**UNIVERSIDADE FEDERAL DE SANTA MARIA**
**CENTRO DE CIÊNCIAS NATURAIS E EXATAS**
**PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA E MODELAGEM QUANTITATIVA**

**Cleiton Tibulo**

# A RELAÇÃO ENTRE A POLUIÇÃO DO AR E AS PRINCIPAIS VARIÁVEIS CLIMATOLÓGICAS NA CIDADE DE CANOAS, RIO GRANDE DO SUL

Santa Maria, RS
2017

**Cleiton Tibulo**

# A RELAÇÃO ENTRE A POLUIÇÃO DO AR E AS PRINCIPAIS VARIÁVEIS CLIMATOLÓGICAS NA CIDADE DE CANOAS, RIO GRANDE DO SUL

Monografia apresentada ao Curso em Estatística e Modelagem Quantitativa, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Especialista em Estatística e Modelagem Quantitativa.**

Orientadora: Profª Dra. Roselaine Ruviaro Zanini

Santa Maria, RS
2017

**Cleiton Tibulo**

# A RELAÇÃO ENTRE A POLUIÇÃO DO AR E AS PRINCIPAIS VARIÁVEIS CLIMATOLÓGICAS NA CIDADE DE CANOAS, RIO GRANDE DO SUL

Monografia apresentada ao Curso em Estatística e Modelagem Quantitativa, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Especialista em Estatística e Modelagem Quantitativa.**

**Aprovado em 26 de maio de 2017:**

Roselaine R. Zanini

**Roselaine Ruviaro Zanini, Dra.**
(Presidente/Orientadora)

**Adriano Mendonça Souza, Dr.**
(UFSM)

**Luciane Flores Jacobi, Dra.**
(UFSM)

Santa Maria, RS
2017

## AGRADECIMENTOS

Agradeço a Deus por me dar saúde, força e disposição para trabalhar e estudar, superando as dificuldades, e por proporcionar a oportunidade de cursar a especialização em estatística.

À minha esposa, Vaneza De Carli Tibulo, agradeço pelo incentivo, auxílio e apoio que sempre me dá para prosseguir com os estudos, além da troca de experiências que me proporciona.

À minha orientadora, professora Roselaine Ruviaro Zanini, pelo conhecimento, competência, incentivo, paciência e amizade transmitida. Pela segunda vez, um especial muito obrigado e espero, em algum momento, poder retribuir tamanha competência.

Aos professores Adriano Mendonça Souza, Luciane Flores Jacobi, Anaelena Bragança de Moraes, pelos diversos conhecimentos repassados, compreensão, ajuda e colaboração dadas ao longo desse período e principalmente por serem exemplos de professores.

Ao meu pai, Ivar Tibulo, e à minha mãe, Valdomira Tibulo, pelos incentivos constantes e por me ensinarem a prosseguir mesmo em situações adversas, a eles devo minha educação.

À minha família, como um todo, pelo carinho e compreensão proferidos.

Agradeço à Universidade Federal de Santa Maria e a todos os professores e funcionários do departamento de Estatística que me auxiliaram nesse período.

Ao Colégio Militar de Santa Maria, que forneceu ferramentas necessárias para que este curso fosse realizado.

A minha colega Steffani Dapper e a FEPAM, por conceder os dados necessários para a realização desta pesquisa.

Por fim, agradeço aos professores que se disponibilizaram a participar da banca, gastando seu precioso tempo a fim de corrigir e tornar este trabalho melhor.

Obrigado a todos!

# RESUMO

## A RELAÇÃO ENTRE A POLUIÇÃO DO AR E AS PRINCIPAIS VARIÁVEIS CLIMATOLÓGICAS NA CIDADE DE CANOAS, RIO GRANDE DO SUL

AUTOR: Cleiton Tibulo
ORIENTADORA: Roselaine Ruviaro Zanini

A poluição do ar, por muitas vezes, não é perceptível, e tem afetado diretamente a saúde das pessoas e é uma das mais graves formas de poluição; principalmente, quando associada a variações climáticas, as consequências podem ser atenuadas ou agravadas. Nesse contexto, a presente pesquisa tem por objetivo utilizar as técnicas de análise multivariada de agrupamento e fatorial para determinar a associação entre as variáveis de poluição do ar e as variáveis climatológicas. Posteriormente, foram ajustados os modelos de série temporal ARMAX, utilizando como variáveis exógenas as variáveis definidas na análise multivariada e os modelos ARIMA, Alisamento Exponencial Simples, Biparamétrico e Holt-Winters, com finalidade de apresentar uma comparação do desempenho das previsões dos modelos, ajustando para as variáveis de poluição do ar. As variáveis climatológicas foram coletadas no Instituto Nacional de Meteorologia (INMET) e os poluentes do ar, na Fundação Estadual de Proteção Ambiental (FEPAM) para o município de Canoas-RS, durante todo o ano de 2014. No agrupamento entre as variáveis formaram-se dois grupos bem definidos: O *cluster* 1 reúne as variáveis mais próximas (O3, Insolação, Evaporação, Temperatura Mínima e Temperatura Máxima), o *cluster* 2, as outras variáveis mais próximas (SO2, CO, PM10). Para o grupo 1 foram ajustados modelos de previsão propostos para variável O3, usando as demais variáveis do grupo como exógenas ao modelo ARMAX. Na comparação dos resultados de previsão foram utilizados os critérios: MAPE, EQM, MAD e SSE. Os resultados mostraram que o modelo SARIMAX (1,1,1[1,1,1])(0,1,1), produziu melhores resultados de previsão em relação aos seus concorrentes SARIMA e da classe de alisamento exponencial. No grupo 2 foram ajustados modelos de previsão para variável PM10, usando as demais variáveis do grupo como exógenas ao modelo ARMAX. Na comparação dos resultados de previsão, o modelo ARX(1,[1,1]) produziu melhores resultados de previsão em relação aos seus concorrentes ARIMA e da classe de alisamento exponencial. Conclui-se que o uso conjunto de técnicas de análise multivariada e séries temporais pode proporcionar melhores resultados para previsão. Assim, destaca-se a ampla possibilidade de utilização das técnicas conjuntas de análise multivariada e séries temporais quando se deseja fazer previsões e descrever um processo temporal com a utilização de variáveis explicativas, podendo ser utilizadas como ferramentas eficientes de apoio nas tomadas de decisão em inúmeras áreas.



**Palavras-chave:** Poluição do ar. Variáveis climatológicas. Modelos de séries temporais. Técnicas de análise multivariada.

## ABSTRACT

### THE RELATION BETWEEN AIR POLLUTION AND THE MAIN CLIMATE VARIABLES IN THE CITY OF CANOAS, RIO GRANDE DO SUL

AUTHOR: CLEITON TIBULO
ADVISOR: ROSELAINE RUVIARO ZANINI

Oftentimes, air pollution is not noticeable, and there are few cities with proper instruments to measure it. Air pollution has directly affected people's health and is one of the most severe forms of pollution; especially when it is associated with climate variations, consequences may be softened or worsened. In this context, the present work aims to use multivariate analysis techniques of grouping and factorial to determine the association between air pollution variables and climate variables. Thereafter, the ARMAX model will be employed, using the defined variables in the multivariate analysis as exogenous variables, and the models of ARIMA temporal series, Simple Exponential Smoothing, Biparametric and Holt-Winters, with the purpose to present a comparison of the performance of the models, adjusting air pollution variables. Data were collected along with the National Institute of Meteorology (INMET) (climate variables) and the State Foundation of Environmental Protection (FEPAM) (air pollutants) referring to the city of Canoas-RS, during the year of 2014. In the grouping among the variables, two well defined groups were formed: *Cluster* 1 puts together the closest variables (O3, Insolation, Evaporation, Minimum and Maximum Temperatures), cluster 2 puts together the closest variables (SO2, CO, PM10). For group 1, prediction models for O3 variable were adjusted, using the other group variables as exogenous to the ARMAX model. In the comparison of the prediction results of the best model, the following criteria were used: MAPE, EQM, MAD and SSE. Results have shown that the SARIMAX model (1,1,1[1,1,1]) (0,1,1), with the exogenous variables inclusion, produced better prediction results in comparison to its opponent ARIMA and the class of exponential smoothing. In group 2, prediction models for the PM10 variable were adjusted, using the other group variables as exogenous to the ARMAX model. In the comparison of the prediction, the ARX model (1,[1,1]), with the inclusion of exogenous variables, produced better prediction results in relation to its opponents ARIMA and the smoothing exponential class. It was possible to conclude that the combined used of multivariate analysis techniques and temporal series provides better prediction results. The vast possibility of usage of combined techniques of multivariate analysis and temporal series stands out, when the intention is to make predictions and describe the temporal process with the usage of explanatory variables, being used as an efficient support tool in the decision making in countless areas.



**Keywords:** Air pollution. Climate variables. Temporal series models. Multivariate analysis models.

## LISTA DE FIGURAS


Figura 1 – Aplicação do método de análise de agrupamentos ..... 19
Figura 2 – Dendograma das variáveis em estudo (2A) e *Scree plot* (2B) ..... 35
Figura 3 – Plano Fatorial – Fator 1 *vs* Fator 2 ..... 36
Figura 4 – Dendograma (4A) e Plano Fatorial – Fator 1 *vs* Fator 2 (4B) ..... 37
Figura 5 – Plano Fatorial – Fator 1 *vs* Fator 2 ..... 38
Figura 6 – Gráfico da série original O3 em Canoas - RS ..... 39
Figura 7 – Gráfico da Série O3 com uma diferença ..... 40
Figura 8 – Gráfico da série O3 em pontos ..... 40
Figura 9 – Função de Autocorrelação (FAC) e Autocorrelação Parcial (FACP) ..... 41
Figura 10 – Função de autocorrelação (FAC) e autocorrelação parcial (FACP) dos resíduos ..... 43
Figura 11 – Modelo SARIMA (1,1,1)(0,1,1) ajustado à série observada ..... 44
Figura 12 – Função de autocorrelação (FAC) e autocorrelação parcial (FACP) dos resíduos do modelo SARIMAX ..... 45
Figura 13 – Modelo SARIMAX (1,1,1[1,1,1])(0,1,1) ajustado à série observada ..... 46
Figura 14 – Modelo de alisamento exponencial simples ..... 48
Figura 15 – Modelo de alisamento biparamétrico ajustado a variável O3 ..... 49
Figura 16 – Holt-Winters aditivo ajustado à variável O3 ..... 50
Figura 17 – Gráfico da série original PM10 em Canoas - RS ..... 51
Figura 18 – Gráfico da série PM10 em pontos ..... 52
Figura 19 – Função de Autocorrelação (FAC) e Autocorrelação Parcial (FACP) ..... 53
Figura 20 – Função de autocorrelação (FAC) e autocorrelação parcial (FACP) dos resíduos ..... 54
Figura 21 – Modelo AR(1) ajustado à série observada (variável PM10) ..... 55
Figura 22 – Função de autocorrelação (FAC) e autocorrelação parcial (FACP) dos resíduos ARX (1,[1,1]) ..... 56
Figura 23 – Modelo ARX (1,[1,1]) ajustado à série observada (variável PM10) ..... 57
Figura 24 – Modelo de alisamento Exponencial Simples (variável PM10) ..... 58
Figura 25 – Modelo de alisamento biparamétrico (variável PM10) ..... 60
Figura 26 – Holt-Winters multiplicativo (variável PM10) ..... 61

## LISTA DE TABELAS


Tabela 1 – Fontes e características de alguns poluentes do ar ..... 15
Tabela 2 – Critérios de avaliação da qualidade de previsão ..... 30
Tabela 3 – Análise descritivas dos poluentes do ar ..... 34
Tabela 4 – Correlação entre os fatores e as variáveis ..... 36
Tabela 5 – Modelos concorrentes SARIMA e os respectivos critérios para a série O3 ..... 42
Tabela 6 – Valores previstos (h=20), segundo o modelo SARIMA (1,1,1)(0,1,1) ..... 43
Tabela 7 – Modelos concorrentes SARIMAX e os respectivos critérios (variável O3) ..... 44
Tabela 8 – Valores previstos (h=20) modelo SARIMAX (1,1,1[1,1,1])(0,1,1) (variável O3) ..... 46
Tabela 9 – Valores previstos (h=20) do modelo de Alisamento exponencial simples (variável O3) ..... 47
Tabela 10 – Valores previstos (h=20) do modelo de Alisamento exponencial biparamétrico (variável O3) ..... 48
Tabela 11 – Valores Previstos (h=20) modelos Holt-Winters (variável O3) ..... 50
Tabela 12 – Modelos concorrentes ARIMA e os respectivos critérios (variável PM10) ..... 53
Tabela 13 – Valores previstos (h=20) segundo o modelo AR(1) (variável PM10) ..... 54
Tabela 14 – Modelos concorrentes ARIMAX e os respectivos critérios ..... 55
Tabela 15 – Valores previstos (h=20) modelo ARX (1,[1,1]) ..... 57
Tabela 16 – Valores previstos (h=20) do modelo de alisamento exponencial simples (variável PM10) ..... 58
Tabela 17 – Valores previstos (h=20) do modelo de alisamento exponencial biparamétrico (variável PM10) ..... 59
Tabela 18 – Valores Previstos (h=20) modelos Holt-Winters ..... 61
Tabela 19 – Comparação dos modelos utilizando os critérios de seleção ..... 62

## LISTA DE ABREVIATURAS E SIGLAS

AR – Modelo Autorregressivo

ARMA – Modelo Autorregressivo e de Médias Móveis

ARIMA – Modelo Autorregressivo Integrado de Médias Móveis

ARMAX – Modelo Autorregressivo e de médias móveis com entrada de variáveis exógenas

FEPAM – Fundação Estadual de Proteção Ambiental Henrique Luiz Roessler

INMET – Instituto Nacional de Meteorologia

MA – Modelo de Médias Móveis

SARIMA – Modelo Autorregressivo Integrado de Médias Móveis Sazonal

CO – Monóxido de Carbono

NOx – Óxidos de Nitrogênio

O3 – Ozônio

PM10 – Material Particulado Inalável (PM10)

O2 – Dióxido de Enxofre

$Z_t$ – Série temporal

$\Phi$ – Parâmetro autorregressivo

$\Phi$ – Parâmetro autorregressivo sazonal

$\Theta$ – Parâmetro de médias móveis

$\Theta$ – Parâmetro de médias móveis sazonal

p – Número de parâmetros autorregressivos

d – Número de diferenciações

q – Número de parâmetros de médias móveis

P – Número de parâmetros autorregressivos sazonais

D – Número de diferenciações sazonais

Q – Número de parâmetros de médias móveis sazonais

$\mu$ – Média de uma variável aleatória

AEHW – Alisamento Exponencial de Holt-Winters

$\varepsilon$ – Erro aleatório

**SUMÁRIO**


**1 INTRODUÇÃO** ..... 11
1.1 PROBLEMÁTICA ..... 12
1.2 OBJETIVOS ..... 12
**1.2.1 Objetivo geral** ..... 12
**1.2.2 Objetivos específicos** ..... 12
1.3 JUSTIFICATIVA ..... 12
1.4 ORGANIZAÇÃO DO TRABALHO ..... 13

**2 REVISÃO DE LITERATURA** ..... 14
2.1 POLUENTES DO AR ..... 14
2.2 ANÁLISE MULTIVARIADA ..... 16
**2.2.1 Análise de agrupamentos** ..... 17
**2.2.2 Análise fatorial** ..... 19
2.3 SÉRIES TEMPORAIS ..... 20
**2.3.1 Modelos Box-Jenkins** ..... 21
*2.3.1.1 Modelo Autorregressivo* ..... 22
*2.3.1.2 Modelo de Médias Móveis* ..... 22
*2.3.1.3 Modelo Autorregressivo e de Médias Móveis* ..... 22
*2.3.1.4 Modelo Autorregressivo, Integrado e de Médias Móveis* ..... 23
*2.3.1.5 Modelo SARIMA* ..... 24
*2.3.1.6 Modelo ARMAX* ..... 24
**2.3.2 Alisamento Exponencial** ..... 25
*2.3.2.1 Alisamento Exponencial Simples* ..... 25
*2.3.2.2 Alisamento Exponencial Biparamétrico* ..... 26
*2.3.2.3 Alisamento Exponencial de Holt-Winters (AEHW)* ..... 27
2.4 CRITÉRIOS DE SELEÇÃO DOS MELHORES MODELOS ..... 29
**2.4.1 Critérios de seleção dos Modelos de Box Jenkins** ..... 29
**2.4.2 Critérios de comparação de modelos** ..... 30

**3 MATERIAIS E MÉTODOS** ..... 31

**4 ANÁLISE DOS RESULTADOS** ..... 34
4.1 ESTATÍSTICA DESCRITIVA ..... 34
4.2 ANÁLISE MULTIVARIADA DE AGRUPAMENTOS ..... 34
4.3 AJUSTE DE MODELOS DE PREVISÃO PARA A VARIÁVEL O3 ..... 38
**4.3.1 Modelos da Classe Box-Jenkins** ..... 39
*4.3.1.1 Ajuste do modelo ARIMA para a variável O3* ..... 39
*4.3.1.2 Ajuste do modelo ARIMAX para a variável O3* ..... 44
**4.3.2 Ajuste de modelos da classe de alisamento exponencial para variável O3** ..... 47
*4.3.2.1 Alisamento exponencial simples* ..... 47
*4.3.2.2 Alisamento exponencial biparamétrico* ..... 48
*4.3.2.3 Alisamento exponencial de Holt-Winters* ..... 49
4.4 AJUSTE DE MODELOS DE PREVISÃO PARA A VARIÁVEL PM10 ..... 50
**4.4.1 Modelos da classe Box-Jenkins** ..... 51
*4.4.1.1 Ajuste do modelo ARIMA para a variável PM10* ..... 51
*4.4.1.2 Ajuste do modelo ARIMAX* ..... 55
**4.4.2 Ajuste de modelos da classe de alisamento exponencial para variável PM10** ..... 57
*4.4.2.1 Alisamento exponencial simples* ..... 58
*4.4.2.2 Alisamento exponencial biparamétrico* ..... 59

*4.4.2.3 Alisamento exponencial de Holt-Winters* ..... 60
4.5 COMPARAÇÃO DAS PREVISÕES DOS MODELOS AJUSTADOS ..... 61
**5 CONCLUSÃO** ..... 63
**REFERÊNCIAS** ..... 65

# 1 INTRODUÇÃO

Os índices de poluição vêm crescendo de forma desordenada, têm causado inúmeras preocupações e possibilitado estudos. Nas cidades de grande e médio porte, tem crescido a preocupação com poluentes que se associam ao ar e, dessa forma, desloca-se facilmente, atingindo inúmeras pessoas em períodos curtos de tempo, de acordo com Gioda e Gioda (2006), Barcellos et al. (2009) e Leite et al. (2011).

Em parte, essa grande concentração de poluição no ar se deve ao acúmulo de veículos, pessoas, indústrias e demais setores que, por estarem concentrados em uma determinada região, acabam contribuindo para os altos índices de poluição. De acordo com Monte, Albuquerque e Reisen (2015), com a industrialização, aliada ao crescimento populacional, especialmente o urbano, vêm aumentando as preocupações relacionadas ao meio ambiente.

O grande temor que vem afligindo os habitantes de regiões com altos índices de poluição é os males que a exposição constante a esses poluentes podem causar à saúde, atingindo em cheio a qualidade de vida das pessoas.

> Somente com modelos matemáticos é possível fazer previsões ou simular campos de concentração em conexão com políticas de limitação da liberação de poluentes em concordância com planos de melhoria da qualidade de vida da população. A introdução da modelagem matemática produz um salto de qualidade na gestão da poluição atmosférica em respeito àquela possível somente através de medidas (MOREIRA, TIRABASSI e MORAES, 2008, p. 1).

A poluição do ar pode se deslocar por grandes distâncias, dessa forma, ela não é restrita a um lugar ou a uma grande cidade. Plantações e animais também são atingidos por essa poluição, principalmente em regiões que se localizam nos arredores de cidades com número considerável de habitantes. Sendo assim, os alimentos também são afetados e contaminados, tornando-se um ciclo perigoso que pode ocasionar inúmeras doenças. Esse fenômeno está associado a variáveis climáticas como vento, umidade do ar, chuva, temperaturas máximas e mínimas e demais variáveis. De acordo com Moreira Tirambassi e Moraes (2008), as condições meteorológicas desempenham um papel importantíssimo no transporte e na difusão dos poluentes. Ainda, segundo os autores, modelos matemáticos constituem uma ferramenta importante para auxiliar nas medidas de concentrações de poluentes e para conhecimento de sua evolução.

