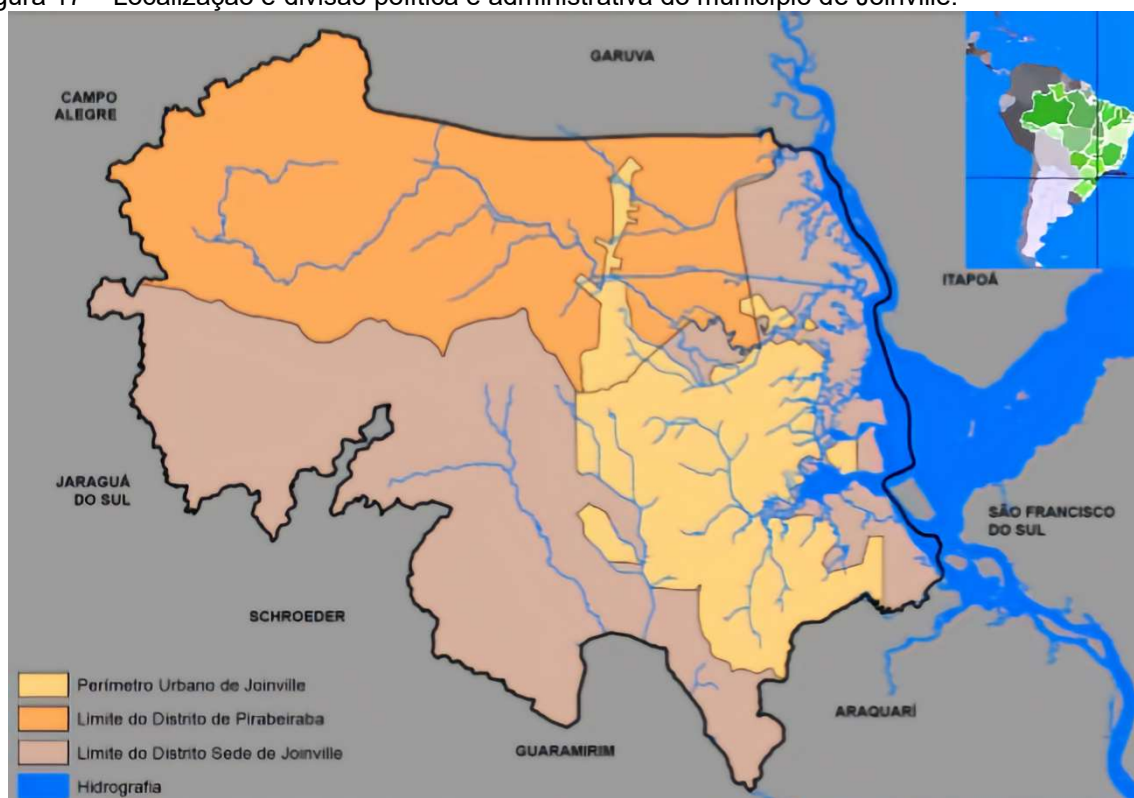
Considerando que em modelos de previsão, o foco não deve estar voltado a uma boa descrição do passado, mas a uma boa previsão dos valores ainda não vistos (KOURENTZES, 2016), propõe-se, que neste trabalho, o modelo com menor valor de MAPE de teste para cada categoria deva ser escolhido, definindo qual método de análise de séries temporais é mais assertivo para a realização das previsões de consumo de água.

Os scripts elaborados no R para cada um dos modelos estão nos Apêndices indicados de A a E, conforme o desenvolvimento do estudo de caso deste trabalho.

#### 4 ESTUDO DE CASO

Localizada na região nordeste do estado de Santa Catarina, Joinville (Figura 17) é formada pelo distrito de Pirabeiraba e pelo distrito sede de Joinville, e possui uma área territorial de 1.135,05 km<sup>2</sup>, dos quais 212,6 km<sup>2</sup> são de área urbana (IBGE, 2018).

Figura 17 – Localização e divisão política e administrativa do município de Joinville.



Fonte: SEPUD (2014).

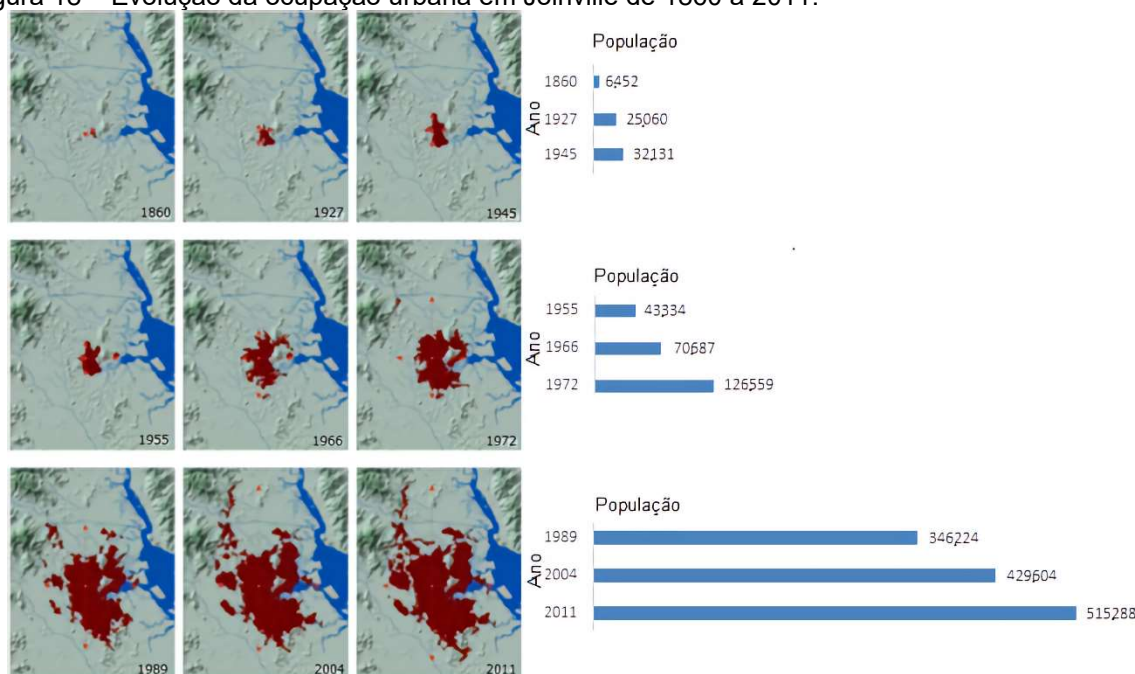
Fundada em 1851, com a chegada dos primeiros imigrantes europeus, a já existente colônia Dona Francisca deu origem ao que hoje é conhecido como município de Joinville, sendo desmembrada politicamente de São Francisco do Sul em 1866. A partir da década de 1970 com o surgimento das primeiras indústrias têxteis e metalúrgicas, o município começou a se desenvolver (SEPUD, 2017).

No início do século XX a inauguração da estrada de ferro ligando São Paulo a São Francisco do Sul, o surgimento da energia elétrica, e o desenvolvimento de um sistema de transporte coletivo aceleraram o desenvolvimento da cidade. Mais tarde, entre as décadas de 1950 e 1980, com o fim da Segunda Guerra Mundial e o corte de

importação de produtos europeus, a cidade se transformou em um dos principais polos industriais da região (SEPUD, 2014).

Segundo dados do Instituto Brasileiro de Geografia e Estatística - IBGE (2018), a partir da década de 50, pela intensificação do desenvolvimento industrial e econômico da cidade, a população urbana do município ultrapassou a população rural, e na década de 60, a cidade passou a apresentar taxas de crescimento demográfico na faixa de 6%, duas vezes maiores que as verificadas no estado e no país. A Figura 18, mostra a evolução da área de ocupação urbana e a população entre os anos de 1860 a 2011.

Figura 18 – Evolução da ocupação urbana em Joinville de 1860 a 2011.



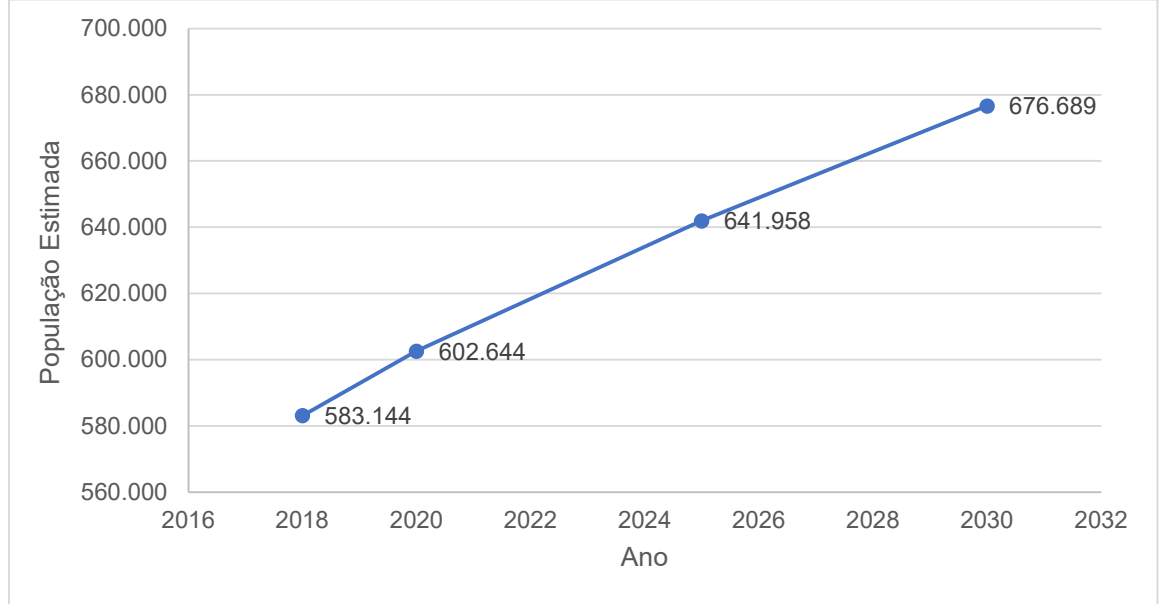
Fonte: adaptado de SEPUD (2017).

A população registrada pelo censo de 2010, realizado pelo IBGE, foi de 515.288 habitantes, sendo 497.788 moradores da área urbana. Em 2016 a população registrada foi de 569.645 habitantes, com 550.272 pessoas vivendo na área urbana. Estimou-se que em 2017 população total da cidade seria de 577.077 habitantes (IBGE, 2018). A Figura 19 mostra as previsões populacionais de 2018 até 2030.

A população do município, de acordo com o censo de 2010, é formada por 50,3% de mulheres e 49,7% de homens. A faixa etária dos habitantes registrada pelo IBGE (2010) mostra que cerca de 50% da população tem idade inferior a 30 anos. Em

relação à escolaridade, apenas pouco mais de 11% da população teve algum acesso a instrução superior (SEPUD, 2017).

Figura 19 – Estimativa populacional de Joinville de 2018 a 2030.



Fonte: adaptado de IBGE (2018); SEPUD (2017).

O perfil de rendimento da população de Joinville em uma amostra de 160 mil residências, mostra que em 2010 mais de 80% dos habitantes ganhava até 3 salários mínimos. Em 2015, o salário médio mensal era de 3,0 salários mínimos. A proporção de pessoas ocupadas em relação à população total era de 40,2% (IBGE, 2018).

Na economia, Joinville é hoje o mais importante polo econômico, tecnológico e industrial do Estado, e o maior parque fabril de Santa Catarina (SEPUD, 2017). Conforme a Prefeitura Municipal de Joinville (PMJ) a cidade concentra grande parte da atividade econômica na indústria, com destaque para os setores metalmecânico, têxtil, plástico, metalúrgico, químico e farmacêutico (PMJ, 2016a). A cidade apresentou em 2015 (Tabela 7) um PIB per capita de R\$ 47,2 mil, em um total de R\$ 26,5 bilhões, com queda de 1,3 bilhões de reais em 2016 (IBGE, 2018).

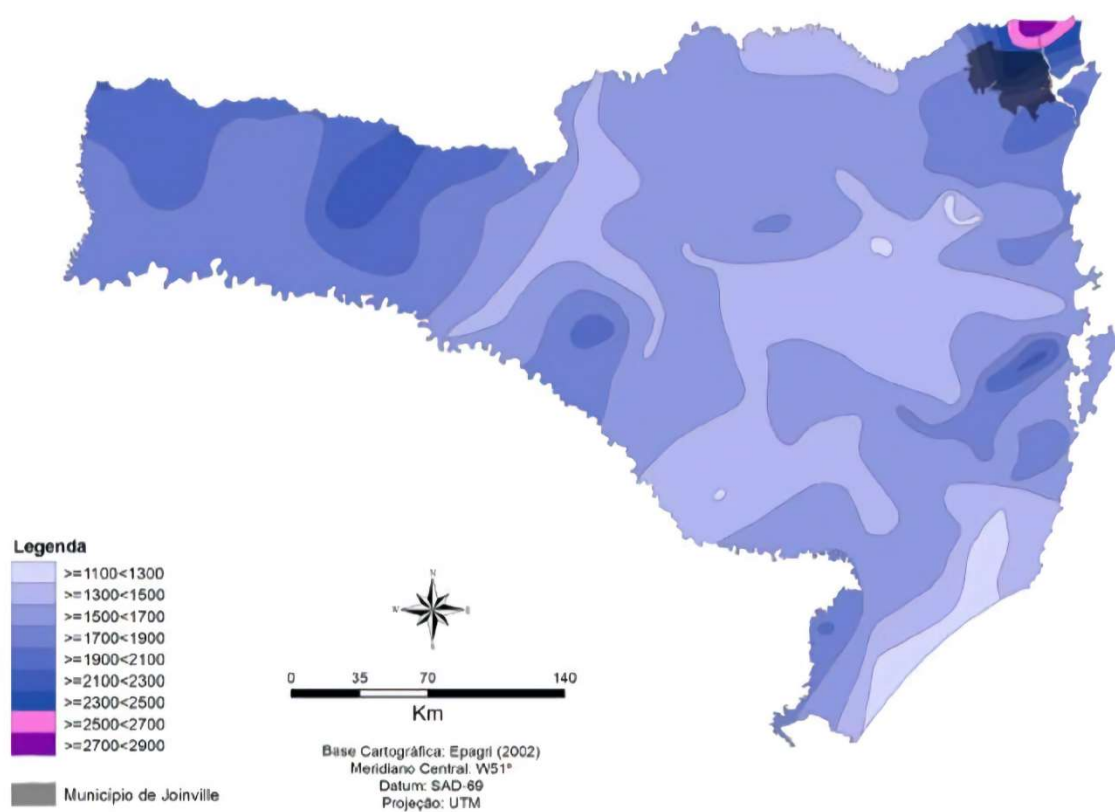
Tabela 6 – Evolução do PIB de Joinville, de 2012 a 2015.

Ano	PIB a preço corrente (bilhões de reais)	PIB per capita (reais)
2012	20,472	38.896
2013	22,049	40.311
2014	25,136	45.323
2015	26,552	47.233
2016	25.217	44.268

Fonte: adaptado de IBGE (2018), DATASUS (2017).

De acordo com a classificação de Köppen, o clima predominante na região é do tipo “mesotérmico, úmido, sem estação seca”. A umidade relativa média anual do ar é de 76,04% (SEPUD, 2017). O regime de chuvas de Joinville varia entre 1700 mm e 2700 mm anualmente (Figura 20), apresentando um dos maiores índices de pluviosidade entre as regiões de Santa Catarina (PMJ, 2016a).

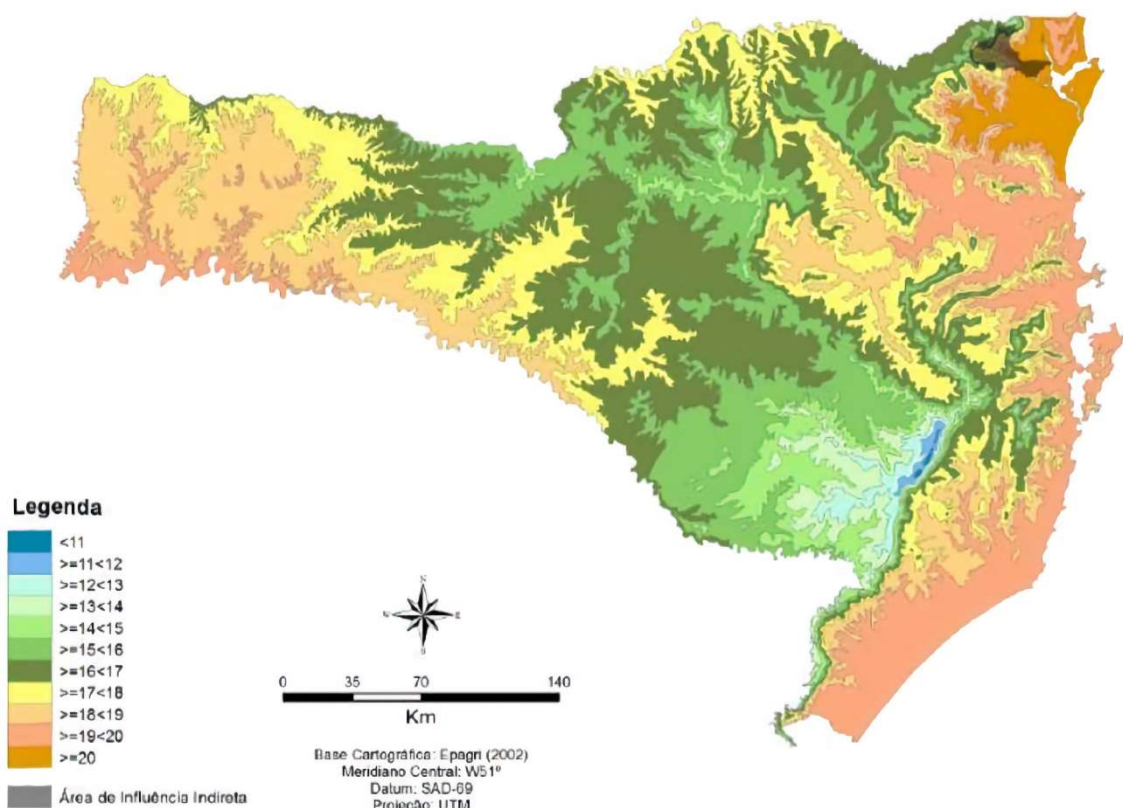
Figura 20 – Distribuição da precipitação total anual no estado de Santa Catarina.



Fonte: PMJ (2016a).

As temperaturas médias anuais na cidade encontram-se acima de 20° C (Figura 21). As maiores temperaturas ocorrem nos primeiros meses do ano, onde as médias mensais apresentam-se acima de 25°C, destacando-se o mês de fevereiro. As temperaturas mais baixas estão relacionadas ao período entre junho a setembro, onde as médias mensais não ultrapassam os 20°C (PMJ, 2016a).

Figura 21 – Distribuição da temperatura média anual no estado de Santa Catarina.



Fonte: PMJ (2016a).

O regime hídrico das bacias hidrográficas do município é favorecido pela associação de sua geomorfologia, condições climáticas e cobertura vegetal. Joinville conta com 7 bacias hidrográficas: Rio Cubatão, Rio Palmital, Vertente Leste, Vertente Sul, Rio Cachoeira, Rio Piraí e Rio Itapocuzinho (Figura 22). A captação das águas para abastecimento da cidade é realizada pela Companhia Águas de Joinville (CAJ) nos Rios Cubatão e Piraí, onde se localizam as 2 estações de tratamento (SEPUD, 2017).



Figura 22 – Bacias hidrográficas de Joinville.

### Todas as bacias hidrográficas de Joinville

Joinville conta com 7 bacias hidrográficas

■ Zona rural  
■ Zona urbana

1 - Rio Cubatão

2 - Rio Palmital

3 - Vertente Leste - Rios Comprido, Guaxanduva, Iriú-Mirim, do Ferro, Iriú-Guaçu e Cubatãozinho

4 - Vertente Sul - Rios Santinho, Velho, Buguaçu e Paranaguamirim

5 - Rio Cachoeira

6 - Rio Piraí

7 - Rio Itapocuzinho

A - Estação de tratamento do Rio Cubatão

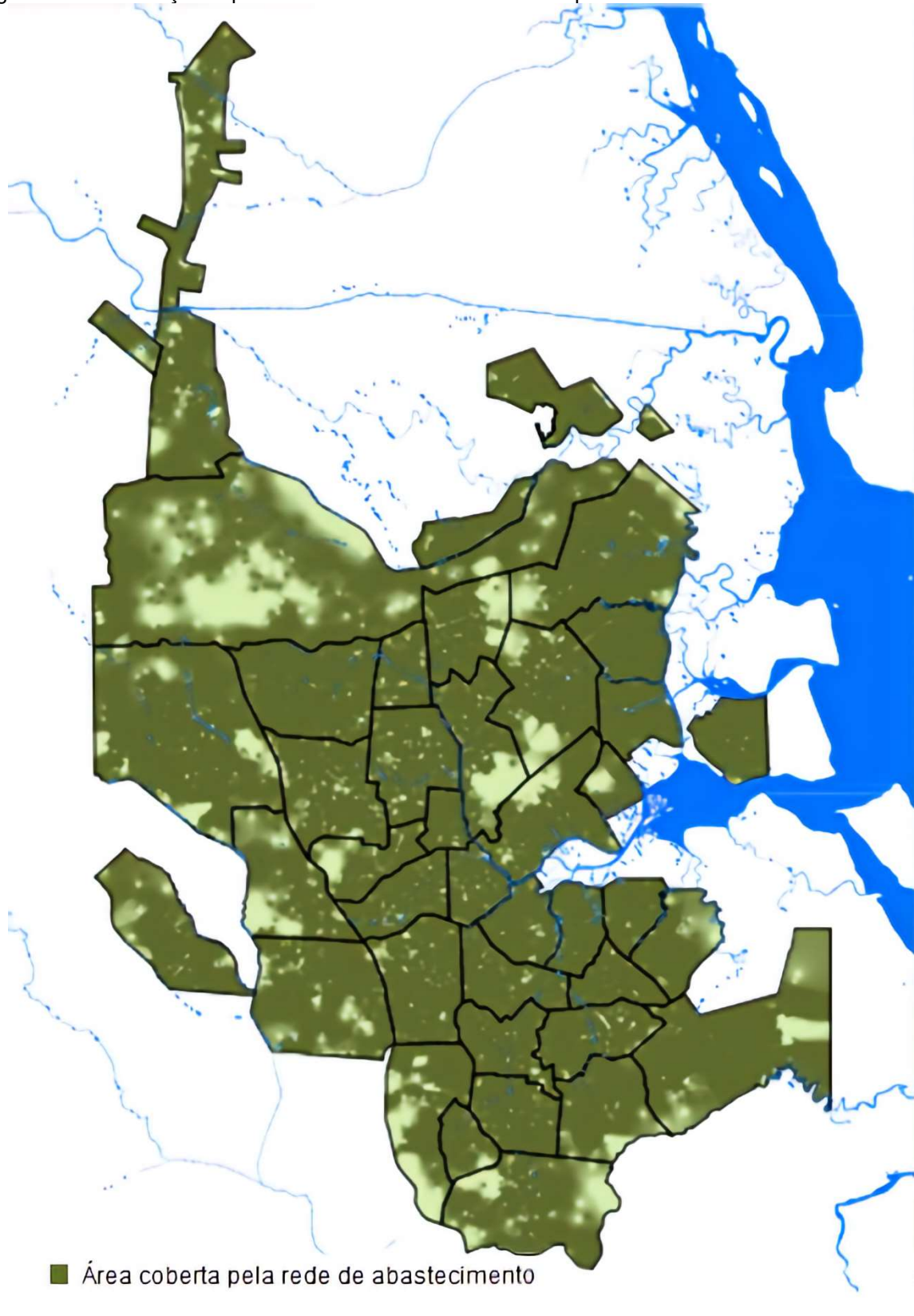
B - Estação de tratamento do Rio Piraí



Fonte: adaptado de SEPUD (2017).

Joinville é abastecida pelos mananciais de água localizados nas bacias hidrográficas do Rio Cubatão e do Rio Piraí. O abastecimento é realizado pela Companhia Águas de Joinville (CAJ), e atende mais de 98% das residências do município (Figura 23) (PMJ, 2016b).

Figura 23 – Distribuição espacial da rede de abastecimento no perímetro urbano de Joinville.



Fonte: adaptado de PMJ (2016b).



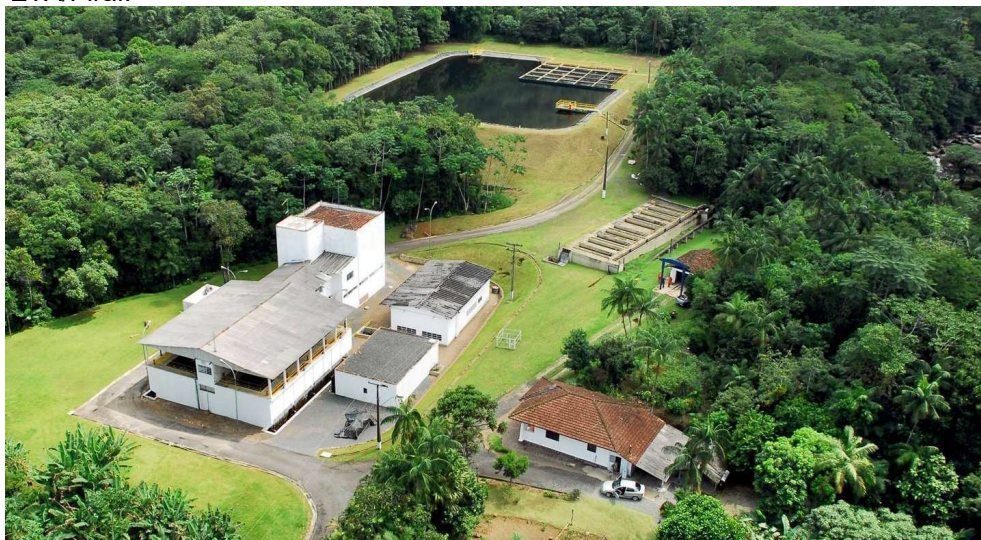
A Companhia possui 11 reservatórios de água tratada distribuídos por toda a cidade, com capacidade nominal de 37 milhões de litros de água (CAJ, 2017). A CAJ possui duas estações de tratamento de água (ETA): a ETA Cubatão (Figura 24) é responsável por 70% da água consumida e a ETA Pirai (Figura 25) pelos outros 30% (SEPUD, 2017).

Figura 24 – ETA/Cubatão.



Fonte: Google (2017).

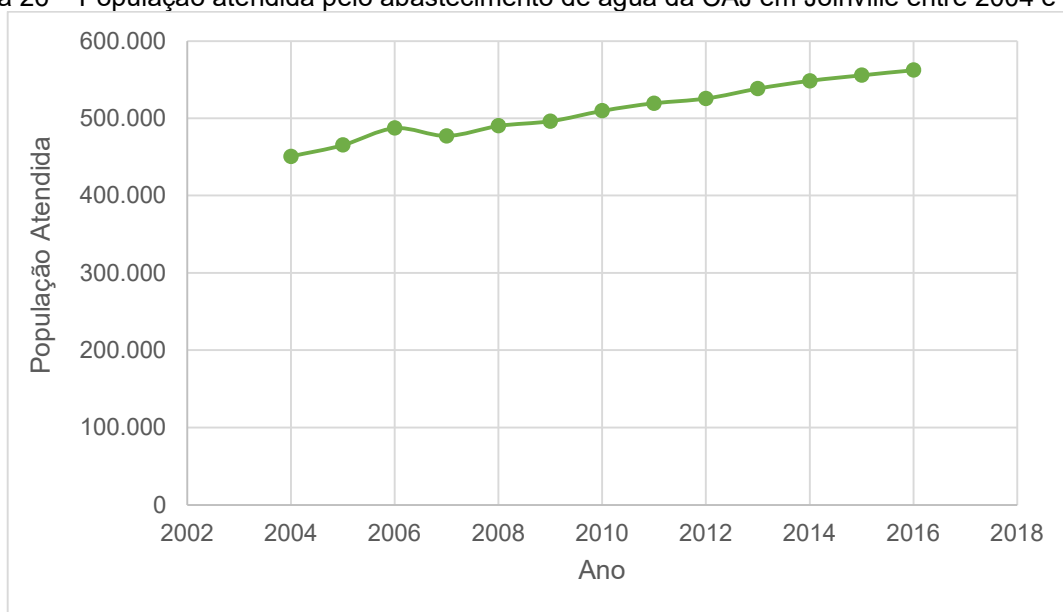
Figura 25 – ETA/Pirai.



Fonte: Google (2017).