Dessa forma, é de suma importância a analise entre as variáveis climatológicas e os principais poluentes do ar, verificando as relações entre eles, por meio do emprego de técnicas de análise multivariada. Os modelos de séries temporais podem auxiliar no entendimento do

comportamento da série de poluentes do ar e realizar previsões futuras a fim de colaborar no planejamento de áreas de fundamental importância como: a saúde, a meteorologia e o desenvolvimento urbano.

## 1.1 PROBLEMÁTICA

A problemática que orienta a presente pesquisa está voltada para identificar as principais características e influências entre os poluentes do ar e as variáveis climatológicas e para verificar se modelos de análise multivariada podem auxiliar na escolha de variáveis explicativas que auxiliem no ajuste de modelos de série temporal, proporcionando melhores resultados de previsão.

## 1.2 OBJETIVOS

### 1.2.1 Objetivo geral

Determinar as variáveis exógenas a serem usadas nos modelos de séries temporais, de forma a proporcionar melhores resultados de previsão.

### 1.2.2 Objetivos específicos

- Aplicar técnicas de análise multivariada aos dados de poluentes do ar e variáveis climatológicas;
- Verificar a relação entre os principais poluentes do ar e as variáveis climatológicas por meio de modelos estatísticos;
- Ajustar o modelo de série temporal ARIMAX, ARIMA e da classe de Alisamento Exponencial à série de dados climatológicos e de poluentes;
- Realizar previsões futuras;
- Comparar os modelos propostos.

## 1.3 JUSTIFICATIVA

Nada é mais importante para qualquer ser humano do que a sua própria saúde. Entretanto, esse bem que nos é dado vem sofrendo com os diversos maus tratos que o

desenvolvimento tecnológico e urbano vem nos causando. Dentre esses males, está a poluição do ar que constantemente vem se agravando em médias e grandes cidades, sendo de fundamental importância compreendê-la e estudar a sua relação com variáveis climatológicas.

> A degradação do ar decorre de um conjunto de fatores como: as taxas de emissões dos poluentes, a localização, a concentração das fontes emissoras, as características físico-químicas dos poluentes, bem como a dispersão destes na atmosfera e a reação química entre eles, ações que são fortemente influenciadas por condições climatológicas (INSTITUTO DE ENERGIA E MEIO AMBIENTE, 2014, p. 3).

Dessa forma, este estudo pretende aplicar análise multivariada e previsão de séries temporais aos dados em estudo. Cabe ressaltar que essas técnicas estatísticas vêm ganhando espaço na sociedade atual e estão sendo aplicadas constantemente para resolução de problemas nas mais diferentes áreas e, se aplicadas conjuntamente, podem auxiliar a determinar planejamentos futuros, principalmente na área de poluentes do ar, que afetam diretamente a todas as pessoas, independente de sexo, idade ou condição financeira.

## 1.4 ORGANIZAÇÃO DO TRABALHO

A presente pesquisa está dividida em cinco capítulos, os quais buscam o entendimento da proposta central, por meio da apresentação da pesquisa desenvolvida:

No capítulo 1, aborda-se a introdução da pesquisa, contendo a importância do trabalho, justificativa e objetivos.

No capítulo 2, reúnem-se informações sobre os poluentes do ar e revisão bibliográfica sobre Análise Multivariada de *Cluster* e fatorial, modelos de séries temporais ARIMA, ARMAX e Alisamento Exponencial Simples, Biparamétrico e de Holt-Winters, com as suas características para emprego no problema.

No capítulo 3, é descrita a metodologia empregada para aplicação dos modelos Multivariados e análise de *Cluster* ARIMA, ARMAX, e Alisamento Exponencial Simples, Biparamétrico e de Holt-Winters.

O capítulo 4 apresenta as avaliações estatísticas com emprego das técnicas de análise multivariada, séries temporais e os resultados obtidos.

Por fim, no quinto capítulo, é apresentada a conclusão geral do trabalho.

# 2 REVISÃO DE LITERATURA

Nos últimos anos, vários estudos vêm desenvolvendo e aperfeiçoando modelos estatísticos cada vez mais sofisticados, com o intuito de melhorar as pesquisas nas mais diversas áreas do conhecimento e facilitar o planejamento e a vida do homem. Dentre eles destacam-se os modelos de Análise Multivariada e Séries Temporais. A aplicação conjunta das técnicas pode trazer ganhos no ajuste e na previsão dos modelos multivariados de séries temporais.

## 2.1 POLUENTES DO AR

Os elevados níveis de poluição do ar têm sido uma preocupação constante, principalmente em cidades com grande circulação de automóveis, caminhões/ônibus, indústrias e outras fontes, associadas às constantes alterações entre níveis máximos e mínimos na temperatura e umidade relativa do ar, que potencializam a concentração de poluentes causando danos à saúde. Os efeitos podem ser, conforme Gioda e Gioda (2006), desde mudanças bioquímicas e fisiológicas a episódios de sinusite, hipertensão, ardência nos olhos, estresse, escamação da pele, perda dos sentidos, dificuldade de respirar, tosse e agravamento de doenças respiratórias e cardíacas.

As alterações de temperatura, umidade e o regime de chuvas podem aumentar os efeitos das doenças respiratórias, assim como alterar as condições de exposição aos poluentes atmosféricos. O tempo de resistência de poluentes na atmosfera aumenta e podem ser transportados a longas distâncias em condições favoráveis de altas temperaturas e baixa umidade, afetando a saúde de populações distantes das fontes geradoras, segundo Barcellos et al. (2009).

Em grandes cidades, o estresse bioclimático causa constante desconforto térmico, chuvas, enchentes, aumento dos índices de concentração de poluentes que, associados, geram danos materiais e sociais (ASSIS, 2011).

Leite et al. (2011) ressaltam que, nos grandes centros urbanos, são, cada vez mais frequentes, os dias em que o ar atinge níveis críticos de poluição, o que pode ser causado pela ausência de ventos ou pelas inversões térmicas, que dependem de fatores da região, como o relevo do entorno do ponto de emissão, as condições meteorológicas e a natureza dos poluentes. Isso significa que esses poluentes podem viajar milhares de quilômetros pela atmosfera, atingindo, assim, comunidades distantes do ponto de emissão.

> A poluição atmosférica pode ser decorrente de diversas atividades econômicas, como também de fenômenos naturais. Independentemente do fator gerador, ela exerce influência no meio ambiente e na qualidade de vida dos seres humanos (DAPPER, 2016, p. 12).

Para Gonçalves et al. (2010), as variáveis meteorológicas chuva, nevoeiro, neve e granizo são os principais agentes dos processos de remoção da poluição da atmosfera, sendo a ocorrência de chuvas o mais importante para o ecossistema.

Para melhor visualização das caraterísticas, principais fontes dos poluentes, seus efeitos para a saúde e meio ambiente, os dados coletados da FEPAM serão sintetizados na Tabela 1.

Pode-se observar, na Tabela 1, que, dentre as principais fontes geradoras de poluição do ar, estão as emissões veiculares e as atividades industriais. Os efeitos à saúde humana estão diretamente relacionados ao sistema nervoso e respiratório, indiretamente pela contaminação do solo e dos alimentos.

Tabela 1 – Fontes e características de alguns poluentes do ar

(continua)

| **Poluente** | **Características** | **Principais Fontes** | **Efeitos sobre o meio ambiente** | **Efeitos sobre a Saúde** |
|---|---|---|---|---|
| Partículas Inaláveis (PM10) | Partículas de material sólido ou líquido que ficam suspensas no ar, na forma de poeira, neblina, aerossol, fumaça, fuligem, etc. Tamanho < 10 micra | Processos de combustão (indústrias e veículos automotores), aerossol secundário (formado na atmosfera), Pólen, aerossol marinho e solo. | Danos à vegetação, redução da visibilidade e contaminação do solo. | Aumento de atendimentos hospitalares e mortes prematuras. Insuficiências respiratórias pela deposição deste poluente nos pulmões. |
| Dióxido de Enxofre (SO2) | Gás incolor, com forte odor, altamente solúvel. Na presença de vapor d'água pode ser transformado a SO3 passando rapidamente a H2SO4, sendo um dos principais constituintes da chuva ácida. É um importante precursor dos sulfatos, um dos principais componentes das partículas inaláveis. No verão, através dos processos fotoquímicos, as reações do SO2 são mais rápidas. | Combustão de combustíveis fósseis (carvão), queima de óleo combustível, refinaria de petróleo, veículos a diesel, vulcões, emissões de reações biológicas. | Pode levar à formação de chuva ácida, causar corrosão aos materiais e danos à vegetação. | Desconforto na respiração, doenças respiratórias, agravamento de doenças respiratórias e cardiovasculares já existentes. Pessoas com asma, doenças crônicas de coração e pulmão são mais sensíveis ao SO2. Irritação ocular. |

(conclusão)

| | | | | |
|---|---|---|---|---|
| Monóxido de Carbono (CO) | Gás incolor, inodoro e insípido. | Combustão incompleta em geral, principalmente em veículos automotores, queimadas e reações fotoquímicas. | - | Causa efeito danoso no sistema nervoso central, com perda de consciência e visão. Exposições mais curtas podem também provocar dores de cabeça e tonturas. |
| Ozônio (O3) | Gás incolor, inodoro nas concentrações ambientais e o principal componente da névoa fotoquímica mais conhecido como *smog*. Composto muito ativo quimicamente. | Não é emitido diretamente à atmosfera, sendo produzido fotoquimicamente pela radiação solar sobre os NOx e compostos orgânicos voláteis (VOCs). | Danos às colheitas, à vegetação natural, plantações agrícolas; plantas ornamentais. Pode danificar materiais devido ao seu alto poder oxidante. | Irritação nos olhos e vias respiratórias, diminuição da capacidade pulmonar. Exposição a altas concentrações pode resultar em sensações de aperto no peito, tosse e chiado na respiração. O O3 tem sido associado ao aumento de admissões hospitalares. |

Fonte: FEPAM (http://www.fepam.rs.gov.br/qualidade/poluentes.asp).

Cabe ressaltar que, para esta pesquisa, foram obtidos dados de monitoramento do município de Canos-RS, da rede automática de monitoramento da FEPAM. Os dados coletados pela rede e utilizados no decorrer do trabalho são: PM10, SO2, O3, CO.

De acordo com os autores citados neste item da pesquisa, as variáveis meteorológicas estão associadas de alguma forma aos poluentes do ar. Essa associação pode ser melhor interpretada pelos modelos de análise multivariados, que serão resumidamente expostos no próximo item deste trabalho.

## 2.2 ANÁLISE MULTIVARIADA

A Análise Multivariada pode auxiliar a determinar as variáveis que mais influenciam a poluição do ar e sua associação com as variáveis climatológicas. Para Gerhardt et al. (2001), a análise multivariada é um conjunto de técnicas estatísticas que trata simultaneamente muitas variáveis.

A técnica estatística de análise multivariada mostra-se, cada vez mais, como uma poderosa ferramenta para análise e avaliação de dados em várias áreas do conhecimento. Por ser multidisciplinar, está inserida em várias áreas do conhecimento, por isso, faz-se necessária a sua aplicação, o seu entendimento e a sua interpretação como ferramenta de pesquisa, de acordo com Vicini (2005). Segundo Kubrusly e Saboia (2006), a utilização de técnicas de

análise multivariada tem se mostrado eficiente, pois permite uma visão mais global dos fenômenos estudados.

"Análise Multivariada corresponde a um grande número de métodos e técnicas que utilizam, simultaneamente, todas as variáveis na interpretação teórica de um conjunto de dados obtidos" (NETO, 2004, p. 3). De acordo com Vicini (2005), o desafio da análise multivariada consiste em interpretar um conjunto de dados e ser capaz de traduzir as informações que estão contidas nele que, até então, não eram percebidos por estarem em um espaço dimensional maior que três.

Com auxílio da análise multivariada, pretende-se determinar a associação entre as variáveis climatológicas e as variáveis de poluição do ar, bem como determinar as principais variáveis que podem ser ajustadas em modelos de séries temporais.

### 2.2.1 Análise de agrupamentos

A técnica de análise de agrupamentos permite verificar a relação de interdependência entre as variáveis, segundo Souza e Silva (2009). Ainda segundo os autores um *cluster* visa agrupar variáveis com características comuns de maior homogeneidade nos grupos e grande heterogeneidade entre os grupos, sem perder informações de todo o conjunto de dados em estudo.

Para Vicini (2005) e Saad (2009), a análise de agrupamentos estuda todo um conjunto de relações interdependentes, sendo que o objetivo é encontrar e separar objetos em grupos similares. A técnica não faz distinção entre variáveis dependentes e independentes, variáveis do tipo causa e efeito.

A análise de agrupamentos (*cluster analysis*) não pressupõe um número nem uma estrutura de conglomerados. Toma como base similaridades (ou proximidades) entre itens (objetos, indivíduos ou variáveis), que podem ser, de alguma forma, medidas. Como resultado, obtém-se certo número de conglomerados no qual se agrupam, devido a sua proximidade, de acordo com Rinaldi, Morabito e Tachibana (2009).

> Quando o interesse é verificar como as amostras se relacionam, ou seja, o quanto estas são semelhantes, segundo as variáveis utilizadas no trabalho, destacam-se dois métodos, que podem ser utilizados: a análise de agrupamento hierárquico e a análise fatorial com análise de componentes principais (VICINI, 2005, p. 11).

Para contornar a influência de diferentes escalas, faz-se necessário padronizar as variáveis, segundo Marchesan, Souza e Menezes (2011) e Saad (2009), ou seja, subtraímos a média do grupo de cada valor e dividimos pelo desvio padrão:

$$Z_{ij} = \frac{X_{ij} - \bar{X}_j}{S_j}$$

Isso garante que as variáveis sejam consideradas com importâncias equivalentes.

A técnica segue as seguintes etapas discutidas em Souza, Schimidt e Turcato (2011), Rinaldi, Morabito e Tachibana (2009), Vicini (2005), Souza e Silva (2009), Marchesan, Souza e Menezes (2011) e Saad (2009):

1) Definir a medida de parecença, as mais usuais são: distância euclidiana, City-Block, Mahalanobis e coeficiente de correlação. As medidas de parecença se dividem em dois grupos (de similaridade, quanto menor as distâncias, mais próximos estão os objetos, exemplo correlação cuja distância é calculada pela equação $d_{ij} = 1 - r_{ij}$) ou (dissimilaridade, quanto maior o valor, mais distantes estão os objetos, exemplo distância euclidiana, City-block).

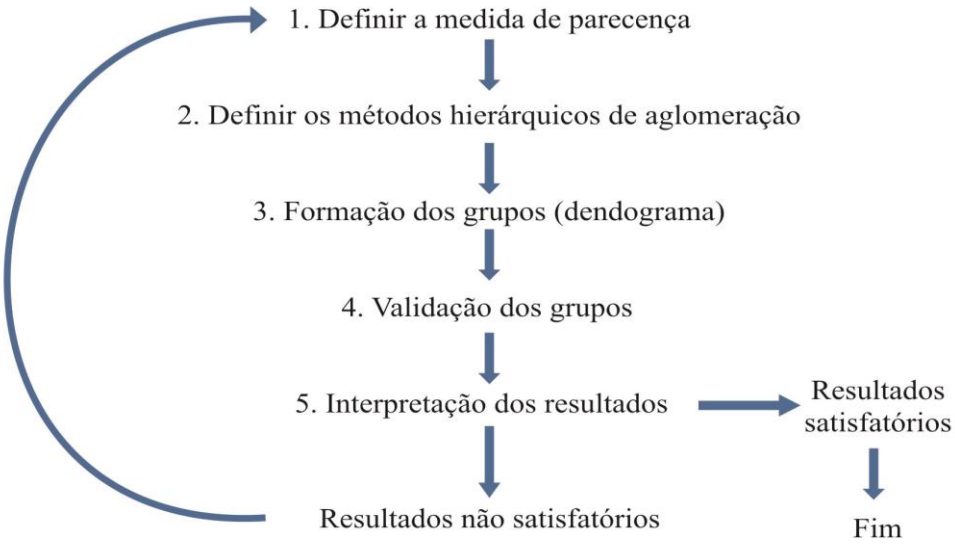
2) Definir os métodos hierárquicos de aglomeração: os agrupamentos são formados a partir da matriz de parecença; os métodos mais usuais são: Ligação Simples (*Single Linkage*), Método de Ligação Completa (*Complete Linkage*), Método da Média das Distâncias (*Average Linkage*), Método do Centroide (*Centroid Method*) e Método de Ward.

3) Formação dos grupos: o número de grupos pode ser definido por meio de algum conhecimento que se tenha sobre os dados, conveniência de análise ou ainda com base nos resultados da análise;

4) Validação dos grupos: ao final do processo de formação de grupos, é importante caracterizá-los. O uso de estatísticas descritivas é recomendado para essa fase da análise;

5) Interpretação dos resultados.

Figura 1 – Aplicação do método de análise de agrupamentos




Fonte: Autor.

De acordo com Souza, Schimidt e Turcato (2011), o dendograma fornece uma inspeção visual em busca de observações atípicas e oferece uma interpretação do agrupamento, garantindo que eles sejam verdadeiramente distintos, cabendo ao pesquisador eliminar variáveis que não se agrupam ou que não diferem significativamente ao longo dos grupamentos.

"Quando o objetivo é agrupar variáveis, as medidas de similaridade geralmente empregadas são medidas correlacionais. Destas, a principal medida é a correlação, muito utilizada por ser de fácil interpretação" (RINALDI, MORABITO e TACHIBANA, 2009, p. 4). Quanto maior a correlação, menor a distância. Utilizando-se como distâncias $d_{ij} = 1 - r_{ij}$, obtém-se a menor distância.

A análise de agrupamentos é um método simples, baseado nos cálculos de distância, que não requerem conhecimento estatístico para a sua aplicação. Devido a isso, devem-se utilizar outros métodos e comparar os resultados para que a análise dos dados seja realizada de forma mais adequada, de acordo com Vicini (2005).

### 2.2.2 Análise fatorial

A análise fatorial procura definir o relacionamento entre as variáveis, de modo simples, e usando fatores que objetivam trabalhar com um número menor de variáveis

originais. Em síntese, o método é utilizado para redução e sumarização dos dados. Portanto, o seu propósito principal é interpretar a estrutura de um conjunto de dados multivariados, a partir da matriz de variância covariância ou da matriz de correlação, segundo Lírio e Souza (2008).

> Os fatores, advindos da transformação linear, possuem a prerrogativa de que as relações existentes dentro de um conjunto de m variáveis seja o reflexo das correlações de cada uma dessas variáveis com p fatores, não correlacionados entre si, sendo o número de fatores menor que o número de variáveis, isto é, $p < m$ (LÍRIO e SOUZA, 2008, p. 23).

Para Rinaldi, Morabito e Tachibana (2009), a análise fatorial busca a redução do número de agrupamentos com pouca perda de representatividade das variáveis. Os autovalores de cada eixo representam o quanto de variância original é explicado pelos seus fatores.

A análise fatorial descobre quais variáveis podem ser reunidas em um fator, permitindo reduzir o número de variáveis iniciais sem perda significativa de informação. A análise fatorial é um conjunto de técnicas que visam tornar os dados observados mais claros, os fatores explicam a variância das variáveis observadas que se revelam pelas correlações entre as variáveis observadas. Em resumo, a AF é um método para determinar o número de fatores existente em um conjunto de dados e serve para determinar quais testes, ou variáveis, pertencem a quais fatores, de acordo com Vicini (2005).

A análise fatorial será utilizada neste trabalho como uma contraprova para análise de *Cluster* confirmando resultados propostos nos agrupamentos e variáveis a serem utilizadas nos modelos multivariados de séries temporais.

Na próxima etapa do trabalho, serão descritos os modelos de séries temporais a serem utilizados neste trabalho.