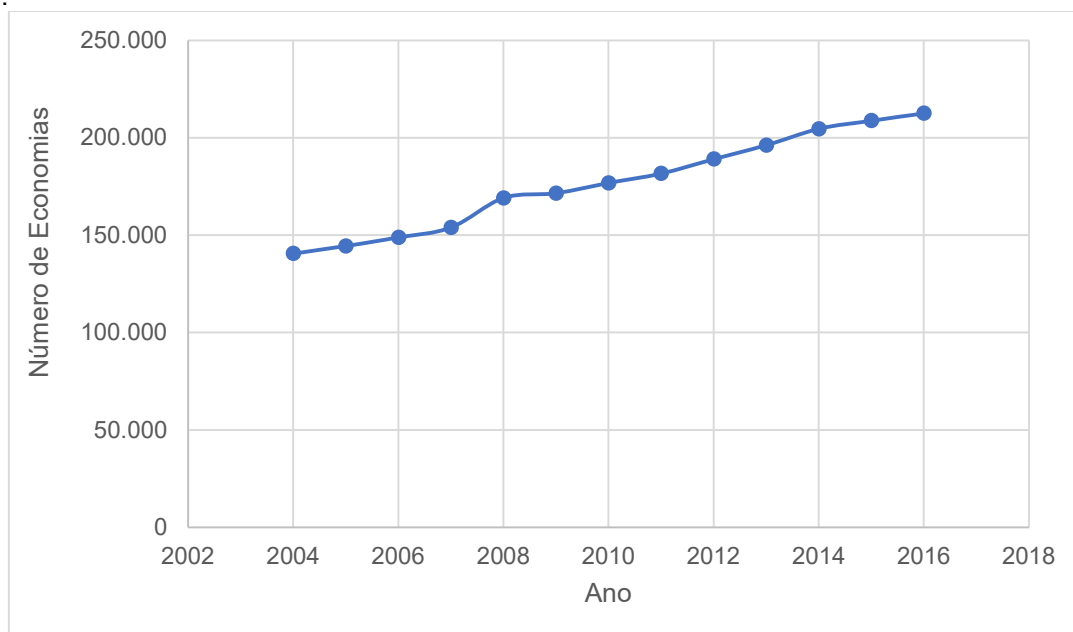
Anualmente, a CAJ apresenta ao Sistema Nacional de Informações sobre Saneamento (SNIS) informações referentes à sua operação. Esses dados podem ser observados nas Figuras 26 e 27.

Figura 26 – População atendida pelo abastecimento de água da CAJ em Joinville entre 2004 e 2016.



Fonte: adaptado de SNIS (2017).

Figura 27 – Número de economias de água ativas registrado pela CAJ em Joinville entre 2004 e 2016.



Fonte: adaptado de SNIS (2017).

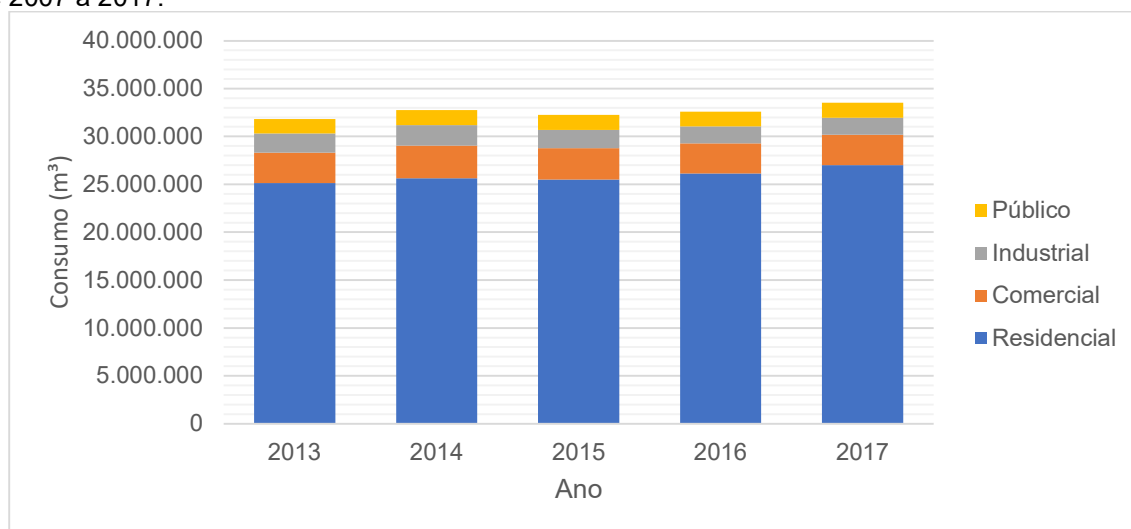
Em relação ao consumo, os dados utilizados neste estudo referem-se ao consumo mensal de água micromedido por economia, no ambiente urbano da cidade de Joinville, no período de janeiro de 2007 a dezembro de 2017, e foram fornecidos pela CAJ, separados em 4 categorias de consumo: residencial, comercial, industrial e público.

A Companhia Águas de Joinville enquadra as economias dentro das quatro categorias de consumo conforme os artigos 13 a 18 da resolução nº 52/2015 elaborada pela Prefeitura Municipal de Joinville (PMJ, 2015).

São consideradas economias residenciais: todos os imóveis destinados exclusivamente para fins de moradia. Na categoria industrial enquadram-se os destinados a atividades de produção e/ou transformação, empreendimentos, conjuntos habitacionais e condomínios em fase de construção que possuem área maior que 750 m<sup>2</sup>. Na categoria comercial, enquadram-se os imóveis destinados ao exercício de atividades comerciais e/ou de serviços, imóveis com ligações de caráter temporário e demais imóveis que não se enquadram nas outras categorias. Na categoria público são consideradas as economias cujos serviços de abastecimento de água são utilizados por órgãos da administração pública municipal, estadual e federal (PMJ, 2015).

Conforme os dados da CAJ (2018) (Figura 28), em todos os anos, a categoria residencial é responsável pela maior parte do consumo de água na cidade de Joinville (na faixa de 80% do volume total consumido), seguida das categorias comercial (cerca de 10%) e industrial (aproximadamente 6%). A categoria público é a que representa a menor taxa de consumo em todos os anos (na faixa de 4%).

Figura 28 – Consumo anual total de água em Joinville, dividido por categoria de consumo, nos anos de 2007 a 2017.

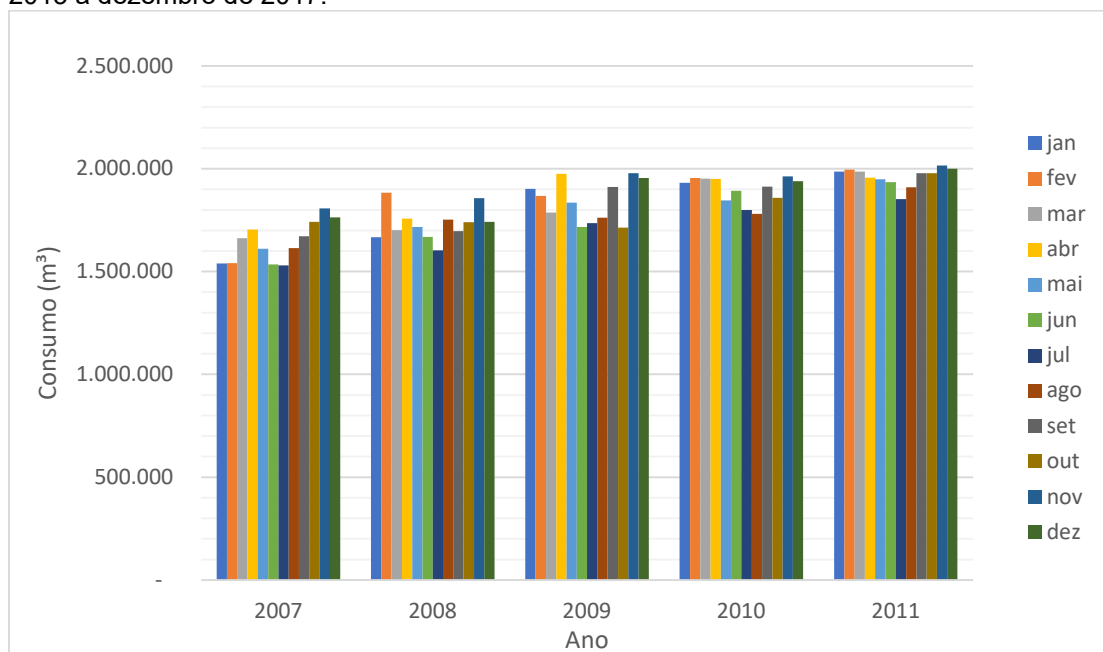


Fonte: adaptado de CAJ (2018).

A Figura 29 mostra o consumo residencial mês a mês no período de janeiro de 2013 a dezembro de 2017. É possível observar que o consumo de água desta

categoria apresenta uma tendência de crescimento. De maneira geral, os maiores valores de consumo são registrados nos meses de verão (dezembro, janeiro e fevereiro), e os menores, nos meses de inverno (junho, julho e agosto). No período, o maior valor de consumo, 2.400.382 m<sup>3</sup>, foi registrado no mês de fevereiro de 2014, e o menor em agosto de 2013 (1.986.054 m<sup>3</sup>). A média de consumo da categoria, no período, é de 2.156.411 m<sup>3</sup> ao mês.

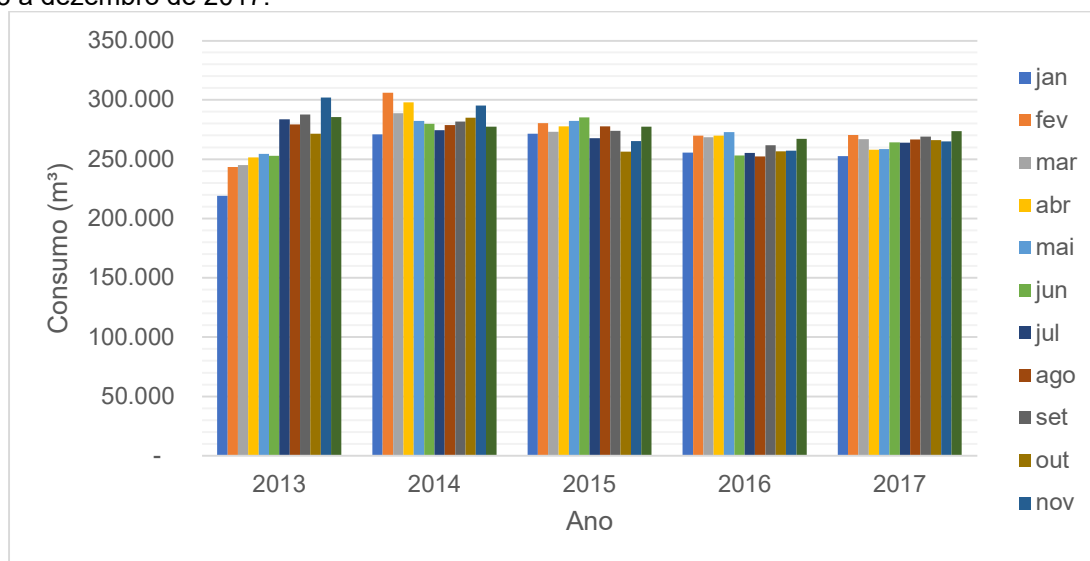
Figura 29 – Consumo mensal de água na categoria residencial em Joinville no período de janeiro de 2013 a dezembro de 2017.



Fonte: adaptado de CAJ (2018).

Os dados de consumo comercial mensal de água em Joinville, no período, são apresentados na Figura 30. É possível observar uma tendência de crescimento do consumo nos anos de 2013 e 2014. A partir daí, é visível um decréscimo do consumo nos anos de 2016 e 2017, quando comparados aos anos anteriores. O maior valor de consumo na categoria no período de janeiro de 2013 a dezembro de 2017 foi registrado em fevereiro de 2014 (305.956 m<sup>3</sup>), e o menor valor consumido ocorreu em janeiro de 2013 (219.056 m<sup>3</sup>). A média de consumo na categoria, no período analisado, foi de 269.964 m<sup>3</sup>.

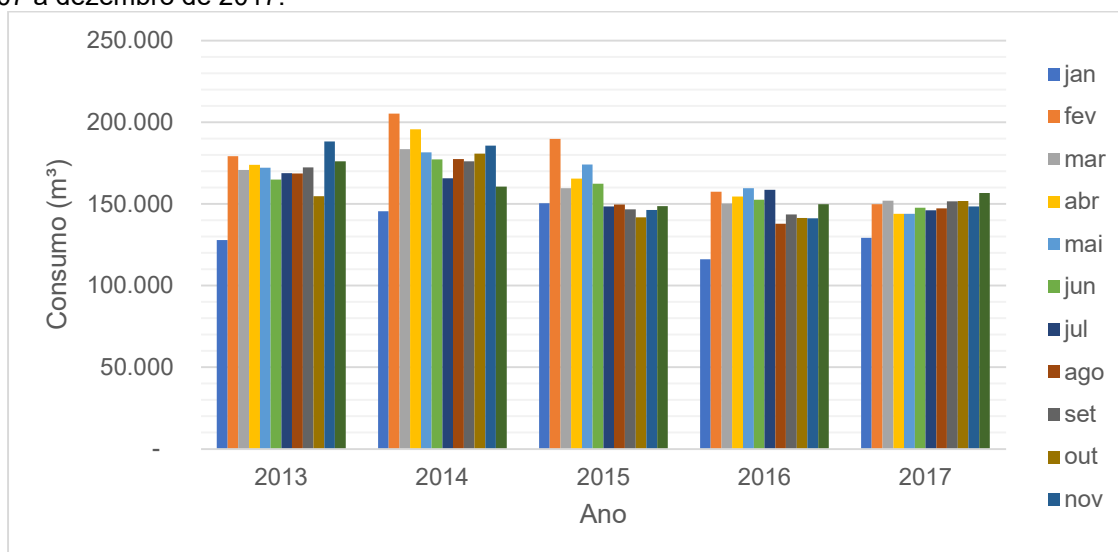
Figura 30 – Consumo mensal de água na categoria comercial em Joinville no período de janeiro de 2013 a dezembro de 2017.



Fonte: adaptado de CAJ (2018).

A Figura 31 apresenta os dados mensais do consumo industrial de água no município. É possível verificar um consumo maior no período de fevereiro de 2013 a junho de 2015 e um decréscimo nos anos seguintes. Pode se observar também que no ano de 2017 o consumo mensal se mostrou mais homogêneo quando comparado aos outros anos. A média do consumo de água na categoria, no período de janeiro de 2013 a dezembro de 2017 foi de 159.463 m³. O menor valor consumido foi de 116.044 m³ em janeiro de 2016, e o maior valor registrado foi de 205.387 m³ em fevereiro de 2014.

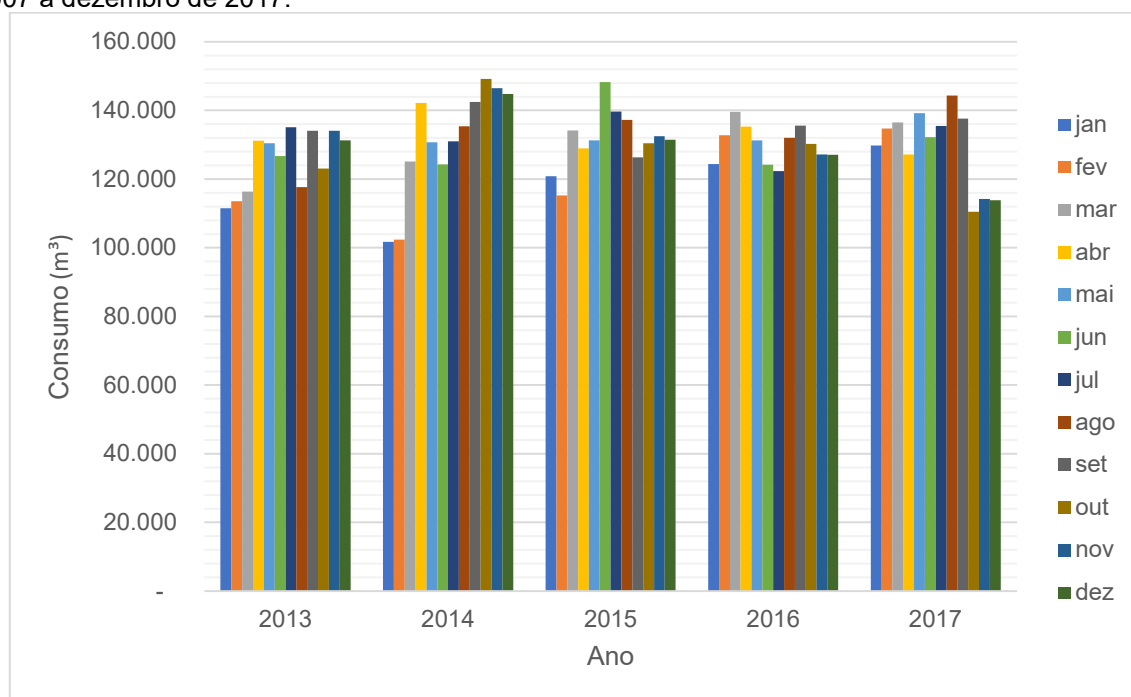
Figura 31 – Consumo mensal de água na categoria industrial em Joinville no período de janeiro de 2007 a dezembro de 2017.



Fonte: adaptado de CAJ (2018).

O comportamento de consumo de água da categoria público pode ser observado na Figura 32. De maneira geral, é possível observar um aumento no consumo em 2014, com uma leve queda nos anos seguintes. Houve um crescimento no primeiro trimestre de cada ano a partir de 2014, e nos demais períodos o consumo se manteve acima dos 120.000 m<sup>3</sup>, com exceção do último trimestre de 2017 que registrou um consumo médio na faixa de 115.000 m<sup>3</sup>. O maior valor de consumo na categoria no período de janeiro de 2013 a dezembro de 2017 foi registrado em outubro de 2014 (149.150 m<sup>3</sup>) e o menor valor consumido ocorreu em janeiro de 2014 (101.699 m<sup>3</sup>). A média de consumo na categoria, no período analisado, foi de 129.571 m<sup>3</sup>.

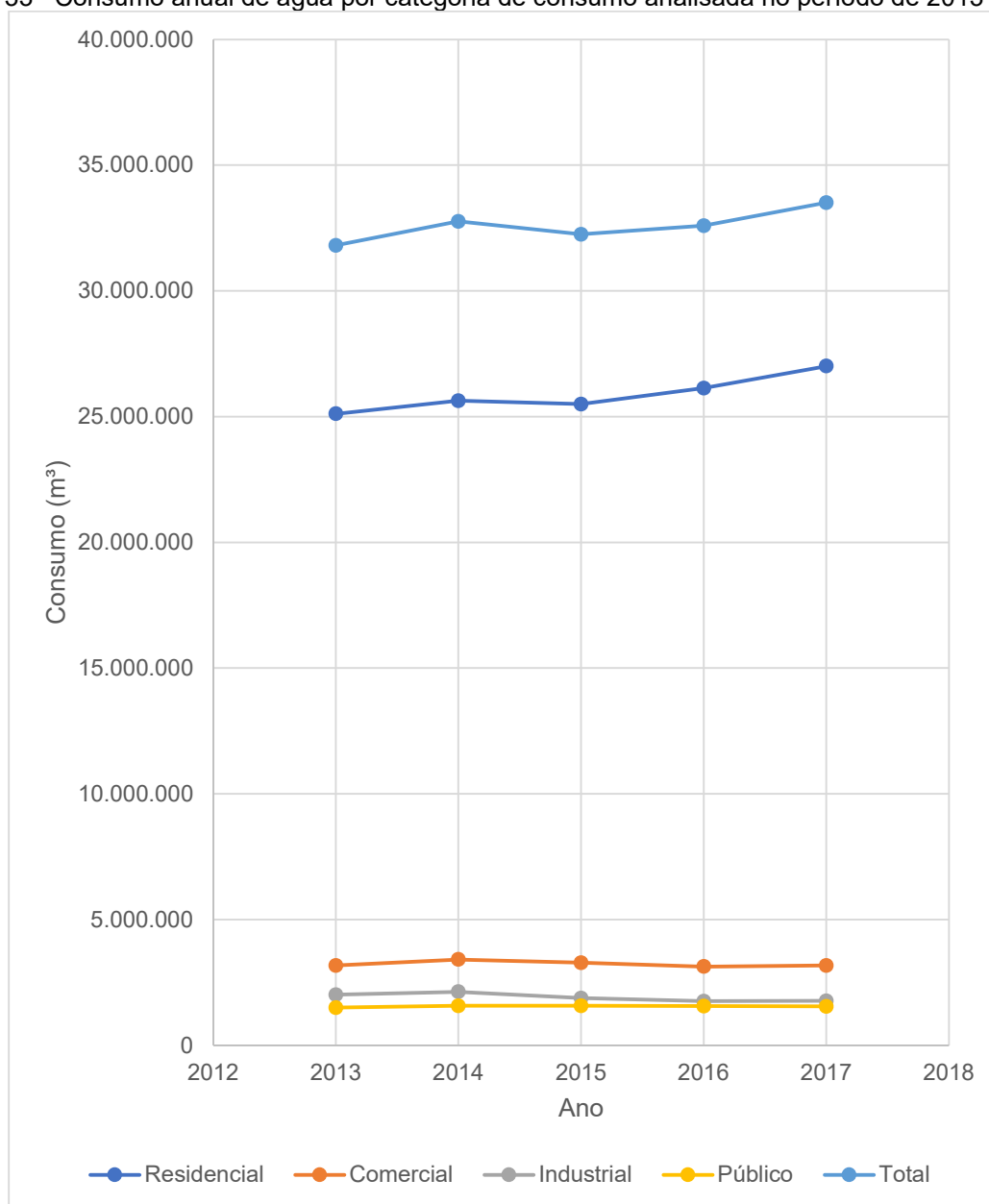
Figura 32 – Consumo mensal de água na categoria público em Joinville no período de janeiro de 2007 a dezembro de 2017.



Fonte: adaptado de CAJ (2018).

A Figura 33 mostra o consumo anual em cada categoria, mais a soma de todas as categorias, identificada como total, entre os anos de 2013 a 2017.

Figura 33 –Consumo anual de água por categoria de consumo analisada no período de 2013 a 2017.



Fonte: adaptado de CAJ (2018).

Observando o consumo de água no município como um todo, é possível verificar que o consumo de água vinha apresentando uma tendência de crescimento até o final de 2014. Durante 2015 houve uma queda no consumo nas categorias, exceto na categoria público. Conforme a CAJ (2017), essa diminuição no consumo foi originada pela crise econômica e maior conscientização da população em relação à economia de água, devido à crise hídrica que ocorreu em alguns estados brasileiros no período. No entanto, a partir de 2016 os valores voltaram a crescer e é esperado que essa tendência se mantenha nos próximos anos (CAJ, 2017).



## 5 RESULTADOS

Os resultados descritos neste capítulo tratam da aplicação da metodologia desenvolvida neste trabalho para estudo do consumo de água da cidade de Joinville. No estudo de caso foi organizado um banco de dados com os consumos mensais micromedidos por economia, separados nas categorias residencial, comercial, industrial, e público, fornecidos pela Companhia Águas de Joinville. Foi incluída também a categoria total, representada pela soma dos consumos de todas as categorias. Os dados para a modelagem foram do período de janeiro de 2013 a dezembro de 2017, totalizando 5 anos de medição, com 60 entradas para cada categoria, respeitando o número mínimo de dados para os modelos ARIMA, que deve ser de 50 observações. Os dados para a realização dos testes de acurácia foram os consumos do período de janeiro a março de 2018.

Na modelagem, foram utilizados 10 modelos, sendo um modelo de suavização exponencial e um modelo ARIMA para cada uma das cinco categorias analisadas. Esses modelos foram identificados como: Residencial com Suavização Exponencial (RSE), Residencial com ARIMA (RA), Comercial com Suavização Exponencial (CSE), Comercial com ARIMA (CA), Industrial com Suavização Exponencial (ISE), Industrial com ARIMA (IA), Público com Suavização Exponencial (PSE), Público com ARIMA (PA), Total com Suavização Exponencial (TSE) e Total com ARIMA (TA).

Após a modelagem, foram analisadas a distribuição normal e a existência de autocorrelação nos resíduos dos modelos. A distribuição normal foi verificada por meio da visualização do gráfico quantil-quantil e por meio dos testes de Shapiro-Wilk e Lilliefors. Neste trabalho, para que a distribuição fosse considerada normal, foi necessário que além da confirmação de normalidade por meio do gráfico quantil-quantil, o p-valor estivesse acima de 5% em pelo menos um dos testes estatísticos realizados.

Neste trabalho, a ausência de autocorrelação foi aceita quando o gráfico da ACF não apresentava pregos acima da linha pontilhada, ou quando havia no máximo um prego tocando ou ultrapassando levemente a linha pontilhada conforme Razali *et al.* (2018), desde que o teste de Box-Pierce apresentasse p-valor acima de 5%.

A previsão do consumo foi realizada com um horizonte de três meses. Os resultados preditos foram comparados com os dados de consumo de cada categoria

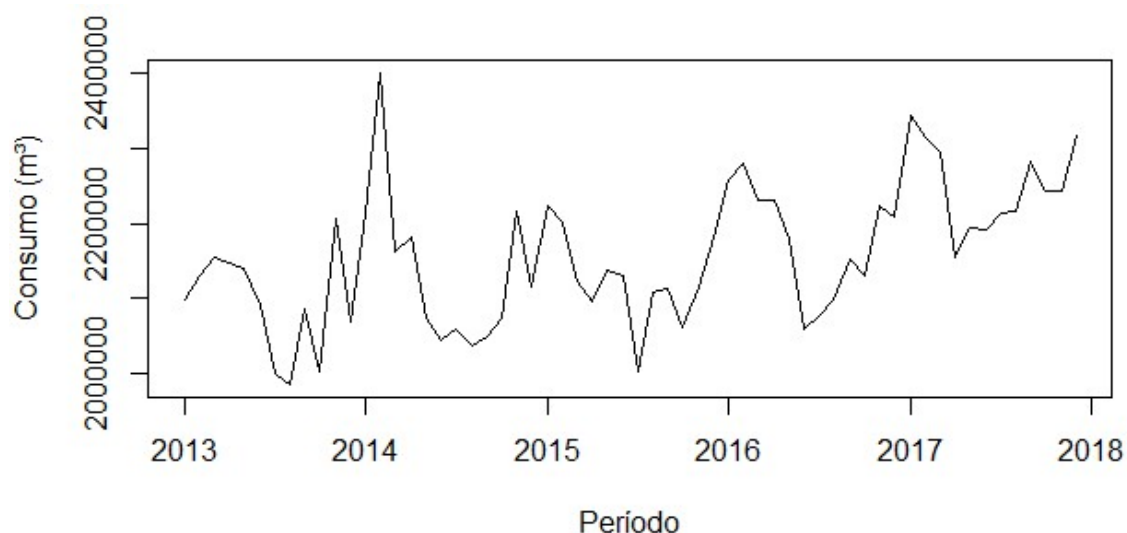


nos meses de janeiro, fevereiro e março de 2018, para verificar a acurácia dos modelos. Na verificação foram comparados os valores do MAPE de treinamento (obtido na modelagem) e do MAPE de teste (obtido quando os valores preditos são comparados com os valores de consumo real. Os modelos com menor valor de MAPE de teste foram escolhidos como mais adequados para a realização das previsões.

## 5.1 CATEGORIA RESIDENCIAL

A série temporal de consumo residencial de água em Joinville é mostrada na Figura 34. É possível observar que a série apresenta uma tendência de crescimento contínua no período analisado. Com os maiores valores de consumo registrados nos meses de verão (fim e início de cada ano), e os menores valores nos meses de inverno.

Figura 34 – Série temporal do consumo residencial de água em Joinville de janeiro de 2013 a dezembro de 2017.



Fonte: a autora (2020).

As estatísticas descritivas da série de consumo residencial de água são apresentadas na tabela 8.

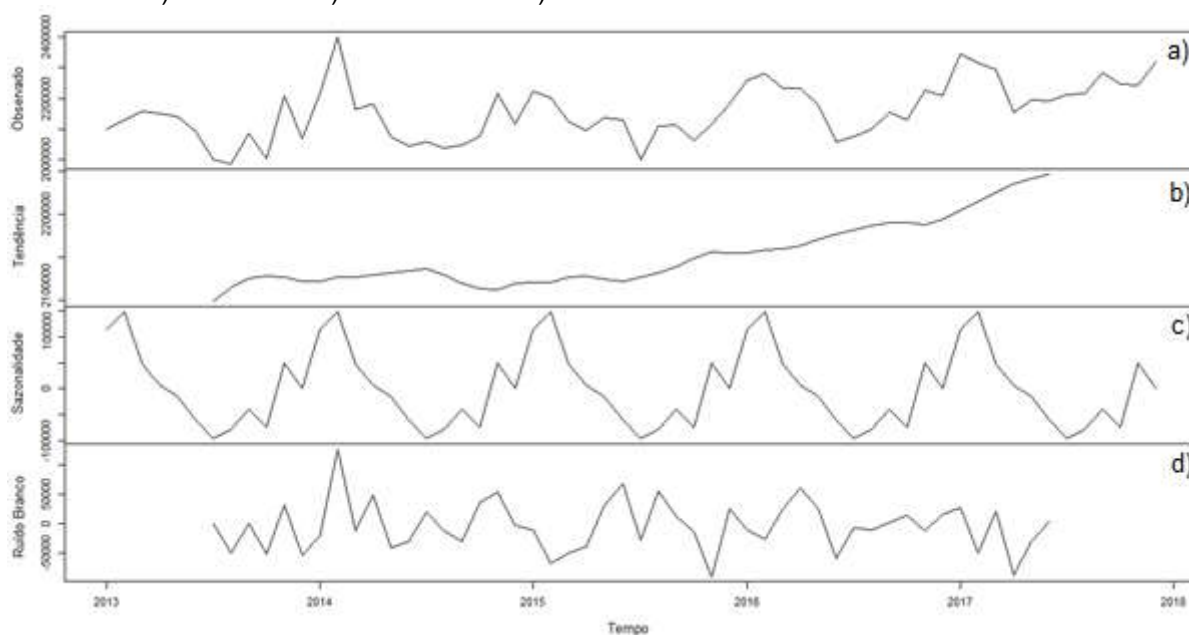
Tabela 7 – Estatísticas descritivas da série temporal de consumo da categoria residencial.