## 2.3 SÉRIES TEMPORAIS

A análise de séries temporais aplicadas a dados de poluição do ar, associados a dados climatológicos, pode trazer um ganho considerável às pesquisas, atrai interesse especial, já que o clima e a poluição interferem diretamente na qualidade de vida dos habitantes de qualquer região, e é uma preocupação constante e crescente.

De acordo com Batista (2009), no estudo de uma série histórica, busca-se investigar o mecanismo gerador da série temporal, fazer previsões de valores futuros, descrever o seu comportamento, avaliar os fatores que influenciam no seu comportamento, buscando relações

de causa e efeito, além de construir um modelo apropriado no intuito de se realizar previsões para embasar tomadas de decisão.

Para Bressan e Lima (2002), Morettin e Toloi (2004), a previsão não constitui um fim em si, mas um meio de fornecer informações e subsídios para uma consequente tomada de decisão, visando atingir determinados objetivos e reduzir as incertezas no processo de tomadas de decisões.

Com auxílio de modelos de séries temporais, pretende-se ajustar modelos para realizar previsões futuras dos principais índices de poluição do ar e utilizar as variáveis climatológicas como variáveis explicativas aos modelos.

A seguir, são apresentados alguns conceitos importantes referentes à classe de modelos de séries temporais desenvolvidos por Box e Jenkins.

### 2.3.1 Modelos Box-Jenkins

De acordo com Reisen, Molinares e Teixeira (2008) e Batista (2009), a metodologia proposta por Box-Jenkins tem sido a mais utilizada para modelagem e previsão de séries temporais e vem sendo amplamente utilizada em diversas áreas do conhecimento.

Esses modelos, genericamente conhecidos por ARIMA ou Autorregressivos Integrados de Médias Móveis, são modelos matemáticos que visam captar o comportamento da correlação seriada ou autocorrelação entre os valores da série temporal e, com base nesse comportamento, realizar previsões futuras, segundo Danfá (2009).

> A filosofia da modelagem Box e Jenkins se utiliza de duas ideias: o princípio da parcimônia e a construção de modelos através de ciclo iterativo. O princípio da parcimônia estabelece a escolha de um modelo com o menor número de parâmetros possíveis, para adequada representação matemática. O ciclo iterativo é uma estratégia de seleção de modelos a ser empreendida até que tenhamos um modelo satisfatório (SOUZA e CARMARGO, 1996, p. 27).

Os autores Box, Jenkins e Reinsel (1994) subdividem os modelos ARIMA em quatro categorias: (i) modelos autorregressivos, identificados como modelos AR(p), que modelam as estruturas de autocorrelação da variável em relação a valores passados; (ii) modelos de médias móveis, chamados de modelos MA(q), que englobam a estrutura de autocorrelação relacionada aos erros de previsão anteriores; (iii) modelos autorregressivos e de médias móveis, simbolizados por ARMA (p,q) e (iv) modelos autorregressivos integrados e de média móvel, chamados de modelos ARIMA (p,d,q), que usam diferenciações para tornar a série estacionária.

*2.3.1.1 Modelo Autorregressivo*

Para Morettin (2008), o modelo é explicado pelas observações passadas, sendo representado pela equação:

$$\tilde{Z}_t = \underbrace{\phi_1 Z_{t-1} + \phi_2 Z_{t-2} +, \dots, + \phi_p Z_{t-p}}_{\text{autorregressivo (p)}} + \varepsilon_t,$$

onde $Z_t$ corresponde às observações históricas da série e consideramos d=q=0;
$\phi_p$ correspondem aos parâmetros autorregressivos;
$\varepsilon_t$ representa os erros que não podem ser explicados pelo modelo.

O modelo autorregressivo (AR) de ordem p é gerado pela combinação das observações passadas mais o erro aleatório ($\varepsilon_t$),

*2.3.1.2 Modelo de Médias Móveis*

De acordo com Morettin (2008), o processo de médias móveis (MA) de ordem q é definido conforme a equação:

$$\tilde{Z}_t = \underbrace{\theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} -, \dots, - \theta_q \varepsilon_{t-q}}_{\text{médias móveis (q)}} + \varepsilon_t,$$

onde $\theta_q$ correspondem às ordens do modelo (MA), $\varepsilon_t$ representa os erros que não podem ser explicados pelo modelo e p=d=0.

O modelo de médias móveis (MA) é explicado pelos erros de previsão de observações passadas e considera que a série em estudo seja estacionária.

*2.3.1.3 Modelo Autorregressivo e de Médias Móveis*

Para Figueredo (2008), os modelos AR, MA e ARMA são modelos de séries lineares estacionárias, que se caracterizam por se desenvolverem no tempo ao redor de uma média constante, possuem uma característica em comum que considera modelos onde a série é estacionária. Sendo assim o modelo ARMA é a combinação dos processos autorregressivos (AR) e de médias móveis (MA), segundo Bueno (2008) e Morettin e Toloi (2004), definindo

assim o modelo misto ARMA (p,q). Os modelos ARMA relacionam os valores futuros com as observações passadas e podem ser representadas por:

$$\tilde{Z}_t = \underbrace{\phi_1 Z_{t-1} + \phi_2 Z_{t-2} +, \dots, + \phi_p Z_{t-p}}_{\text{autorregressivo (p)}} + \varepsilon_t - \underbrace{\theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} -, \dots, - \theta_q \varepsilon_{t-q}}_{\text{médias móveis (q)}},$$

São indicados para séries que não podem ser explicadas puramente pelos modelos AR ou MA.

*2.3.1.4 Modelo Autorregressivo, Integrado e de Médias Móveis*

Esses modelos, genericamente conhecidos por ARIMA ou Autorregressivos Integrados de Médias Móveis são modelos matemáticos que visam captar o comportamento da correlação seriada ou autocorrelação entre os valores da série temporal e, com base nesse comportamento, realizar previsões futuras, segundo Danfá (2009). De acordo com Batista (2009), essa classe de modelos vem sendo largamente utilizada.

Segundo Werner e Ribeiro (2003), quando uma série temporal apresenta média e variância dependentes do tempo é porque ela não é estacionária. Como a maioria dos procedimentos de análise estatística de séries temporais supõe que estas sejam estacionárias, será necessário transformá-las caso ainda não sejam. O processo mais simples de transformação consiste em tomar diferenças sucessivas da série original até obter uma série estacionária.

O número (d) de diferenças simples necessárias para tornar a série estacionária é denominado ordem de integração. De acordo com Batista (2009), se a série se torna estacionária após d diferenças, é chamada integrada (I) de ordem d.

Sendo assim, o modelo ARMA integrado passa a ser denominado de modelo ARIMA. A inclusão do termo ordem de integração permite que sejam utilizados os modelos ARIMA (p,d,q), dados pela equação:

$$\Delta^d \tilde{Z}_t = \underbrace{\phi_1 Z_{t-1} + \phi_2 Z_{t-2} +, \dots, + \phi_p Z_{t-p}}_{\text{autorregressivo (p)}} + \varepsilon_t - \underbrace{\theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} -, \dots, - \theta_q \varepsilon_{t-q}}_{\text{médias móveis (q)}}$$

em que os parâmetros $\phi_1, \dots, \phi_p$ são os parâmetros referentes à parte autorregressiva, $\theta_1, \dots, \theta_q$ são os parâmetros de médias móveis, $\varepsilon_t$é o erro que não pode ser estimado pelo modelo, e d representa o número de diferenças aplicadas na série.

*2.3.1.5 Modelo SARIMA*

Em estudos de séries temporais, é comum encontrar sazonalidade, o que remete à necessidade de utilizar modelos que incorporem essa importante característica temporal, de acordo com Reisen, Molinares e Teixeira (2008). De acordo com Morettin e Toloi (2004), os modelos ARIMA foram ampliados para os modelos SARIMA (p,d,q)(P,D,Q) representados por:

$$\nabla^{D}\Delta^{d}\tilde{Z}_{t} = \underbrace{\phi_1 Z_{t-1} + \phi_2 Z_{t-2} +, \dots, +\phi_p Z_{t-p}}_{\text{autorregressivo (p)}} + \varepsilon_t - \underbrace{\theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} -, \dots, -\theta_q \varepsilon_{t-q}}_{\text{médias móveis (q)}}$$

$$+ \underbrace{\Phi_1 Z_{t-1s} + \Phi_2 Z_{t-2s} +, \dots, +\Phi_P Z_{t-P}}_{\text{autorregressivo sazonal (P)}} - \underbrace{\Theta_1 \varepsilon_{t-1s} - \Theta_2 \varepsilon_{t-2s} -, \dots, \Theta_Q \varepsilon_{\varepsilon t-Q}}_{\text{médias móveis sazonal (Q)}}$$

em que (p,d,q) são as ordens do modelo referentes à parte ordinal ou simples, enquanto que (P,D,Q) são referentes às ordens do modelo da parte sazonal. Os parâmetros $\Phi_1, \dots, \Phi_P$ são os referentes à parte autorregressiva; $\Theta_1, \dots, \Theta_Q$ são os parâmetros de médias móveis referentes à parte sazonal; $\varepsilon_t$ é o erro que não pode ser estimado do modelo; d representa número de diferenças da parte ordinal, e D representa o número de diferenças da parte sazonal.

*2.3.1.6 Modelo ARMAX*

Segundo Martínez e Zamprogno (2003), Silva (2011) e Ozaki e Admi (2012), o modelo ARMAX (autorregressivo com médias móveis e entrada de variáveis exógenas) baseia-se em um modelo ARMA, em que é possível combiná-lo com uma estrutura de regressão (variáveis explicativas) na forma:

$$\tilde{Z}_t = \underbrace{\phi_1 Z_{t-1} + \phi_2 Z_{t-2} +, \dots, +\phi_p Z_{t-p}}_{\text{autorregressivo (p)}} + \underbrace{\beta_1 Y_{t-1} \beta_2 Y_{t-2} +, \dots + \beta_r Y_{t-r}}_{\text{variável independente}} +$$

$$\varepsilon_t - \underbrace{\theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} -, \dots, -\theta_q \varepsilon_{t-q}}_{\text{médias móveis (q)}},$$

onde $Z_t$ é a variável dependente, $y_t$ é a variável independente, $\varepsilon_t$ é ruído branco, $\phi_i$ e $\theta_i$ são os parâmetros desconhecidos, $\beta_i$ é o coeficiente da variável exógena, e $\varepsilon_t$ é a sequência de resíduos do modelo. Essa estrutura pode considerar a inclusão de mais variáveis explicativas

(exógenas). Quando a série é não estacionária, sendo necessário aplicar diferenças na série, o modelo ARMAX passa a se denominar ARIMAX.

Conforme Chechi e Sanches (2013) e Lizzi (2012), o modelo SARIMAX é uma extensão do modelo ARIMA/SARIMA onde podem se incluir variação sazonal e variáveis externas explicativas.

Desse modelo podem derivar outros modelos múltiplos que possibilitam a análise de mais de uma variável correlacionada a previsão como ARIX (p,d,0,r[]), IMAX (0,d,p,r[]) ARMAX (p,0,q,r[]), ARX (p,0,0,r[]), MAX (0,0,q,r[]),com a diferença que, além da estimação das ordens dos coeficientes dos modelos, é preciso estimar os coeficientes das variáveis exógenas, de acordo com Andrade (2009) e Silva (2011).

Ainda há de se considerar que, em qualquer modelo da classe Box Jenkins, a estrutura residual deve ser analisada, isto é, média zero, variância constante e erros não correlacionados com qualquer realização da própria série (autocorrelação igual a zero) para que haja um bom ajuste de modelo, segundo Bueno (2008).

### 2.3.2 Alisamento Exponencial

Uma das grandes classes de modelos de previsão que estuda o padrão de qualquer ruído contido em séries de tempo e que, através dele, prevê valores futuros é o de alisamento exponencial. A grande popularidade atribuída ao método é devido a grande popularidade atribuída ao método advém de sua simplicidade e eficiência, segundo Morettin e Toloi (1987).

A seguir são apresentados alguns conceitos importantes referentes à classe de modelos de Alisamento Exponencial.

#### *2.3.2.1 Alisamento Exponencial Simples*

O AES pode ser descrito da seguinte maneira, de acordo com Morettin e Toloi (2004) e Morettin e Toloi (1987):

$$\bar{Z}_t = \propto Z_t + (1-\propto)\bar{Z}_{t-1}, \text{ para t = 1,2,..., n}$$

onde: $Z_t$ é o valor no instante t; $\bar{Z}_t$ é o valor exponencial suavizado, $\bar{Z}_{t-1}$ é o valor exponencial suavizado no instante t – 1; $\propto$ é a constante de suavização assumindo valores no intervalo $0 \leq \propto \leq 1$;

Por substituição, pode-se escrever a expressão anterior:

$$\bar{Z}_t = \propto Z_t + (1-\propto)Z_{t-1} + \propto (1-\propto)^2 Z_{t-2}$$

O que significa que o método é uma média ponderada que dá pesos para as variáveis mais recentes. A previsão de todos os valores futuros é dada pelo último valor exponencial suavizado, isto é,

$$\hat{Z}_t(h) = \bar{Z}_{t-1}$$
$$\hat{Z}_t(h) = \propto Z_t + (1-\propto)\hat{Z}_{t-1}(h+1)$$

onde: $\hat{Z}_t(h)$ é à previsão h passos a frente e $\hat{Z}_{t-1}$ é a previsão imediatamente anterior. O AES é um método muito usado devido ao fácil entendimento, grande flexibilidade da constante de suavização e à necessidade de armazenar somente $Z_t$, $\bar{Z}_t$ e $\propto$.

*2.3.2.2 Alisamento Exponencial Biparamétrico*

De acordo com Morettin e Toloi (2004) e Morettin e Toloi (1987), o método é indicado para séries com tendências lineares positivas ou negativas. O método é similar ao AES, a diferença é que, em vez de suavizar só o nível, utiliza uma nova constante para a tendência.

Os valores das constantes nível e tendência, no instante t, serão estimados por:

$$\bar{Z}_t = AZ_t + (1-A)\left(\bar{Z}_{t-1} + \hat{T}_{t-1}\right), \qquad 0 < A < 1$$
$$\hat{T}_t = C(\bar{Z}_t - \bar{Z}_{t-1}) + (1-C)\hat{T}_{t-1}, 0 < C < 1$$

respectivamente, A e C são as constantes para o nível e tendência, sendo $\hat{Z}_t$.o fator do nível e $\hat{T}_t$ o fator da tendência.

A equação de previsão é dada por:

$$\hat{Z}_t(h) = \bar{Z}_t + \hat{T}_t h, \text{ com h= 1,2,....}$$

A previsão é feita adicionando-se ao valor básico, $\bar{Z}_t$, a tendência multiplicada pelo número de passos à frente (h) desejado.

*2.3.2.3 Alisamento Exponencial de Holt-Winters (AEHW)*

O método de AEHW é um dos mais utilizados para a previsão de curto prazo, devido à sua simplicidade, baixo custo de operação, boa precisão e capacidade de ajuste automático rápido a mudanças na série. Pode ter duas variações de construção, o modelo multiplicativo e o modelo aditivo, de acordo com o comportamento das componentes da série. Cada um dos procedimentos (aditivo ou multiplicativo) requer a determinação de três constantes de alisamento (A, C e D), associadas a cada uma das componentes para sazonalidade, nível e tendência, segundo Lúcio et al. (2010) e Nascimento (2013).

De acordo com Veríssimo et al. (2012) e Morettin e Toloi (2004), o modelo aditivo caracteriza-se pelo padrão sazonal não se alterar à medida que o nível da série muda, indicado para casos em que a série apresenta comportamento homogêneo. A equação geral do modelo é dada por:

$$Z_t = \mu_t + T_t + F_t + \varepsilon_t,$$

para t=1,...,N, onde N é tamanho da série.

As três equações de suavização são apresentadas abaixo:

$$\hat{F}_t = D(Z_t - \overline{Z}_t) + (1-D)\hat{F}_{t-s} \quad 0 < D < 1;$$
$$\overline{Z}_t = A(Z_t - \hat{F}_{t-s}) + (1-A)(\overline{Z}_{t-1} + \hat{T}_{t-1}),\ 0 < A < 1;$$
$$\hat{T}_t = C(\overline{Z}_t - \overline{Z}_{t-1}) + (1-C)\hat{T}_{t-1},\ 0 < C < 1,$$

onde: A é a constante de suavização do nível, C é a constante de suavização da tendência e D é a constante da componente sazonal, $\hat{T}_t$ é a estimativa do fator da tendência, $\hat{F}_t$ é a estimativa do fator da sazonal e $\overline{Z}_t$ é a estimativa do fator do nível.

A equação de previsão(h) para o modelo aditivo é dada por:

$$\hat{Z}_t(h) = \hat{Z}_t + \hat{T}_t\, h + \hat{F}_{t+h-s}$$

com h = 1, 2,...,s.

O modelo multiplicativo se caracteriza por ter a sazonalidade da série afetada pela tendência. É indicado quando a amplitude da série em estudo varia de acordo com a sua evolução ou aumenta à medida que o nível da série aumenta. É representado pela seguinte equação, segundo Veríssimo et al. (2012) e Morettin e Toloi (2004):

$$Z_t = \mu_t T_t + F_t + \varepsilon_t$$

para t=1,...,N onde N é tamanho da série.

As três equações de suavização para nível tendência e sazonalidade são similares às anteriores, conforme se pode ver a seguir:

$$\hat{F}_t = D\left(\frac{Z_t}{\overline{Z}_t}\right) + (1-D)\hat{F}_{t-s},\ 0 < D < 1, t = s+1,..., N;$$

$$\overline{Z}_t = A\left(\frac{Z_t}{\hat{F}_{t\text{-}s}}\right) + (1-A)(\hat{Z}_{t-1} + \hat{T}_{t-1}),\ 0 < A < 1, t = s+1,..., N;$$

$$\hat{T}_t = C(\overline{Z}_t - \overline{Z}_{t-1}) + (1-C)\hat{T}_{t-1},\ 0 < C < 1, t = s+1,...., N;$$

onde: A é a constante de suavização do nível, C é a constante de suavização da tendência, D é a constante de suavização da componente sazonal, $\hat{T}_t$ é a estimativa do fator da tendência, $\hat{F}_t$ é a estimativa do fator da sazonal e $\overline{Z}_t$ é a estimativa do fator do nível, s corresponde ao valor do intervalo sazonal e N o número de observações.

A equação de previsão do modelo multiplicativo é dada por:

$$\hat{Z}_t(h) = (\hat{Z}_t + \hat{T}_t\, h)\hat{F}_{t+h-s}$$

com h = 1, 2,...,s.

O modelo AEHW pode ser empregado para séries que apresentam comportamento geral, sendo de fácil aplicação. A determinação das constantes de suavização (A, C, D) é realizada de modo a tornar mínima a soma dos quadrados dos erros ajustados.

A seguir são apresentados os critérios que foram considerados na seleção do melhor modelo ajustado.

## 2.4 CRITÉRIOS DE SELEÇÃO DOS MELHORES MODELOS

Como um dos principais objetivos da análise de séries temporais é gerar previsões de valores futuros, é importante utilizar alguns critérios de eficiência do modelo que considerem o grau de precisão de suas previsões futuras, segundo Souto et al. (2007).

Um bom modelo é o que mais se aproxima dos dados reais e, para Sousa (2012), não existem modelos perfeitos, pois não há como identificar e controlar todas as variáveis envolvidas no fenômeno estudado, obtendo-se valores aproximados para os resultados projetados pelo modelo ajustado.

Existem vários critérios disponíveis na literatura, sendo alguns apresentados a seguir.

### 2.4.1 Critérios de seleção dos Modelos de Box Jenkins

Os critérios a serem utilizados para verificar o ajuste do modelo autorregressivo integrado e de médias móveis são: AIC e BIC, relatados abaixo.

O critério de AIC (*Akaike Information Criteria*) proposto por Akaike (1974):

$$\mathrm{AIC} = \ell n(\hat{\sigma}_e^2) + \frac{2(p+q)}{n}$$

e o critério BIC (*Bayesian Information Criterion*) desenvolvido por Akaike (1978):

$$\mathrm{BIC} = \ell n(\hat{\sigma}_e^2) + \frac{(p+q)\ell n(n)}{n},$$

onde $\hat{\sigma}_e^2$ é a variância estimada dos erros, n é o tamanho da amostra e (p,q) é o número de parâmetros de cada componente AR e MA. Os critérios levam em conta a variância do erro, o tamanho da amostra n e os valores de (p,q).