	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo
m³	1.986.054	2.090.619	2.150.993	2.156.411	2.215.871	2.400.832

Fonte: a autora (2020).

A série temporal não possuía dados faltantes e o consumo apresentou distribuição normal, com p-valor de 0,6436 no teste de Shapiro-Wilk. Foi observada também a existência de sazonalidade no consumo (Figura 35(c)).

Figura 35 – Decomposição da série de dados de consumo residencial de água em Joinville. a) Observado. b) Tendência. c) Sazonalidade. d) Ruído Branco.



Fonte: a autora (2020).

### 5.1.1 Suavização exponencial

O primeiro modelo produzido para a categoria foi por meio da suavização exponencial. Identificado como RSE, o modelo definido pelo *software* foi do tipo (M, A, A) com erro multiplicativo, tendência aditiva e sazonalidade aditiva. As equações 17 a 21 foram utilizadas para o desenvolvimento do modelo (MIRANDA, 2009).

$$e_t = (y_t - \mu_t) / \mu_t \quad (17)$$

$$(18)$$

$$\mu_t = l_{t-1} + b_{t-1} + S_{t-m}$$

$$l_t = \mu_t(1 + \alpha e_t) - S_{t-m} \quad (19)$$

$$b_t = \mu_t(1 + \beta^* e_t) - (l_{t-1} + S_{t-m}) \quad (20)$$

$$S_t = \mu_t(1 + \gamma e_t) - (l_{t-1} + b_{t-1}) \quad (21)$$

Onde:

$e_t$  é o erro de previsão um passo à frente;

$y_t$  é o dado observado no período  $t$ ;

$\mu_t$  é o valor esperado de  $y_t$ ;

$l_t$  é o nível da série no período  $t$ ;

$b_t$  é o crescimento da série no período  $t$ ;

$S_t$  é a sazonalidade da série no período  $t$ ;

$m$  é o tamanho do ciclo sazonal;

$\alpha$  é o parâmetro de suavização para o nível;

$\beta^* = \alpha\beta$ ;

$\beta$  é o parâmetro de suavização para o crescimento;

$\gamma$  é o parâmetro de suavização para a sazonalidade.

A tabela 9 apresenta os valores dos parâmetros de nível, tendência, sazonalidade e amortecimento, e o erro padrão ( $\sigma$ ) obtidos para o modelo.

Tabela 8 – Parâmetros de suavização e erro padrão do modelo RSE.

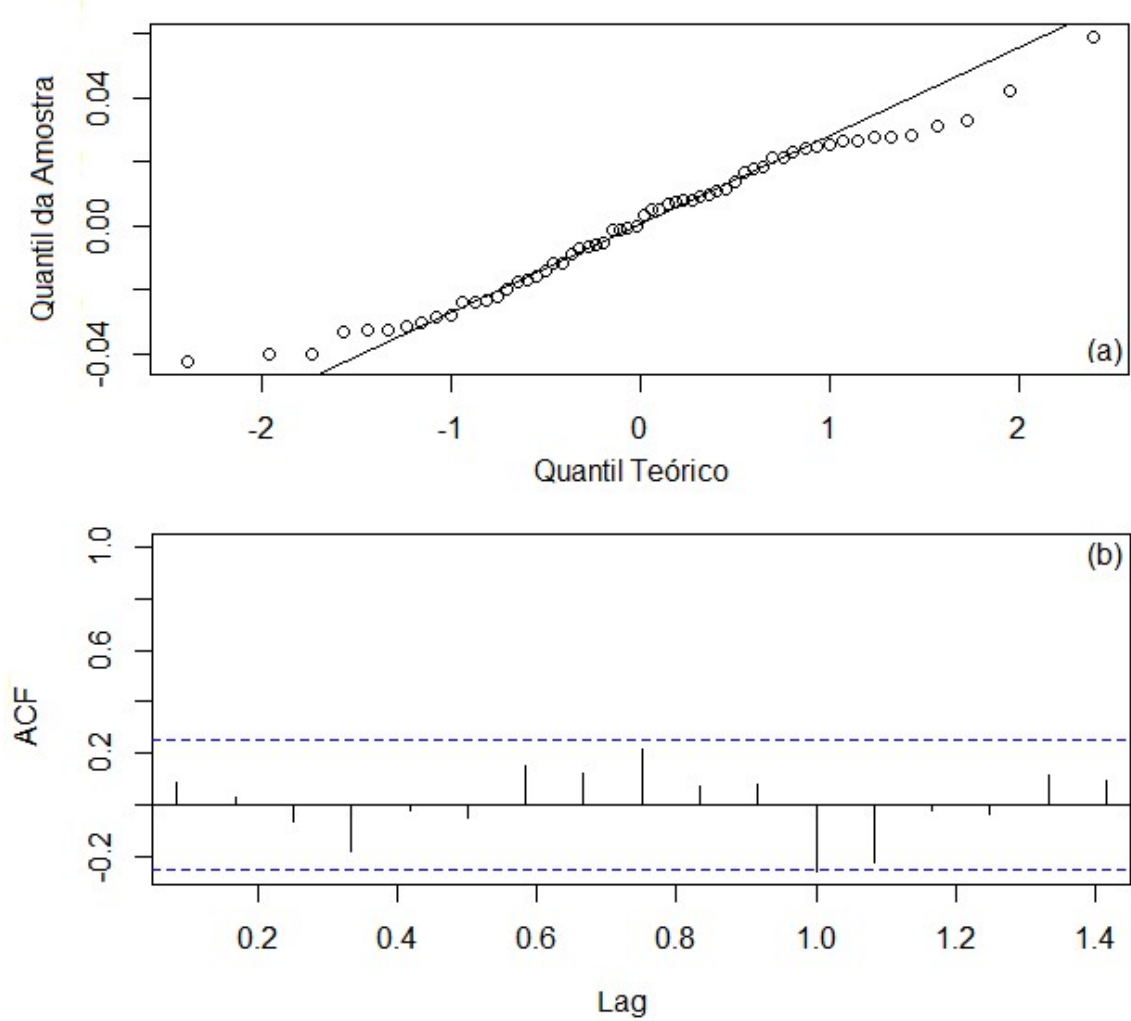
Parâmetros	RSE
$\alpha$	0,1374
$\beta$	1,00E-04
$\gamma$	1,00E-04
$\sigma$	0,0269

Fonte: a autora (2020).

Os resultados obtidos nos testes de aceitação do modelo (Figura 36) indicaram que os resíduos do modelo RSE possuem distribuição normal, com p-valor obtido no teste de Shapiro-Wilk igual a 0,2374 e 0,6918 no teste de Lilliefors. O gráfico da ACF

(Figura 36(b)) mostra um pequeno prego tocando a linha pontilhada no *lag* 1.0 que poderia indicar autocorrelação, mas considerando os resultados do teste de Box-Pierce (p-valor igual a 0,5351) assume-se que os resíduos não são autocorrelacionados.

Figura 36 – Análise dos resíduos do modelo RSE. (a) Gráfico quantil-quantil. (b) ACF.



Fonte: a autora (2020).

A Tabela 10 apresenta as medidas de erro e o AIC obtidos na modelagem.

Tabela 9 – Erros e AIC obtidos no modelo RSE.

Modelo	RMSE	MAE	MPE	MAPE	MASE	AIC
RSE	50.222	42.144,	-0,02264	1,9458	0,5775	1.577,1

Fonte: a autora (2020).

### 5.1.2 ARIMA

O segundo modelo da categoria residencial, identificado como RA, foi desenvolvido com a aplicação do método ARIMA Sazonal. O modelo definido pelo *software* foi do tipo ARIMA (1,0,0)(1,1,0)[12] com *drift*. A equação 22 foi utilizada no desenvolvimento do modelo (BOX; JENKINS, 1976).

$$\psi(B)\Psi(B^s)\nabla^D\nabla^dZ_t = c + \theta(B)\Theta(B^s)a_t \quad (22)$$

Onde:

$\psi$  é o coeficiente autorregressivo  $(1 - \psi_1 B - \dots - \psi_p B^p)$ ;

$B$  é o operador não sazonal de defasagem;

$\Psi$  é o coeficiente autorregressivo sazonal  $(1 - \Psi_1 B - \dots - \Psi_p B^p)$ ;

$B^s$  é o operador sazonal de defasagem;

$s$  é o período da sazonalidade;

$\nabla^D$  é o coeficiente de integração sazonal, definido como  $\nabla^D = (1 - B)^D$ ;

$\nabla^d$  é o coeficiente de integração, definido como  $\nabla^d = (1 - B)^d$ ;

$Z_t$  é a observação presente da série;

$c$  é uma constante;

$\theta$  é o coeficiente de médias móveis  $(1 - \theta_1 B - \dots - \theta_q B^q)$ ;

$\Theta$  é o coeficiente de médias móveis sazonal  $(1 - \theta_1 B - \dots - \theta_q \Theta^q)$ ;

$a_t$  é o erro ou ruído branco.

Os valores dos coeficientes obtidos na modelagem podem ser observados na Tabela 11.

Tabela 10 – Coeficientes do modelo RA.

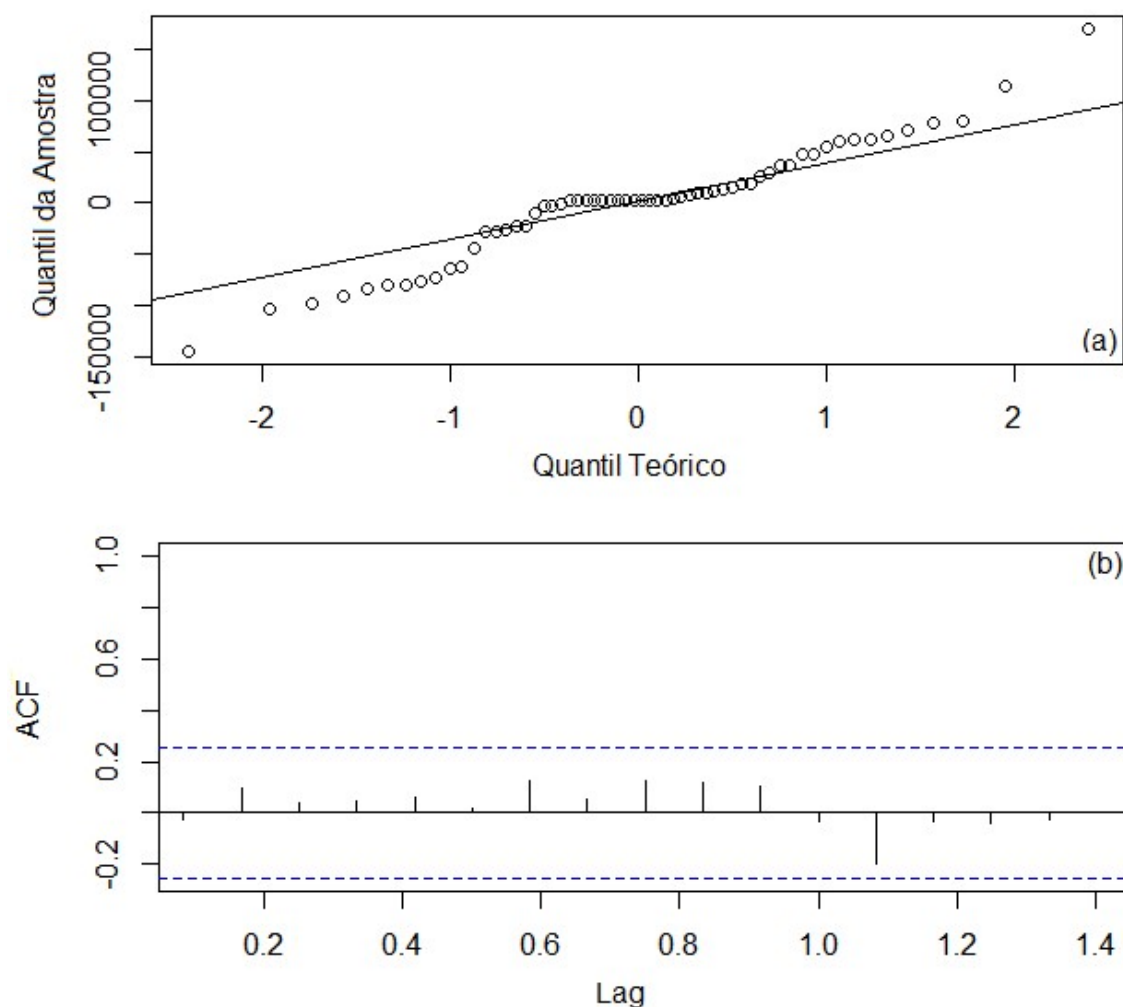
Modelo	AR1	SAR1	DRIFT
RA	0,2808	-0,5928	3.003,0037

Fonte: a autora (2020).

Após a modelagem foram verificadas a distribuição normal e a autocorrelação dos resíduos do modelo (Figura 37). Os resultados dos testes de normalidade

indicaram que os resíduos do modelo não possuem distribuição normal. No teste de Shapiro-Wilk o p-valor calculado foi de 0,03008 e o teste de Lilliefors apontou p-valor de 0,0001031. A falta de normalidade da distribuição também pode ser vista na Figura 37(a).

Figura 37 – Análise dos resíduos do modelo RA. (a) Gráfico quantil-quantil. (b) ACF.



Fonte: a autora (2020).

O modelo não apresenta indícios de que seus resíduos sejam autocorrelacionados, como mostra a ausência de pregos no gráfico da ACF (Figura 37(b)) e o resultado do teste de Box-Pierce (p-valor igual a 0,9891). Apesar de não possuir autocorrelação, a ausência de normalidade na distribuição dos resíduos indica que o modelo RA não é adequado para a realização das previsões do consumo de água residencial na cidade de Joinville.

A Tabela 12 mostra os valores de erro e o AIC calculados pelo modelo RA.

Tabela 11 – Erros e AIC obtidos no modelo RA.

Modelo	RMSE	MAE	MPE	MAPE	MASE	AIC
RA	55.207,9	38.916,9	-0.0144	1,7882	0,53330	1.208,4

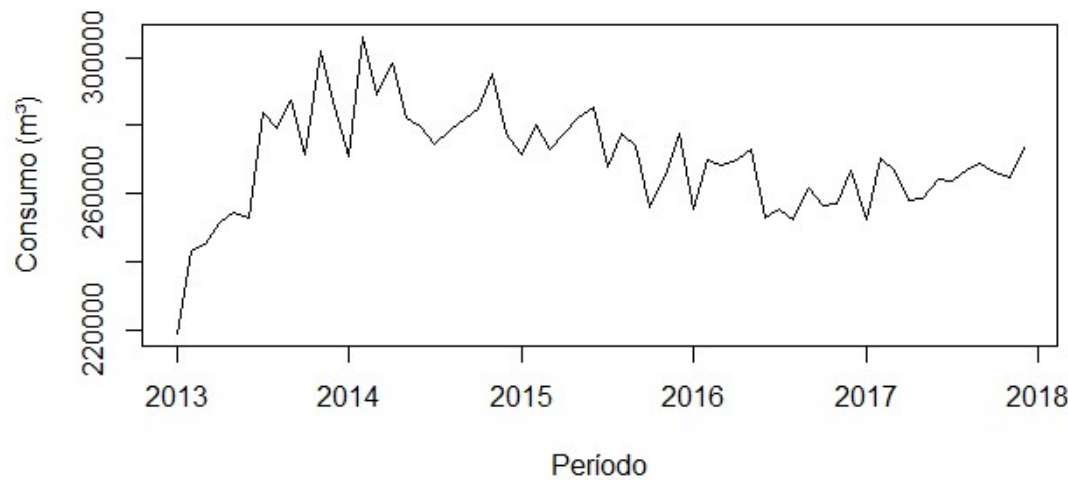
Fonte: a autora (2020).

Os modelos de suavização exponencial e ARIMA desenvolvidos para a categoria, apresentam algumas diferenças. Nos testes estatísticos os dois métodos produziram modelos sem autocorrelação. No entanto, o modelo RA não apresentou distribuição normal dos resíduos, e foi descartado, mesmo apresentando MAPE de treinamento menor (1,7882%) quando comparado ao modelo RSE (MAPE de treinamento igual a 1,9646%). Considerando os resultados obtidos na modelagem, a acurácia foi testada e a escolha dos modelos encontra-se discutida no subitem 5.6.

5.2CATEGORIA COMERCIAL

A série de dados do consumo comercial de água em Joinville no período de janeiro de 2013 a dezembro de 2017 é mostrada na Figura 38. Essa categoria apresenta uma tendência de crescimento até o início de 2014, e após esse período aponta um declínio de consumo até o começo de 2017, quando volta a crescer.

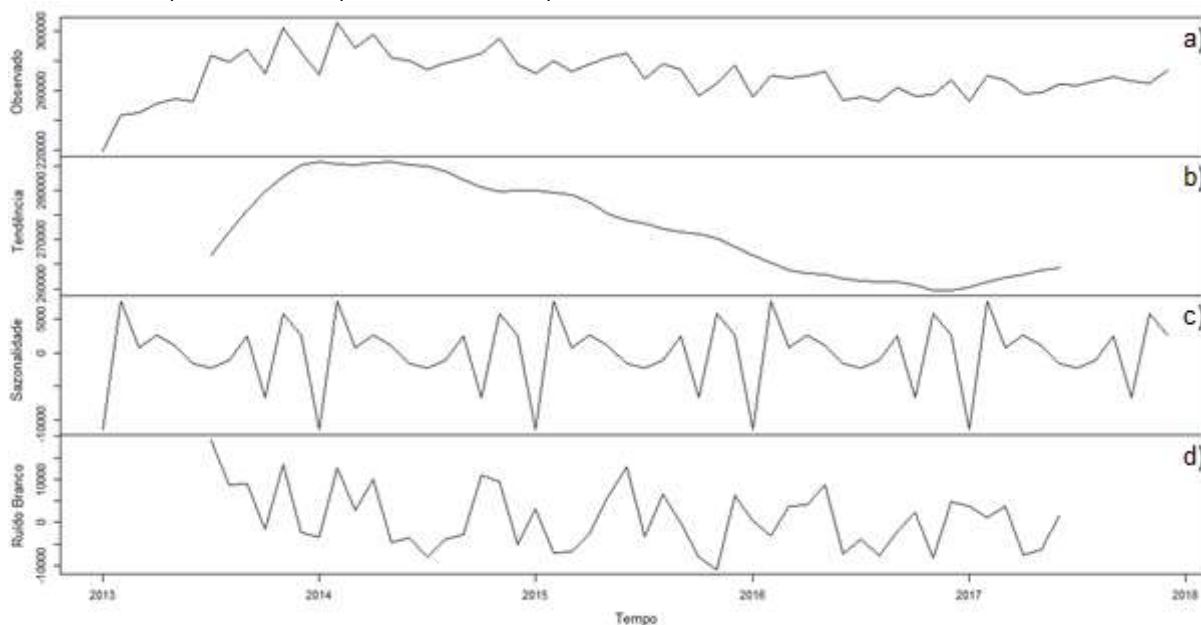
Figura 38 – Série temporal do consumo comercial de água em Joinville de janeiro de 2013 a dezembro de 2017.



Fonte: a autora (2020).

Nas verificações iniciais a sazonalidade da série não se mostra tão pronunciada, como no caso do consumo residencial, mas é possível observar que existem os maiores e menores picos de consumo (Figura 39(c)) nos meses de verão. Há um aumento no consumo no período que precede as festas de final de ano e férias escolares, queda no período de recesso de algumas empresas da região e aumento na retomada das atividades.

Figura 39 – Decomposição da série de dados de consumo comercial de água em Joinville. a) Observado. b) Tendência. c) Sazonalidade. d) Ruído Branco.



Fonte: a autora (2020).

A série de dados de consumo comercial de água em Joinville apresenta distribuição normal, com p-valor no teste de Shapiro-Wilk de 0,3271. As estatísticas descritivas do consumo de água da categoria podem ser observadas na Tabela 13.

Tabela 12 – Estatísticas descritivas da série temporal de consumo da categoria comercial.

	<b>Mínimo</b>	<b>1º Quartil</b>	<b>Mediana</b>	<b>Média</b>	<b>3º Quartil</b>	<b>Máximo</b>
<b>m³</b>	219.056	258.318	270.075	269.964	279.400	305.956

Fonte: a autora (2020).



### 5.2.1 Suavização exponencial

O primeiro modelo produzido para o consumo comercial foi desenvolvido por intermédio do método de suavização exponencial. O modelo definido pelo *software* foi do tipo (M, Ad, N) com erro multiplicativo, tendência aditiva amortecida e sem sazonalidade, identificado como CSE. As equações 23 a 26 foram utilizadas para o desenvolvimento do modelo (MIRANDA, 2009).

$$e_t = (y_t - \mu_t) / \mu_t \quad (23)$$

$$\mu_t = l_{t-1} + \phi b_{t-1} \quad (24)$$

$$l_t = \mu_t(1 + \alpha e_t) \quad (25)$$

$$b_t = (\mu_t + \beta^* e_t) - l_{t-1} \quad (26)$$

Onde:

$e_t$  é o erro de previsão um passo à frente;

$y_t$  é o dado observado no período  $t$ ;

$\mu_t$  é o valor esperado de  $y_t$ ;

$l_t$  é o nível da série no período  $t$ ;

$\phi$  é o parâmetro de amortecimento do crescimento;

$b_t$  é o crescimento da série no período  $t$ ;

$\alpha$  é o parâmetro de suavização para o nível;

$\beta^* = \alpha\beta$ ;

$\beta$  é o parâmetro de suavização para o crescimento.

A Tabela 14 mostra os valores dos parâmetros de suavização de nível, tendência, amortecimento do crescimento, e o erro padrão obtidos para o modelo.

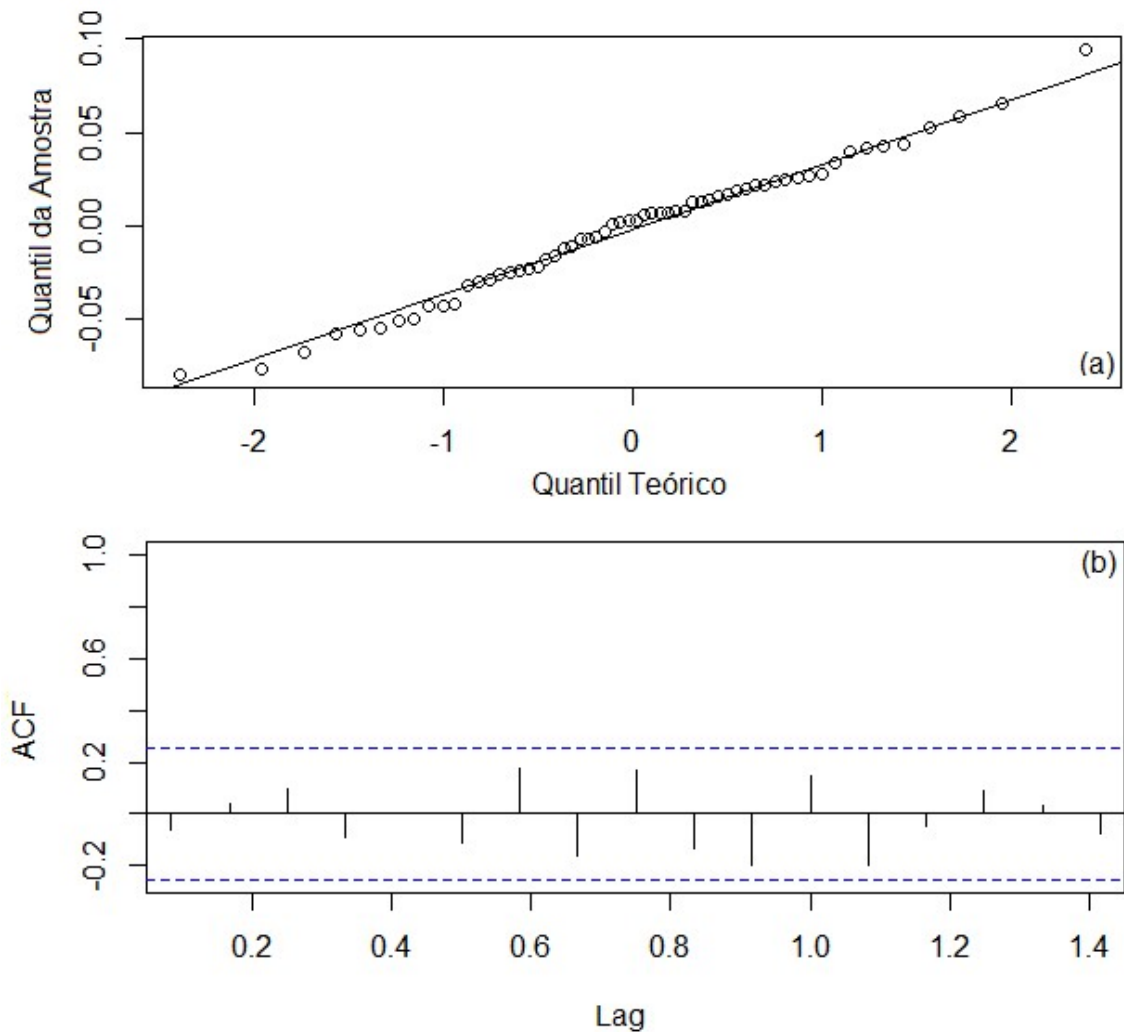
Tabela 13 – Parâmetros de suavização e erro padrão do modelo CSE.

Parâmetros	CSE
$\alpha$	0,2849
$\beta$	0,0395
$\phi$	0,9177
$\sigma$	0,0376

Fonte: a autora (2020).

Nos testes de aceitação do modelo os resultados indicaram que o modelo CSE é adequado para a previsão do consumo de água comercial em Joinville. Os testes de normalidade de Shapiro-Wilk e Lilliefors atestam que seus resíduos possuem distribuição normal, com p-valor igual a 0,8447 e 0,4635, respectivamente. A presença de normalidade na distribuição também pode ser atestada pela observação do gráfico quantil-quantil na Figura 40(a).

Figura 40 – Análise dos resíduos do modelo CSE. (a) Gráfico quantil-quantil. (b) ACF.



Fonte: a autora (2020).

Os resíduos do modelo CSE não são autocorrelacionados, conforme o teste de Box-Pierce (p-valor igual a 0,6571), resultado corroborado pela observação do gráfico da ACF (Figura 40(b)). Os valores dos erros e do AIC do modelo CSE são apresentados na Tabela 15.

Tabela 14 – Erros e AIC obtidos no modelo CSE.

<b>Modelo</b>	<b>RMSE</b>	<b>MAE</b>	<b>MPE</b>	<b>MAPE</b>	<b>MASE</b>	<b>AIC</b>
<b>CSE</b>	9.831,29	7.825,3	-0,3131	2,8938	0,5440	1.359,45

Fonte: a autora (2020).

### 5.2.2 ARIMA

O modelo ARIMA sazonal para a categoria comercial, identificado como CA, foi desenvolvido por intermédio da Equação 27 (BOX; JENKINS, 1976), e definido pelo *software* como do tipo ARIMA (1,1,0)(1,0,0)[12].

$$\psi(B)\Psi(B^s)\nabla^D\nabla^dZ_t = c + \theta(B)\Theta(B^s)a_t \quad (27)$$

Onde:

$\psi$  é o coeficiente autorregressivo  $(1 - \psi_1 B - \dots - \psi_p B^p)$ ;

$B$  é o operador não sazonal de defasagem;

$\Psi$  é o coeficiente autorregressivo sazonal  $(1 - \Psi_1 B - \dots - \Psi_p B^p)$ ;

$B^s$  é o operador sazonal de defasagem;

$s$  é o período da sazonalidade;

$\nabla^D$  é o coeficiente de integração sazonal, definido como  $\nabla^D = (1 - B)^D$ ;

$\nabla^d$  é o coeficiente de integração, definido como  $\nabla^d = (1 - B)^d$ ;

$Z_t$  é a observação presente da série;

$c$  é uma constante;

$\theta$  é o coeficiente de médias móveis  $(1 - \theta_1 B - \dots - \theta_q B^q)$ ;

$\Theta$  é o coeficiente de médias móveis sazonal  $(1 - \Theta_1 B - \dots - \Theta_q B^q)$ ;

$a_t$  é o erro ou ruído branco.

Os valores dos coeficientes não sazonais e sazonais obtidos na modelagem podem ser observados na Tabela 16.

Tabela 15 – Coeficientes do modelo CA.

<b>Modelo</b>	<b>AR1</b>	<b>SAR1</b>
<b>CA</b>	-0,4289	0,3104

Fonte: a autora (2020).

Após a modelagem foram verificadas a distribuição normal e a autocorrelação dos resíduos do modelo por meio da observação gráfica e dos testes estatísticos. Para o modelo CA, o valor obtido no teste de Shapiro-Wilk indica que os resíduos possuem distribuição normal, com p-valor igual a 0,1635, condição que também pode ser verificada na Figura 41(a) e no teste de Lilliefors (p-valor igual a 0,1758). Os resíduos do modelo não possuem autocorrelação, conforme o teste de Box-Pierce (p-valor igual a 0,7988) e como mostra a Figura 41(b).

A Tabela 17 apresenta os valores dos erros e do AIC obtidos no modelo CA.

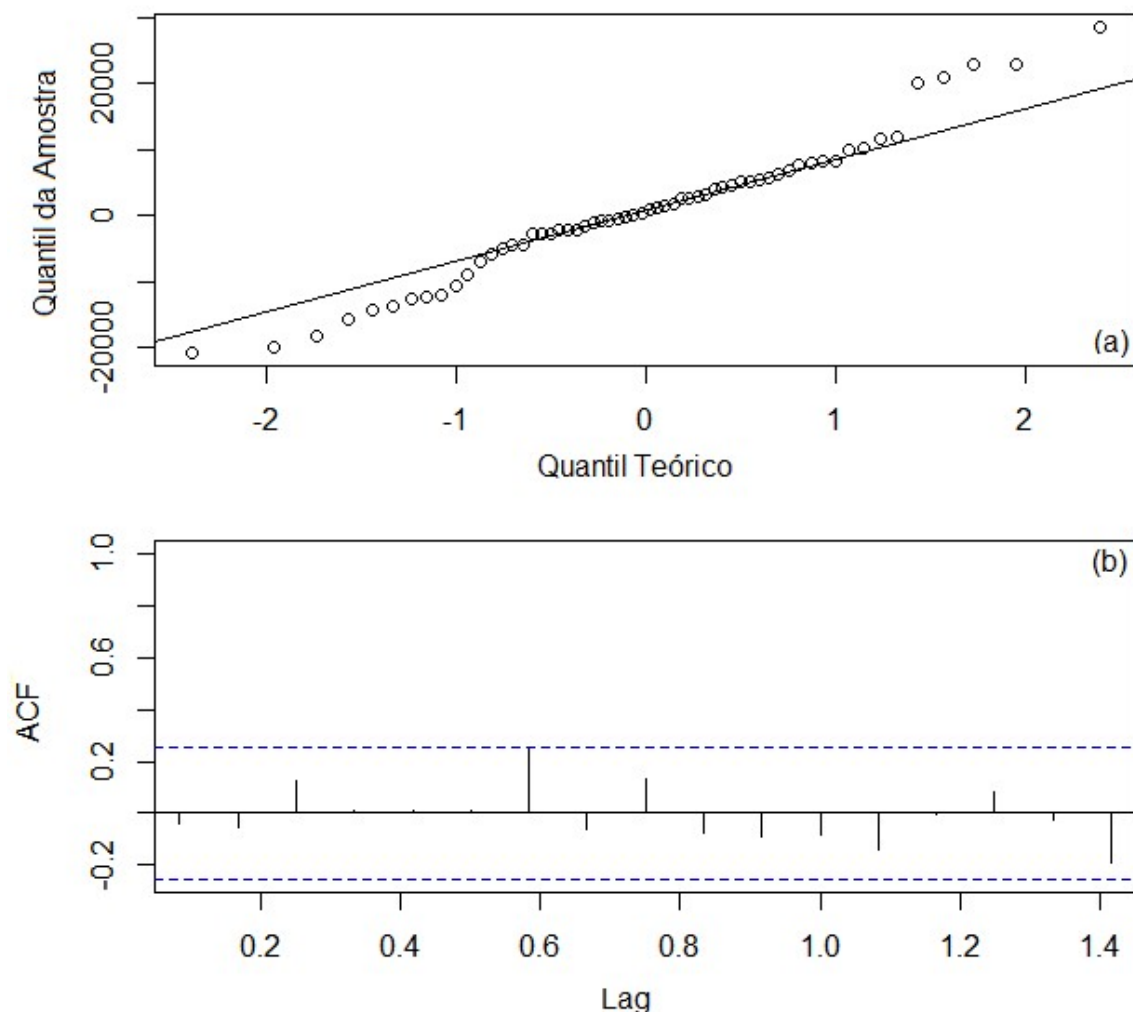
Tabela 16 – Erros e AIC obtidos no modelo CA.

<b>Modelo</b>	<b>RMSE</b>	<b>MAE</b>	<b>MPE</b>	<b>MAPE</b>	<b>MASE</b>	<b>AIC</b>
<b>CA</b>	10.294,3	7.680,55	0,2782	2,8324	0,5340	1.266,1

Fonte: a autora (2020).

Na comparação dos modelos, o método de suavização exponencial não captou a sazonalidade da série, sendo do tipo (M, Ad, N) enquanto o modelo ARIMA propôs um parâmetro autorregressivo para a sazonalidade. Na etapa de treinamento os MAPEs obtidos nos dois modelos foram baixos e muito semelhantes, também abaixo de 3%. A escolha dos modelos e as previsões realizadas podem ser observadas nos subitens 5.6 e 5.7, respectivamente.

Figura 41 – Análise dos resíduos do modelo CA. (a) Gráfico quantil-quantil. (b) ACF.

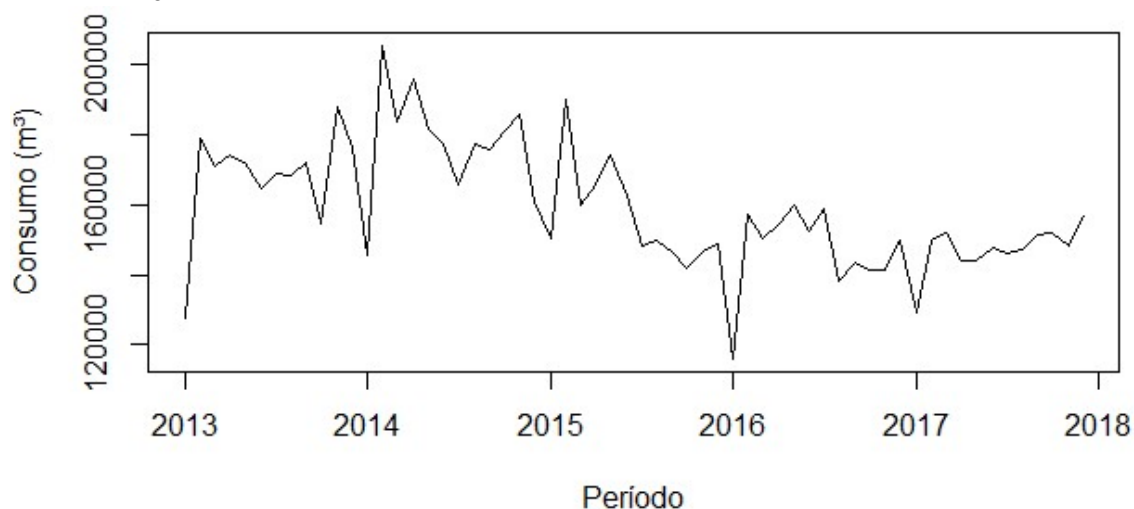


Fonte: a autora (2020).

### 5.3 CATEGORIA INDUSTRIAL

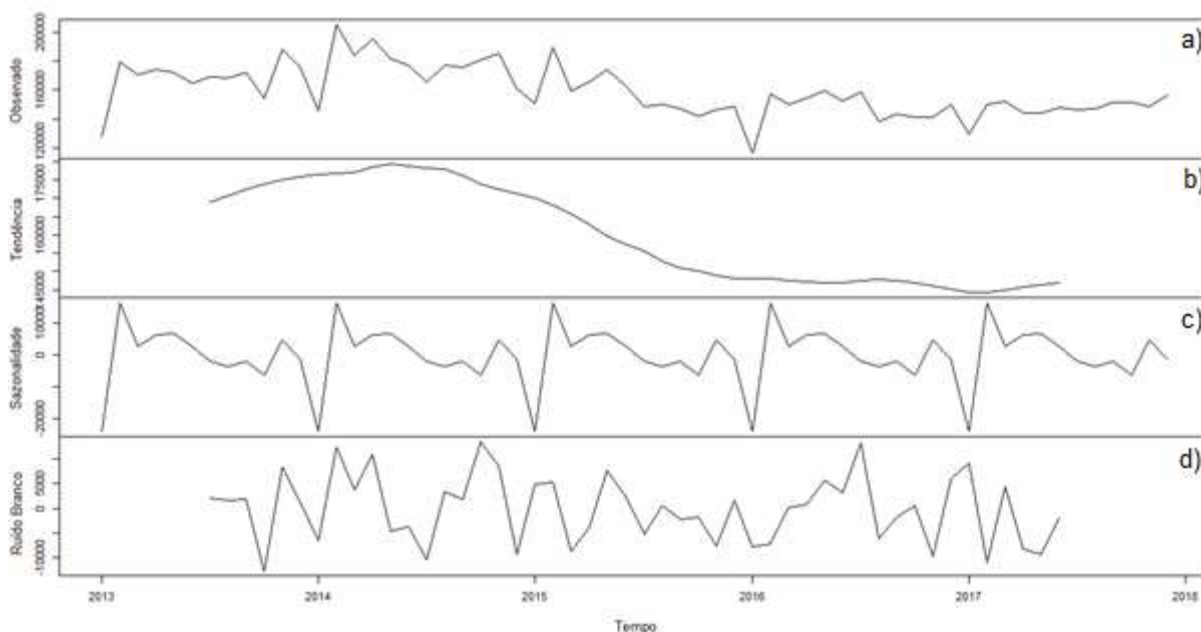
A série temporal de consumo de água da categoria industrial em Joinville no período de janeiro de 2013 a dezembro de 2017 é apresentada na Figura 42. É possível observar que nos anos iniciais da série havia uma tendência de crescimento no consumo, substituída por uma tendência de queda a partir de 2015 (Figura 43(b)), possivelmente explicada pela crise econômica que diminuiu o PIB do município em 1,3 bilhões de reais (IBGE, 2018).

Figura 42 – Série temporal do consumo industrial de água em Joinville de janeiro de 2013 a dezembro de 2017.



Fonte: a autora (2020).

Figura 43 – Decomposição da série de dados de consumo industrial de água em Joinville. a) Observado. b) Tendência. c) Sazonalidade. d) Ruído Branco.



Fonte: a autora (2020).

Nas análises iniciais, foi observado que a série de consumo possui sazonalidade (Figura 43(c)) com uma diminuição substancial do consumo no período que coincide com as férias de verão. Os dados apresentam distribuição normal, com p-valor no teste de Shapiro-Wilk de 0,2776. As estatísticas descritivas do consumo de água da categoria podem ser observadas na Tabela 18.

Tabela 17 – Estatísticas descritivas da série temporal de consumo da categoria industrial.

	<b>Mínimo</b>	<b>1º Quartil</b>	<b>Mediana</b>	<b>Média</b>	<b>3º Quartil</b>	<b>Máximo</b>
<b>m³</b>	116.044	147.503	155.638	159.463	172.715	205.387

Fonte: a autora (2020).

### 5.3.1 Suavização exponencial

O modelo desenvolvido para a categoria com aplicação do método de suavização exponencial foi identificado com ISE. O *software* definiu o modelo do tipo (A, N, A) com erro aditivo, sem tendência e com sazonalidade aditiva como mais adequado para prever o consumo na série de dados analisada. As equações 28 a 30 descrevem o modelo (MIRANDA, 2009).

$$e_t = y_t - \mu_t \quad (27)$$

$$\mu_t = l_{t-1} + S_{t-m} \quad (28)$$

$$l_t = (\mu_t + \alpha e_t) - S_{t-m} \quad (29)$$

$$S_t = (\mu_t + \alpha e_t) - l_{t-1} \quad (30)$$

Onde:

$e_t$  é o erro de previsão um passo à frente;

$y_t$  é o dado observado no período  $t$ ;

$\mu_t$  é o valor esperado de  $y_t$ ;

$l_t$  é o nível da série no período  $t$ ;

$S_t$  é a sazonalidade da série no período  $t$ ;

$m$  é o tamanho do ciclo sazonal;

$\alpha$  é o parâmetro de suavização para o nível.

Os valores dos parâmetros de nível, sazonalidade e o erro padrão obtidos para o modelo ISE podem ser observados na Tabela 19.

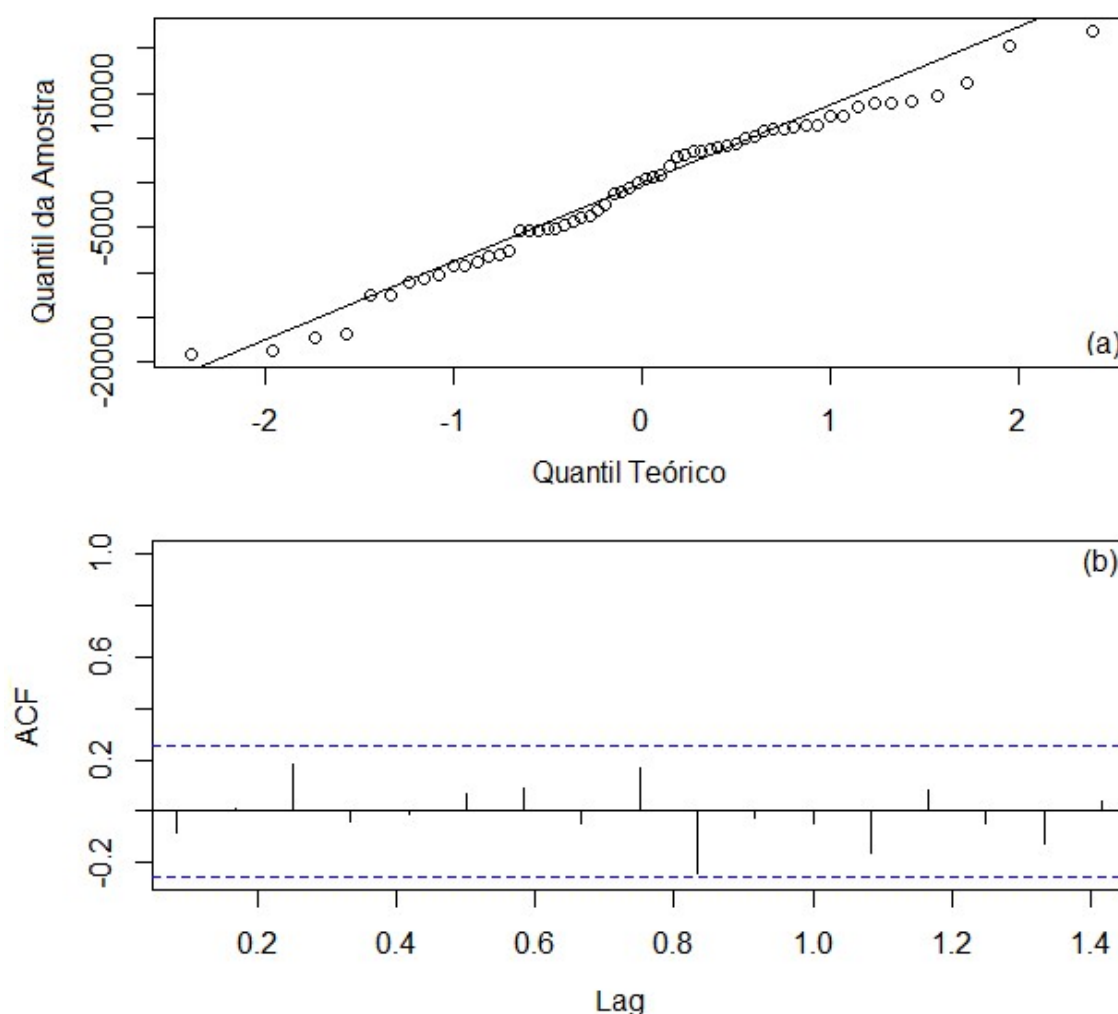
Tabela 18 – Parâmetros de suavização e erro padrão do modelo ISE.

Parâmetros	ISE
$\alpha$	0,4198
$\gamma$	1,00E-04
$\sigma$	9.520,3

Fonte: a autora (2020).

Na etapa de análise dos resíduos os testes estatísticos indicaram que o modelo ISE é adequado para a previsão do consumo de água industrial no município. Há normalidade na distribuição dos resíduos, que pode ser observada no gráfico quantil-quantil (Figura 44(a)) e atestada pelos resultados dos testes de Shapiro-Wilk (p-valor igual a 0,3166) e Lilliefors (p-valor igual a 0,1352). O teste de Box-Pierce indicou que os dados não possuem autocorrelação (Figura 44(b)), com p-valor igual a 0,81.

Figura 44 – Análise dos resíduos do modelo ISE. (a) Gráfico quantil-quantil. (b) ACF.



Fonte: a autora (2020).



A Tabela 20 mostra os valores dos erros e do AIC encontrados no modelo ISE.

Tabela 19 – Erros e AIC obtidos no modelo ISE.

<b>Modelo</b>	<b>RMSE</b>	<b>MAE</b>	<b>MPE</b>	<b>MAPE</b>	<b>MASE</b>	<b>AIC</b>
<b>ISE</b>	8.335,9	6.914,6	-0.7152	4,4318	0,4926	1.359,1

Fonte: a autora (2020).

### 5.3.2 ARIMA

Identificado como IA, o modelo ARIMA sazonal para a categoria industrial foi definido pelo *software* como do tipo ARIMA (0,1,1)(0,1,1)[12], conforme Equação 31 (BOX; JENKINS, 1976).

$$\psi(B)\Psi(B^s)\nabla^D\nabla^dZ_t = c + \theta(B)\Theta(B^s)a_t \quad (31)$$

Onde:

$\psi$  é o coeficiente autorregressivo  $(1 - \psi_1 B - \dots - \psi_p B^p)$ ;

$B$  é o operador não sazonal de defasagem;

$\Psi$  é o coeficiente autorregressivo sazonal  $(1 - \Psi_1 B - \dots - \Psi_p B^p)$ ;

$B^s$  é o operador sazonal de defasagem;

$s$  é o período da sazonalidade;

$\nabla^D$  é o coeficiente de integração sazonal, definido como  $\nabla^D = (1 - B)^D$ ;

$\nabla^d$  é o coeficiente de integração, definido como  $\nabla^d = (1 - B)^d$ ;

$Z_t$  é a observação presente da série;

$c$  é uma constante;

$\theta$  é o coeficiente de médias móveis  $(1 - \theta_1 B - \dots - \theta_q B^q)$ ;

$\Theta$  é o coeficiente de médias móveis sazonal  $(1 - \Theta_1 B - \dots - \Theta_q B^q)$ ;

$a_t$  é o erro ou ruído branco.

São apresentados na Tabela 21 os valores dos coeficientes não sazonais e sazonais obtidos na modelagem.

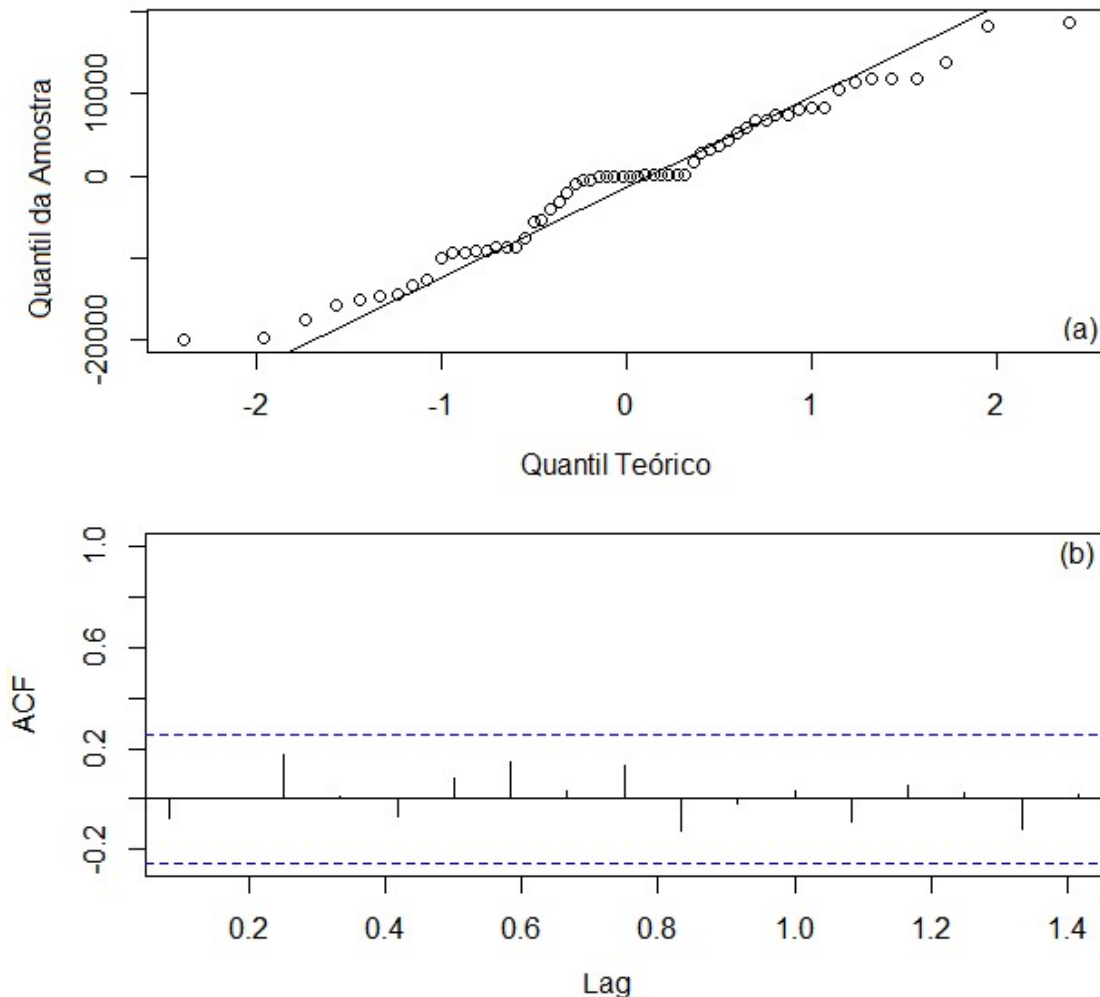
Tabela 20 – Coeficientes do modelo IA.

Modelo	MA1	SMA1
IA	-0,5298	-0,5047

Fonte: a autora (2020).

Em relação à distribuição, o resultado do teste de Shapiro-Wilk indica que os resíduos do modelo IA possuem distribuição normal com p-valor de 0,294. O teste de Lilliefors aponta que os resíduos são normalmente distribuídos apenas ao nível de 1% de significância (p-valor igual a 0,03374), abaixo do valor adotado neste trabalho ( $\alpha$  igual a 5%), como mostra a Figura 45(a)

Figura 45 – Análise dos resíduos do modelo IA. (a) Gráfico quantil-quantil. (b) ACF.



Fonte: a autora (2020).

Considerando que a distribuição dos resíduos foi considerada normal em pelo menos um dos testes estatísticos neste trabalho assumiu-se que os resíduos são

normalmente distribuídos. Na análise de autocorrelação, a Figura 45(b) valida o resultado observado no teste de Box-Pierce (p-valor igual a 0,881), e considera-se que os resíduos do modelo não são autocorrelacionados. Os valores dos erros e do AIC do modelo IA são mostrados na Tabela 22.

Tabela 21 – Erros e AIC obtidos no modelo IA.

<b>Modelo</b>	<b>RMSE</b>	<b>MAE</b>	<b>MPE</b>	<b>MAPE</b>	<b>MASE</b>	<b>AIC</b>
<b>IA</b>	9.338,22	7.258,26	-0,6174	4,6471	0,5171	1.014,3

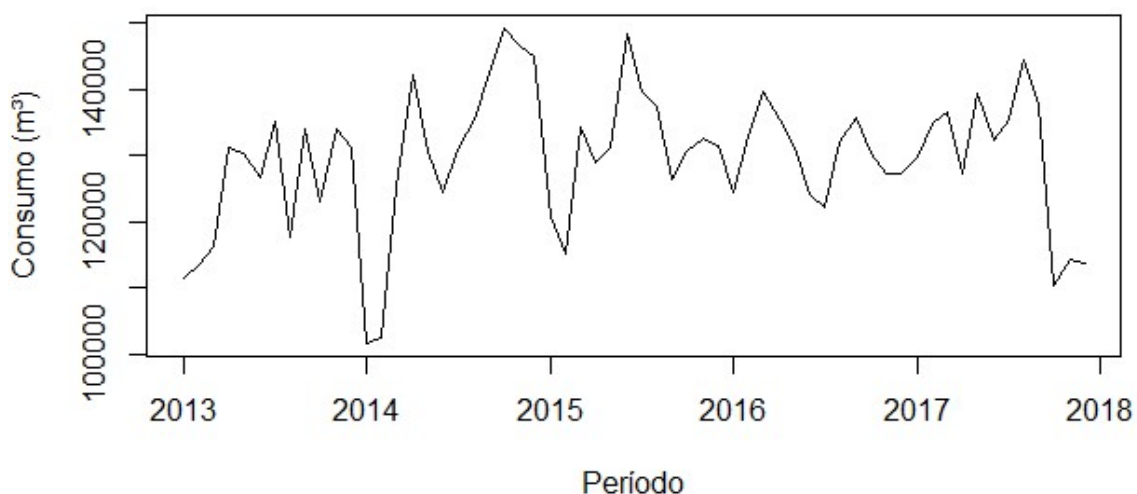
Fonte: a autora (2020).

Algumas semelhanças podem ser vistas nos dois modelos da categoria industrial. Os dois métodos de previsão resultaram em modelos com resíduos normalmente distribuídos e sem autocorrelação, com MAPEs de treinamento abaixo de 5%. A sazonalidade da série foi captada pelo modelo de suavização exponencial, e classificada como aditiva. No modelo ARIMA é possível observar um parâmetro de média móvel para a sazonalidade. O modelo de suavização não considerou a tendência da série. Tendo em vista os resultados obtidos na modelagem, a acurácia dos modelos foi testada e a escolha do modelo, bem como a análise das previsões encontram-se nos subitens 5.6 e 5.7.

#### 5.4 CATEGORIA PÚBLICO

O comportamento de consumo da categoria público pode ser observado na Figura 46 e, diferentemente das outras categorias, o consumo público de água não apresenta uma tendência muito pronunciada, ou um comportamento regular. É possível observar um consumo baixo no início de 2014, e um crescimento entre 2014 e 2015, quando comparado ao ano anterior. Entre meados de 2015 e 2017 a série não apresenta grande variação nos dados, mas mostra uma queda acentuada a partir de outubro de 2017.

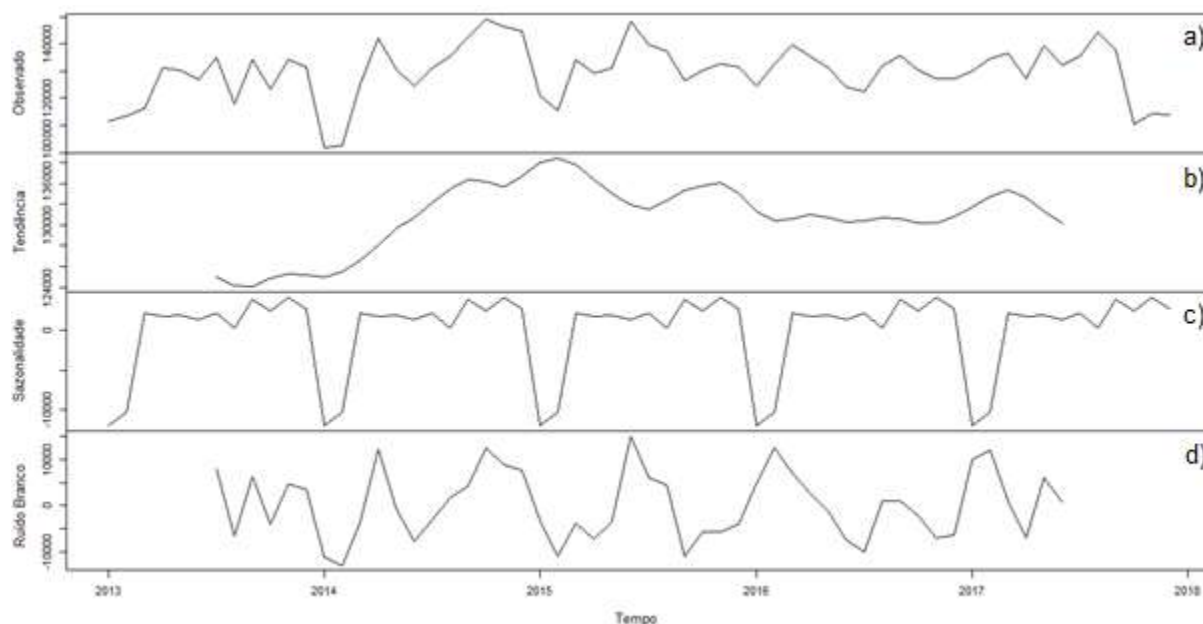
Figura 46 – Série temporal do consumo público de água em Joinville de janeiro de 2007 a dezembro de 2017.