Nesse sentido, vários modelos concorrentes são ajustados e, a partir dos valores dos critérios citados, é possível escolher o modelo mais adequado para realizar previsões h passos à frente. Quanto menores forem esses valores, melhor é a adequação do modelo para realizar previsões.

Além disso, a condição de ruído branco dos resíduos é considerada, ou seja, faz-se necessário que os erros do modelo devem ter média zero e variância constante, além de serem descorrelacionados ao longo do tempo, indicando que o modelo está tão bem ajustado restando apenas a parte aleatória que não pode ser explicada pelo modelo.

### 2.4.2 Critérios de comparação de modelos

Os critérios de avaliação da qualidade das previsões a serem utilizados para os modelos descritos acima são: Erro Quadrático Médio (EQM), Soma dos Quadrados dos Erros (SSE), Média Absoluta dos Erros (MAD) e o Erro Absoluto Médio Percentual (MAPE), descritos na Tabela 2, de acordo com Tidre, Biase e Silva (2013), Esquivel, Senna e Gomes (2012), Barbosa e Chaves (2012) e Feliciano (2009).

Tabela 2 – Critérios de avaliação da qualidade de previsão

| Critério | Equação |
|---|---|
| Erro quadrático médio | $EQM = \frac{\sum_{t=1}^{n}\left(Z_t - \hat{Z}_t\right)^2}{n}$ |
| Soma dos quadrados dos Erros | $SSE = \sum_{t=1}^{n}(Z_t - \hat{Z}_t)^2$ |
| Média absoluta dos erros | $MAD = \frac{\sum_{t=1}^{n}\left\|Z_t - \hat{Z}_t\right\|}{n}$ |
| Erro Absoluto Médio Percentual | $MAPE = \frac{\sum_{t=1}^{n}\left\|\frac{Z_t - \hat{Z}_t}{Z_t}\right\|}{n}.100\%$ |

onde:
$Z_t$ é o valor da série no instante t;
$\hat{Z}_t$ é o valor previsto para o instante t;
n é o número período de tempos.

Quanto menores os valores encontrados nos critérios (EQM), (SSE), (MAD) e o (MAPE), mais próximos dos valores reais estão os valores previstos pelos modelos ajustados.

# 3 MATERIAIS E MÉTODOS

Diante dos objetivos deste estudo, fez-se uma pesquisa exploratória e descritiva, pois, segundo Gil (2002), tais pesquisas têm como objetivo principal o aprimoramento de ideias, levantamento bibliográfico, análise de exemplos e estabelecimento de relação entre variáveis.

Quanto aos procedimentos técnicos, Gil (2002) caracteriza como:

- Pesquisa Bibliográfica: desenvolvida com base em material já elaborado, constituído principalmente de livros e artigos científicos e

- Estudo de caso: estudo teórico aplicado a uma determinada situação, gerando possíveis hipóteses a estudos futuros.

Para realização deste estudo, foram considerados os dados da série histórica, fornecidos pela FEPAM, das medições horárias dos poluentes do ar realizadas no município de Canoas, no ano de 2014 e 2015. Devido ao fato de haver grande número de dados faltantes em 2015, foi utilizada apenas a série histórica de 2014, a qual se refere às medições dos seguintes poluentes: material particulado (PM10), ozônio (O3), dióxido de enxofre (SO2) e monóxido de carbono (CO). Os dados disponíveis foram transformados em médias diárias. As variáveis climáticas, Umidade Relativa Média do Ar (%), Temperatura Mínima (°C), Temperatura Máxima (°C), Temperatura Média (°C), Precipitação (mm), Insolação (h), Evaporação (mm) e Velocidade do Vento (m/s) são dados coletados diariamente pelo Instituo Nacional de Meteorologia (INMET), estação meteorológica Porto Alegre (OMM: 83967), localizada a 16,7 Km do município de Canoas.

Os dados coletados foram tratados em técnicas de análise multivariada e, após definidas as similaridades, foram usados como subsídios para ajuste de modelos de previsão de séries temporais que permitem a inclusão de variáveis explicativas. Como uma das finalidades foi verificar se as variáveis exógenas produzem melhores resultados de previsão, modelos concorrentes também foram ajustados para fins de comparação.

Para verificar a interação entre as variáveis de poluentes do ar e as climatológicas, foram utilizadas as técnicas de análise multivariada de *cluster* e fatorial, obedecendo a seguinte ordem:

1) Padronizar as variáveis;
2) Definir a medida de parecença e do método de aglomeração hierárquico;
3) Construir os Dendogramas e *Scree plot*, definindo os agrupamentos análise de *Cluster* e pontos de corte;

4) Determinar os fatores;
5) Realizar a análise fatorial para confirmar os resultados encontrados na análise de *cluster*;
6) Exclusão das variáveis que não se agrupam por falta de homogeneidade em seus grupos, se for o caso;
7) Refazer as análises após eliminar as variáveis;
8) Definir os grupos e as variáveis que podem ser usadas como variáveis exógenas aos modelos de previsão ARIMAX.

Os modelos de séries temporais a serem ajustados e testados são: os modelos ARIMAX e ARIMA e os modelos de Alisamento Exponencial Simples, Biparamétrico e Holt-Winters. Para realizar uma comparação do modelo mais apropriado, foram reservadas as 20 últimas observações das séries históricas em estudo.

Para análise dos modelos de séries temporais, foram consideradas as seguintes etapas:

1) Realizar uma análise gráfica da série original (verificar indícios de tendência e sazonalidade);
2) Através de testes estatísticos, confirmar ou descartar a presença de tendência e sazonalidade;
3) Empregar transformações numéricas na série, caso necessário, a fim de tornar a série estacionária para o modelo ARIMA e ARMAX;
4) Construir gráficos da função de autocorrelação (FAC) e autocorrelação parcial (FACP) para obter uma indicação de modelos a serem estimados ARIMA e ARMAX;
5) Ajustar os modelos ARIMA, ARMAX, Alisamento Exponencial Simples, Biparamétrico e de Holt-Winters;
6) Escolher os melhores modelos, por meio dos critérios de seleção AIC e BIC, para os modelos Box-Jenkins;
7) Utilizar os critérios de seleção EQM, MAD e SSE para escolha do melhor modelo de Alisamento Exponencial de Holt-Winters aditivo ou multiplicativo;
8) Analisar a normalidade dos resíduos dos modelos por meio de testes estatísticos;
9) Verificar a condição de ruído branco (média zero, variância constante e erros descorrelacionados);
10) Realizar previsões 20 passos à frente com os modelos ARIMA, ARMAX, Alisamento Exponencial Simples, Biparamétrico e de Holt-Winters;

11) Comparar o resultado das previsões dos melhores modelos ARIMA, ARMAX, Alisamento Exponencial Simples, Biparamétrico e de Holt-Winters, com os valores da série histórica reservados para esse fim, usando os critérios: MAPE, EQM, MAD e SSE;
12) Identificar o modelo que melhor se ajuste à série de dados;
13) Como última etapa, verificar se as técnicas de análise multivariada auxiliam na escolha de variáveis exógenas e se essas varáveis melhoram as previsões dos modelos de séries temporais;
14) Para realizar o ajuste dos modelos, foram utilizados os softwares R e o STATISTICA 9.1.

# 4 ANÁLISE DOS RESULTADOS

Neste capítulo são apresentados os resultados dos testes estatísticos e das técnicas de análise multivariada e séries temporais descritos no capítulo 2.

## 4.1 ESTATÍSTICA DESCRITIVA

Inicialmente são apresentados as principais medidas descritivas dos principais poluentes do ar.

Os valores das médias e desvios padrões das variáveis, coeficiente de variação e valor máximo e mínimo, utilizados neste estudo, estão apresentados na Tabela 3. Observa-se que há um alto coeficiente de variação, principalmente nas variáveis CO e SO2, indicando que os dados em estudo são heterogêneos, ou seja, demonstram um comportamento com variabilidade considerável durante o ano, o que remete a necessidade de estudos dessas variáveis para entender os seus comportamentos.

Tabela 3 – Análise descritivas dos poluentes do ar

| **Variáveis** | **Média** | **Desvio-Padrão** | **Coeficiente de variação** | **Mínimo** | **Máximo** |
|---|---|---|---|---|---|
| CO (μg/m$^{3)}$ | 0,06 | 0,063 | 105 | 0,03 | 0,42 |
| O3 (μg /m$^{3)}$ | 25,77 | 10,22 | 39,6 | 2,3 | 55,86 |
| PM10 (μg /m$^{3)}$ | 31,43 | 17,62 | 56,06 | 4,34 | 133,37 |
| SO2 (μg /m$^{3)}$ | 1,33 | 1,73 | 130 | 0,65 | 21,05 |

CO: Monóxido de carbono; O3: Ozônio; PM10: Material particulado; SO2: Dióxido de enxofre.

Na sequência do trabalho foi realizado uma análise multivariada, conforme descrito anteriormente.

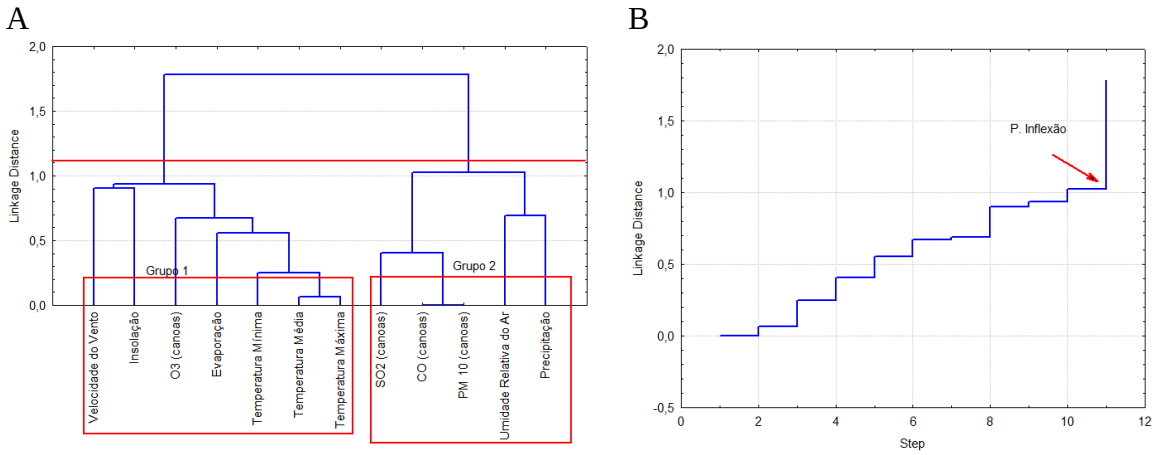
## 4.2 ANÁLISE MULTIVARIADA DE AGRUPAMENTOS

Com a finalidade de verificar a existência de relações entre as variáveis climáticas e poluentes do ar, realizou-se, então, a Análise Agrupamentos.

Nesta pesquisa, utiliza-se como processo de aglomeração hierárquica o método de ligação completo (na busca de maior homogeneidade dentro de cada grupo), sendo a métrica de similaridade a distância Person r (Correlação).

Na Figura 2A mostra-se o comportamento do dendograma com todas as variáveis, na qual se pode identificar a formação de dois grupos, os quais possuem as variáveis de maior relevância dentro do conjunto em estudo. Na Figura 2B demonstra-se o método gráfico sugerido por Cattel (1966), quando a curva (Ponto de inflexão da curva) passa a ser praticamente paralela ao eixo das abscissas, excluem-se as componentes que restam, pois possuem pouca informação, definindo o ponto de corte para a formação dos grupos.

Figura 2 – Dendograma das variáveis em estudo (2A) e *Scree plot* (2B)




Fonte: Autor.

Realizando-se um corte na altura 1,2 de acordo com o método gráfico (Figura 1B), obtém-se o grupo 1 com 7 variáveis e o grupo 2 com 5 variáveis.

O *cluster* 1, representado pelo grupo 1, reúne as variáveis mais próximas: Velocidade do Vento, Insolação, O3, Evaporação, Temperatura Média e Temperatura Máxima e Mínima. O *cluster* 2, representado pelo grupo 2, reúne as variáveis mais próximas: SO2, CO, PM10, Umidade Relativa do Ar e Precipitação. Identificam-se dois grupos bem definidos, nos quais percebe-se que a variável O3 é heterogênea em relação às outras variáveis de poluentes em estudo.

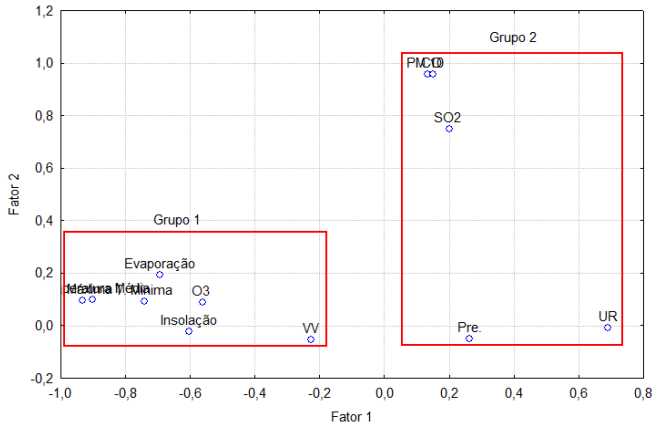
Para a composição de cada grupo, verifica-se a importância de cada variável por meio da matriz de correlações (análise fatorial), apresentadas na Tabela 4.

Tabela 4 – Correlação entre os fatores e as variáveis

| Variável | Factor 1 | Factor 2 |
|---|---|---|
| Precipitação | 0,26 | -0,049 |
| Temperatura Máxima | -0,93 | 0,094 |
| Temperatura Mínima | -0,74 | 0,092 |
| Insolação | -0,60 | -0,02 |
| Evaporação | -0,69 | 0,19 |
| Temperatura Média | -0,90 | 0,09 |
| Umidade Relativa do Ar | 0,69 | -0,008 |
| Velocidade do Vento | -0,22 | -0,05 |
| PM 10 | 0,13 | 0,96 |
| SO2 | 0,20 | 0,75 |
| O3 | -0,56 | 0,089 |
| CO | 0,15 | 0,96 |

Os resultados mostram a relação dos fatores com as variáveis originais, as quais, na Figura 3 apresenta-se no plano fatorial a relação das componentes principais com as variáveis originais e de que forma essas variáveis se agrupam, confirmando os resultados dos agrupamentos formados e apresentados no dendograma da Figura 3. Entretanto, verifica-se baixa correlação entre Velocidade do Vento no Grupo 1 e UR e Precipitação com o Grupo 2.

Figura 3 – Plano Fatorial – Fator 1 *vs* Fator 2




Fonte: Autor.

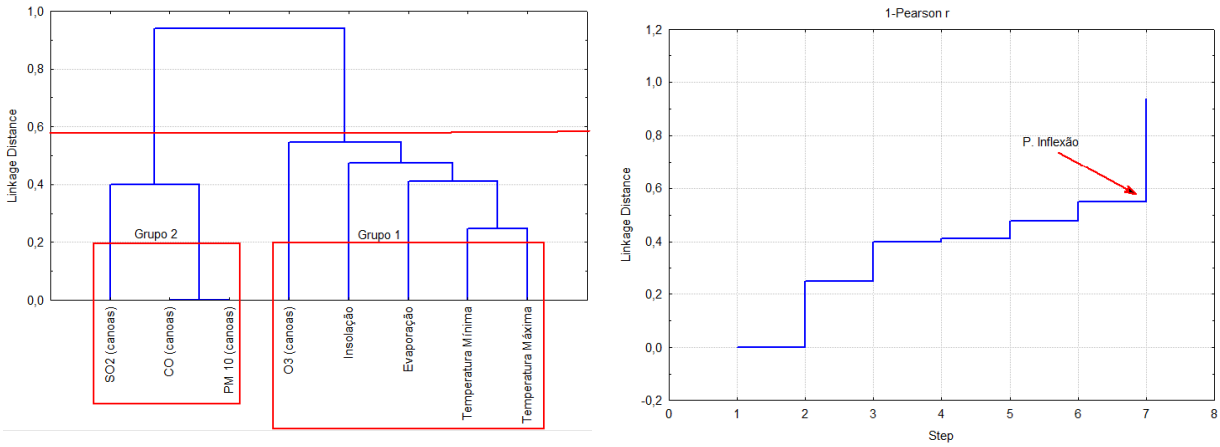
Como o intuito é utilizar as variáveis climatológicas para auxiliar nas previsões das varáveis de poluentes do ar, é necessário achar o máximo de homogeneidade em seus grupos. Dessa forma, serão excluídas as variáveis Precipitação e Umidade Relativa do Ar no Grupo 2 e Velocidade do Vento no Grupo 1, tendo em vista possuírem baixa correlação

(homogeneidade dentro dos seus grupos), a variável Temperatura Média será excluída pela proximidade com a Temperatura Máxima e Mínima, possuindo a mesma carga de informação.

Na Figura 4 demonstra-se a nova formação dos grupos no dendograma (Figura 4A) e o *Scree plot* (Figura 4B), que define o ponto de corte e confirma as componentes dos grupos 1 e 2.

Figura 4 – Dendograma (4A) e Plano Fatorial – Fator 1 *vs* Fator 2 (4B)

A B

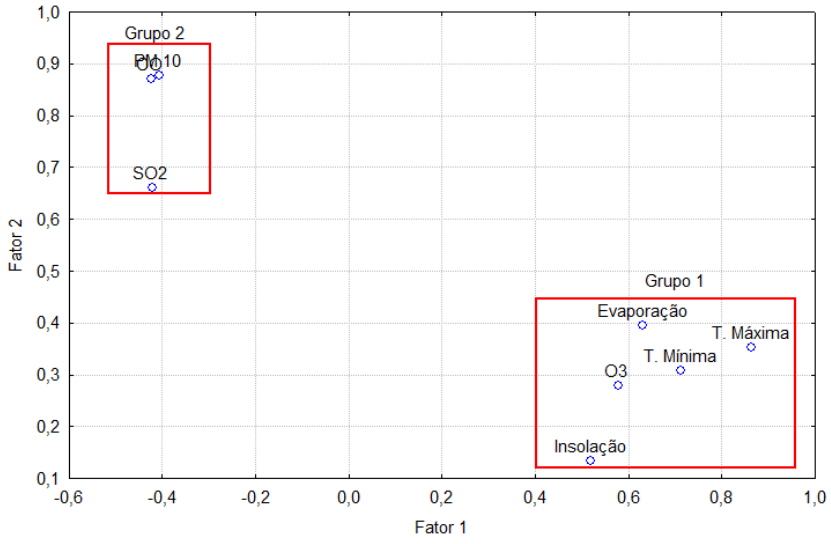



Fonte: Autor.

Realizando-se um corte na altura 0,6 de acordo com o método gráfico de Cattel (1966), obtêm-se os grupos finais definidos.

O *cluster* 1, representado pelo grupo 1, reúne as variáveis mais próximas: O3, Insolação, Evaporação, Temperatura Mínima e Temperatura Máxima. O *cluster* 2, representado pelo grupo 2, reúne as variáveis mais próximas: SO2, CO, PM10. O Plano Fatorial (Figura 5), também corrobora e confirma a formação dos grupos.

Figura 5 – Plano Fatorial – Fator 1 *vs* Fator 2




Fonte: Autor.

Com os grupos definidos, o próximo passo foi definir as variáveis de poluição do ar, que serão modeladas para a previsão, e as variáveis utilizadas como exógenas aos modelos.

Devido às características de agrupamentos e por ser um dos principais poluentes do ar, no grupo 1, serão construídos modelos de previsão para a variável O3 - Ozônio e como variáveis explicativas: Insolação, Evaporação, Temperatura Mínima e Temperatura Máxima. No grupo 2, como as variáveis SO2, CO, PM10 não são influenciadas diretamente por variáveis climáticas e são fortemente correlacionadas, serão construídos modelos para a variável PM10 - Material Particulado Inalável, tendo em vista essa variável ser um dos principais poluentes a causarem danos à saúde e as varáveis SO2- Dióxido de Enxofre e CO - monóxido de carbono foram usadas como exógenas ao modelo.

Com os grupos definidos pela análise multivariada de *cluster* e fatorial, o próximo passo foi ajustar os modelos de séries temporais para as variáveis O3 e PM10.