Fonte: a autora (2020).

É possível perceber que a série de dados possui sazonalidade (Figura 47(c)) com uma queda acentuada no consumo nos meses que coincidem com as férias de verão.

Figura 47 – Decomposição da série de dados de consumo público de água em Joinville. a) Observado. b) Tendência. c) Sazonalidade. d) Ruído Branco.



Fonte: a autora (2020).

Nas análises iniciais os dados de consumo público de água em Joinville apresentaram uma distribuição normal muito próxima ao limite do nível de significância

adotado (5%), com p-valor no teste de Shapiro-Wilk de 0,05446. As estatísticas descritivas do consumo público de água em Joinville encontram-se na Tabela 23.

Tabela 22 – Estatísticas descritivas da série temporal de consumo da categoria público.

	<b>Mínimo</b>	<b>1º Quartil</b>	<b>Mediana</b>	<b>Média</b>	<b>3º Quartil</b>	<b>Máximo</b>
<b>m³</b>	101.699	124.346	131.235	129.571	135.412	149.150

Fonte: a autora (2020).

#### 5.4.1 Suavização exponencial

O primeiro modelo desenvolvido para o consumo público em Joinville foi identificado como PSE, e fez uso do método de suavização exponencial. O modelo definido pelo *software* como o mais adequado para prever o consumo na categoria foi do tipo (M, N, N) com erro multiplicativo, sem tendência e sem sazonalidade, descrito pelas equações 35 a 37 (MIRANDA, 2009).

$$e_t = (y_t - \mu_t)\mu_t \quad (35)$$

$$\mu_t = l_{t-1} \quad (36)$$

$$l_t = \mu_t(1 + \alpha e_t) \quad (37)$$

Onde:

$e_t$  é o erro de previsão um passo à frente;

$y_t$  é o dado observado no período  $t$ ;

$\mu_t$  é o valor esperado de  $y_t$ ;

$l_t$  é o nível da série no período  $t$ ;

$\alpha$  é o parâmetro de suavização para o nível.

Na Tabela 24 são mostrados os valores do parâmetro de suavização de nível e o erro padrão obtidos para o modelo PSE.

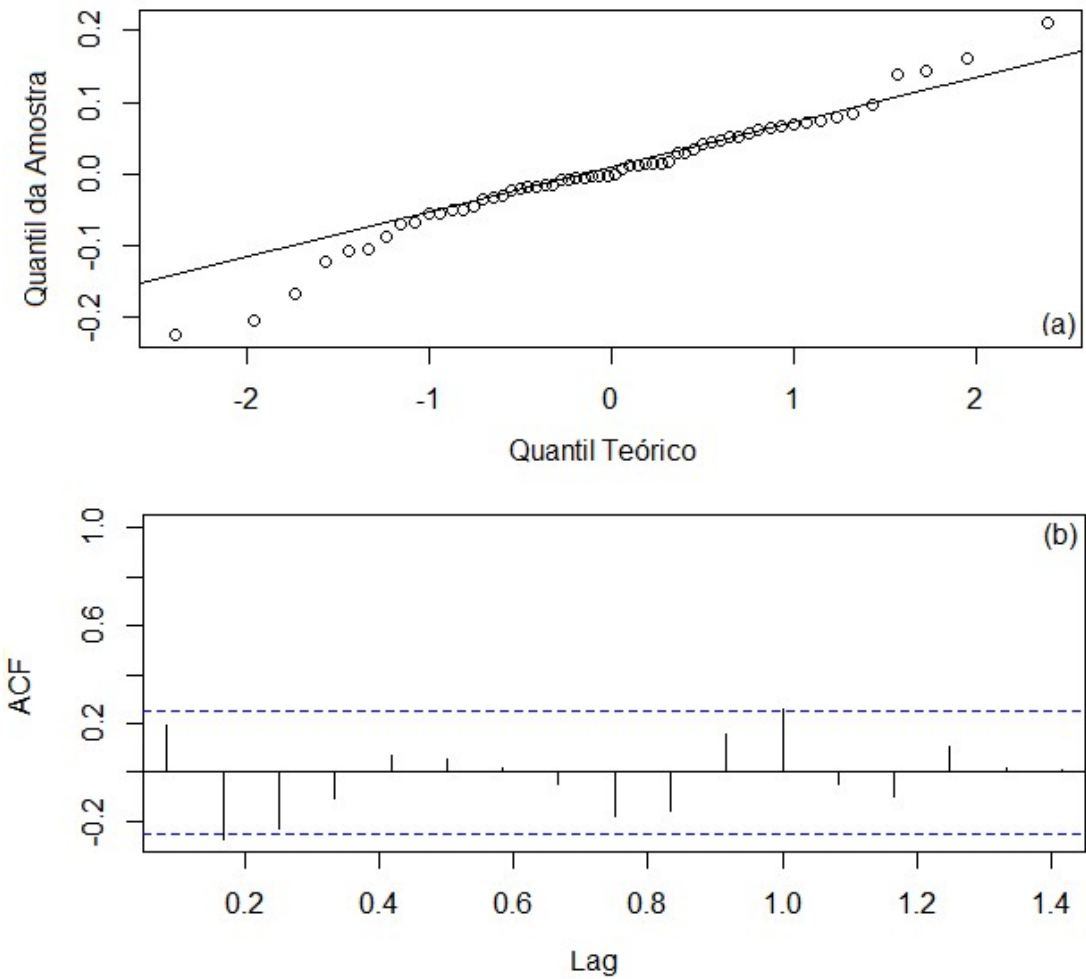
Tabela 23 – Parâmetros de suavização e erro padrão do modelo PSE.

Parâmetros	PSE
$\alpha$	0,5594
$\sigma$	0,08

Fonte: a autora (2020).

Os resultados da análise dos resíduos, mostram que os resíduos do modelo PSE possuem distribuição normal (Figura 48 (a)), com p-valor obtido no teste de Shapiro-Wilk igual a 0,1399 e 0,2326 no teste de Lilliefors. Apesar do resultado do teste de Box-Pierce (p-valor igual a 0,4297, o gráfico da ACF (Figura 48 (b)) mostra dois pregos que ultrapassam o limite da linha pontilhada (*lags* 0.2 e 1.0) indicando que pode haver autocorrelação nos resíduos do modelo de suavização exponencial. Desse modo, o modelo PSE não pode ser considerado adequado para a realização de previsões do consumo de água para a categoria público em Joinville no período de análise.

Figura 48 – Análise dos resíduos do modelo PSE. (a) Gráfico quantil-quantil. (b) ACF.



Fonte: a autora (2020).

Na Tabela 25 são apresentados os valores dos erros e do AIC obtidos no modelo PSE.

Tabela 24 – Erros e AIC obtidos no modelo PSE.

<b>Modelo</b>	<b>RMSE</b>	<b>MAE</b>	<b>MPE</b>	<b>MAPE</b>	<b>MASE</b>	<b>AIC</b>
<b>PSE</b>	10.092,7	7.366,61	-0,4095	5,8639	0,7312	1.358,9

Fonte: a autora (2020).

#### 5.4.2 ARIMA

Identificado como PA, o segundo modelo desenvolvido para a categoria público, pelo método ARIMA Sazonal, foi definido pelo *software* como sendo do tipo ARIMA (0,0,1)(1,0,0)[12] com média diferente de zero, conforme a Equação 39 (BOX; JENKINS, 1976).

$$\psi(B)\Psi(B^s)\nabla^D\nabla^dZ_t = c + \theta(B)\Theta(B^s)a_t \quad (39)$$

Onde:

$\psi$  é o coeficiente autorregressivo  $(1 - \psi_1 B - \dots - \psi_p B^p)$ ;

$B$  é o operador não sazonal de defasagem;

$\Psi$  é o coeficiente autorregressivo sazonal  $(1 - \Psi_1 B - \dots - \Psi_p B^p)$ ;

$B^s$  é o operador sazonal de defasagem;

$s$  é o período da sazonalidade;

$\nabla^D$  é o coeficiente de integração sazonal, definido como  $\nabla^D = (1 - B)^D$ ;

$\nabla^d$  é o coeficiente de integração, definido como  $\nabla^d = (1 - B)^d$ ;

$Z_t$  é a observação presente da série;

$c$  é uma constante;

$\theta$  é o coeficiente de médias móveis  $(1 - \theta_1 B - \dots - \theta_q B^q)$ ;

$\Theta$  é o coeficiente de médias móveis sazonal  $(1 - \Theta_1 B - \dots - \Theta_q B^q)$ ;

$a_t$  é o erro ou ruído branco.

Os valores dos coeficientes do modelo PA são mostrados na Tabela 26.

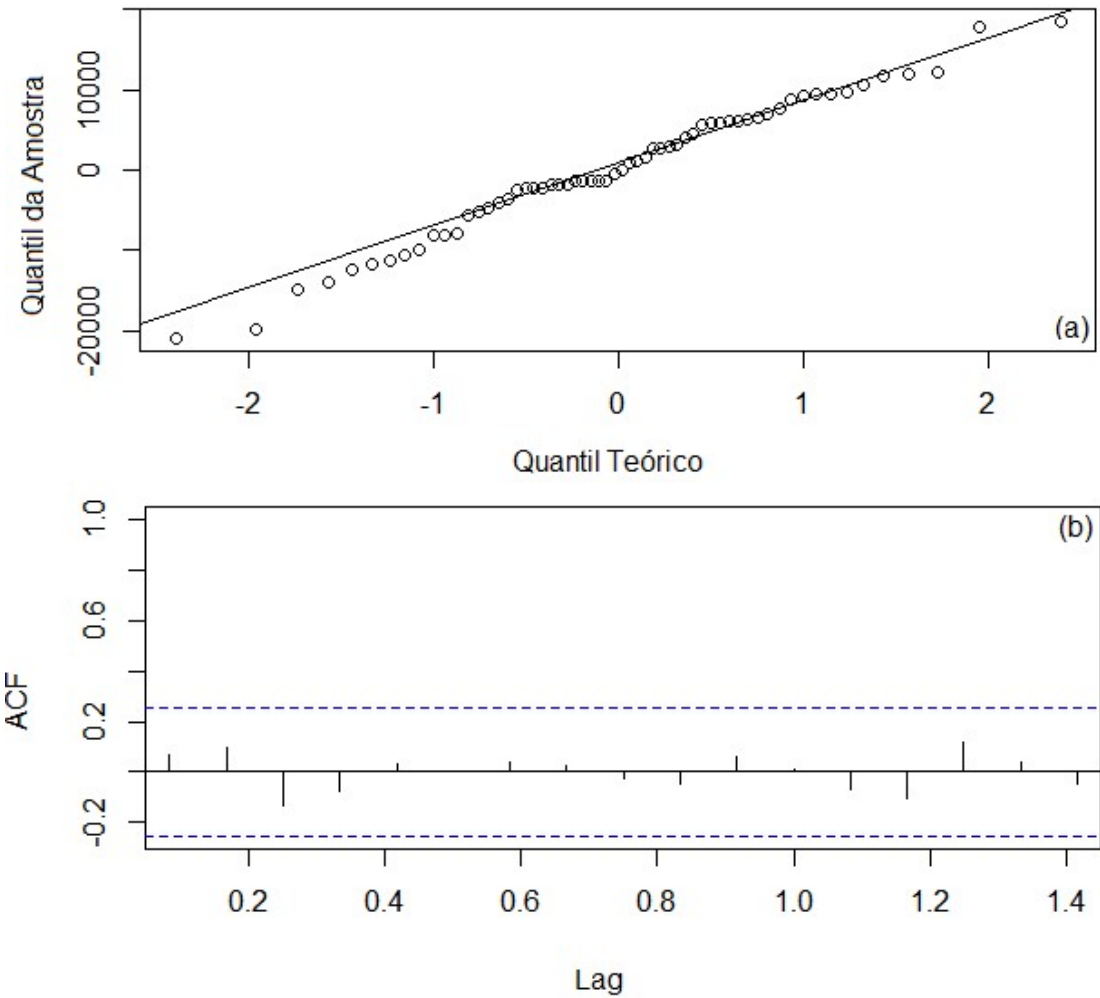
Tabela 25 – Coeficientes do modelo RA.

Modelo	MA1	SAR1	MÉDIA
PA	0.5087	0,3180	129.037,647

Fonte: a autora (2020).

Nos testes estatísticos os resultados apontam que os resíduos do modelo PA possuem distribuição normal (Figura 49(a)) com p-valor de 0,6039 no teste de Shapiro-Wilk e de 0,2426 no teste de Lilliefors. Na análise de autocorrelação, considera-se que os resíduos do modelo não são autocorrelacionados (Figura 49(b)) com p-valor igual a 0,9918 no teste de Box-Pierce.

Figura 49 – Análise dos resíduos do modelo PA. (a) Gráfico quantil-quantil. (b) ACF.



Fonte: a autora (2020).



Os valores dos erros e do AIC obtidos no modelo PA são apresentados na Tabela 27.

Tabela 26 – Erros e AIC obtidos no modelo PA.

<b>Modelo</b>	<b>ME</b>	<b>RMSE</b>	<b>MAE</b>	<b>MPE</b>	<b>MAPE</b>	<b>MASE</b>	<b>AIC</b>
<b>PA</b>	255.734	8.447,19	6.737,63	-0,2720	5,3161	0,6688	1.264,84

Fonte: a autora (2020).

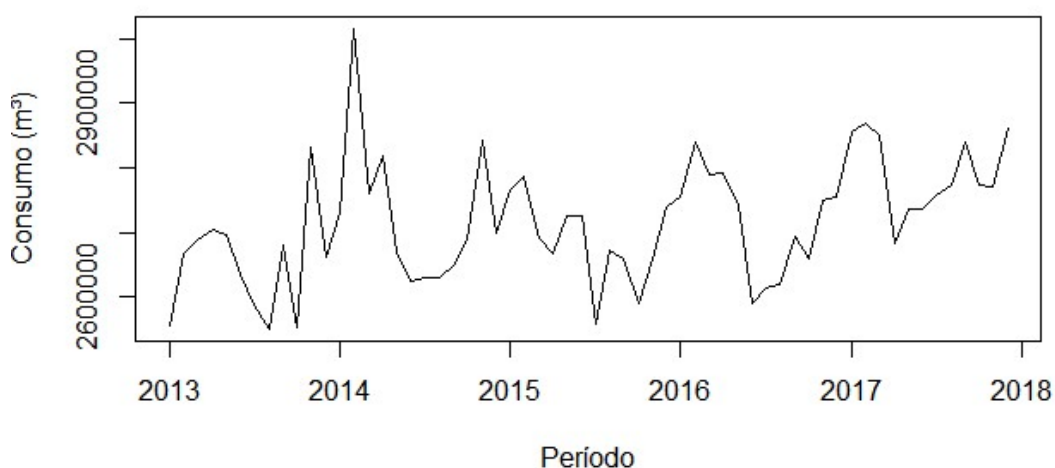
É possível observar algumas diferenças entre os modelos produzidos para a categoria. O modelo desenvolvido por meio de suavização exponencial é um modelo SES (suavização exponencial simples), sem tendência e sem sazonalidade, com erros multiplicativos. O modelo ARIMA é do tipo  $(0,0,1)(1,0,0)$  [12] com média diferente de zero, e apresenta um parâmetro autorregressivo para a sazonalidade. Os dois modelos produzidos possuem resíduos com distribuição normal, e os MAPEs obtidos são baixos, e giram em torno dos 5%. Entretanto, os resíduos do modelo desenvolvido por intermédio de suavização exponencial, diferentemente dos resíduos do modelo ARIMA, são autocorrelacionados e, por isso, o modelo PSE não foi aceito para a realização das previsões.

## 5.5 CATEGORIA TOTAL

A categoria total, analisada neste estudo, contempla a soma de todas as categorias de consumo, e é apresentada na Figura 50.

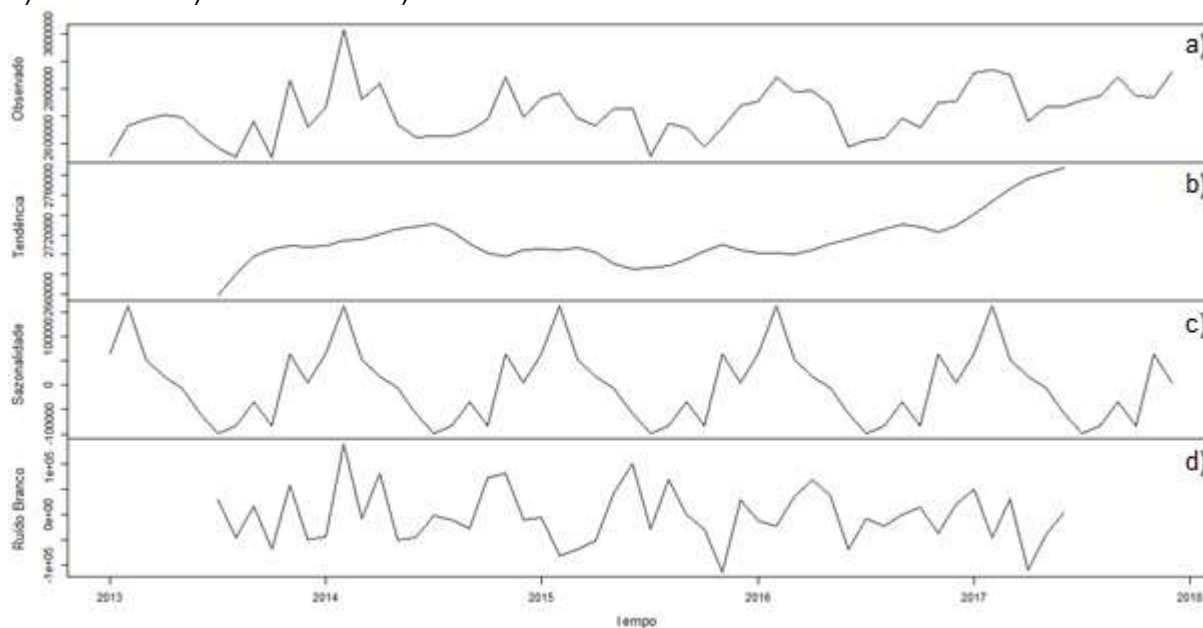
Inicialmente verificou-se que os dados da série apresentam distribuição normal, com p-valor de 0,2272 no teste de Shapiro-Wilk. A série apresenta uma tendência de crescimento até metade de 2014, com queda no consumo até meados de 2015, quando volta a crescer, mantendo a tendência até o final do período analisado (Figura 51(b)). Foi observada também a existência de sazonalidade nos dados (Figura 51(c)), com diminuição de consumo nos meses mais frios e grandes consumos sendo registrados nos meses de verão. As estatísticas descritivas da série podem ser observadas na Tabela 28.

Figura 50 – Série temporal do consumo total de água em Joinville de janeiro de 2013 a dezembro de 2017.



Fonte: a autora (2020).

Figura 51 – Decomposição da série de dados de consumo total de água em Joinville. a) Observados. b) Tendência. c) Sazonalidade. d) Ruído Branco.



Fonte: a autora (2020).

Tabela 27 – Estatísticas descritivas da série temporal de consumo da categoria total.

	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo
<b>m³</b>	2.551.517	2.659.113	2.702.161	2.715.408	2.771.253	3.014.547

Fonte: a autora (2020).

### 5.5.1 Suavização exponencial

O primeiro modelo produzido para o consumo total de água em Joinville foi desenvolvido por meio do método de suavização exponencial. Identificado como TSE, o modelo definido pelo *software* foi do tipo (A, N, N) com erro aditivo, sem tendência e sem sazonalidade. As equações 40 a 42 foram utilizadas para o desenvolvimento do modelo (MIRANDA, 2009).

$$e_t = y_t - \mu_t \quad (40)$$

$$\mu_t = l_{t-1} \quad (41)$$

$$l_t = \mu_t + \alpha e_t \quad (42)$$

Onde:

$e_t$  é o erro de previsão um passo à frente;

$y_t$  é o dado observado no período  $t$ ;

$\mu_t$  é o valor esperado de  $y_t$ ;

$l_t$  é o nível da série no período  $t$ ;

$\alpha$  é o parâmetro de suavização para o nível.

Os valores do parâmetro de suavização de nível e o erro padrão obtidos para o modelo TSE são mostrados na Tabela 29.

Tabela 28 – Parâmetros de suavização e erro padrão do modelo TSE.

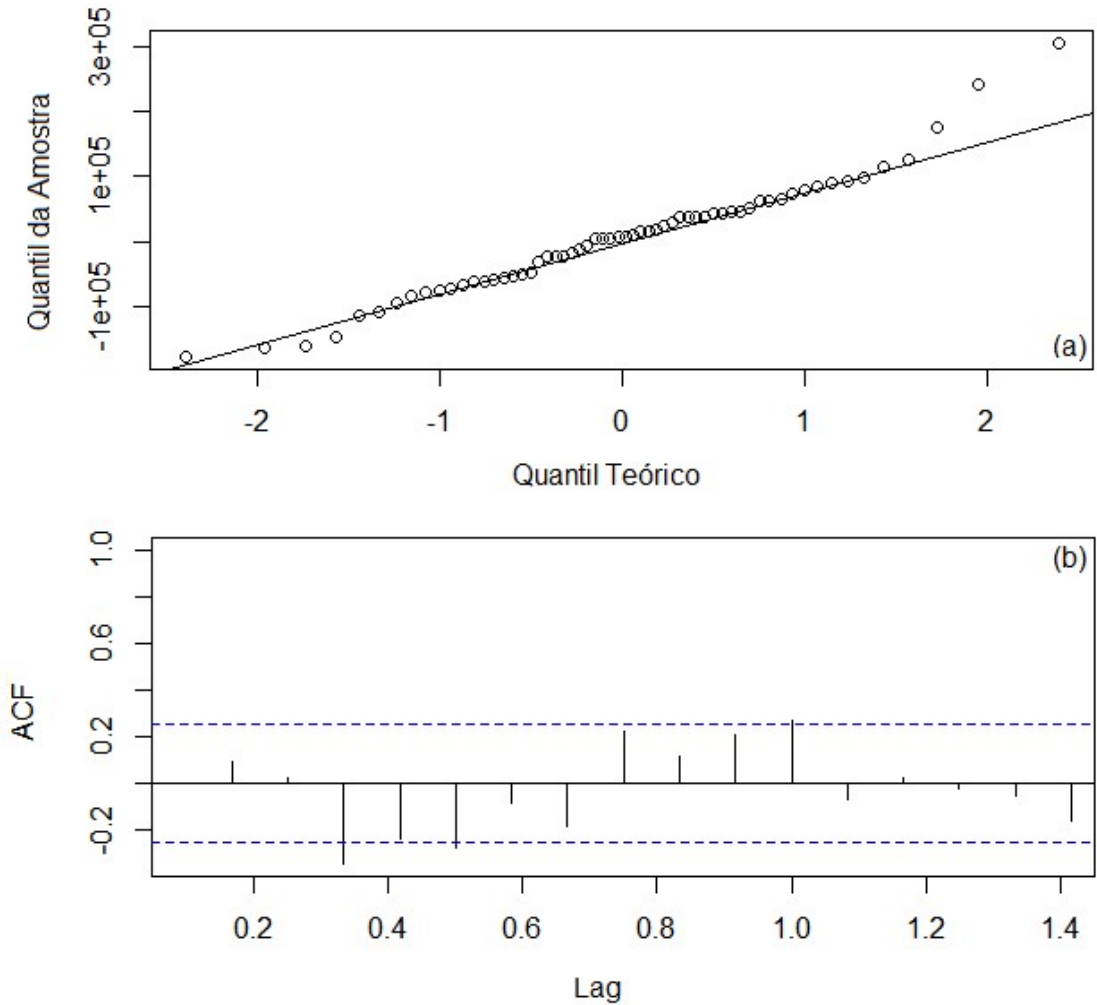
Parâmetros	TSE
$\alpha$	0,5207
$\sigma$	90.542,73

Fonte: a autora (2020).

Na etapa de análise dos resíduos, os testes estatísticos indicaram que os resíduos do modelo TSE possuem distribuição normal (Figura52(a)), com p-valor obtido no teste de Lilliefors de 0,5768 e de 0,0736 no de Shapiro-Wilk. Os resíduos,

entretanto, são autocorrelacionados, conforme pode ser observado na Figura 52(b) com três pregos no gráfico da ACF e p-valor no teste de Box -Pierce de 0,03948

Figura 52 – Análise dos resíduos do modelo TSE. (a) Gráfico quantil-quantil. (b) ACF.



Fonte: a autora (2020).

Considerando os resultados dos testes de autocorrelação o modelo TSE foi considerado inadequado para a realização das previsões de consumo total de água na cidade de Joinville no período analisado. Os erros e o AIC obtidos no modelo TSE são apresentados na Tabela 30.

Tabela 29 – Erros e AIC obtidos no modelo TSE.

Modelo	RMSE	MAE	MPE	MAPE	MASE	AIC
TSE	89.020,9	67.356,17	0,12282	2,4708	0,7943	1.616,26

Fonte: a autora (2020).

### 5.5.2 ARIMA

O modelo ARIMA sazonal produzido para a categoria total foi identificado como TA, e definido pelo *software* como sendo do tipo ARIMA (1,1,1)(1,0,1)[12]. A equação 43 foi utilizada para o desenvolvimento do modelo (BOX; JENKINS, 1976).

$$\psi(B)\Psi(B^s)\nabla^D\nabla^dZ_t = c + \theta(B)\Theta(B^s)a_t \quad (43)$$

Onde:

$\psi$  é o coeficiente autorregressivo  $(1 - \psi_1 B - \dots - \psi_p B^p)$ ;

$B$  é o operador não sazonal de defasagem;

$\Psi$  é o coeficiente autorregressivo sazonal  $(1 - \Psi_1 B - \dots - \Psi_p B^p)$ ;

$B^s$  é o operador sazonal de defasagem;

$s$  é o período da sazonalidade;

$\nabla^D$  é o coeficiente de integração sazonal, definido como  $\nabla^D = (1 - B)^D$ ;

$\nabla^d$  é o coeficiente de integração, definido como  $\nabla^d = (1 - B)^d$ ;

$Z_t$  é a observação presente da série;

$c$  é uma constante;

$\theta$  é o coeficiente de médias móveis  $(1 - \theta_1 B - \dots - \theta_q B^q)$ ;

$\Theta$  é o coeficiente de médias móveis sazonal  $(1 - \theta_1 B - \dots - \theta_q \Theta^q)$ ;

$a_t$  é o erro ou ruído branco.

Os valores dos coeficientes não sazonais e sazonais obtidos para o modelo TA podem ser observados na Tabela 31.

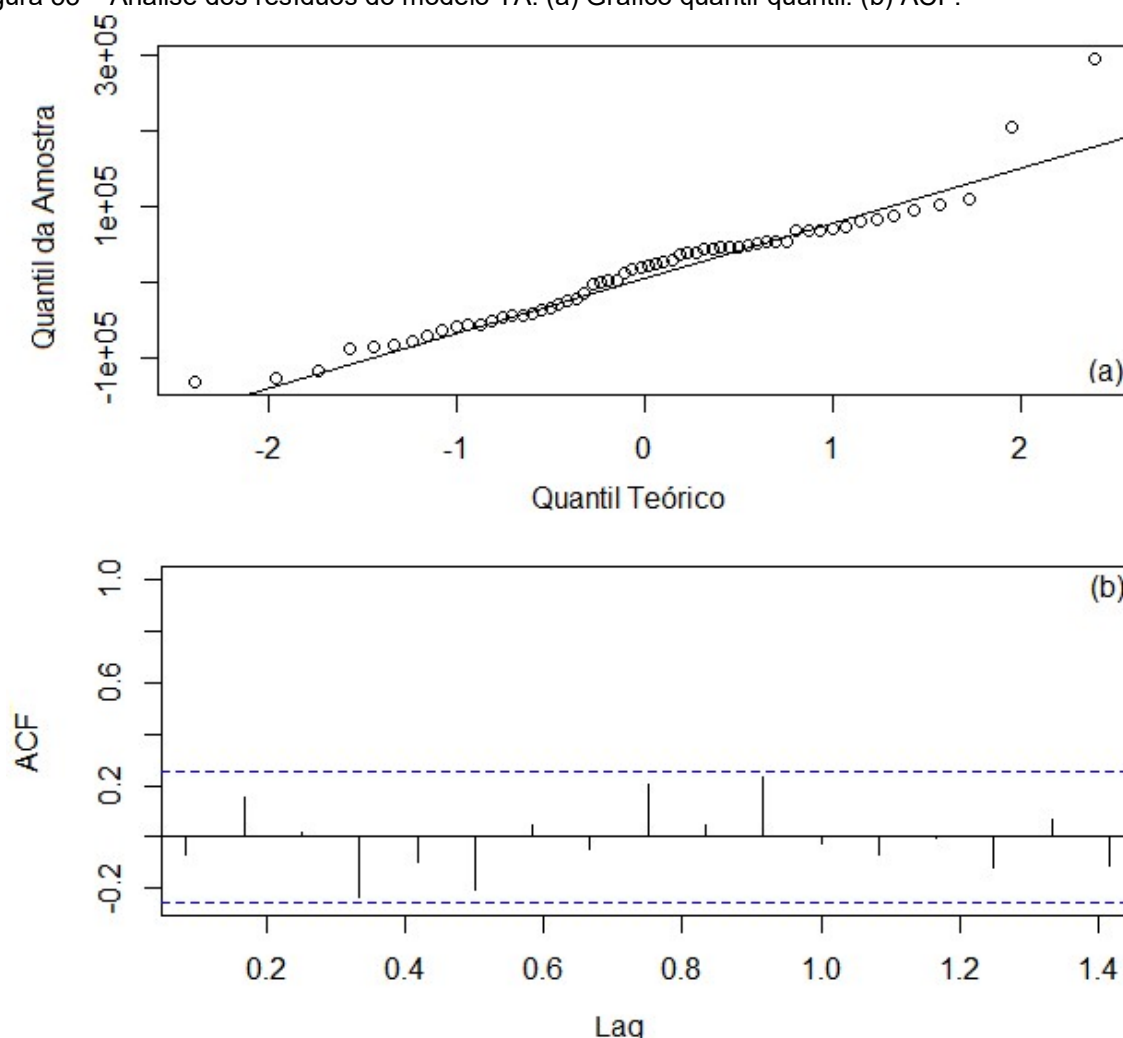
Tabela 30 – Coeficientes do modelo TA.

Modelo	AR1	MA1	SAR1	SMA1
TA	0,2926	-0,8954	0,8556	-0,5909

Fonte: a autora (2020).

Após a modelagem foram verificadas a distribuição normal e a existência de autocorrelação entre resíduos do modelo por meio da observação gráfica e dos testes estatísticos. Em relação a distribuição normal (Figura 53(a)), o valor obtido no teste de Lilliefors indica que os resíduos são normalmente distribuídos, com p-valor igual a 0,4886, enquanto o teste de Shapiro-Wilk aponta que os resíduos não possuem distribuição normal (p-valor igual a 0,007236). Na análise da autocorrelação, o resultado obtido no teste de Box-Pierce (p-valor igual a 0,7686) não indica autocorrelação entre os resíduos, como pode ser observado na Figura 53(b). A Tabela 32 apresenta os valores dos erros e do AIC obtidos no modelo TA.

Figura 53 – Análise dos resíduos do modelo TA. (a) Gráfico quantil-quantil. (b) ACF.



Fonte: a autora (2020).

Tabela 31 – Erros e AIC obtidos no modelo TA.

Modelo	RMSE	MAE	MPE	MAPE	MASE	AIC
TA	76.346,87	59.778,17	0,3863	2,1856	0,7049	1.510,06

Fonte: a autora (2020).

Comparando os modelos desenvolvidos para a categoria do consumo total de água em Joinville foram encontradas algumas diferenças. O modelo de suavização exponencial TSE foi um modelo SES com erros multiplicativos (M, N, N), que não considerou tendência nem sazonalidade. Os resíduos são normalmente distribuídos, mas autocorrelacionados. Já o modelo ARIMA desenvolvido foi do tipo  $(1,1,1)(1,0,1)[12]$  com um parâmetro autorregressivo e um parâmetro de médias móveis para a sazonalidade. Os resíduos do modelo TA foram considerados normalmente distribuídos e sem autocorrelação. Os resultados obtidos na modelagem desta e das demais categorias são discutidos a seguir.

## 5.6 ESCOLHA DOS MODELOS

Após a modelagem foram realizados os testes de acurácia nos modelos aceitos, incluindo os valores de consumo de cada categoria para o período cujas previsões foram simuladas. Os AICs, os MAPEs obtidos na modelagem (MAPEs de treinamento) e MAPEs de teste (Tabela 33) foram comparados e os modelos foram escolhidos.

Tabela 32 – Resumo dos resultados obtidos na modelagem.

<b>Categoria</b>	<b>Método</b>	<b>Id</b>	<b>Tipo</b>	<b>MAPE de Treinamento</b>	<b>MAPE de Teste</b>	<b>AIC</b>
<b>Residencial</b>	Suavização exponencial	RSE	(M, A, A)	1,95%	2,17%	1.577,14
<b>Comercial</b>	Suavização exponencial	CSE	(M, Ad, N)	2,89%	3,17%	1.359,45
	ARIMA	CA	$(1,1,0)(1,0,0)[12]$	2,83%	2,64%	1.266,09
<b>Industrial</b>	Suavização exponencial	ISE	(A, N, A)	4,43%	7,71%	1.359,06
	ARIMA	IA	$(0,1,1)(0,1,1)[12]$	4,65%	5,41%	1.014,03
<b>Público</b>	ARIMA	PA	$(0,0,1)(1,0,0)[12]$ com média diferente de zero	5,32%	17,99%	1.264,84
<b>Total</b>	ARIMA	TA	$(1,1,1)(1,0,1)[12]$	2,19%	0,63%	1.510,06

Fonte: a autora (2020).

Na categoria residencial, o modelo de suavização exponencial aplicado mostrou bons resultados na etapa de análise dos resíduos que apresentaram distribuição normal nos dois testes realizados. Em relação à autocorrelação, o gráfico

da ACF mostrou um prego tocando a linha pontilhada, mas considerando o resultado do teste de Box-Pierce que indicava ausência de autocorrelação com p-valor de 0,5351, o modelo foi aceito, conforme os critérios definidos na metodologia deste trabalho.

Já modelo ARIMA (RA) não foi considerado adequado por não possuir resíduos normalmente distribuídos em nenhum dos testes realizados. Dessa maneira, o modelo RSE foi escolhido, com erro percentual absoluto médio (MAPE) de treinamento de 1,95%, MAPE de teste de 2,17%.

Na categoria comercial os dois modelos desenvolvidos pela aplicação dos métodos se mostraram adequados para prever o consumo da categoria no município de Joinville. Na modelagem o modelo CSE apresentou um MAPE de 2,89%, enquanto o modelo CA calculou um MAPE de 2,83%. Considerando os resultados da fase de testes, o modelo CA foi considerado o mais adequado para a realização das previsões, calculando um MAPE de teste menor (2,64%) que o modelo CSE (3,17%).

Na categoria industrial, os dois modelos foram considerados qualificados por apresentarem resíduos normalmente distribuídos, atestados por pelo menos um dos testes de normalidade realizados, e com ausência de autocorrelação. No modelo ISE o MAPE de treinamento calculado foi de 4,43%, o MAPE de teste, de 7,71%. Nessa categoria, o modelo ARIMA (IA) foi considerado o mais adequado para a previsão, com MAPE de treinamento de 4,65% e MAPE de teste de 5,41%, mesmo apresentando indícios de ausência de normalidade na distribuição dos resíduos no teste de Lilliefors.

Na categoria público, o modelo PSE, produzido pelo método de suavização exponencial não foi aceito para a previsão por conter dois pregos ultrapassando a linha limítrofe no gráfico da ACF, apesar de o teste de Box-Pierce considerar ausência de autocorrelação com p-valor de 0,4297, e os testes de normalidade indicarem distribuição normal dos resíduos. Nesta categoria, o modelo PA foi considerado qualificado para a realização das previsões, por possuir resíduos normalmente distribuídos e sem autocorrelação. Na etapa de treinamento o MAPE obtido foi de 5,35%. Entretanto, na etapa da verificação da acurácia, o MAPE de teste calculado foi de 17,99%, indicando que o modelo ARIMA não foi tão efetivo para captar a estrutura da série, tendo melhor desempenho na adequação à curva dos dados observados do que na previsão propriamente dita.



Na categoria total, apenas o modelo TA foi considerado adequado para a realização da previsão. Na modelagem, o modelo desenvolvido por meio de suavização exponencial (TSE) apresentou resíduos com três pregos no gráfico da ACF, indicando autocorrelação, corroborada pelo resultado do teste de Box-Pierce (p-valor igual a 0,039) e por isso foi desconsiderado. O MAPE de treinamento obtido no modelo TA foi de 2,16% e o de teste, de 0,61%, o menor entre todos os modelos desenvolvidos neste trabalho.

Neste trabalho, nos modelos de suavização exponencial o principal problema encontrado foi a presença de autocorrelação nos resíduos, enquanto nos modelos desenvolvidos pelo método ARIMA a ausência de distribuição normal dos resíduos foi o motivo da não aceitação. Entretanto, é importante observar que os problemas de distribuição normal e autocorrelação encontrados nos resíduos são peculiaridades de cada modelo, não havendo relação entre a classe de problema e o método escolhido, já que a adequabilidade do modelo é função do comportamento da série em estudo (CHECHI; BAYER, 2011).

Comparando os resultados obtidos, os modelos desenvolvidos pelo método ARIMA tiveram melhor desempenho em quatro das cinco categorias analisadas. Apenas na categoria residencial o método de suavização exponencial produziu um modelo mais assertivo. O modelo ARIMA da categoria residencial poderia ser melhorado com a aplicação de transformações matemáticas nos dados da série antes da modelagem, com o propósito de tornar o padrão mais consistente em todo o conjunto de dados (HYNDMAN; ATHANASOPOULOS, 2013).

Como o objetivo do trabalho foi realizar a modelagem automática, com o mínimo de interferência do previsor, as transformações não foram realizadas. Já que todas as séries possuíam distribuição normal, assumiu-se que o *software* realizaria as diferenciações necessárias para tornar a série estacionária.

Os resultados deste trabalho foram bastante semelhantes aos obtidos por Razali *et al.* (2018), quando na comparação entre os métodos de suavização exponencial e ARIMA para previsão do gasto mensal em consumo de água, na Universidade Tun Hussein Onn Malaysia, o método ARIMA se mostrou mais assertivo, com MAPE menor. Nos modelos ARIMA foram realizadas diferenciações na série.

No trabalho de Chechi e Bayer (2011), na comparação dos dois métodos de previsão, o modelo ARIMA também foi o mais acurado para a previsão das

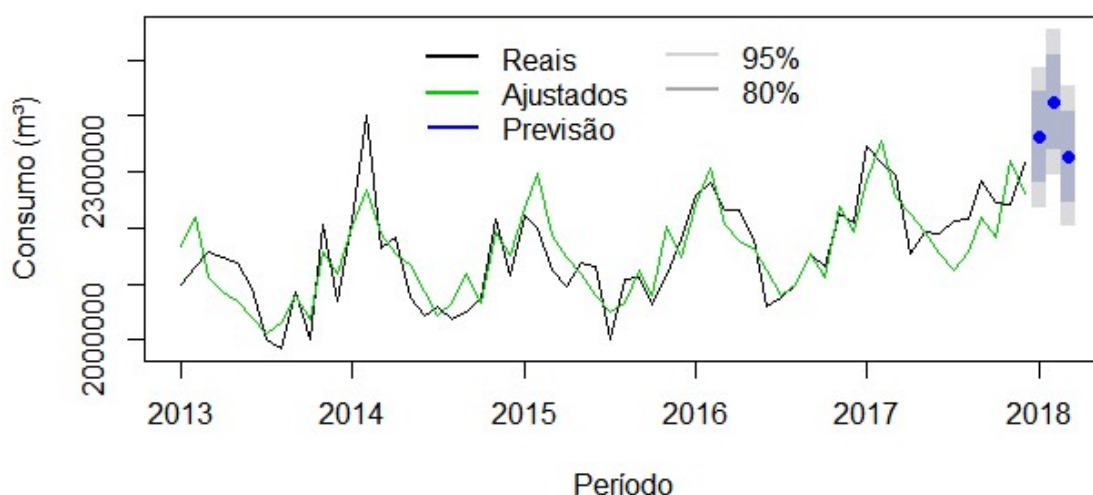
temperaturas médias mensais na cidade de Erechim. Todos os modelos ARIMA desenvolvidos pelos autores apresentaram sazonalidade.

Ainda em outras áreas de estudo, na comparação entre os dois métodos, os trabalhos de Newaz (2008) e Assis *et al.* (2010) também reasseguram a superioridade do método ARIMA sobre o de suavização exponencial para a realização de previsões de curto prazo.

## 5.7 PREVISÃO

As previsões foram realizadas para todas as categorias com um horizonte de 3 meses. Na categoria residencial o modelo utilizado para a realização das previsões foi o modelo RSE. Na Figura 54 é possível observar a função de dados de consumo residencial do modelo (em preto), a função ajustada de previsão (em verde) e os intervalos de confiança de 80% (cinza escuro) e 95% (cinza claro). Os pontos azuis representam as previsões para o período de janeiro a março de 2018.

Figura 54 – Gráfico de previsão de consumo de água na categoria residencial em Joinville para o modelo RSE.

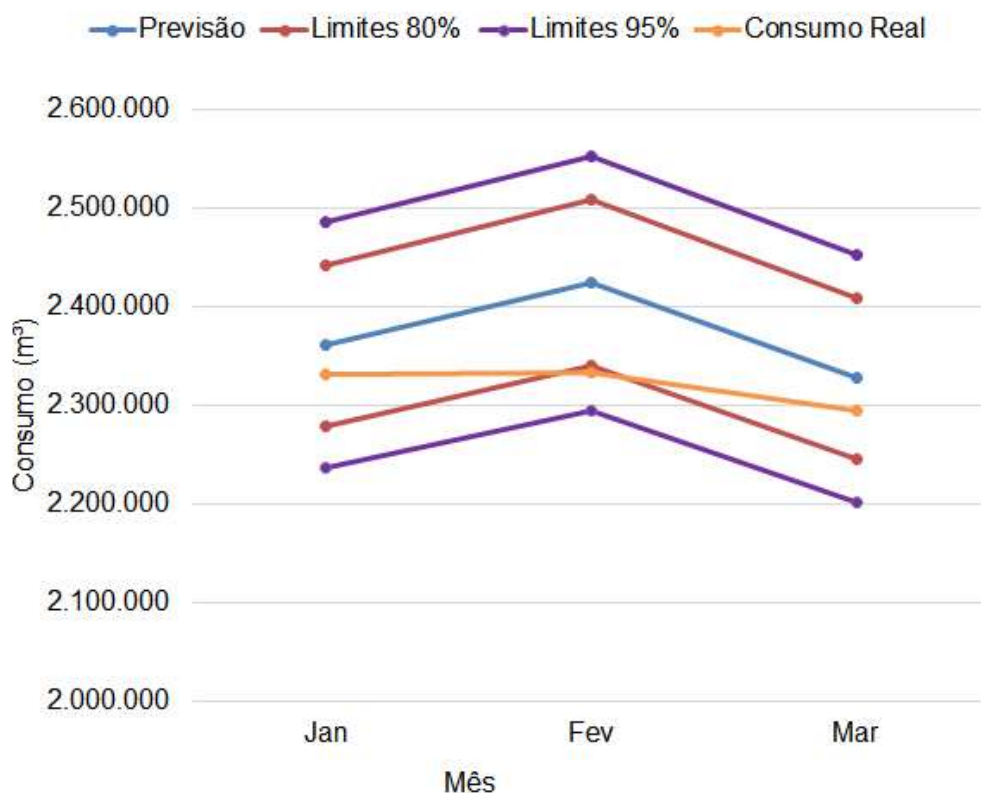


Fonte: a autora (2020).

Nas previsões realizadas para a categoria residencial esperava-se um consumo no trimestre de cerca de 7,11 milhões de m<sup>3</sup>, com aumento de consumo no mês de janeiro, quando comparado a dezembro de 2017, um crescimento ainda maior no mês de fevereiro com queda a partir do mês de março (Figura 54). Observando o

consumo real, nota-se que o maior valor de consumo foi registrado no mês de fevereiro e o menor no mês de março (Figura 55).

Figura 55 – Curvas de previsão e de demanda de água na categoria residencial em Joinville, em m<sup>3</sup>, pelo modelo RSE, no período de janeiro a março de 2018.



Fonte: a autora (2020).

A diferença média entre os valores previstos e os registrados (Tabela 34) foi de 2,11%, com maior discrepância no mês de fevereiro, quando o modelo estimou um consumo maior em cerca de 91 mil m<sup>3</sup>, deixando o valor de consumo observado abaixo do limite inferior do intervalo de confiança de 80% (Figura 55).

Tabela 33 – Consumos de água residencial previstos e observados em Joinville, para o ano de 2018.

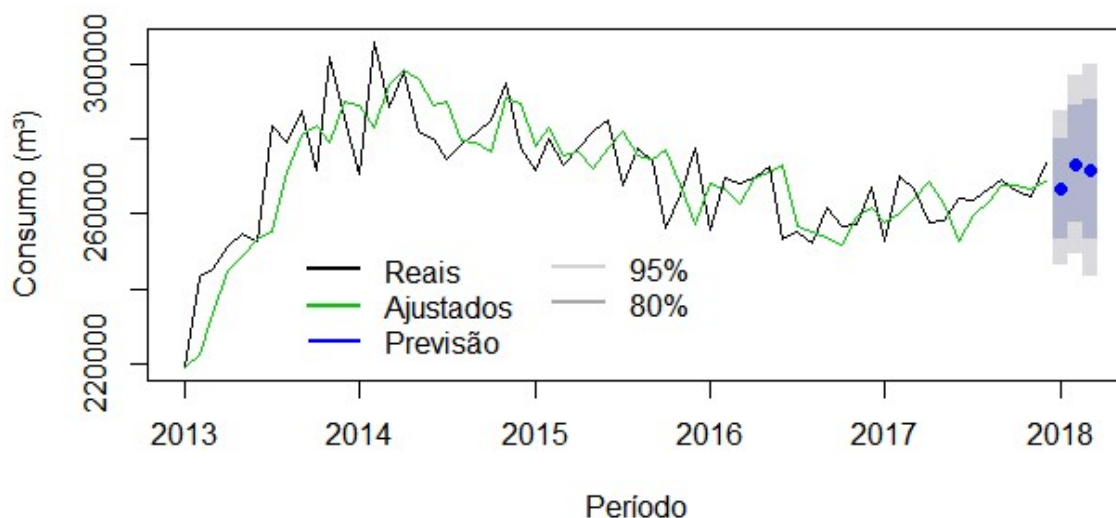
<b>Categoria</b>	<b>Modelo</b>	<b>Mês</b>	<b>Previsão (m³)</b>	<b>Consumo real (m³)</b>	<b>Diferença</b>
<b>Residencial</b>	<b>RSE</b>	Jan	2.360.909	2.332.073	1,22%
		Fev	2.424.287	2.333.264	<b>3,75%</b>
		Mar	2.327.116	2.295.320	1,37%
		<b>Diferença Média</b>			<b>2,11%</b>

Fonte: a autora (2020).

Na categoria comercial o modelo definido para a realização das previsões foi o modelo CA. A função de dados de consumo comercial do modelo (em preto), a função

ajustada de previsão (em verde) e os intervalos de confiança de 80% (cinza escuro) e 95% (cinza claro) podem ser observados na Figura 56. Os pontos azuis representam as previsões para o período de janeiro a março de 2018, cujos valores são mostrados na Tabela 35, totalizando para o trimestre um consumo de cerca de 812 mil m<sup>3</sup>.

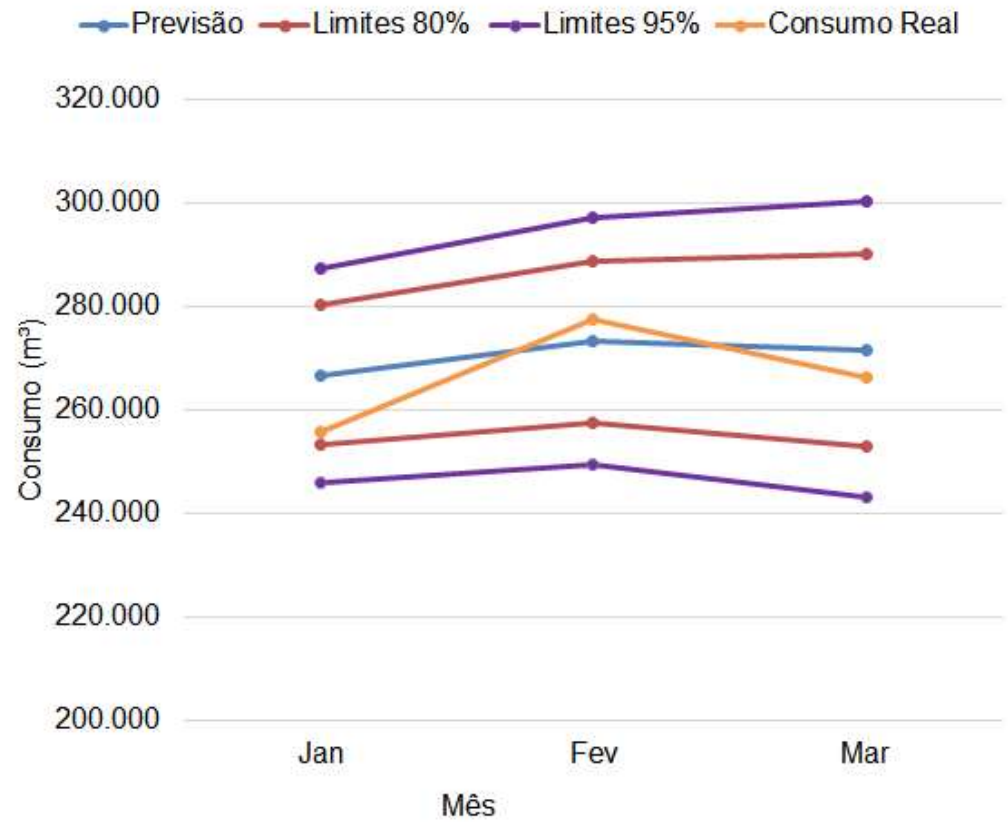
Figura 56– Gráfico de previsão de consumo de água na categoria comercial em Joinville para o modelo CA.



Fonte: a autora (2020).

Na categoria comercial a Figura 56 mostra que na previsão, a demanda deveria ser menor em janeiro e maior no mês de fevereiro, como ocorreu com o consumo real (Figura 57). A diferença média observada entre os valores consumidos e os preditos foi de 2,57% (Tabela 35), com destaque para o mês de fevereiro onde o modelo subestimou o consumo da categoria em cerca de 4 mil m<sup>3</sup>. Em janeiro o modelo previu um valor de consumo de aproximadamente 11 mil m<sup>3</sup> maior que o registrado. Todos os valores consumidos ficaram dentro dos intervalos de confiança ajustados pelo modelo.

Figura 57 – Curvas de previsão e de demanda de água na categoria comercial em Joinville, em m³, pelo modelo CA, no período de janeiro a março de 2018.



Fonte: a autora (2020).

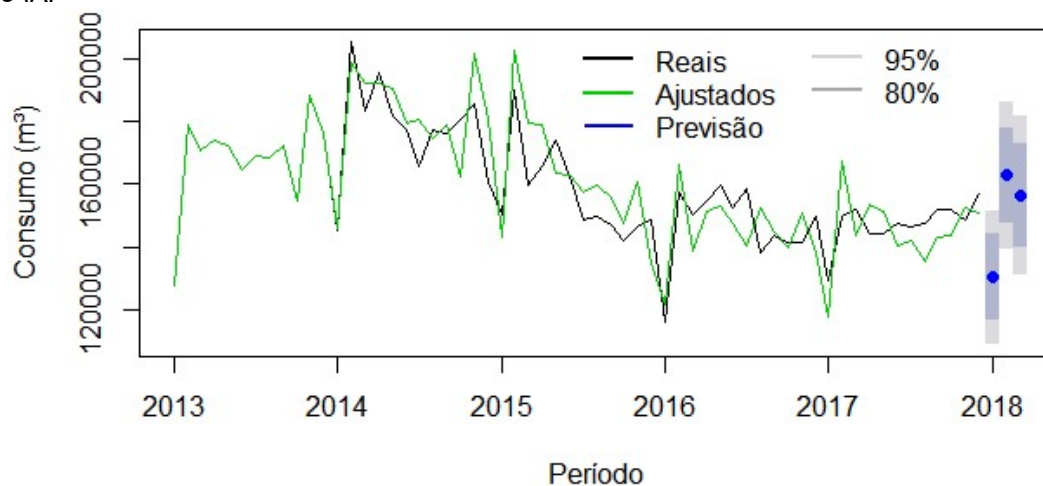
Tabela 34 – Consumos de água comercial previstos e observados em Joinville, para o ano de 2018.

Categoria	Modelo	Mês	Previsão (m³)	Consumo real (m³)	Diferença
Comercial	CA	Jan	266.767	255.708	4,15%
		Fev	273.312	277.463	-1,52%
		Mar	271.784	266.217	2,05%
		Diferença Média			2,57%

Fonte: a autora (2020).

O modelo escolhido para prever o consumo no primeiro trimestre de 2018 na categoria industrial foi o IA. A Figura 58 mostra a função de dados de consumo industrial (em preto) e a curva ajustada de previsão (em verde), os intervalos de confiança de 80% (cinza escuro) e 95% (cinza claro) para o modelo. Os pontos azuis são as previsões para os meses de janeiro, fevereiro e março de 2018. A previsão de consumo para o trimestre foi de aproximadamente 449 mil m³.

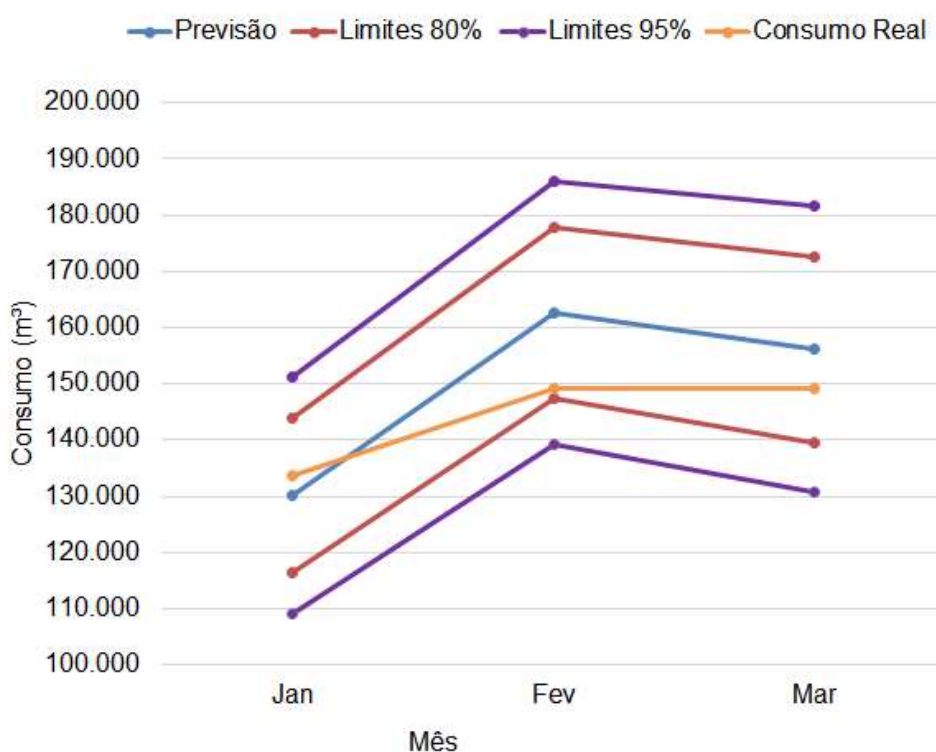
Figura 58 – Gráfico de previsão de consumo de água na categoria industrial em Joinville para o modelo IA.



Fonte: a autora (2020).

Da mesma maneira como o modelo de previsão da categoria comercial, as previsões para a categoria industrial indicaram que o menor consumo deveria ser esperado para o mês de janeiro e o maior consumo, para o mês de fevereiro, como mostra a Figura 58. Entretanto, na Figura 59 é possível observar que a maior demanda de água se deu no mês de março.

Figura 59 – Curvas de previsão e de demanda de água na categoria industrial em Joinville, em  $m^3$ , pelo modelo IA, no período de janeiro a março de 2018.



Fonte: a autora (2020).

As previsões realizadas na categoria industrial indicaram uma diferença média de 5,12% quando comparadas aos valores de consumo registrados (Tabela 36). A maior discrepância ocorreu no mês de fevereiro, quando o modelo superestimou o consumo em aproximadamente 14,5 mil m³. No mês de janeiro, o consumo foi subestimado pelo modelo em cerca de 3,5 mil m³. Todos os valores consumidos ficaram dentro dos intervalos de confiança de 80% e 95%.