## 4.3 AJUSTE DE MODELOS DE PREVISÃO PARA A VARIÁVEL O3

Todos os modelos de previsão foram ajustados para uma previsão de 20 passos à frente, ou seja, dias que não serão contabilizados no ajuste dos modelos. Os modelos de previsão buscam não apenas descrever valores futuros, mas que eles possam auxiliar em tomadas de decisões. Para tanto, faz-se necessário testar modelos univariados e multivariados que melhor possam representar as variáveis.

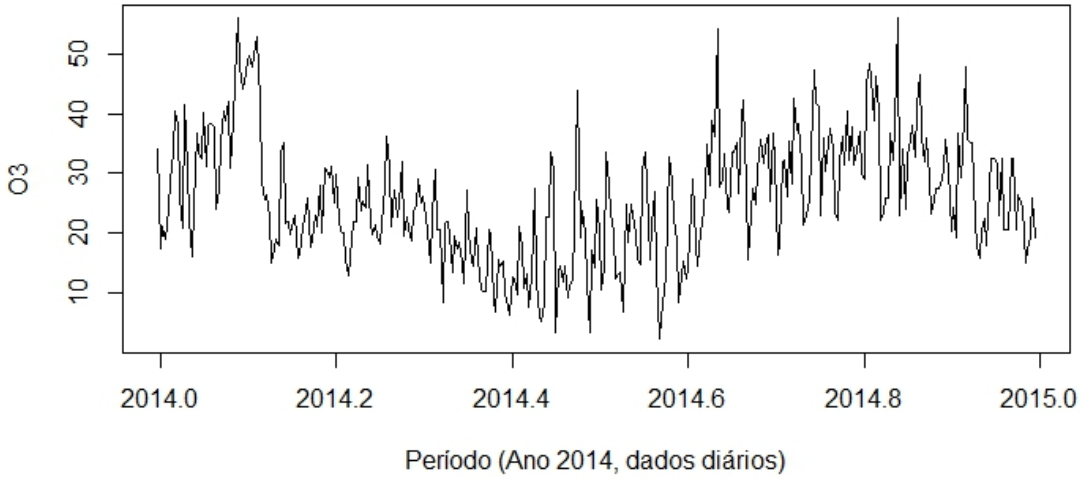
### 4.3.1 Modelos da Classe Box-Jenkins

Nesta parte do trabalho, foram testados modelos da classe Box-Jenkins, com o intuito de verificar se as variáveis exógenas selecionadas pelos métodos de análise multivariada auxiliam para uma previsão de melhor qualidade.

#### *4.3.1.1 Ajuste do modelo ARIMA para a variável O3*

O primeiro modelo de previsão ajustado foi o modelo para a variável O3 – Ozônio. Para dar-se início ao ajuste do modelo, observa-se, na Figura 6, o comportamento da série original, buscando identificar suas componentes, como tendência e/ou sazonalidade.

Figura 6 – Gráfico da série original O3 em Canoas - RS

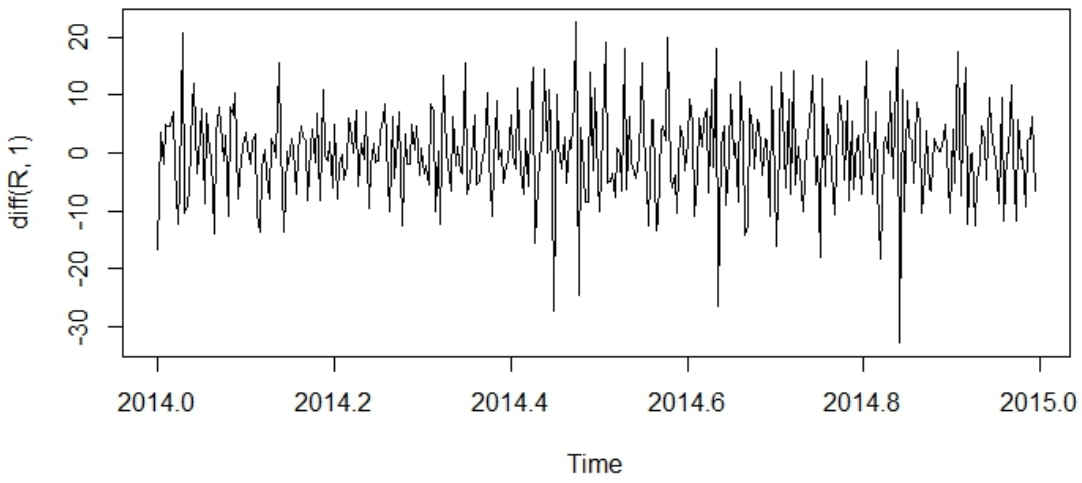



Fonte: Autor.

Analisando a Figura 6, verifica-se que a série possui indícios de sazonalidade e tendência, sendo necessário aplicar diferenciações na série original.

Após a etapa inicial de análise gráfica dos dados, foram realizados testes estatísticos para confirmar as hipóteses levantadas na análise gráfica. De acordo com o teste de sequências p-valor < 0,05, a série apresenta indícios de não ser estacionária.

Figura 7 – Gráfico da Série O3 com uma diferença



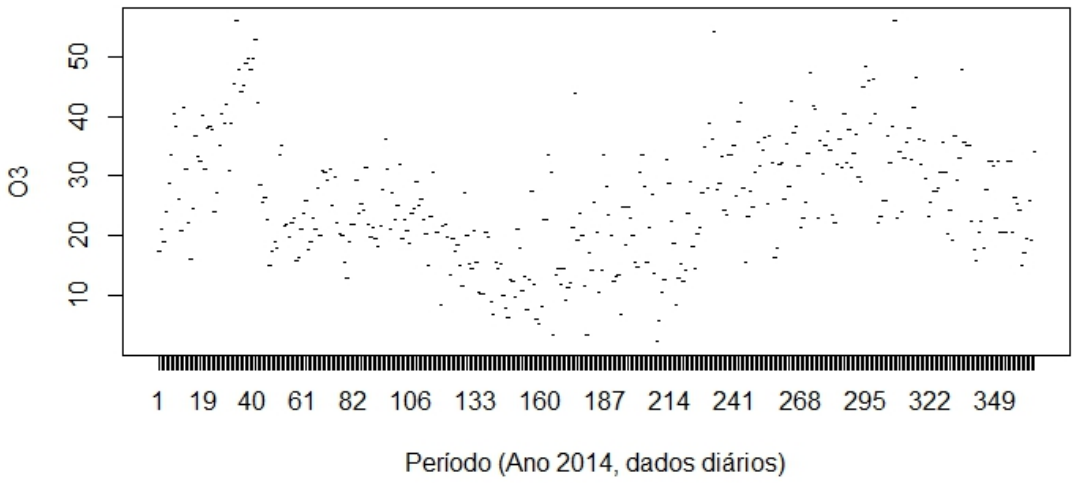

Fonte: Autor.

Na Figura 7 demonstra-se a série O3, com diferença de ordem 1, sendo essa diferença necessária para estabilizar a série.

Para avaliar a sazonalidade, foi realizado o teste Kruskal-Wallis, sendo rejeitada a hipótese nula $H_0$ (p-valor < 0,001) e aceita-se a hipótese alternativa $H_1$ de presença de efeito sazonal na série.

Confirmando o resultado do teste Kruskal-Wallis, o teste de Friedman também rejeitou a hipótese nula $H_0$ (p-valor < 0,001), confirmando a presença de sazonalidade na série em estudo.

Figura 8 – Gráfico da série O3 em pontos



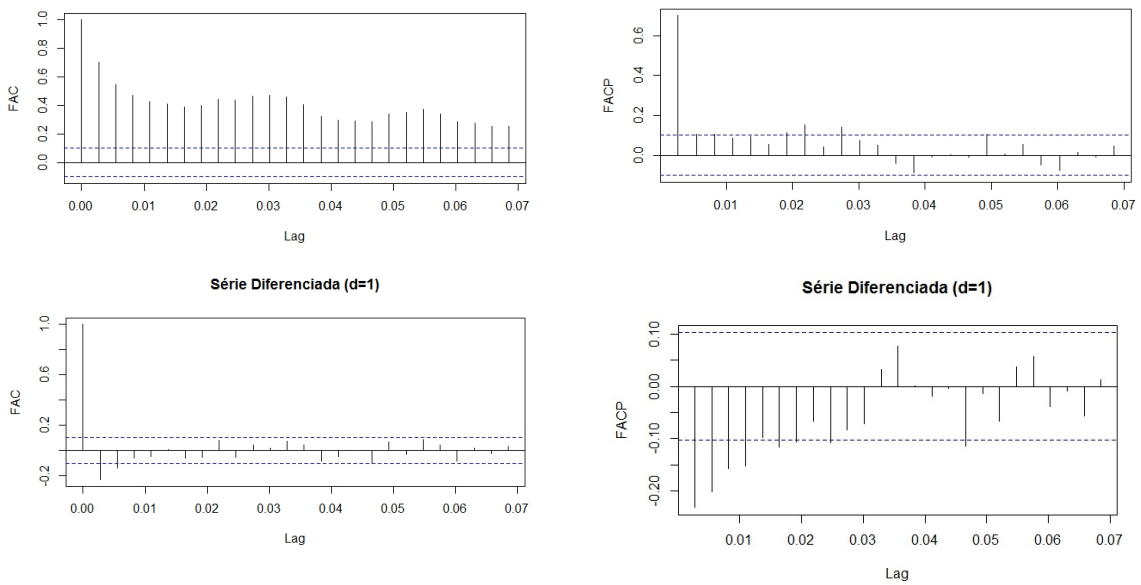

Fonte: Autor.

Na Figura 8 facilita a identificação dos períodos com menor ou maior concentração de O3. Observa-se, entre os dias 106 e 241, períodos de menores concentrações de O3, o que coincide com os meses de abril, maio, junho, julho e agosto, períodos de menores temperaturas; nos meses iniciais e finais do ano de 2014, observa-se um aumento da concentração de O3.

Ao término da etapa de análise inicial da série O3, identificou-se a presença de sazonalidade e tendências na série. A próxima etapa foi a construção dos gráficos de autocorrelação (FAC) e de autocorrelação parcial.

Os gráficos das funções de autocorrelação (FAC) e de autocorrelação parcial (FACP), expressos na Figura 9, demonstram as correlações seriais no domínio do tempo. Com base em suas frequências estimadas, é possível verificar sazonalidades e ciclos na série. Na Figura 9, apresentam-se os correlogramas das séries, sem diferença e com diferença de ordem 1.

Figura 9 – Função de Autocorrelação (FAC) e Autocorrelação Parcial (FACP)




Fonte: Autor.

A FAC em nível corrobora com os testes estatísticos que a série não é estacionária, demostrando fortes correlações entre os *lag*, sendo necessário aplicar diferenciações para estabilizar a série. Ao se analisar os ACF e PACF com diferença de ordem 1, percebem-se características de um modelo estacionário autorregressivo de ordem 1 a 3 e de médias móveis de ordem 1 ou 2. Os testes estatísticos e a análise gráfica indicam que um modelo do tipo SARIMA deve ser considerado. Essa análise é sumária e serve para dar início ao processo de modelagem, embora vários modelos de ordens diferentes serão considerados e testados.

Na Tabela 5, estão expostos alguns modelos concorrentes ajustados para a série O3, com os respectivos valores de AIC e BIC, os parâmetros do modelo, a média e a variância dos resíduos referentes a cada modelo significativo.

Tabela 5 – Modelos concorrentes SARIMA e os respectivos critérios para a série O3

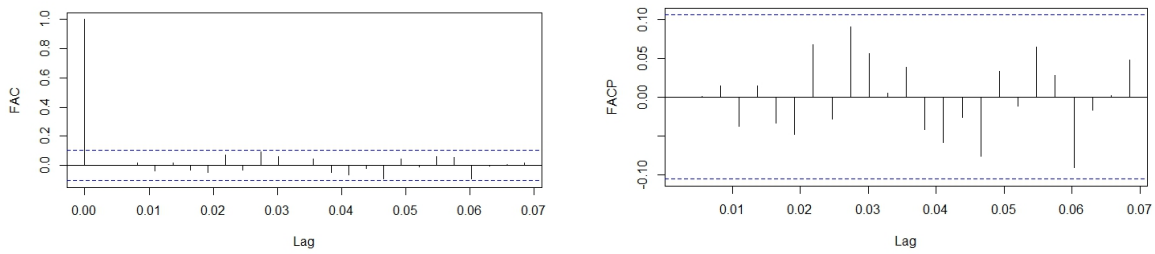
| **SARIMA (p,d,q)(P,D,Q)** | **Parâmetros** | **Critérios de seleção AIC** | **BIC** | **Média/Variância dos resíduos** |
|---|---|---|---|---|
| (1,1,1)(0,1,1) | $\phi_1 = 0{,}4849$<br>$\emptyset_1 = -0{,}9115$<br>$\Theta_1 = -0{,}9332$ | 2298,33 | 2313,55 | $\mu \cong 0{,}0495$<br>$\sigma^2 = 48{,}92$ |
| (2,1,1)(2,1,1) | $\phi_1 = 0{,}4723$<br>$\phi_2 = -0{,}0333$<br>$\emptyset_1 = -0{,}9147$<br>$\Phi_1 = -0{,}053$<br>$\Phi_2 = -0{,}0376$<br>$\Theta_1 = 0{,}1621$ | 2464,23 | 2491,51 | $\mu \cong 0{,}0951$<br>$\sigma^2 = 48{,}92$ |
| (2,1,2)(2,0,0) | $\phi_1 = 0{,}4718$<br>$\phi_2 = -0{,}312$<br>$\emptyset_1$ =-0,9111<br>$\Phi_1 = 0{,}1089$<br>$\Phi_2 = -0{,}0522$ | 2462,35 | 2485,73 | $\mu \cong 0{,}0914$<br>$\sigma^2 = 48{,}94$ |

Considerando os três melhores modelos significativos ajustados, analisou-se a condição de ruído branco (média zero, variância constante, erros não correlacionados). Na análise dos resíduos dos modelos SARIMA, observou-se que a média dos modelos pode ser considerada como zero, conforme Tabela 5.

Os testes de autocorrelação dos resíduos Ljung-Box (p-valor = 0,9949) e Box-Pierce (p-valor = 0,9949), do melhor modelo SARIMA (1,1,1)(0,1,1), de acordo com os critérios de seleção AIC e BIC, indicam que os erros do modelo ajustado são descorrelacionados ao longo do tempo.

Analisando a Figura 10, gráfico de ACF e PACF, pode se perceber que não há autocorrelação significativamente diferente de zero em nenhuma defasagem, portanto, os dados são independentes e demonstram um bom ajuste do modelo, confirmando os resultados dos testes de Ljung-Box e Box-Pierce.

Figura 10 – Função de autocorrelação (FAC) e autocorrelação parcial (FACP) dos resíduos




Fonte: Autor.

Os testes de normalidade dos resíduos Jarque-Bera (p-valor = 0,0977) e Shapiro-Wilk (p-valor = 0,1139) demonstram que os resíduos do modelo SARIMA (1,1,1)(0,1,1) seguem distribuição normal.

Considerando a análise residual e os critérios AIC e BIC, o modelo SARIMA (1,1,1)(0,1,1), com um parâmetro autorregressivo (*p*), um parâmetro de diferenças (*d*) a fim de tornar a série estacionária, um parâmetro de médias móveis (*q*) na parte ordinal, um parâmetro de diferenças (D) na parte sazonal e um parâmetro de médias móveis (*Q*) na parte sazonal atende a todas as condições de ruído branco e normalidade dos resíduos e concluem a última etapa da escolha do modelo para realizar previsões futuras.

Nessa etapa do trabalho, foram obtidas as previsões da série O3 para um horizonte de 20 passos à frente, ou seja, dias que não foram considerados no ajuste do modelo.

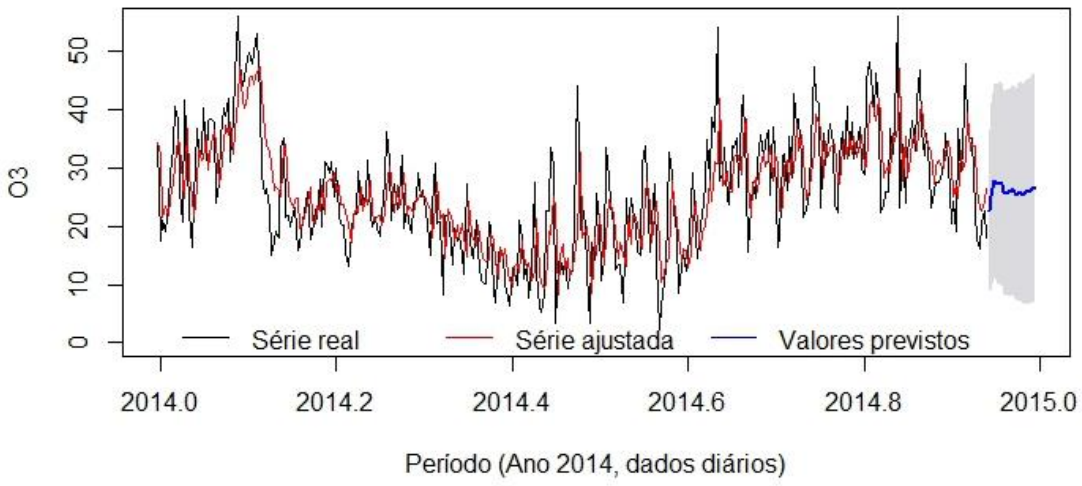
As previsões para o modelo ajustado SARIMA (1,1,1)(0,1,1) podem ser observadas na Tabela 6.

Tabela 6 – Valores previstos (h=20), segundo o modelo SARIMA (1,1,1)(0,1,1)

| **Período Dias** | **Valores previstos** | **Série real** | **Período Dias** | **Valores previstos** | **Série real** |
|---|---|---|---|---|---|
| 346 | 21,8097 | 27,6696 | 356 | 27,6193 | 32,4 |
| 347 | 24,7150 | 32,4 | 357 | 26,5745 | 20,6522 |
| 348 | 26,0552 | 32,4 | 358 | 25,8671 | 26,4217 |
| 349 | 26,60 | 31,6762 | 359 | 26,6137 | 25,2435 |
| 350 | 24,4477 | 22,8833 | 360 | 26,9073 | 24,3217 |
| 351 | 26,4326 | 32,4 | 361 | 26,9447 | 15,0087 |
| 352 | 26,5411 | 20,6522 | 362 | 24,5462 | 17,1087 |
| 353 | 25,4727 | 20,6522 | 363 | 26,4117 | 19,6478 |
| 354 | 22,4134 | 20,6522 | 364 | 26,4624 | 25,8 |
| 355 | 25,6128 | 32,4 | 365 | 25,3660 | 19,3349 |

Na Figura 11, é demonstrada a série real, ajustada, e os valores previstos para os limites de previsão com intervalo de confiança de 95%.

Figura 11 – Modelo SARIMA (1,1,1)(0,1,1) ajustado à série observada




*4.3.1.2 Ajuste do modelo ARIMAX* para a variável O3

Para o ajuste do modelo ARIMAX, foi utilizada a variável endógena O3, e foram utilizadas as variáveis exógenas Insolação, Evaporação, Temperatura Mínima e Temperatura Máxima, com o intuito de melhorar as previsões da variável em estudo.

Dentre os diversos modelos testados na Tabela 7, estão apresentados os três modelos com melhor desempenho nos critérios de seleção AIC e BIC.