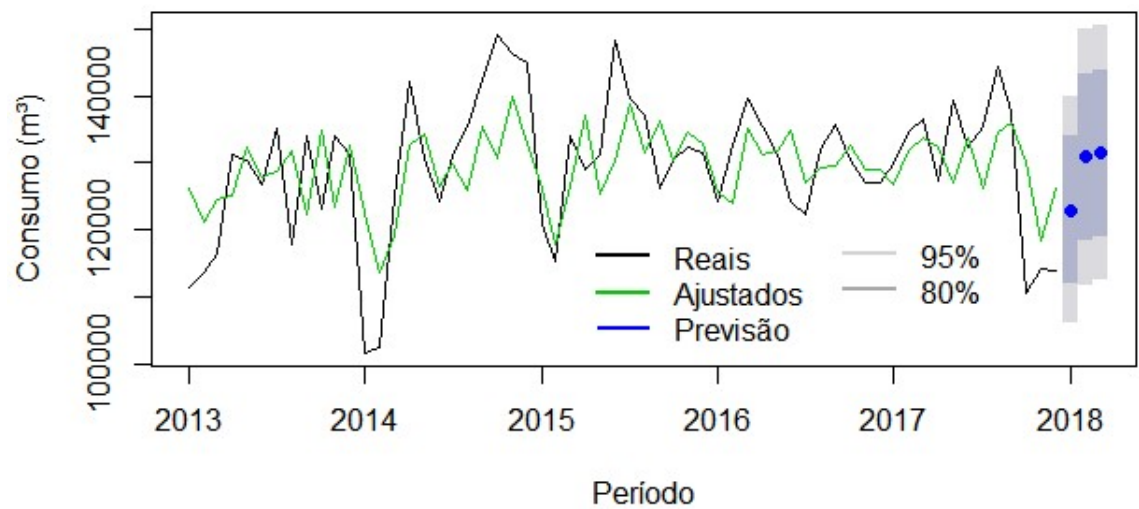
Tabela 35 – Consumos de água industrial previstos e observados em Joinville, para o ano de 2018.

Categoria	Modelo	Mês	Previsão (m³)	Consumo real (m³)	Diferença
Industrial	IA	Jan	130.155	133.696	-2,72%
		Fev	162.682	149.207	8,28%
		Mar	156.093	149.292	4,36%
Diferença Média					5,12%

Fonte: a autora (2020).

Na categoria público o modelo utilizado para a realização das previsões foi o modelo PA. Na Figura 60 são apresentadas a função de dados de consumo da categoria público (em preto), a função ajustada de previsão (em verde), os intervalos de confiança de 80% (cinza escuro) e 95% (cinza claro) para o modelo. As previsões para os meses de janeiro, fevereiro e março de 2018 são representadas pelos pontos azuis. Para o trimestre foi estimado um consumo de aproximadamente 385 mil m³.

Figura 60 – Gráfico de previsão de consumo de água na categoria público em Joinville para o modelo PA.



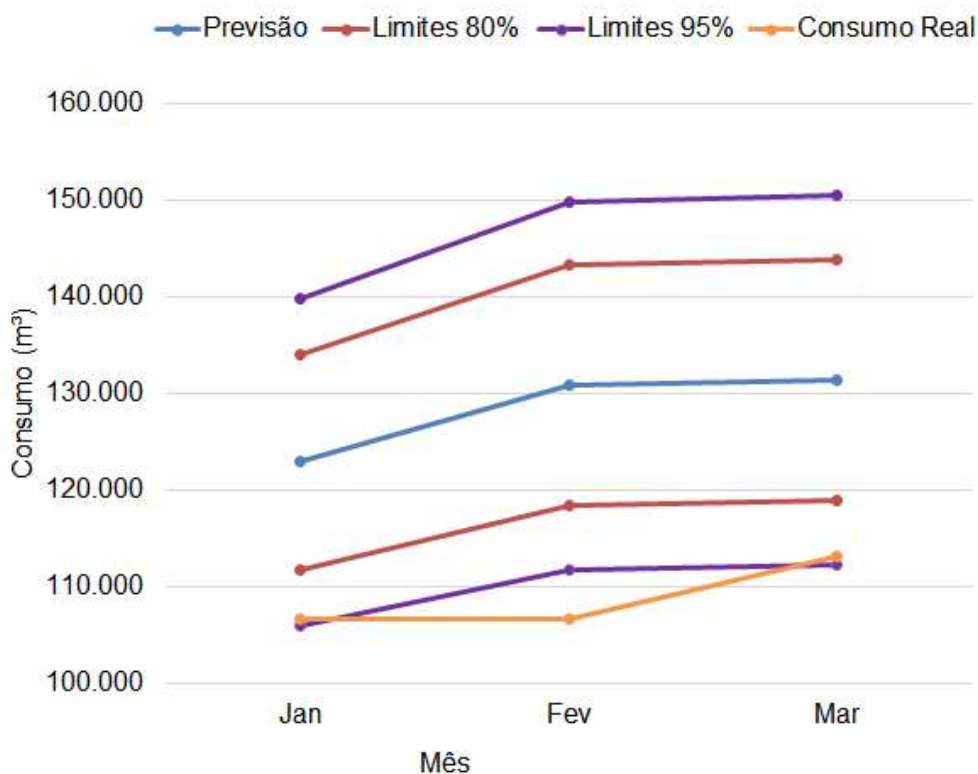
Fonte: a autora (2020).

As previsões para a categoria público indicaram que no primeiro trimestre de 2018 o menor consumo deveria ocorrer em janeiro e o maior consumo seria esperado



para o mês de março, como mostra a Figura 60. No entanto, o menor consumo registrado para a categoria ocorreu no mês de fevereiro (Figura 61).

Figura 61 – Curvas de previsão e de demanda de água na categoria público em Joinville, em m<sup>3</sup>, pelo modelo PA, no período de janeiro a março de 2018.



Fonte: a autora (2020).

No período de previsão houve uma queda no consumo que não foi captada pelo modelo, que estimou valores para o primeiro trimestre de 2018 muito próximos à média de consumo da amostra de treinamento. Comparativamente, a diminuição no consumo foi tão expressiva que no mês de fevereiro o valor registrado não estava contido nos intervalos de confiança do modelo, e nos meses de janeiro e março, os consumos observados não estavam dentro do intervalo de confiança de 80% (Figura 61).

O modelo da categoria público foi o que apresentou maior diferença entre a previsão e o consumo real. O modelo ARIMA PA apresentou MAPE de teste de aproximadamente 18%, o maior entre todos os modelos desenvolvidos neste trabalho, e a diferença média entre os valores preditos e observados foi de 15,18% (Tabela 37). Esta diferença, de cerca de 19 mil m<sup>3</sup>, pode ser considerada bastante representativa uma vez que a média de consumo da categoria é de cerca de 130 mil m<sup>3</sup>.



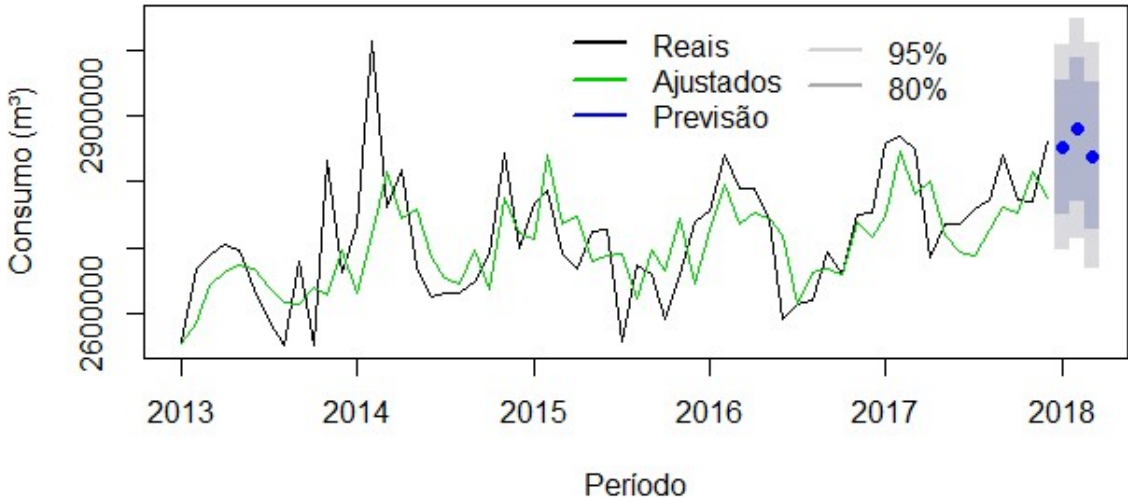
Tabela 36 – Consumos de água pública previstos e observados em Joinville, para o ano de 2018.

<b>Categoria</b>	<b>Modelo</b>	<b>Mês</b>	<b>Previsão (m³)</b>	<b>Consumo real (m³)</b>	<b>Diferença</b>
<b>Público</b>	PA	Jan	122.934	106.708	13,20%
		Fev	130.836	106.720	<b>18,43%</b>
		Mar	131.422	113.138	13,91%
		<b>Diferença Média</b>			15,18%

Fonte: a autora (2020).

Na categoria total o modelo definido para realizar as previsões de consumo de água foi o modelo TA. Na Figura 62 pode-se observar os pontos azuis, que representam as previsões para os meses de janeiro, fevereiro e março de 2018, a função de dados de consumo total (em preto) e a curva ajustada de previsão (em verde), os intervalos de confiança de 80% (cinza escuro) e 95% (cinza claro) para o modelo. De acordo com o modelo TA era esperado um consumo no período de aproximadamente 8,57 milhões de m³.

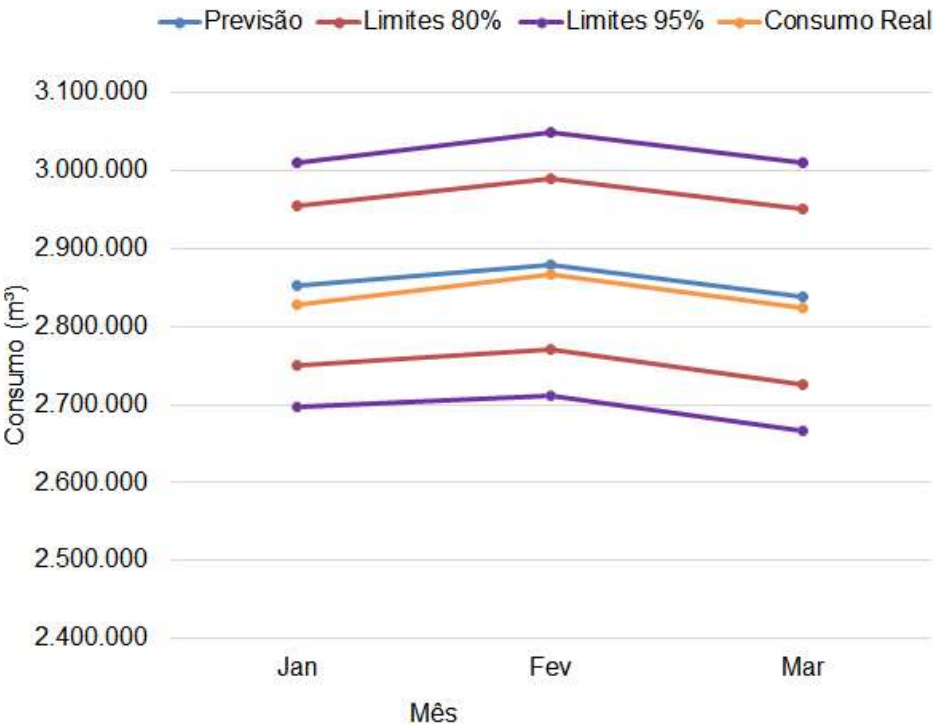
Figura 62 – Gráfico de previsão de consumo de água na categoria total em Joinville para o modelo TA.



Fonte: a autora (2020).

Assim como nas categorias residencial, comercial e industrial, as previsões para a categoria total indicaram que no primeiro trimestre de 2018 o maior consumo deveria ser esperado para o mês de fevereiro, como ocorreu com os valores de consumo observados no período (Figura 63).

Figura 63 – Curvas de previsão e de demanda de água na categoria total em Joinville, em m³, pelo modelo TA, no período de janeiro a março de 2018.



Fonte: a autora (2020).

Nas previsões realizadas para a categoria total (Tabela 38), a diferença média entre os valores previstos e os registrados foi de 0,63%, com maior acurácia no mês de fevereiro, quando o modelo superestimou o consumo em cerca de 13,5 mil m³ (0,47%). A maior diferença foi observada no mês de janeiro com uma previsão de consumo 0,88% maior que o consumo registrado. No modelo TA todos os valores observados ficaram dentro dos intervalos de confiança de 80% e 95% ajustados.

Tabela 37 – Consumos de água total previstos e observados em Joinville, para o ano de 2018.

Categoria	Modelo	Mês	Previsão (m³)	Consumo real (m³)	Diferença
Total	TA	Jan	2.853.297	2.828.185	0,88%
		Fev	2.880.177	2.866.654	0,47%
		Mar	2.839.161	2.823.967	0,54%
		Diferença Média			0,63%

Fonte: a autora (2020).

Na comparação dos resultados da previsão, o modelo que apresentou maior diferença entre os valores previstos e o consumo real foi o modelo ARIMA da categoria público, com diferença no mês de fevereiro em torno de 18%. O modelo cujas

previsões ficaram mais próximas da realidade foi o modelo ARIMA da categoria total, que superestimou em apenas 0,47% o consumo do mês de fevereiro de 2018.

Nesta ótica, os resultados do modelo TA para o mês de fevereiro são bastante interessantes, considerando que três das quatro categorias analisadas isoladamente apresentaram maior diferença entre os consumos preditos e registrados neste mês.

Considerando a soma das previsões de consumo categorizado (categorias residencial, comercial, industrial e público), o total previsto para o trimestre é de 8.758.298 m<sup>3</sup>, cerca de 186 mil m<sup>3</sup> (2,16%) maior que a previsão obtida no modelo TA (8.572.635 m<sup>3</sup>), cujos MAPE de teste e diferença média entre os valores previstos e registrados foram de 0,63%.

Em comparação com os valores de consumo registrados, a previsão de demanda realizada por categoria superestimou em 239.492 m<sup>3</sup> o consumo, enquanto a previsão do modelo TA superestimou o consumo real em 53.829 m<sup>3</sup>.

Esses resultados podem indicar que, para a cidade de Joinville, a realização de previsões sem categorização do consumo poderia ser mais assertiva. No entanto, é importante ressaltar, que os resultados obtidos neste trabalho se aplicam apenas às previsões para os meses iniciais de 2018, para a cidade de Joinville. Se o horizonte de previsão fosse alterado, os resultados poderiam apresentar mudanças significativas.

Dessa maneira, a escolha de realizar a previsão com o consumo dividido em categorias deve ficar a critério do pesquisador, dependendo da disponibilidade de dados e do interesse em se estudar o comportamento de cada tipo de consumidor.

## 6 CONSIDERAÇÕES FINAIS

De maneira geral, os resultados mostram que a metodologia desenvolvida neste trabalho pode ser aplicada para a previsão de curto prazo do consumo urbano (podendo apresentar variação na escolha do método, dependendo com comportamento da série de consumo), desde que haja disponibilidade de dados com o número mínimo de observações requeridas para a aplicação do método ARIMA.

Outras vantagens da metodologia desenvolvida neste trabalho é a aplicação de métodos clássicos de previsão; o uso de *software* livre; facilidade no treinamento de pessoal; e facilidade de adaptação para outras situações em que se deseja prever o consumo de água.

No estudo de caso da cidade de Joinville, para a categoria residencial o método ARIMA não se mostrou adequado para a modelagem, ajustando um modelo com resíduos normalmente distribuídos, mas autocorrelacionados. Para a realização das previsões o modelo escolhido foi o produzido pelo método de suavização exponencial (RSE) que obteve MAPE de treinamento de 1,95% e MAPE de teste de 2,17%, indicando que o método é adequado não só se ajustando aos valores passados, mas também sendo preciso na previsão de valores futuros.

Na categoria comercial os métodos de suavização exponencial e ARIMA produziram bons modelos, com seus resíduos apresentando distribuição normal e sem autocorrelação. O melhor modelo da categoria foi o desenvolvido pelo método ARIMA (CA), com um MAPE de treinamento de 2,83% e MAPE de teste de 2,64%, mostrando que o modelo foi capaz de se adequar à série de dados e também gerar boas previsões. O mesmo aconteceu com a categoria industrial, onde o modelo IA, desenvolvido por meio do método ARIMA, apresentou MAPE de treinamento de 4,65% e MAPE de teste de 5,41%.

Na categoria público, na comparação entre os modelos PA e PSE foi observado que o modelo PSE não era adequado para a realização das previsões por possuir indícios de autocorrelação nos resíduos. Já o modelo PA apresentou bons resultados quando se trata da fase do ajuste da curva de previsão aos valores passados, com MAPE de 5,32%, mas na realização das previsões, o MAPE de teste calculado foi três vezes maior (17,99%). Essa diferença pode ser explicada por uma queda de consumo no primeiro trimestre de 2018 que não foi captada pelo modelo ARIMA.

Na modelagem do consumo total, assim como na categoria público, o modelo ajustado pelo método de suavização exponencial (TSE) apresentou resíduos autocorrelacionados e não foi utilizado para a realização da previsão. O modelo TA apresentou MAPE de treinamento de 2,19% e MAPE de teste de 0,63% indicando que o modelo captou o comportamento da série e foi capaz de produzir previsões bastante precisas, com menor MAPE de teste entre todos os modelos ajustados em todas as categorias.

Na análise da soma das previsões categorizadas e as previsões realizadas pelo modelo da categoria total, observa-se uma diferença da ordem de 2,16%, cerca de 1,5% maior que o previsto pelo modelo TA. Esses resultados podem indicar que para a cidade de Joinville a previsão do consumo urbano (sem categorias de consumo) seria mais assertiva. No entanto, a categorização ou não do consumo pode ser definida, sem grandes prejuízos na previsão, de acordo com os objetivos almejados na sua realização e pela disponibilidade de dados.

Quando se analisa a indicação de um método específico para a realização da modelagem do consumo de água em Joinville, o método ARIMA mostrou indícios de ser mais assertivo, pois produziu modelos com bons resultados em quatro das cinco categorias estudadas, tanto na verificação da adequação do modelo como na avaliação das medidas de erro de previsão.

Com relação às previsões, o estudo apontou tendência de crescimento em todas as categorias analisadas, quando comparadas ao consumo de água no mesmo período de 2017. Nas categorias residencial, comercial, industrial e total, os maiores valores de consumo previstos pelos modelos encontram-se no mês de fevereiro, e apenas na categoria público o maior valor previsto foi para o mês de março.

Por se tratarem de previsões de curto prazo, as informações resultantes desse trabalho podem ser aplicadas no gerenciamento e otimização dos sistemas operacionais (DONKOR *et al.*, 2014), auxiliando a identificação de possíveis falhas de rede como vazamentos e rompimento de tubulações (BRENTAN *et al.*, 2017), e na tomada de decisões, como por exemplo, a implementação de restrições regulatórias para o uso da água em épocas de estresse hídrico ou seca (TIWARI; ADAMOWSKI, 2015).

Em função dos resultados obtidos neste trabalho, conclui-se que os objetivos foram atingidos, com algumas limitações, principalmente ligadas ao uso das funções automáticas, que limitaram uma exploração mais profunda dos métodos de previsão.

Dessa maneira se fazem necessárias como sugestões para trabalhos futuros: a definição manual de alguns parâmetros dos modelos e manipulação dos dados; a realização do monitoramento dos erros de previsão e da verificação da acurácia, considerando a alimentação dos modelos com dados mais recentes; e a definição de um valor de erro máximo admitido, para determinar quando uma nova modelagem para as séries de consumo deve ser realizada.

É importante salientar que a adoção de qualquer modelo estatístico de previsão pressupõe uma escolha adequada deste e de seus parâmetros, uma vez que o grau de acurácia da previsão de demanda depende em parte destas escolhas.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ADAMOWSKI, J.; CHAN, H. F.; PRASHER, S. O. OZGA-ZIELINSKI, B. SLIUSARIEVA, A. Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. **Water Resources Research**, v. 48, 2012.
- ADAMOWSKI, J.; KARAPATAKI, C. Comparison of multivariate regression and artificial neural networks for peak urban water-demand forecasting: evaluation of different ANN learning algorithms. **ASCE Journal of Hydrologic Engineering**, v. 15, p. 729-743, 2010.
- ANA. AGÊNCIA NACIONAL DE ÁGUAS. **Conjuntura dos recursos hídricos no Brasil**. Brasília, 2017. 100p. Disponível em: < <http://conjuntura.ana.gov.br/>>. Acesso em: 20 mai. 2020.
- AL-ZHARANI, M. A.; ABO-MONASAR, A. Urban residential water demand prediction based on artificial neural networks and time series models. **Water Resources Management**, v. 29, p. 3651-3662, 2015.
- AMPONSAH, S. K.; OTOO, D.; TODOKO, C. A. K. Time series analysis of water consumption in the Hohoe municipality of the Volta region, Ghana. **International Journal of Applied Mathematical Research**, v. 4, n. 2, p. 393-403, 2015.
- ARANDIA, E.; BA, A.; ECK, B. McKEENA, S. Tailoring seasonal time series models to forecast short-term water demand. **Journal of Water Resources Planning and Management**, v. 142, n. 3. 2016.
- ASSIS, K.; AMRAN, A.; REMALI, Y.; AFFENDY, H. A Comparison of univariate time series methods for forecasting cocoa bean prices. **Trends in Agricultural Economics**, v.3, p.207-215, 2010.
- BABEL, M. S.; GUPTA, A. Das; PRADHAN, P. A multieconometric approach for domestic water demand modeling: an application to Kathmandu, Nepal. **Water Resources Management**, v. 21, p. 573-589, 2007.
- BABEL, M. S.; SHINDE, V. R. Identifying Prominent Explanatory Variables for Water Demand Prediction Using Artificial Neural Networks: A Case Study of Bangkok. **Water Resources Management**, v. 25, p. 1653-1676, 2011.
- BACH, P. M.; RAUCH, W. MIKKELSEN, P. S.; McCARTHY, D. T., DELETIC, A. A critical review of integrated urban water modelling – urban drainage and beyond. **Environmental Modelling & Software**, v. 54, p. 88-107, 2014.
- BAI, Y. WANG, P. LI, C. XIE J. WANG Y. A multi-scale relevance vector regression approach for daily urban water demand forecasting. **Journal of Hydrology**, v. 517, p. 236-245, 2014.

BEECHER, J. A.; CHESTNUT, T. W. Declining Water Sales and Utility Revenues: A framework for understanding and adapting. **National Water Rates Summit.**, Racine, Wisconsin, Alliance for Water Efficiency, 2012. 22 p.

BERTOLAZZI, L. CUSTÓDIO, D. A. Análise econômica de soluções para reduzir o consumo de água potável – estudo de caso de uma residência em Joinville – SC. **Revista Gestão & Sustentabilidade Ambiental**, v. 9, n. esp., p. 186-206, 2020.

BEZ, R. **Uma Análise do Setor de Abastecimento de Água por Categoria de Consumo: O Caso de Santa Catarina.** Florianópolis, 2012. 91 p. Dissertação. (Programa de Pós-Graduação em Economia) – Universidade Federal de Santa Catarina, Florianópolis, 2012.

BILLAH, B.; KING, M. L.; SNYDER, R. D.; KOEHLE, A. B. (2006). Exponential smoothing model selection for forecasting. **International Journal of Forecasting**, v.22, n. 2, p. 239-247, 2006.

BILLINGS, B.; JONES, C. **Forecasting Urban Water Demand.** 2 ed. Denver: American Waterworks Association. 2008. 340 p.

BILLINGS, R.; AGTHE, D. State-space versus multiple regression for forecasting urban water demand. **Journal of Water Resource Planning and Management**, v. 124, n. 2, p. 113-117, 1998.

BOX, G. E. P.; JENKINS, G. M. **Time series analysis forecasting and control.** San Francisco: H. Day, 1976.

BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. **Time series analysis: forecasting and control.** 4 ed. New Jersey: John Wiley & Sons, Inc. 2008. 756 p.

BRADFORD, W.; BRIDGEMAN, J.; GATERELL, M. A review of the 1892 water demand forecasts for Birmigham. **Engineering Hystory and Heritage**, v. 164, p. 39-49, 2010.

BRASIL. **Decreto nº 7.217**, de 21 de junho de 2010. Regulamenta a Lei nº 11.445, de 5 de janeiro de 2007, que estabelece diretrizes nacionais para o saneamento básico, e dá outras providências. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/\\_Ato2007-2010/2010/Decreto/D7217.htm](http://www.planalto.gov.br/ccivil_03/_Ato2007-2010/2010/Decreto/D7217.htm)>. Acesso em 18 jun. 2018.

BRASIL. **Lei Federal nº 9.433**, de 8 de janeiro de 1997. Institui a Política Nacional de Recursos Hídricos, cria o Sistema Nacional de Gerenciamento de Recursos Hídricos, regulamenta o inciso XIX do art. 21 da Constituição Federal, e altera o art. 1º da Lei nº 8.001, de 13 de março de 1990, que modificou a Lei nº 7.990, de 28 de dezembro de 1989. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/LEIS/L9433.htm](http://www.planalto.gov.br/ccivil_03/LEIS/L9433.htm)>. Acesso em 01 abr. 2017.



BRASIL. Ministério do Meio Ambiente, Secretaria de Recursos Hídricos. **Caderno setorial de recursos hídricos: saneamento**. Brasília: MMA, 2006. Disponível em: <<http://www.mma.gov.br/publicacoes/agua/category/42recursoshidricos>>. Acesso em 01 abr. 2017.

BRENTAN, B. M.; LUVIZOTTO JR., E. HERRERA, M.; IZQUIERDO, J.; PEREZGARCIA, R. Hybrid regression model for near real-time urban water demand forecasting. **Journal of Computational and Applied Mathematics**, v. 309, p. 532-541, 2017.

BROWN, R. G., **Statistical Forecasting for Inventory Control**, New York: McGraw-Hill, 1959.

CABRAL, P.; AUGUSTO, G.; TEWOLDE, M.; ARAYA, Y. Entropy in urban systems. **Entropy**, v. 15, p. 5223-5236, 2013.

CAIADO, J. Performance of combined double seasonal univariate time series models for forecasting water demand. **Journal of Hydrologic Engineering**, v. 15, n. 3, p. 515-222, 2010.

CAJ. Companhia Águas de Joinville. **Plano de Vendas 2018/20**. Joinville, 2017.

CAJ. Companhia Águas de Joinville. **Consumo urbano de água micromedido por economia em Joinville**. Dados de janeiro de 2007 a março de 2018. Joinville, 2018

CHATFIELD, C. **The Analysis of Time Series, An Introduction** 6 ed. Boca Raton. Chapman & Hall/CRC, 2005. 357 p.

CHECHI, L.; BAYER, F. M. Modelos univariados de séries temporais para previsão das temperaturas médias mensais de Erechim, RS. **Revista Brasileira de Engenharia Ambiental**, v. 16, n. 12, p. 1321-1329, 2012.

CRIBARI NETO, F. O comportamento estocástico do produto no Brasil. **Pesquisa e Planejamento Econômico**, v. 2, p. 381-402, 1990.

COMUNELLO, A. C.; DAMINELLI, L. M.; BORTOLOTTI, S. L. V. Análise e previsão da demanda de água no município de Medianeira-PR usando o modelo de Holt-Wintersthematic. **Revista Eletrônica Científica Inovação e Tecnologia**, v. 2, n. 8, p. 10-19, 2013.

DANIELL, K. A.; RINAUDO, J. D.; CHAN, N. W. W.; NAUGES, C.; GRAFTON, Q. Understanding and managing urban water in transition. In: GRAFTON, Q; DANIELL, K. A; NAUGES, C; RINAUDO, J. D.; CHAN, N. W. W. **Global Issues in Water Policy: Understanding and Managing Urban Water in Transition**. London: Springer, v. 15, 2015, p. 1-32.

DATASUS. Ministério da Saúde. **TABNET**. Disponível em: <<http://tabnet.datasus.gov.br/cgi/tabcgi.exe?ibge/cnv/pibmunbSC.def>>. Acesso: 08 abr. 2018.

DEWS. Department of Energy and Water Supply. **Planning Guidelines for Water Supply and Sewerage**. Queensland Government, 2010.

DONKOR, A.; S.M. ASCE; MAZZUCHI, T. A.; SOYER, R. ROBERSON, J. A; P.E. Urban water demand forecasting: review of methods and models. **Journal of Water Resources Planning and Management**, v. 140, p. 146-159, 2014.

EHLERS, R. S. **Análise de séries temporais**. 2009. Disponível em: <<http://www.icmc.usp.br/~ehlers/stemp/stemp.pdf>>. Acesso em: 02 jun. 2018.