Tabela 7 – Modelos concorrentes SARIMAX e os respectivos critérios (variável O3)

(continua)

| **SARIMAX (p,d,q,r[])(P,D,Q)** | **Parâmetros** | **Critérios de seleção** AIC | BIC | **Média/Variância dos resíduos** |
|---|---|---|---|---|
| (1,1,1[1,1,1]))(0,1,1) | $\phi_1 = 0{,}4634$<br>$\emptyset_1 = -0{,}9178$<br>$\Theta_1 = -0{,}9312$<br>T. max.= 0,3093<br>T. min.= -0,1263<br>Evap. = -0,1262 | 2207,67 | 2234,31 | $\mu \cong 0{,}0064$<br>$\sigma^2 = 50{,}67$ |
| (1,1,1[1,1,1]))(1,1,1) | $\phi_1 = 0{,}4611$<br>$\emptyset_1 = -0{,}9225$<br>$\Phi_1 = 0{,}1161$<br>$\Theta_1 = -0{,}9996$<br>T. max.=0,3144<br>T. min.=0,05117<br>Evap. = -0,1324 | 2206,59 | 2237,03 | $\mu \cong 0{,}0020$<br>$\sigma^2 = 49{,}19$ |

(conclusão)

| | | | | |
|---|---|---|---|---|
| (1,1,2[1,1,1]))(0,1,1) | $\phi_1 = 0{,}3873$<br>$\emptyset_1 = -08267$<br>$\emptyset_2 = -0{,}0759$<br>$\Theta_1 = 0{,}9328$<br>T. max.= 0,3134<br>T. min.= 0,0928<br>Evap. = -0,1277 | 2209,50 | 2239,69 | $\mu \cong 0{,}075$<br>$\sigma^2 = 50{,}58$ |

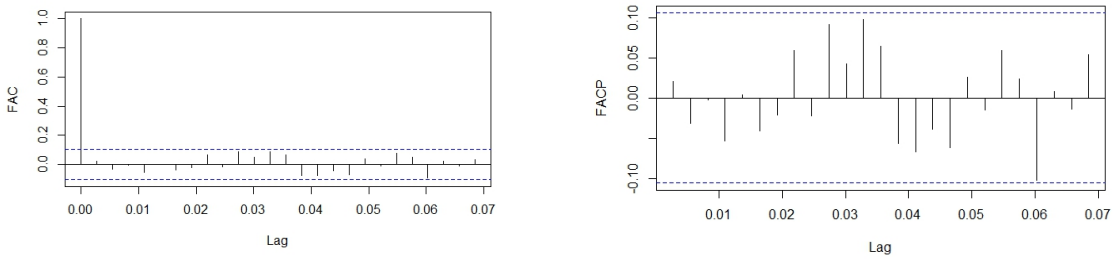
Entre as variáveis exógenas encontradas no item 4.2, por falta de ajuste na construção dos modelos, eliminou-se, nesta etapa, a variável Insolação, porque a série possui muitas observações com a variável de valor zero.

Considerando os três melhores modelos significativos ajustados, analisou-se a condição de ruído branco. Na análise dos resíduos dos modelos SARIMAX, observou-se que a média dos modelos pode ser considerada como zero, conforme Tabela 7.

Os testes de autocorrelação dos resíduos Ljung-Box (p-valor = 0,7046) e Box-Pierce (p-valor = 0,7079), do melhor modelo SARIMAX (1,1,1[1,1,1])(0,1,1), de acordo com os critérios de seleção AIC e BIC, indicam que os erros do modelo ajustado são descorrelacionados ao longo do tempo.

Analisando a Figura 12, gráfico de ACF e PACF, percebe-se que não há autocorrelação significativamente diferente de zero em nenhuma defasagem, portanto, os dados são independentes e demonstram um bom ajuste do modelo, confirmando os resultados dos testes de Ljung-Box e Box-Pierce.

Figura 12 – Função de autocorrelação (FAC) e autocorrelação parcial (FACP) dos resíduos do modelo SARIMAX




Fonte: Autor.

Os testes de normalidade dos resíduos Jarque-Bera (p-valor = 0,02841) e Shapiro-Wilk (p-valor = 0,1801) demonstram que os resíduos do modelo SARIMAX (1,1,1[1,1,1]))(0,1,1) seguem distribuição normal.

Considerando a análise residual e os critérios AIC e BIC, o modelo SARIMAX (1,1,1[1,1,1]))(0,1,1), com um parâmetro autorregressivo (*p*), um parâmetro de diferenças (*d*), um parâmetro de médias móveis (*q*) na parte ordinal, um parâmetro de diferenças (D), um parâmetro de médias móveis (*Q*) na parte sazonal e um parâmetro das variáveis exógenas Temperatura Máxima, Temperatura Mínima e Evaporação, o modelo atende a todas as condições de ruído branco e normalidade dos resíduos e concluem a última etapa da escolha do modelo para realizar previsões futuras.

Nessa etapa do trabalho, foram obtidas as previsões da série O3 para um horizonte de 20 passos à frente, ou seja, dias que não foram considerados no ajuste do modelo.

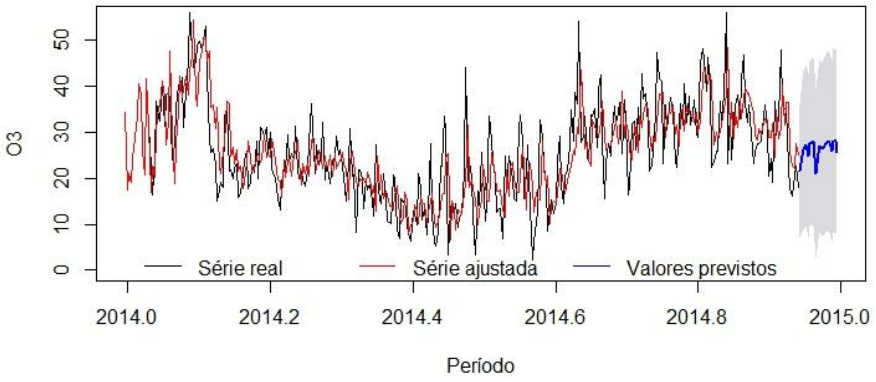
As previsões para o modelo ajustado SARIMAX (1,1,1[1,1,1]))(0,1,1) podem ser observadas na Tabela 8.

Tabela 8 – Valores previstos (h=20) modelo SARIMAX (1,1,1[1,1,1])(0,1,1) (variável O3)

| **Período Dias** | **Valores previstos** | **Série real** | **Período Dias** | **Valores previstos** | **Série real** |
|---|---|---|---|---|---|
| 346 | 22,3937 | 27,6696 | 356 | 27,0345 | 32,4 |
| 347 | 24,8042 | 32,4 | 357 | 26,3240 | 20,6522 |
| 348 | 26,0555 | 32,4 | 358 | 26,4295 | 26,4217 |
| 349 | 27,0960 | 31,6762 | 359 | 27,4279 | 25,2435 |
| 350 | 24,8259 | 22,8833 | 360 | 27,7277 | 24,3217 |
| 351 | 27,4755 | 32,4 | 361 | 26,7357 | 15,0087 |
| 352 | 27,8548 | 20,6522 | 362 | 24,8630 | 17,1087 |
| 353 | 25,9532 | 20,6522 | 363 | 26,9579 | 19,6478 |
| 354 | 20,8950 | 20,6522 | 364 | 28,1132 | 25,8 |
| 355 | 25,2372 | 32,4 | 365 | 25,5714 | 19,3349 |

Na Figura 13, é demonstrada a série ajustada e os valores previstos para os limites de previsão com intervalo de confiança de 95%.

Figura 13 – Modelo SARIMAX (1,1,1[1,1,1])(0,1,1) ajustado à série observada




Fonte: Autor.

### 4.3.2 Ajuste de modelos da classe de alisamento exponencial para variável O3

Os modelos da classe de alisamento exponencial são modelos mais simples de serem aplicados, não necessitam de softwares específicos e possuem menos exigências em seu uso. São métodos automáticos aplicados diretamente à série O3, sendo de fundamental importância sua aplicação para fins de comparação com os métodos mais sofisticados.

Os modelos a serem ajustados são: Alisamento Exponencial Simples, Alisamento Exponencial Biparamétrico e Alisamento Exponencial de Holt-Winters.

#### *4.3.2.1 Alisamento exponencial simples*

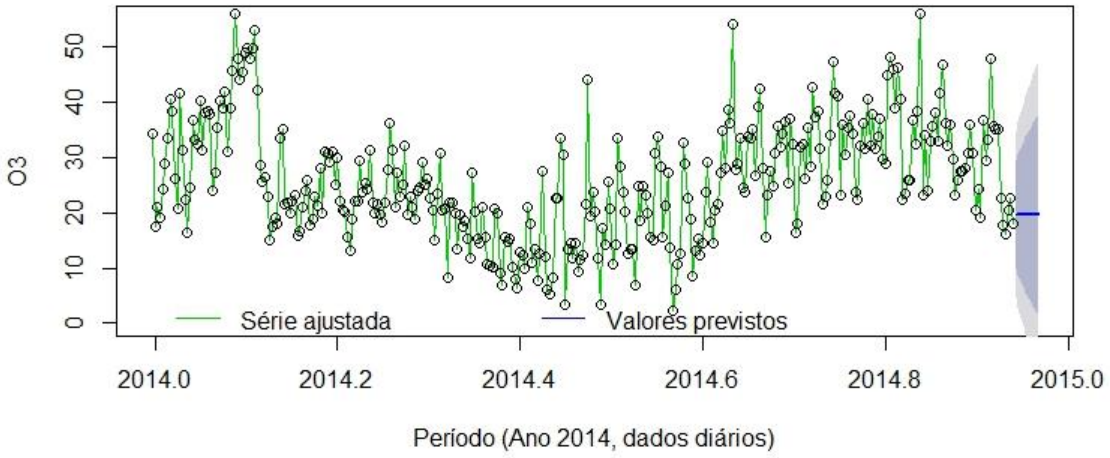
No ajuste do modelo de alisamento exponencial simples, encontrou-se o parâmetro de suavização alfa $\alpha = 0{,}5249$. A previsão de todos os valores futuros da série é dada pela última constante de alisamento. Os valores previstos em comparação com os valores reais estão expostos na Tabela 9.

Tabela 9 – Valores previstos (h=20) do modelo de Alisamento exponencial simples (variável O3)

| Período Dias | Valores previstos | Série real | Período Dias | Valores previstos | Série real |
|---|---|---|---|---|---|
| 346 | 19,6183 | 27,6696 | 356 | 19,6183 | 32,4 |
| 347 | 19,6183 | 32,4 | 357 | 19,6183 | 20,6522 |
| 348 | 19,6183 | 32,4 | 358 | 19,6183 | 26,4217 |
| 349 | 19,6183 | 31,6762 | 359 | 19,6183 | 25,2435 |
| 350 | 19,6183 | 22,8833 | 360 | 19,6183 | 24,3217 |
| 351 | 19,6183 | 32,4 | 361 | 19,6183 | 15,0087 |
| 352 | 19,6183 | 20,6522 | 362 | 19,6183 | 17,1087 |
| 353 | 19,6183 | 20,6522 | 363 | 19,6183 | 19,6478 |
| 354 | 19,6183 | 20,6522 | 364 | 19,6183 | 25,8 |
| 355 | 19,6183 | 32,4 | 365 | 19,6183 | 19,3349 |

Na Figura 14, é demonstrada a série ajustada e os valores previstos para os limites de previsão com intervalo de confiança de 95%.

Figura 14 – Modelo de alisamento exponencial simples




Fonte: Autor.

A série representada pela cor verde é a série ajustada, as circunferências de cor preta, representam a série O3 e, em azul, o valor encontrado para as previsões futuras.

*4.3.2.2 Alisamento exponencial biparamétrico*

Esse modelo usa técnicas similares ao alisamento exponencial simples, suavizando diretamente os valores da tendência, permitindo usar constantes com diferentes valores para o nível e tendência.

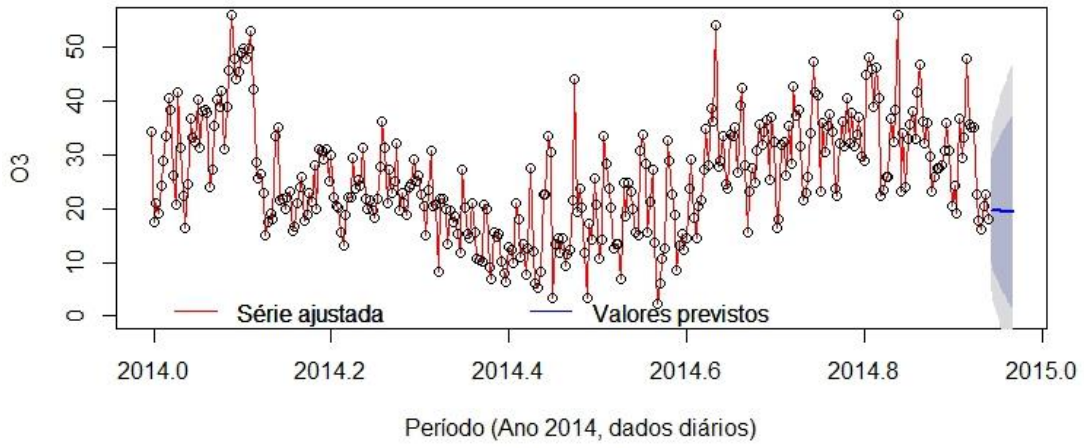
O valor da constante para o nível é de A = 0,5254 e, para a tendência beta C = 0,0001. Os valores previstos em comparação com os valores reais estão expostos na Tabela 10.

Tabela 10 – Valores previstos (h=20) do modelo de Alisamento exponencial biparamétrico (variável O3)

| **Período Dias** | **Valores previstos** | **Série real** | **Período Dias** | **Valores previstos** | **Série real** |
|---|---|---|---|---|---|
| 346 | 19,5743 | 27,6696 | 356 | 19,3285 | 32,4 |
| 347 | 19,5520 | 32,4 | 357 | 19,3061 | 20,6522 |
| 348 | 195296 | 32,4 | 358 | 19,2838 | 26,4217 |
| 349 | 19,5072 | 31,6762 | 359 | 19,2415 | 25,2435 |
| 350 | 19,4849 | 22,8833 | 360 | 19,2614 | 24,3217 |
| 351 | 19,4625 | 32,4 | 361 | 19,2392 | 15,0087 |
| 352 | 19,4402 | 20,6522 | 362 | 19,2681 | 17,1087 |
| 353 | 19,4179 | 20,6522 | 363 | 19,1944 | 19,6478 |
| 354 | 19,3955 | 20,6522 | 364 | 19,1721 | 25,8 |
| 355 | 19,3732 | 32,4 | 365 | 19,1497 | 19,3349 |

Na Figura 15, encontra-se a série ajustada e os valores previstos para os limites de previsão com intervalo de confiança de 95%.

Figura 15 – Modelo de alisamento biparamétrico ajustado a variável O3




Fonte: Autor.

A série representada pela cor vermelha é a série ajustada, as circunferências de cor preta, representa a série O3 e, em azul, o valor encontrado para as previsões futuras.

*4.3.2.3 Alisamento exponencial de Holt-Winters*

O método é adequado para séries que apresentam tendências e/ou sazonalidade. De acordo com as características da série, foi ajustado um modelo aditivo ou multiplicativo e, posteriormente, foi feita a comparação, utilizando os critérios de seleção.

No ajuste dos modelos, os valores das constantes de suavização que melhor se ajustaram ao modelo aditivo foram: A=0,6576 (constante de suavização do nível); C=0,0000 (constante de suavização da tendência) e D=0,8467 (constante de suavização da sazonalidade) e, para o multiplicativo, foi: A=0,4993 (constante de suavização do nível); C=0,0000 (constante de suavização da tendência) e D=0,3865 (constante de suavização da sazonalidade).

Os resultados das previsões do modelo aditivo e multiplicativo 20 passos à frente estão na Tabela 11, assim como os respectivos valores da série real correspondentes a cada período.

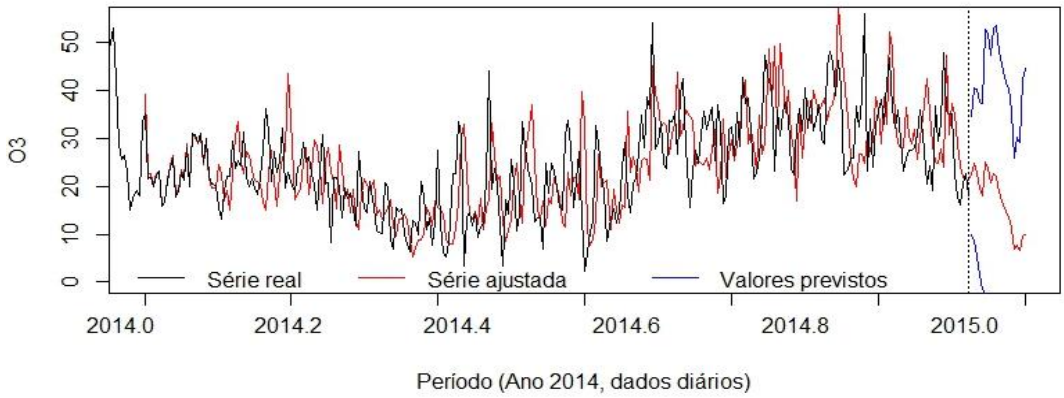
Tabela 11 – Valores Previstos (h=20) modelos Holt-Winters (variável O3)

| Período Dias | Modelo Aditivo | Modelo Multiplicativo | Série real | Período Dias | Modelo Aditivo | Modelo Multiplicativo | Série real |
|---|---|---|---|---|---|---|---|
| 346 | 21,2232 | 22,2510 | 27,6696 | 356 | 24,8862 | 18,9418 | 32,4 |
| 347 | 31,1759 | 24,7385 | 32,4 | 357 | 22,7884 | 17,0318 | 20,6522 |
| 348 | 29,8230 | 22,7836 | 32,4 | 358 | 15,9230 | 14,8120 | 26,4217 |
| 349 | 23,0855 | 19,4903 | 31,6762 | 359 | 12,3161 | 13,6622 | 25,2435 |
| 350 | 19,1876 | 17,8912 | 22,8833 | 360 | 13,0540 | 13,5230 | 24,3217 |
| 351 | 29,7178 | 25,1075 | 32,4 | 361 | 11,1227 | 11,1227 | 15,0087 |
| 352 | 28,3080 | 23,8071 | 20,6522 | 362 | 6,9275 | 7,5886 | 17,1087 |
| 353 | 23,3051 | 20,664 | 20,6522 | 363 | 6,6697 | 9,669 | 19,6478 |
| 354 | 28,6942 | 22,5708 | 20,6522 | 364 | 9,5953 | 10,5953 | 25,8 |
| 355 | 29,9092 | 22,0078 | 32,4 | 365 | 9,7629 | 9,7629 | 19,3349 |

Utilizando os critérios de seleção descritos na metodológia, o modelo aditivo ajustou-se de forma mais adequada à série histórica em relação ao modelo multiplicativo.

Na Figura 16, podem ser observados os valores reais e previstos, além do intervalo de previsão (h=20) do modelo Holt-Winters aditivo.

Figura 16 – Holt-Winters aditivo ajustado à variável O3




Fonte: Autor.

## 4.4 AJUSTE DE MODELOS DE PREVISÃO PARA A VARIÁVEL PM10

Seguindo a análise inicial dos grupos formados pela análise multivariada, foi ajustado o modelo de previsão para a variável PM10 devido. A ordem de modelos a ser seguida é mesma do item 4.3. As variáveis exógenas utilizadas foram as variáveis SO2, CO, selecionadas pelos modelos de análise multivariada.

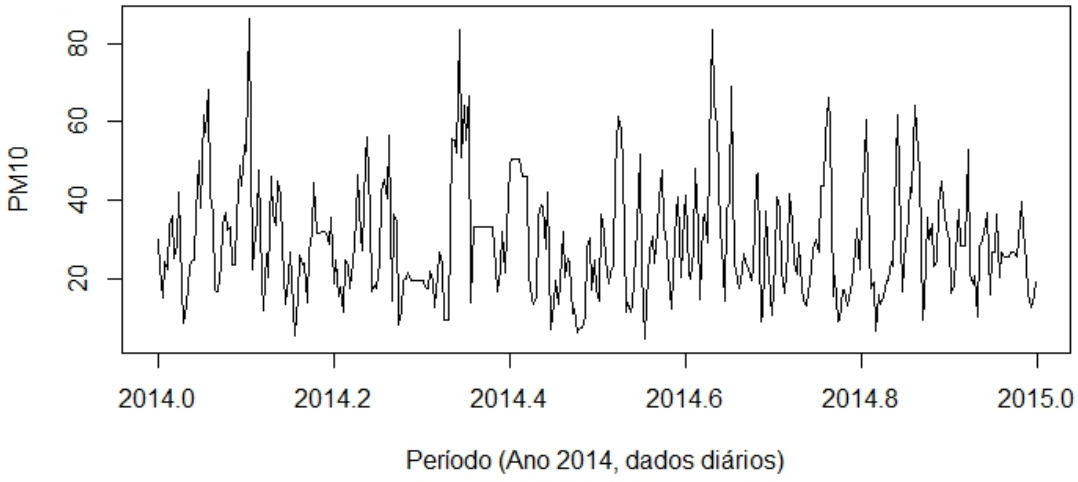
### 4.4.1 Modelos da classe Box-Jenkins

Nesta parte do trabalho, serão testados modelos não automáticos da classe Box-Jenkis, univariados e multivariados, com o intuito de verificar se as variáveis exógenas selecionadas pelos métodos de análise multivariada auxiliam para uma previsão de melhor qualidade da variável PM10.

#### *4.4.1.1 Ajuste do modelo ARIMA para a variável PM10*

O primeiro modelo de previsão a ser ajustado será o modelo ARIMA para variável PM10. Para dar início ao ajuste do modelo, observa-se, na Figura 17, o comportamento da série original, buscando identificar suas componentes, como tendência e/ou sazonalidade.

Figura 17 – Gráfico da série original PM10 em Canoas - RS



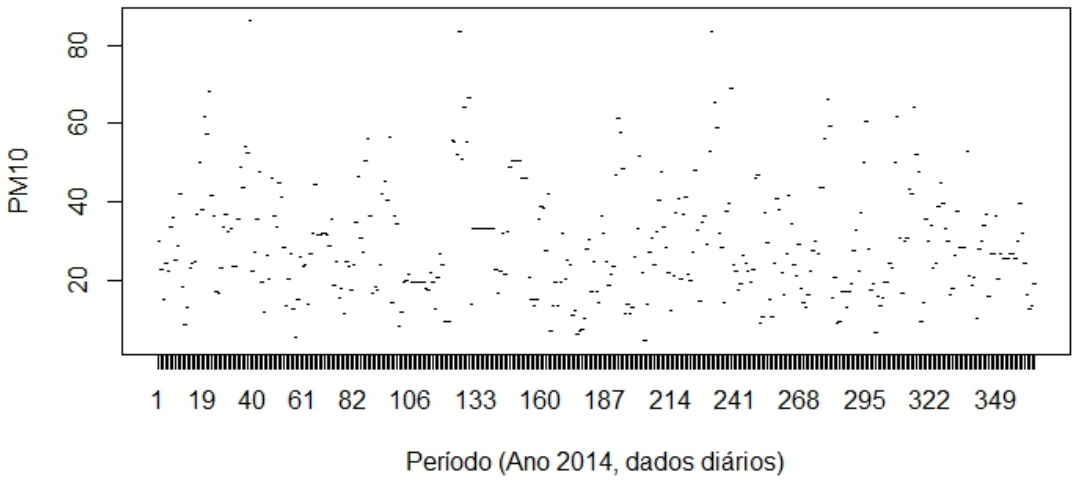

Analisando a Figura 17, verifica-se que a série não possui indícios de tendência, não sendo necessário aplicar diferenciações na série original; entretanto, para confirmar a análise gráfica, serão aplicados testes estatísticos para verificar a condição de estacionariedade e sazonalidade.

Após a etapa inicial de análise gráfica dos dados, foram realizados testes estatísticos para confirmar as hipóteses de estacionariedade e sazonalidade levantadas na análise gráfica. De acordo com o teste de sequências p-valor > 0,05, a série não possui indícios tendência significativa ao nível de $\alpha = 5\%$.

Para avaliar a sazonalidade, foi realizado o teste Kruskal-Wallis, não sendo detectado indício de sazonalidade ao nível de significância $\propto = 5$.

Confirmando o resultado do teste Kruskal-Wallis, o teste de Friedman também rejeitou a presença de sazonalidade na série em estudo.

Figura 18 – Gráfico da série PM10 em pontos



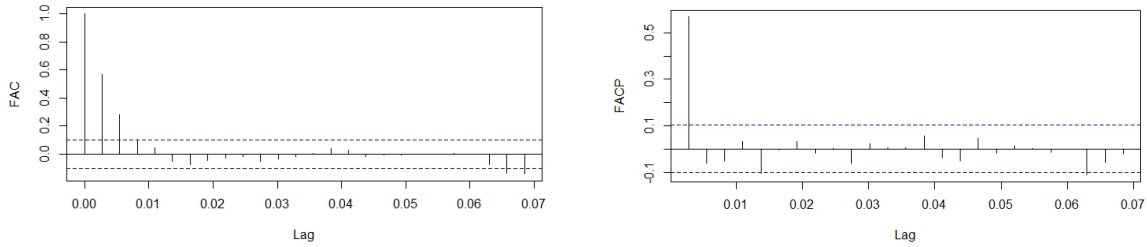

Fonte: Autor.

No gráfico da Figura 18, não se identificam períodos com menor ou maior concentração de PM10, sendo as observações heterogêneas durante o ano, ou seja, não há variabilidade para as diferentes estações climáticas, comprovando a análise de agrupamentos que não vincula o PM10 a varáveis climáticas.

Ao término da etapa de análise inicial da série PM10, não se identificou a presença de sazonalidade e tendência, a próxima etapa, então, é a construção dos gráficos de autocorrelação (FAC) e de autocorrelação parcial.

Os gráficos das funções de autocorrelação (FAC) e de autocorrelação parcial (FACP), expressos na Figura 19, demonstram as correlações seriais no domínio do tempo. Com base em suas frequências estimadas, é possível verificar sazonalidades e ciclos na série. Na Figura 19, apresentam-se os correlogramas das séries, sem diferença.

Figura 19 – Função de Autocorrelação (FAC) e Autocorrelação Parcial (FACP)




Fonte: Autor.

Ao se analisar os ACF e PACF, percebem-se características de um modelo estacionário, autorregressivo de ordem 1 e de médias móveis de ordem 1 ou 2. Os testes estatísticos e a análise gráfica indicam que um modelo do tipo ARMA deve ser considerado. Essa análise é sumária e serve para dar início ao processo de modelagem, embora vários modelos de ordem diferentes foram considerados e testados.

Na Tabela 12, estão expostos alguns modelos concorrentes ajustados para a série a PM10, com os respectivos valores de AIC e BIC, os parâmetros do modelo, a média e a variância dos resíduos referentes a cada modelo significativo.

Tabela 12 – Modelos concorrentes ARIMA e os respectivos critérios (variável PM10)

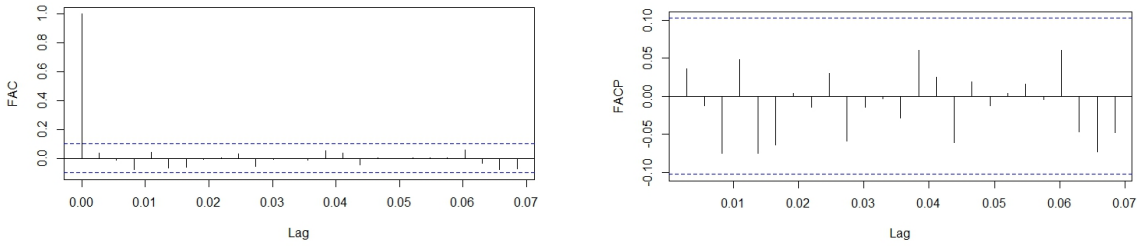
| **ARIMA (p,d,q)** | **Parâmetros** | **Critérios de seleção** | | **Média/Variância dos resíduos** |
|---|---|---|---|---|
| | | **AIC** | **BIC** | |
| (1,0,0) | $\phi_1 = 0{,}5861$<br>C = 29,6131 | 2850,11 | 2861,39 | $\mu \cong 0{,}0002$<br>$\sigma^2 = 43{,}014$ |
| (2,0,0) | $\phi_1 = 0{,}6022$<br>$\phi_2 = -0{,}0608$<br>c = 29,2913 | 2863,33 | 2868,93 | $\mu \cong 0{,}0022$<br>$\sigma^2 = 42{,}48$ |
| (1,1,1) | $\phi_1 = 0{,}5719$<br>$\emptyset_1$ =-0,9999 | 2852,35 | 2864,39 | $\mu \cong 0{,}0002$<br>$\sigma^2 = 42{,}48$ |

Considerando os três melhores modelos significativos ajustados, analisou-se a condição de ruído branco (média zero, variância constante, erros não correlacionados). Na análise dos resíduos dos modelos ARIMA, observou-se que a média dos modelos pode ser considerada como zero, conforme Tabela 12.

Os testes de autocorrelação dos resíduos Ljung-Box (p-valor = 0,4884) e Box-Pierce (p-valor = 0,4903) do melhor modelo AR (1), de acordo com os critérios de seleção AIC e BIC, indicam que os erros do modelo ajustado são descorrelacionados ao longo do tempo.

Analisando a Figura 20, gráfico de ACF e PACF dos resíduos, podemos perceber que não há autocorrelação significativamente diferente de zero em nenhuma defasagem, portanto os dados são independentes e demonstram um bom ajuste do modelo, confirmando os resultados dos testes de Ljung-Box e Box-Pierce.

Figura 20 – Função de autocorrelação (FAC) e autocorrelação parcial (FACP) dos resíduos




Fonte: Autor.

Considerando a análise residual e os critérios AIC e BIC, o modelo AR(1), com um parâmetro autorregressivo, atende a todas as condições de ruído branco e normalidade dos resíduos e concluem a última etapa da escolha do modelo para realizar previsões futuras.

Nessa etapa do trabalho, foram obtidas as previsões da série PM10 para um horizonte de 20 passos à frente, ou seja, dias que não foram considerados no ajuste do modelo.

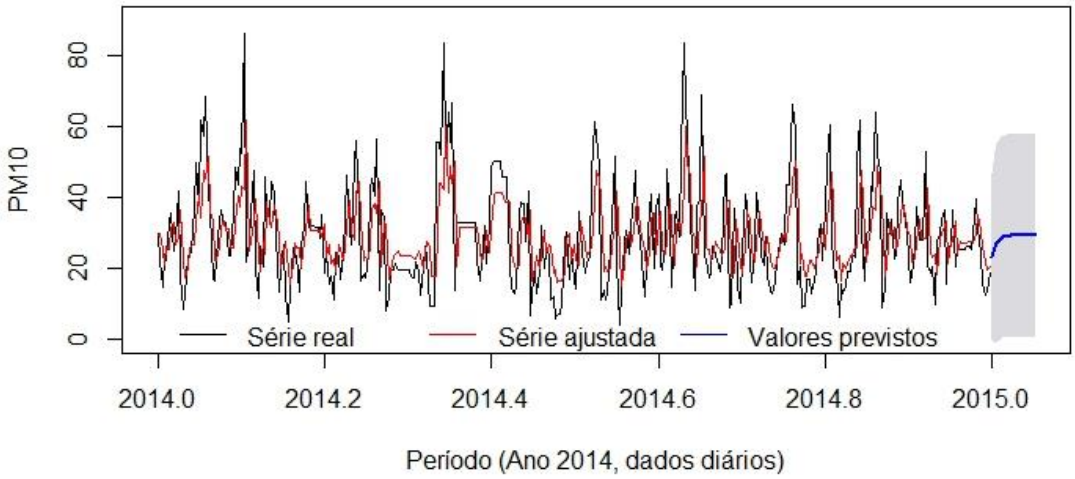
As previsões para o modelo ajustado AR(1) podem ser observadas na Tabela 13.

Tabela 13 – Valores previstos (h=20) segundo o modelo AR(1) (variável PM10)

| **Período Dias** | **Valores previstos** | **Série real** | **Período Dias** | **Valores previstos** | **Série real** |
|---|---|---|---|---|---|
| 346 | 33,5623 | 15,8478 | 356 | 29,6559 | 28,2435 |
| 347 | 31,8565 | 26,5609 | 357 | 29,6374 | 28,1856 |
| 348 | 30,8874 | 26,5609 | 358 | 29,6374 | 28,2875 |
| 349 | 30,3370 | 36,4952 | 359 | 29,6269 | 28,1686 |
| 350 | 30,3370 | 20,3666 | 360 | 29,6217 | 28,0212 |
| 351 | 30,0247 | 26,5608 | 361 | 29,6170 | 28,7018 |
| 352 | 29,8467 | 25,5 | 362 | 29,6115 | 28,4224 |
| 353 | 29,8467 | 25,5 | 363 | 29,6134 | 28,1264 |
| 354 | 29,7458 | 25,5 | 364 | 29,6133 | 27,9802 |
| 355 | 29,6885 | 26,5608 | 365 | 29,6132 | 28,0648 |

Na Figura 21, é demonstrada a série real, ajustada, e os valores previstos para os limites de previsão com intervalo de confiança de 95%.

Figura 21 – Modelo AR(1) ajustado à série observada (variável PM10)




Fonte: Autor.

*4.4.1.2 Ajuste do modelo ARIMAX*

Para o ajuste do modelo ARIMAX, a variável endógena a ser usada é a PM10 e as variáveis exógenas, serão utilizados o CO e o SO2, com o intuito de melhorar as previsões da variável em estudo. Observa-se que o modelo de análise multivariada não definiu variáveis climatológicas neste grupo pela falta de homogeneidade.

Dentre os diversos modelos testados, na Tabela 14, estão apresentados os três modelos com melhor desempenho nos critérios de seleção AIC e BIC.

Tabela 14 – Modelos concorrentes ARIMAX e os respectivos critérios

(continua)

| ARIMAX (p,d,q,r[]) | Parâmetros | Critérios de seleção AIC | BIC | Média/Variância dos resíduos |
|---|---|---|---|---|
| (1,[1,1]) | $\phi_1 = 0{,}5588$<br>C = 26,8307<br>SO2= 0,4484<br>CO= 34,38 | 2701,38 | 2720,60 | $\mu \cong 0{,}002$<br>$\sigma^2 = 43{,}31$ |
| (2,[1,1]) | $\phi_1 = 0{,}5762$<br>$\phi_2 = -0{,}0296$<br>C = 26,9305<br>SO2= 0,4557<br>CO= 32,7264 | 2703,10 | 2726,16 | $\mu \cong 0{,}0020$<br>$\sigma^2 = 49{,}19$ |

(conclusão)

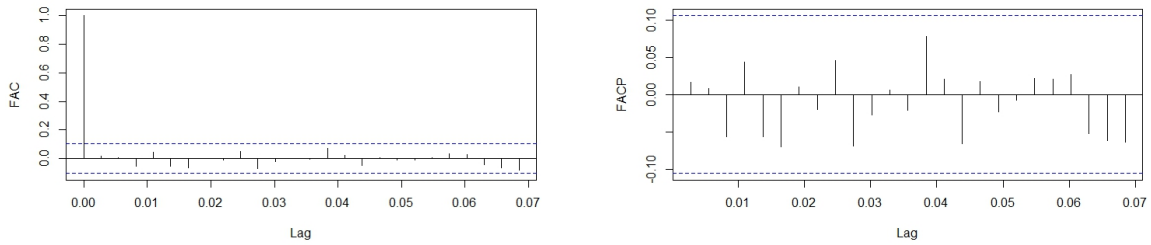
| | | | | |
|---|---|---|---|---|
| (1,0,1[1,1]) | $\phi_1 = 0{,}5289$<br>$\emptyset_1 = 0{,}0445$<br>C = 26,9303<br>SO2= 0,4544<br>CO= 33,002 | 2703,14 | 2726,2 | $\mu \cong 0{,}075$<br>$\sigma^2 = 43{,}21$ |

Considerando os três melhores modelos significativos ajustados, analisou-se a condição de ruído branco (média zero, variância constante, erros não correlacionados). Na análise dos resíduos dos modelos ARX, observou-se que a média dos modelos pode ser considerada como zero, conforme Tabela 14.

Os testes de autocorrelação dos resíduos Ljung-Box (p-valor = 0,7506) e Box-Pierce (p-valor = 0,7517), do melhor modelo ARX (1,[1,1]), de acordo com os critérios de seleção AIC e BIC, indicam que os erros do modelo ajustado são descorrelacionados ao longo do tempo.

Na análise gráfica (Figura 22), ACF e PACF dos resíduos do modelo ARX, percebe-se que não há autocorrelação significativamente diferente de zero em nenhuma defasagem, portanto, os dados são independentes e demonstram um bom ajuste do modelo, confirmando os resultados dos testes de Ljung-Box e Box-Pierce.

Figura 22 – Função de autocorrelação (FAC) e autocorrelação parcial (FACP) dos resíduos ARX (1,[1,1])




Fonte: Autor.

Considerando a análise residual e os critérios AIC e BIC, o modelo ARX (1,[1,1]), com um parâmetro autorregressivo (*p)* e um parâmetro das variáveis exógenas Monóxido de Carbono e Dióxido de Enxofre, o modelo atende a todas as condições de ruído branco e normalidade dos resíduos e concluem a última etapa da escolha do modelo para realizar previsões futuras.

Nessa etapa do trabalho, foram obtidas as previsões da série PM10 para um horizonte de 20 passos à frente, ou seja, dias que não foram considerados no ajuste do modelo.

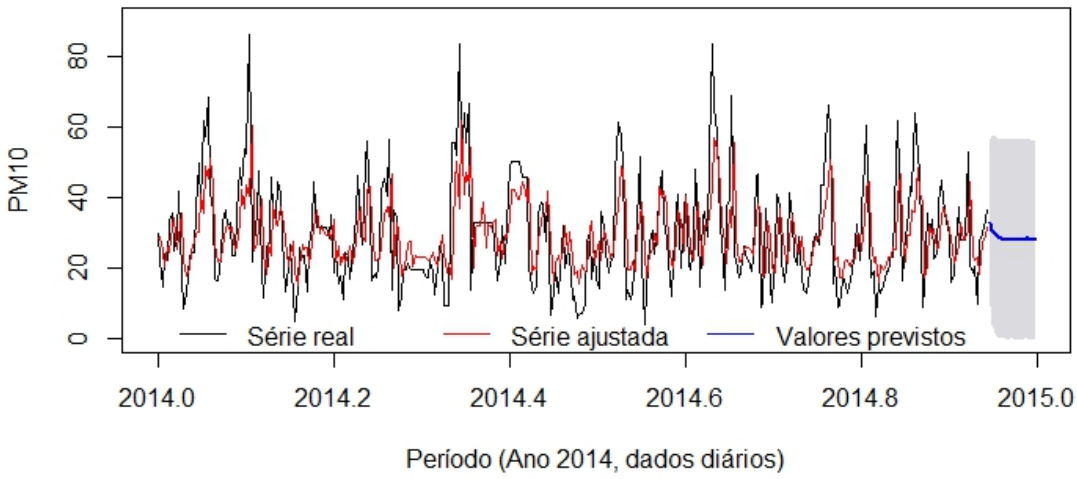
As previsões para o modelo ajustado ARX (1[1,1]) podem ser observadas na Tabela 15.

Tabela 15 – Valores previstos (h=20) modelo ARX (1,[1,1])

| Período Dias | Valores previstos | Série real | Período Dias | Valores previstos | Série real |
|---|---|---|---|---|---|
| 346 | 30,8640 | 15,8478 | 356 | 29,2710 | 28,2435 |
| 347 | 29,7015 | 26,5609 | 357 | 29,2797 | 28,1856 |
| 348 | 29,3233 | 26,5609 | 358 | 29,2847 | 28,2875 |
| 349 | 29,3233 | 36,4952 | 359 | 29,2875 | 28,1686 |
| 350 | 28,4414 | 20,3666 | 360 | 29,2891 | 28,0212 |
| 351 | 28,4863 | 26,5608 | 361 | 29,29 | 28,7018 |
| 352 | 28,3213 | 25,5 | 362 | 29,2906 | 28,4224 |
| 353 | 28,2580 | 25,5 | 363 | 29,2909 | 28,1264 |
| 354 | 28,2226 | 25,5 | 364 | 29,2911 | 27,9802 |
| 355 | 28,2545 | 26,5608 | 365 | 29,2912 | 28,0648 |

A Figura 23 demonstra a série ajustada e os valores previstos para os limites de previsão com intervalo de confiança de 95%.

Figura 23 – Modelo ARX (1,[1,1]) ajustado à série observada (variável PM10)




Fonte: Autor.