GALE, R. P. ZHANG, M. J. What is the P-value anyway? **Bone Marrow Transplantation**, v. 51, p. 1439-1440, 2016.

GARDINER, V.; HERRINGTON, P. The basis and practice of water demand forecasting. In: GARDINER, V.; HERRINGTON, P. **Water demand forecasting: proceedings of a workshop**. Norwich: Taylor & Francis e-Library, 2003, p. 7-16.

GARDNER JR, E. S.; McKENZIE, E. Forecasting trends in time series. **Management Science**, v. 31, p. 1237–1246, 1985

GARDNER JR., E. S. Exponential smoothing: the state of the art – part II. **International Journal of Forecasting**, v. 22, n. 4, p. 637-666, 2006.

GHIASSI, M. ZIMBRA, D. K. SAIDANE, H. Urban water demand forecasting with a dynamic artificial neural network model. **Journal of Water Resources Planning and Management**, v. 134, n. 2, p. 138-146, 2008.

GOOGLE. Google Earth Pro. Version 7.3.1.4507. 2018. Nota **Joinville**. Disponível em: <<https://www.google.com.br/intl/pt-BR/earth/>>. Acesso em: 18 ago. 2017.

GROSS, J. LIEGGES, U. nortest. Tests for Normality. R package version 1.0-4, 2015. Disponível em: < <https://cran.r-project.org/web/packages/nortest/index.html>>. Acesso em 25 jul 2019.

GUAJARATI, D. **Econometria básica**. Rio de Janeiro: Elsevier, 2006. 812 p.

HENNING, E. **Arima Modelos Regressivos de Médias Móveis**. 12 jun. 2017. 89 p. Notas de Aula.

HERRERA, M.; TORGO, L.; IZQUIERDO, J.; PÉREZ-GARCÍA, R. Predictive models for forecasting hourly urban water demand. **Journal of Hydrology**, v. 387, p. 141-150, 2010.

HOFFMANN, M.; WORTHINGTON, A.; HIGGS, H. Urban water demand with fixed volumetric charging in a large municipality: the case of Brisbane, Australia. **The Australian Journal of Agricultural and Resource Economics**, v. 50, p. 347-359, 2006.

HOLT, C.C. **Forecasting seasonals and trends by exponentially weighted moving averages**, ONR Memorandum, v. 52, Pittsburgh: Carnegie Institute of Technology, 1957.

HOMWONGS, C.; SASTRI, T. FOSTER III, J. W. Adaptive forecasting of hourly municipal water consumption. **Journal of Water Resources Planning and Management**, v. 120, n. 6, p.888-905, 1994.

HOWE, C. W.; LINAWEAVER, F. P. The impact of price on residential water demand and its relation to systems design. **Water Resources Research**, v. 3, n. 1, p. 13-22, 1967.

HYNDMAN, R. J.; ATHANASOPOULOS, G. **Forecasting: Principles and Practice**. 2 ed. Melbourne: OText. 2013. 292p.

HYNDMAN, R. J. KHANDAKAR, Y. Automatic time series forecasting: the forecast package for R. **Journal of Statistical Software**, v. 27, n. 3, p. 1-22. 2008.

HYNDMAN, R. ATHANASOPOULOS, G. BERGMEIR, C. CACERES, G. CHHAY, L. O'HARA-WILD, M. PETROPOULOS, F. RAZBASH, S. WANG, E. YASMEEN, F. **\_forecast: Forecasting functions for time series and linear models\_**. R package version 8.4, 2018. Disponível em: < <http://pkg.robjhyndman.com/forecast> >. Acesso em 25 jul 2019.

HYNDMAN, R. J.; KOEHLER, A. B. ORD, J. K. SNYDER, R. D. **Forecasting with Exponential Smoothing: The State Space Approach**. New York: Springer Verlag NY, 2008. 380p.

HYNDMAN, R. J.; KOEHLER, A. B. ORD, J. K. SNYDER, R. D. Prediction intervals for exponential smoothing using two new classes of state space models. **Journal of Forecasting**, v. 24, p. 17-37, 2005.

IBGE. Instituto Brasileiro de Geografia e Estatística. **CENSO 2010**. Disponível em: <https://censo2010.ibge.gov.br/>. Acesso: 10 mar. 2018

IBGE. Instituto Brasileiro de Geografia e Estatística. **Cidades**. Disponível em: <<https://cidades.ibge.gov.br/brasil/sc/joinville>>. Acesso em: 18 abr. 2018.

JAIN, A.; VARSHNEY, A.; JOSHI, U. Short-term water demand forecast modelling at IIT Kanpur using artificial neural networks. **Water Resources Management**, v. 15, p 299-321. 2001.

JOWITT, P. W.; XU, C. Demand forecasting for water distribution systems. **Civil Engineering System**, v. 9, p. 105-121, 1992.

KWIATKOWSKI, D. PHILLIPS, P. C. B. SCHIMIDT, P. SHIN, Y. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? **Journal of Econometrics**, n. 54, p. 159-178, 1992.

KOURENTZES, N. **The difference between in-sample fit and forecast performance**. 29 Nov 2016. Disponível em: <  
<https://kourentzes.com/forecasting/2016/11/29/the-difference-between-in-sample-fit-and-forecast-performance/>> . Acesso em: 04 out 2019.

MAIDMENT, D. R.; PARZEN, E. Time patterns of water use in six Texas cities. **Journal of Water Resource Planning and Management**, v. 110, n. 1, p. 90-106, 1984.

MAKRIDAKIS, A. G.; WHEELWRIGHT, S. C; HYNDMAN, R. J. **Forecasting Methods and Applications**, 3 ed. New York: John Willey & Sons, Inc. 1997. 656 p.

MARLOW, D. R.; MOGLIA, M.; COOK, S.; BEALE, D. J. Towards sustainable urban water management: a critical reassessment. **Water Research**, v. 47, p. 7150-7161, 2013.

MERRETT, S. The demand for water: four Interpretation. **Water International**, v. 29, n. 1, p. 27-29, 2004.

MIRANDA, R. G. **Espaço admissível para os parâmetros do modelo de suavização exponencial com dupla sazonalidade aditivo**. Florianópolis, 2009. 94 p. Tese (Doutorado em Engenharia de Produção) – Universidade Federal de Santa Catarina, Florianópolis, 2009.

MOHAMED, M. M.; AL-MUALLA, A. A. Water demand forecasting in Umm Al-Quwain (UAE) using the IWR-MAIN specific forecasting model. **Water Resources Management**, v. 24, p. 4093-4120, 2010.

NASCIMENTO, R. M. do; OENING, A. P.; MARCILIO, D. C.; AOKI, A. R.; ROCHA JR, E. P.; SCHIOCHET, J. M. Algoritmo de Detecção e Correção de *Outliers* para Previsão de Carga. In: Simpósio Brasileiro de Sistemas Elétricos, 4. 2012. Goiânia. Anais...Goiânia: Disponível em: <  
<http://www.swge.inf.br/anais/SBSE2012/PDFS/ARTIGOS/97036.PDF> > Acesso em: 18 set. 2018.

NEWAZ, M. K. Comparing the performance of time series models for forecasting exchange rate. **BRAC University Journal**, v.5, p.55-65, 2008.

ODAN, F. K.; REIS, L. F. R. Hybrid water demand forecasting model associating artificial neural network with Fourier series. **Journal of Water Resources Planning and Management**, v. 138, p. 245-256, 2012.

OLIVEIRA, P. J. A. de. **Previsão de demanda de água em sistemas urbanos de abastecimento através de modelos ARIMA sazonais**. Campo Grande, 2012. 56 p.

Dissertação (Pós-Graduação em Tecnologias Ambientais) – Universidade Federal de Mato Grosso do Sul, Campo Grande, 2012.

ONU. ORGANIZAÇÃO DAS NAÇÕES UNIDAS. Department of Economic and Social affairs, Population Division, **World Urbanization Prospects: the 2014 Revision**. Disponível em: <<https://esa.un.org/unpd/wup/Country-Profiles/>> ,Acesso em: 15 jun. 2017.

PARKINSON, J. N.; GOLDENFUM, J. A.; TUCCI, A; E. M. **Integrate Urban Water Management: Humid Tropics**. Urban Water Series – Unesco-IHP. Boca Raton: Taylor & Francis Group,2010. 69 p.

PEGELS, C. C. Exponential forecasting: some new variations. **Management Science**, v. 15, n. 5, p. 311-315, 1969.

PEREIRA FILHO, N. de A. **Teste de MonteCarlo de Normalidade Univariado**. Lavras, 2013. 56 p. Tese (Doutorado em Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras, 2013.

PERZS, S. G. The rural exodus in the context of economic crisis, globalization and reform in Brazil. **International Migration Review**, v. 34, p. 842-881, 2000.

PMJ. Prefeitura Municipal de Joinville. **Área Urbana Consolidada de Joinville. Vol. II. Diagnóstico Socioambiental**. 2016a. 156 p.

PMJ. Prefeitura Municipal de Joinville. **Área Urbana Consolidada de Joinville. Vol. I. Metodologia de Identificação e Delimitação**. 2016b. 156 p.

PMJ. Prefeitura Municipal de Joinville. Conselho Municipal dos Serviços de água e Esgoto. **Resolução nº 52/2015**. Estabelece normas para a relação comercial da prestação e utilização dos serviços de abastecimento de água e esgotamento sanitário. Disponível em: < <https://www.aguasdejoinville.com.br/wp-content/uploads/2018/05/RESOLUCAO-52.2015-AMAE-Comercial-dos-Servicos-Categorias.pdf>> . Acesso em 01 ago 2018.

QI, C.; CHANG, N. B. System dynamics modeling for municipal water demand estimation in an urban region under uncertain economic impacts. **Journal of Environmental Management**, v. 92, p. 1628-1641, 2011.

R CORE TEAM. **R: a language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria, 2019. Disponível em: <<http://www.R-project.org>>. Acesso em: 10 mar. 2020.

RAZALI, N. M.; WAH, Y. B. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. **Journal of statistical Modeling and Analytics**, v. 2, n. 1, p. 21-33, 2011.

RINAUDO, J. D. Long-term water demand forecasting. In: GRAFTON, Q; DANIELL, K. A; NAUGES, C.; RINAUDO, J. D.; CHAN, N. W. W. **Global Issues in Water**

**Policy: Understanding and Managing Urban Water in Transition.** London: Springer, v. 15, 2015, p. 239-268.

ROMANO, G.; SALVATI, N.; GUERRINI, A. Estimating the determinants of residential water demand in Italy. **Water**, v. 6, p. 2929-2945, 2014.

ROMANO, M.; KAPELAN, Z. Adaptive water demand forecasting for near realtime management smart water distribution systems. **Enviromental Modelling & Software**, v. 60, p. 265-276, 2014.

SANTOS, C.C. **Previsão de Demanda de Água na Região Metropolitana de São Paulo com Redes Neurais Artificiais e Condições Socioambientais e Meteorológicas.** São Paulo, 2011. 121p. Tese (Doutorado em Engenharia Civil) – Universidade de São Paulo, São Paulo, 2011.

SEBRI, M. Forecasting urban water demand: A meta-regression analysis. **Journal of Environmental Management**, v. 183, p. 777-785, 2016.

SEPUD. Secretaria de Planejamento Urbano e Desenvolvimento Sustentável. **Joinville Cidade em Dados.** 2017. 76 p.

SEPUD. Secretaria de Planejamento Urbano e Desenvolvimento Sustentável. **Joinville Cidade em Dados.** 2014. 148 p.

SETTI, A.A.; LIMA, J.E.F.W.; CHAVES, A.G.M.; PEREIRA, I.C. **Introdução ao gerenciamento de recursos hídricos.** Brasília, Agência Nacional de Energia Elétrica; Agência Nacional de Águas, 2001.

SHARVELLE, S.; DOZIER, A.; ARABI, M.; REICHEL, B. A geospatially-enabled web tool for urban water demand forecasting and assessment of alternative urban water management strategies. **Environmental Modelling & Software**, v. 97, p. 213-228, 2017.

SMITH, J. A. A model of daily municipal water use for short-term forecasting. **Water Resources Research**, v. 24, n. 2, p. 201-206, 1988.

SMITH, R. Forecasting industrial demand for water. In: GARDINER, V.; HERRINGTON, P. **Water demand forecasting: proceedings of a workshop.** Norwich: Taylor & Francis e-Library, 2003, p; 57-68.

SNIS. Sistema Nacional de Informações Sobre Saneamento. **Série Histórica.** Disponível em: < <http://app.cidades.gov.br/serieHistorica/>>. Acesso em: 15 mar. 2018.

STIGSDOTTER, E. K.; CORAZON, S. S.; SIDENIUS, U. KRISTIANSEN, J.; GRAHN, P. It is not all bad for the grey city – a crossover study on physiological and psychological restoration in a forest and an urban environment. **Health & Place**, v. 46, p. 145-154, 2017.

STOKER, P.; ROTHFEDER, R. Drivers of urban water use. **Sustainable Cities and Society**, v.12, p. 1-8, 2014.

TAYLOR, J. W. Exponential smoothing with a damped multiplicative trend. **International Journal of Forecasting**, v. 19, n. 4, p. 715– 725, 2003.

TIWARI, M. K.; ADAMOWSKI, J. F. Medium-term urban water demand forecasting with limited data using an ensemble wavelet-bootstrap machine-learning approach. **Journal of Water Resources Planning and Management**, v. 141, n. 2, 2015.

TRAPLETTI, A.; HORNIK, K.; LEBARON, B. **tseries. Time Series Analysis and Computational Finance**. R package version 0.10-47, 2019. Disponível em: < <https://cran.r-project.org/web/packages/tseries/tseries.pdf>>. Acesso em 25 jul 2019

TSUTIYA, M. T. **Abastecimento de Água**. 3 ed. São Paulo: Departamento de Engenharia Hidráulica e Sanitária da Escola Politécnica da Universidade de São Paulo. 2013. 659 p.

TUCCI, C. E. M. Águas Urbanas. **Estudos Avançados**, São Paulo, v. 22, n. 66, p. 97-112, 2008.

WENTZ, E. A.; GOBER, P. Determinants of small-area water consumption for the city of Phoenix, Arizona. **Water Resources Management**, v. 21, p. 1849-1863, 2007.

WILK, M. B. GNANADESIKAN, R. Probability plotting methods for the analysis of data. **Biometrika**, v. 55, n. 1, p. 1-17, 1968.

WU, L.; ZHOU, H. Urban water demand forecasting based on HP filter and fuzzy neural network. **Journal of Hydroinformatics**, v. 2, p. 172-184, 2010.

YALÇINTAŞ, M.; BULU, M.; KÜÇÜKVAR, M.; SAMADI, H. A framework for sustainable urban water management through demand and supply forecasting: the case of Istanbul. **Sustainability**, v. 7, p. 11050-11067, 2015.

ZHAI, Y.; WANG, J.; TENG, Y.; ZUO, R. Water demand forecasting of Beijing using the time series forecasting method. **Journal of Geographical Science**, v. 22(5), p. 919-932, 2012.

ZHOU, S.; MCMAHON, T.; WALTON, A.; LEWIS, J. Forecasting daily urban water demand: A case study of Melbourne. **Journal of Hydrology**, v. 236, p 153164, 2000.

## APÊNDICE A – Modelagem da Categoria Residencial

#Tratamento dos dados

```
x=c(2098064,2131560,2155893,2148904,2139178,2092157,2000241,1986054,2086
004,2002469,2207489,2068330,2213972,2400832,2162789,2182637,2074590,2044
090,2058373,2038225,2048551,2074917,2216076,2115768,2223332,2200935,2124
300,2096102,2137916,2130628,2001614,2108760,2112368,2061940,2116044,2181
326,2259042,2280324,2231722,2231215,2178884,2059325,2076815,2097688,2153
082,2130012,2224171,2209526,2345104,2314853,2294366,2154050,2194208,2191
145,2212914,2215803,2282421,2244625,2243165,2317752,2332073,2333264,2295
320)
```

```
library(forecast)
```

```
library(tseries)
```

```
library(nortest)
```

```
y=ts(x, frequency=12, start=2013)
```

```
y1=window(y,start=c(2013),end=c(2017,12))
```

```
y2=window(y,start=c(2018,1))
```

```
y1
```

```
tsoutliers(y1)
```

```
plot(y1,xlab="Período", ylab= "Consumo (m³)" )
```

```
summary(y1)
```

```
shapiro.test(y1)
```

```
qqnorm(y1)
```

```
qqline(y1)
```

```
decompose(y1)
```

```
plot(decompose(y1))
```

#Suavização Exponencial

```
modelo1=ets(y1)
```

```
summary(modelo1)
```

```
qqnorm(modelo1$residuals)
```



```

qqline(modelo1$residuals)
shapiro.test(modelo1$residuals)
lillie.test(modelo1$residuals)
acf(modelo1$residuals,xlim=c(0.1,1.4))
Box.test(modelo1$residuals, lag=24)
f1<-forecast(modelo1, h=3)
plot(f1,xlab="Período", ylab= "Consumo (m³)" )
lines(fitted(modelo1),col=3)
legend("top", legend=c("Reais", "Ajustados", "Previsão", "95%", "80%"),
      lty=c(1), col=c(1,3,"blue","lightgray","darkgray"), lwd=2, bty="n", pch=c(NA))
accuracy(f1,y2)
summary(f1)

```

# SARIMA

```

modelo2=auto.arima(y1)
summary(modelo2)
qqnorm(modelo2$residuals)
qqline(modelo2$residuals)
shapiro.test(modelo2$residuals)
lillie.test(modelo2$residuals)
acf(modelo2$residuals,xlim=c(0.1,1.4))
Box.test(modelo2$residuals, lag=24)
f2<-forecast(modelo2,h=3)
plot(f2,xlab="Período", ylab= "Consumo (m³)" )
lines(fitted(modelo2),col=3)
legend("bottomri", legend=c("Reais", "Ajustados", "Previsão", "95%", "80%"),
      lty=c(1), col=c(1,3,"blue","lightgray","darkgray"), lwd=2, bty="n", pch=c(NA))
summary(f2)
accuracy(f2,y2)

```

## APÊNDICE B – Modelagem da Categoria Comercial

#Tratamento dos dados

```
x=c(219056,243378,245053,251417,254555,252988,283621,279224,287595,271603
,302003,285442,270884,305956,288859,298051,282281,279928,274467,278797,28
1717,284937,295179,277503,271537,280290,273185,277663,282312,285252,2676
60,277583,273999,256277,265241,277474,255532,269774,268426,269844,272870,
253179,255392,252453,261819,256543,257196,267082,252636,270306,266850,25
7887,258461,264303,263808,266599,269106,266172,264893,273722,255708,2774
63,266217)
```

```
library(forecast)
```

```
library(tseries)
```

```
library(nortest)
```

```
y=ts(x, frequency=12, start=2013)
```

```
y1=window(y,start=c(2013),end=c(2017,12))
```

```
y2=window(y,start=c(2018,1))
```

```
y1
```

```
tsoutliers(y1)
```

```
plot(y1,xlab="Período", ylab= "Consumo (m³)" )
```

```
summary(y1)
```

```
shapiro.test(y1)
```

```
qqnorm(y1)
```

```
qqline(y1)
```

```
decompose(y1)
```

```
plot(decompose(y1))
```

#Suavização Exponencial

```
modelo3=ets(y1)
```

```
summary(modelo3)
```

```
qqnorm(modelo3$residuals)
```

```
qqline(modelo3$residuals)
```

```

shapiro.test(modelo3$residuals)
lillie.test(modelo3$residuals)
acf(modelo3$residuals,xlim=c(0.1,1.4))
Box.test(modelo3$residuals, lag=24)
f3<-forecast(modelo3, h=3)
plot(f3,xlab="Período", ylab= "Consumo (m³)" )
lines(fitted(modelo3),col=3)
legend("top", legend=c("Reais", "Ajustados", "Previsão", "95%", "80%"),
      lty=c(1), col=c(1,3,"blue","lightgray","darkgray"), lwd=2, bty="n", pch=c(NA))
accuracy(f3,y2)
summary(f3)

```

#### # SARIMA

```

modelo4=auto.arima(y1)
summary(modelo4)
qqnorm(modelo4$residuals)
qqline(modelo4$residuals)
shapiro.test(modelo4$residuals)
lillie.test(modelo4$residuals)
acf(modelo4$residuals,xlim=c(0.1,1.4))
Box.test(modelo4$residuals, lag=24)
f4<-forecast(modelo4,h=3)
plot(f4,xlab="Período", ylab= "Consumo (m³)" )
lines(fitted(modelo4),col=3)
legend("bottomri", legend=c("Reais", "Ajustados", "Previsão", "95%", "80%"),
      lty=c(1), col=c(1,3,"blue","lightgray","darkgray"), lwd=2, bty="n", pch=c(NA))
summary(f4)
accuracy(f4,y2)

```

## APÊNDICE C – Modelagem da Categoria Industrial

#Tratamento dos dados

```
x=c(127811,179145,170793,174030,172147,164842,168923,168566,172277,154643
,188159,176044,145447,205387,183544,195770,181594,177343,165697,177402,17
6005,180792,185628,160669,150430,189803,159701,165441,174128,162451,1483
88,149672,146768,141884,146381,148608,116044,157379,150177,154554,159644,
152564,158721,137943,143656,141331,141178,149844,129319,149777,152050,14
4042,143893,147595,146108,147226,151640,151787,148367,156632,133696,1492
07,149292)
```

```
library(forecast)
```

```
library(tseries)
```

```
library(nortest)
```

```
y=ts(x, frequency=12, start=2013)
```

```
y1=window(y,start=c(2013),end=c(2017,12))
```

```
y2=window(y,start=c(2018,1))
```

```
y1
```

```
tsoutliers(y1)
```

```
plot(y1,xlab="Período", ylab= "Consumo (m³)" )
```

```
summary(y1)
```

```
shapiro.test(y1)
```

```
qqnorm(y1)
```

```
qqline(y1)
```

```
decompose(y1)
```

```
plot(decompose(y1))
```

#Suavização Exponencial

```
modelo5=ets(y1)
```

```
summary(modelo5)
```

```
qqnorm(modelo5$residuals)
```

```
qqline(modelo5$residuals)
```

```

shapiro.test(modelo5$residuals)
lillie.test(modelo5$residuals)
acf(modelo5$residuals,xlim=c(0.1,1.4))
Box.test(modelo5$residuals, lag=24)
f5<-forecast(modelo5, h=3)
plot(f5,xlab="Período", ylab= "Consumo (m³)" )
lines(fitted(modelo5),col=3)
legend("top", legend=c("Reais", "Ajustados", "Previsão", "95%", "80%"),
      lty=c(1), col=c(1,3,"blue","lightgray","darkgray"), lwd=2, bty="n", pch=c(NA))
accuracy(f5,y2)
summary(f5)

```

#### # SARIMA

```

modelo6=auto.arima(y1)
summary(modelo6)
qqnorm(modelo6$residuals)
qqline(modelo6$residuals)
shapiro.test(modelo6$residuals)
lillie.test(modelo6$residuals)
acf(modelo6$residuals,xlim=c(0.1,1.4))
Box.test(modelo6$residuals, lag=24)
f6<-forecast(modelo6,h=3)
plot(f6,xlab="Período", ylab= "Consumo (m³)" )
lines(fitted(modelo6),col=3)
legend("bottomri", legend=c("Reais", "Ajustados", "Previsão", "95%", "80%"),
      lty=c(1), col=c(1,3,"blue","lightgray","darkgray"), lwd=2, bty="n", pch=c(NA))
summary(f6)
accuracy(f6,y2)

```

## APÊNDICE D – Modelagem da Categoria Público

#Tratamento dos dados

```
x=c(111496,113516,116353,131213,130389,126711,135054,117673,134080,123086
,134047,131266,101699,102372,125135,142154,130734,124319,130952,135403,14
2428,149150,146475,144818,120789,115272,134157,128924,131256,148244,1396
28,137216,126295,130475,132441,131490,124355,132749,139613,135315,131256,
124225,122294,132026,135606,130277,127140,127112,129783,134692,136537,12
7208,139212,132251,135437,144347,137573,110445,114236,113854,106708,1067
20,113138)
```

```
library(forecast)
```

```
library(tseries)
```

```
library(nortest)
```

```
y=ts(x, frequency=12, start=2013)
```

```
y1=window(y,start=c(2013),end=c(2017,12))
```

```
y2=window(y,start=c(2018,1))
```

```
y1
```

```
tsoutliers(y1)
```

```
plot(y1,xlab="Período", ylab= "Consumo (m³)" )
```

```
summary(y1)
```

```
shapiro.test(y1)
```

```
qqnorm(y1)
```

```
qqline(y1)
```

```
decompose(y1)
```

```
plot(decompose(y1))
```

#Suavização Exponencial

```
modelo7=ets(y1)
```

```
summary(modelo7)
```

```
qqnorm(modelo7$residuals)
```

```
qqline(modelo7$residuals)
```

```

shapiro.test(modelo7$residuals)
lillie.test(modelo7$residuals)
acf(modelo7$residuals,xlim=c(0.1,1.4))
Box.test(modelo7$residuals, lag=24)
f7<-forecast(modelo7, h=3)
plot(f7,xlab="Período", ylab= "Consumo (m³)" )
lines(fitted(modelo7),col=3)
legend("top", legend=c("Reais", "Ajustados", "Previsão", "95%", "80%"),
      lty=c(1), col=c(1,3,"blue","lightgray","darkgray"), lwd=2, bty="n", pch=c(NA))
accuracy(f7,y2)
summary(f7)

```

#### # SARIMA

```

modelo8=auto.arima(y1)
summary(modelo8)
qqnorm(modelo8$residuals)
qqline(modelo8$residuals)
shapiro.test(modelo8$residuals)
lillie.test(modelo8$residuals)
acf(modelo8$residuals,xlim=c(0.1,1.4))
Box.test(modelo8$residuals, lag=24)
f8<-forecast(modelo8,h=3)
plot(f8,xlab="Período", ylab= "Consumo (m³)" )
lines(fitted(modelo8),col=3)
legend("bottomri", legend=c("Reais", "Ajustados", "Previsão", "95%", "80%"),
      lty=c(1), col=c(1,3,"blue","lightgray","darkgray"), lwd=2, bty="n", pch=c(NA))
summary(f8)
accuracy(f8,y2)

```

## APÊNDICE E – Modelagem da Categoria Total

#Tratamento dos dados

```
x=c(2556427,2667599,2688092,2705564,2696269,2636698,2587839,2551517,2679
956,2551801,2831698,2661082,2732002,3014547,2760327,2818612,2669199,2625
680,2629489,2629827,2648701,2689796,2843358,2698758,2766088,2786300,2691
343,2668130,2725612,2726575,2557290,2673231,2659430,2590576,2660107,2738
898,2754973,2840226,2789938,2790928,2742654,2589293,2613222,2620110,2694
163,2658163,2749685,2753564,2856842,2869628,2849803,2683187,2735774,2735
294,2758267,2773975,2840740,2773029,2770661,2861960,2828185,2866654,2823
967)
```

```
library(forecast)
```

```
library(tseries)
```

```
library(nortest)
```

```
y=ts(x, frequency=12, start=2013)
```

```
y1=window(y,start=c(2013),end=c(2017,12))
```

```
y2=window(y,start=c(2018,1))
```

```
y1
```

```
tsoutliers(y1)
```

```
plot(y1,xlab="Período", ylab= "Consumo (m³)" )
```

```
summary(y1)
```

```
shapiro.test(y1)
```

```
qqnorm(y1)
```

```
qqline(y1)
```

```
decompose(y1)
```

```
plot(decompose(y1))
```

#Suavização Exponencial

```
modelo9=ets(y1)
```

```
summary(modelo9)
```

```
qqnorm(modelo9$residuals)
```



```

qqline(modelo9$residuals)
shapiro.test(modelo9$residuals)
lillie.test(modelo9$residuals)
acf(modelo9$residuals,xlim=c(0.1,1.4))
Box.test(modelo9$residuals, lag=24)
f9<-forecast(modelo9, h=3)
plot(f9,xlab="Período", ylab= "Consumo (m³)" )
lines(fitted(modelo9),col=3)
legend("top", legend=c("Reais", "Ajustados", "Previsão", "95%", "80%"),
      lty=c(1), col=c(1,3,"blue","lightgray","darkgray"), lwd=2, bty="n", pch=c(NA))
accuracy(f9,y2)
summary(f9)

```

# SARIMA

```

modelo10=auto.arima(y1)
summary(modelo10)
qqnorm(modelo10$residuals)
qqline(modelo10$residuals)
shapiro.test(modelo10$residuals)
lillie.test(modelo10$residuals)
acf(modelo10$residuals,xlim=c(0.1,1.4))
Box.test(modelo10$residuals, lag=24)
f10<-forecast(modelo10,h=3)
plot(f10,xlab="Período", ylab= "Consumo (m³)" )
lines(fitted(modelo10),col=3)
legend("bottomri", legend=c("Reais", "Ajustados", "Previsão", "95%", "80%"),
      lty=c(1), col=c(1,3,"blue","lightgray","darkgray"), lwd=2, bty="n", pch=c(NA))
summary(f10)
accuracy(f10,y2)

```