### 4.4.2 Ajuste de modelos da classe de alisamento exponencial para variável PM10

Nesta etapa do trabalho, buscar-se-á ajustar os modelos de: Alisamento exponencial simples, Alisamento exponencial biparamétrico e alisamento exponencial de Holt-Winters para a série de dados da variável PM10.

*4.4.2.1 Alisamento exponencial simples*

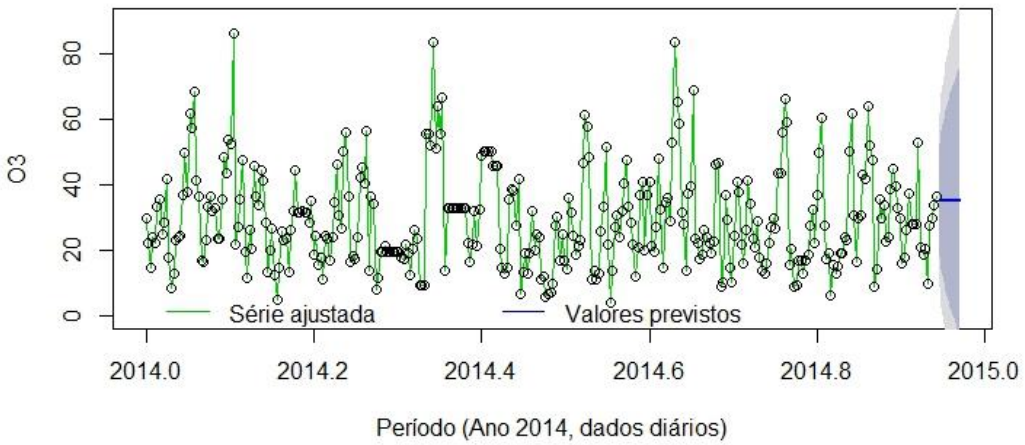
No ajuste do modelo de alisamento exponencial simples, encontrou-se o parâmetro de suavização alfa α = 0,7176. A previsão de todos os valores futuros da série é dada pela última constante de alisamento. Os valores previstos em comparação com os valores reais estão expostos na Tabela 16.

Tabela 16 – Valores previstos (h=20) do modelo de alisamento exponencial simples (variável PM10)

| Período Dias | Valores previstos | Série real | Período Dias | Valores previstos | Série real |
|---|---|---|---|---|---|
| 346 | 35,3896 | 15,8478 | 356 | 35,3896 | 28,2435 |
| 347 | 35,3896 | 26,5609 | 357 | 35,3896 | 28,1856 |
| 348 | 35,3896 | 26,5609 | 358 | 35,3896 | 28,2875 |
| 349 | 35,3896 | 36,4952 | 359 | 35,3896 | 28,1686 |
| 350 | 35,3896 | 20,3666 | 360 | 35,3896 | 28,0212 |
| 351 | 35,3896 | 26,5608 | 361 | 35,3896 | 28,7018 |
| 352 | 35,3896 | 25,5 | 362 | 35,3896 | 28,4224 |
| 353 | 35,3896 | 25,5 | 363 | 35,3896 | 28,1264 |
| 354 | 35,3896 | 25,5 | 364 | 35,3896 | 27,9802 |
| 355 | 35,3896 | 26,5608 | 365 | 35,3896 | 28,0648 |

Na Figura 24, é demonstrada a série ajustada e os valores previstos para os limites de previsão com intervalo de confiança de 95%.

Figura 24 – Modelo de alisamento Exponencial Simples (variável PM10)




Fonte: Autor.

A série ajustada representada pela cor verde, as circunferências de cor preta é a série PM10 e, em azul, o valor encontrado para as previsões futuras.

*4.4.2.2 Alisamento exponencial biparamétrico*

O modelo de alisamento exponencial biparamétrico é indicado para séries com tendência, permitindo usar constantes com diferentes valores para o nível e tendência. Mesmo não sendo detectada tendência na série, o modelo deve ser considerado, caso produza resultados satisfatórios.

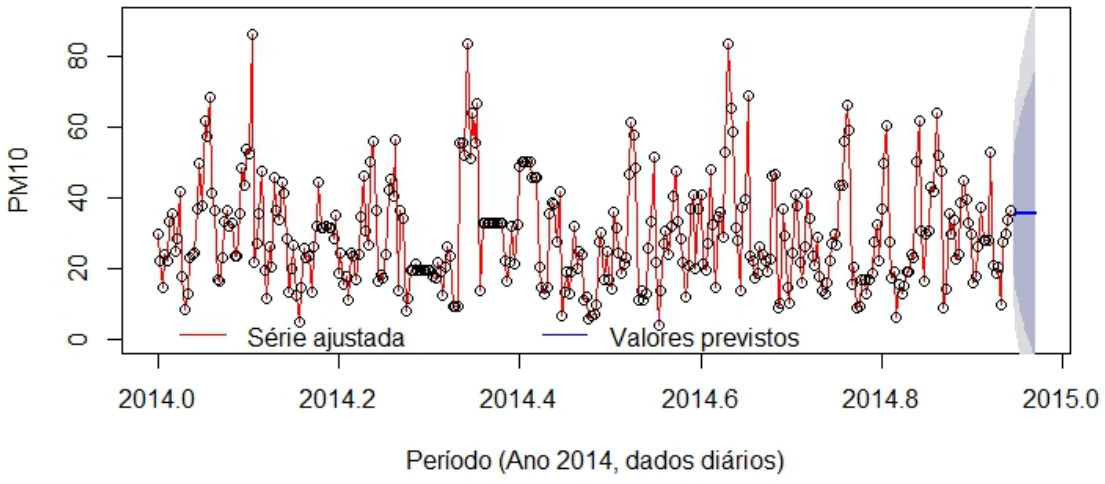
O valor da constante para o nível A = 0,7182 e para constante da tendência C = 0,00001. Conforme já havia sido testada, a série não possui tendência significativa, o que justifica o baixo valor da constante. Os valores previstos em comparação com os valores reais estão expostos na Tabela 17.

Tabela 17 – Valores previstos (h=20) do modelo de alisamento exponencial biparamétrico (variável PM10)

| **Período Dias** | **Valores previstos** | **Série real** | **Período Dias** | **Valores previstos** | **Série real** |
|---|---|---|---|---|---|
| 346 | 35,4206 | 15,8478 | 356 | 35,6654 | 28,2435 |
| 347 | 35,4451 | 26,5609 | 357 | 35,6899 | 28,1856 |
| 348 | 35,4696 | 26,5609 | 358 | 35,7144 | 28,2875 |
| 349 | 35,4940 | 36,4952 | 359 | 35,7389 | 28,1686 |
| 350 | 35,5185 | 20,3666 | 360 | 35,7634 | 28,0212 |
| 351 | 35,5430 | 26,5608 | 361 | 35,7878 | 28,7018 |
| 352 | 35,5675 | 25,5 | 362 | 35,8123 | 28,4224 |
| 353 | 35,5920 | 25,5 | 363 | 35,8368 | 28,1264 |
| 354 | 35,6165 | 25,5 | 364 | 35,8613 | 27,9802 |
| 355 | 35,6409 | 26,5608 | 365 | 35,8858 | 28,0648 |

Na Figura 25, encontra-se a série ajustada e os valores previstos para os limites de previsão com intervalo de confiança de 95%.

Figura 25 – Modelo de alisamento biparamétrico (variável PM10)




Fonte: Autor.

A cor vermelha representa a série ajustada, as circunferências de cor preta, a série PM10 e, em azul, o valor encontrado para as previsões futuras.

*4.4.2.3 Alisamento exponencial de Holt-Winters*

O método demonstra bom desempenho para séries com tendências e/ou sazonalidade. Para a série PM10, serão testadas as duas variações de ajuste desse modelo, o aditivo e o multiplicativo.

No ajuste dos modelos, os valores das constantes de suavização que melhor se ajustaram ao modelo aditivo foram: A=0,5447 (constante de suavização do nível); C=0,0000 (constante de suavização da tendência) e D=0,5980 (constante de suavização da sazonalidade) e, para o multiplicativo, foi: A=0,2725 (constante de suavização do nível); C=0,0000 (constante de suavização da tendência) e D=0,3650 (constante de suavização da sazonalidade).

Os resultados das previsões do modelo aditivo e multiplicativo 20 passos à frente estão na Tabela 18, assim como os respectivos valores da série real correspondente a cada período.

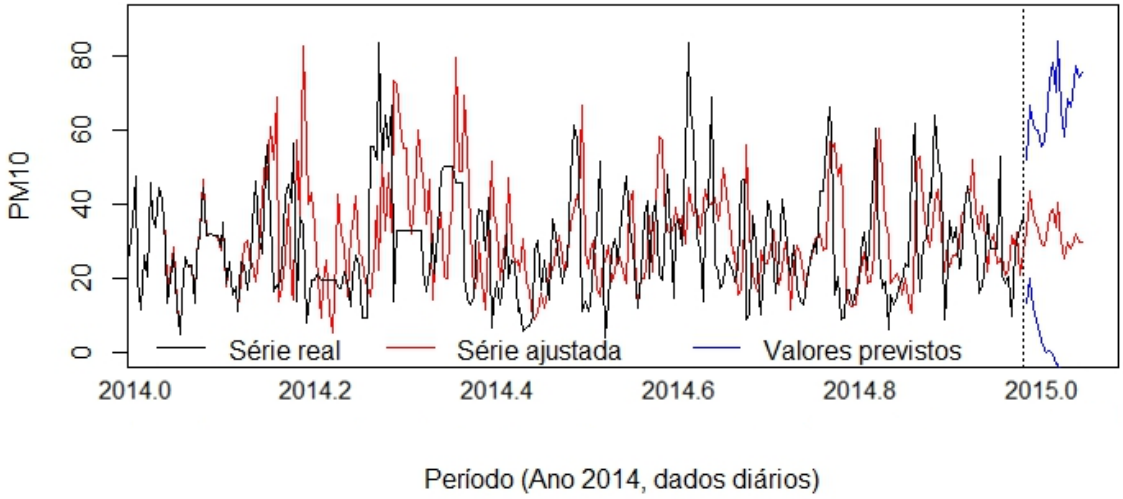
Tabela 18 – Valores Previstos (h=20) modelos Holt-Winters

| Período Dias | Modelo Aditivo | Modelo Multiplicativo | Série real | Período Dias | Modelo Aditivo | Modelo Multiplicativo | Série real |
|---|---|---|---|---|---|---|---|
| 346 | 36,4311 | 32,7009 | 15,8478 | 356 | 46,5064 | 33,7086 | 28,2435 |
| 347 | 47,4882 | 43,4444 | 26,5609 | 357 | 43,7696 | 40,4262 | 28,1856 |
| 348 | 43,9362 | 39,4902 | 26,5609 | 358 | 47,2163 | 30,1283 | 28,2875 |
| 349 | 44,6481 | 35,5683 | 36,4952 | 359 | 32,0132 | 25,4157 | 28,1686 |
| 350 | 45,8332 | 33,6424 | 20,3666 | 360 | 30,5539 | 29,8124 | 28,0212 |
| 351 | 47,5279 | 29,3325 | 26,5608 | 361 | 34,9582 | 28,0886 | 28,7018 |
| 352 | 50,5888 | 29,0958 | 25,5 | 362 | 30,2494 | 28,8398 | 28,4224 |
| 353 | 49,0787 | 31,0256 | 25,5 | 363 | 34,6589 | 32,0430 | 28,1264 |
| 354 | 48,9284 | 36,9654 | 25,5 | 364 | 39,9036 | 29,9797 | 27,9802 |
| 355 | 47,3024 | 38,9468 | 26,5608 | 365 | 38,7421 | 30,1085 | 28,0648 |

Utilizando os critérios de seleção MAPE, EQM, MAD e SSE, o modelo multiplicativo ajustou-se de forma mais adequada à série histórica em relação ao modelo aditivo.

Na Figura 26, podem ser observadas as séries com valores reais e previstos, além do intervalo de previsão (h=20) do modelo Holt-Winters multiplicativo.

Figura 26 – Holt-Winters multiplicativo (variável PM10)




Fonte: Autor.

## 4.5 COMPARAÇÃO DAS PREVISÕES DOS MODELOS AJUSTADOS

Um dos objetivos de uma modelagem de séries temporais é fazer previsões de valores futuros. Para tanto, se faz necessário verificar a eficiência e o bom ajuste do modelo. Conforme proposto na metodologia, foram reservadas as últimas vinte observações da série

original para verificar a qualidade do ajuste do modelo e comparar os modelos propostos dentro de cada classe considerada.

Os critérios a serem utilizados para comparação das previsões dos modelos propostos estão descritos na seção 2.4.2 da revisão bibliográfica. Na Tabela 19, estão expostos os resultados encontrados para os modelos que melhor se ajustaram à série de dados.

Tabela 19 – Comparação dos modelos utilizando os critérios de seleção

| **Modelo** | **Variável** | **MAPE(%)** | **EQM** | **MAD** | **SSE** |
|---|---|---|---|---|---|
| SARIMA(1,1,1)(0,1,1) | O3 | 20,26% | 34,50 | 5,87 | 651,71 |
| SARIMAX(1,1,1[1,1,1])(0,1,1) | O3 | 17,3% | 28,44 | 5,33 | 635,88 |
| ALS | O3 | 23,12% | 29,88 | 6,16 | 1207,90 |
| ALB | O3 | 24,73% | 30,81 | 5,93 | 1249,3 |
| HW Aditivo | O3 | 26,5% | 57,53 | 7,58 | 1387,6 |
| AR(1) | PM10 | 8,8% | 6,22 | 2,38 | 595,63 |
| ARX(1,[1,1]) | PM10 | 5,5% | 2,29 | 1,5 | 400,95 |
| ALS | PM10 | 24,49% | 54,60 | 7,38 | 1728,41 |
| ALB | PM10 | 28,01% | 61,63 | 7,85 | 1812,01 |
| HW Multipicativo | PM10 | 14,1% | 14,34 | 3,75 | 1417,30 |

Considerando os resultados dos critérios propostos, o modelo ARIMAX demonstra melhor desempenho em todos os critérios utilizados em relação aos modelos concorrentes ARIMA e da classe de alisamento exponencial, tanto para a série O3 como para PM10, demonstrando que as previsões se aproximam de forma mais adequada à série real em estudo quando se utilizam variáveis explicativas definidas por modelos de análise multivariada.

# 5 CONCLUSÃO

O objetivo desta pesquisa não foi apenas analisar a relação entre poluentes do ar e as variáveis meteorológicas, mas demonstrar a importância do estudo conjunto de modelos estatísticos como análise multivariada e séries temporais, suas principais implicações, vantagens e desvantagens.

Ao final deste estudo, observaram-se as principais relações entre os poluentes do ar e variáveis climatológicas com a formação de dois grupos: a variável O3 associa-se às variáveis climatológicas Isolação, Evaporação, Temperatura Mínima e Máxima (grupo 1), e as variáveis PM10, SO2 e CO formam o grupo 2 e não estão diretamente sendo influenciadas pelas variáveis climáticas analisadas. A análise de agrupamentos em conjunto com a análise fatorial possibilitou definir as variáveis exógenas a serem utilizadas nos modelos de previsão multivariados, demonstrando ser uma prática simples e eficiente para seleção de variáveis.

As classes de modelos de previsão utilizadas e comparadas para dados de poluição do ar, séries O3 e PM10 demonstraram desempenhos diferentes de previsão. Os modelos automáticos da classe de alisamento exponencial não se ajustaram adequadamente às séries em estudo, tanto para variável O3 quanto para a variável PM10. Dessa forma, não se recomenda o uso destes modelos para essas variáveis em estudo. Já os modelos não automáticos da classe ARIMA e ARIMAX demonstraram desempenho satisfatório às séries em estudo e podem ser empregados para as variáveis de poluentes do ar, desde que sejam seguidos todos os pressupostos exigidos pelos modelos.

Em comparação entre os melhores modelos encontrados para a série O3, SARIMA (1,1,1)(0,1,1) e o SARIMAX (1,1,1[1,1,1])(0,1,1) e para série PM10, AR(1) e ARX(1,[1,1]), nos dois casos, o modelo que permite a inclusão de variáveis exógenas ARIMAX demonstrou melhor desempenho de previsão. Dessa forma, percebe-se a importância do uso de modelos multivariados para selecionar as variáveis exógenas e também de modelos de séries temporais que permitam a inclusão de variáveis explicativas, sendo o uso dessas técnicas de fundamental importância para obtenção de previsões mais precisas.

A previsão de poluentes do ar pode auxiliar no planejamento de diversos órgãos públicos e privados ligados a setores como saúde, mobilidade urbana, educação, entre outros. Enfim, o alerta oportuno de altos índices de poluição permite planejamentos mais apropriados, principalmente para áreas sensíveis como a saúde. Medidas preventivas podem ser tomadas com base em informações seguras, como: restringir fluxo de veículos, manter crianças e idosos em locais fechados ou com uso de máscaras, fiscalizar os principais

geradores de poluição, entre outros pontos. Entretanto, é necessário que os dados coletados sejam precisos e em tempo hábil para se efetuar a modelagem necessária e para que se reproduzam cenários próximos ao real.

Diante dessas análises, destaca-se a ampla possibilidade de utilização destas e de outras técnicas adequadas de análise multivariada e séries temporais nas mais diversas áreas em que se tem por objetivo descrever um processo temporal ou quando se deseja fazer previsões, podendo ser utilizadas como ferramenta de apoio nas tomadas de decisão.

Em estudos futuros, pretende-se verificar a relação de poluentes do ar e variáveis meteorológicas de mais cidades, bem como empregar outros modelos de séries temporais.

# REFERÊNCIAS


AKAIKE, H. A Bayesian analysis of the minimum AIC procedure. **Annals of the Institute of Statistical Mathematics,** v. 30, p. 9-14, 1978.

_____. A new look at the statistical model identification. **IEEE Transactions on Automatic Control,** v. 19, p. 716-723, 1974.

ANDRADE, W. L. S. **Estimação de Modelos ARIMA/ARIMAX e Aplicação em Inferência de Propano.** 2009, 127f. Dissertação (Mestrado em Engenharia Elétrica e Computação), Universidade Federal do Rio Grande do Norte (UFRN), Natal 2009.

ASSIS, E. S. Estudo integrado da ilha de calor em áreas urbanas e sua contribuição ao planejamento: O caso de Belo Horizonte, MG. **Fórum Patrimônio, Mudanças climáticas e o impacto das cidades,** v. 4, n. 1, 2011.

BARBOSA, C. M.; CHAVES, C. A. Um estudo sobre gerenciamento da demanda nos sistemas de planejamento e controle da energia de produção. **VIII Congresso Nacional de Excelência em Gestão,** de 08 a 09 de junho, São Paulo, 2012.

BARCELLOS, C. et al. Mudanças climáticas e ambientais e as doenças infecciosas: cenários e incertezas para o Brasil. **Epidemiologia e Serviços de Saúde,** v. 18, n. 3, p. 285-304, 2009.

BATISTA, A. L. F. **Modelos de séries temporais e redes neurais na previsão de vazão.** 2009, 79f. Dissertação de Mestrado (Mestrado em Engenharia de Sistemas) – Universidade Federal de Lavras, Lavras, 2009.

BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. **Time series analysis:** forecasting and control, 3. ed. Prentice Hall: New Jersey, 1994, 589 p.

BRESSAN, A. A.; LIMA, J. E. Modelos de previsão de preços aplicados aos contratos futuros de boi gordo da BM&F. **Nova Economia,** Belo Horizonte, v. 12, n 1, p. 117-140, 2002.

BUENO, R. L. S. **Econometria de séries temporais.** São Paulo: Cengage Learning, 2008, 299 p.

CHECHI, L.; SANCHES, F. O. Análise de uma série temporal de precipitação para Erechim (RS) e um possível método de previsão climática. **Ambiência Guarapuava,** v. 9, p. 43-45, 2013.

DANFÁ, S. **Distribuição espacial da precipitação pluvial e sua erosividade para Guiné-Bissau.** 2009, 104f. Dissertação (Mestrado em Engenharia Agrícola) Universidade Federal de Lavras do Sul (UFLA), Lavras, 2009.

DAPPER, S. N. **Estudo da associação dos efeitos da poluição do ar na saúde da população de Canoas/RS.** 2016, 98f. Dissertação (Mestrado em Engenharia de Produção) Universidade Federal de Santa Maria (UFSM), Santa Maria, 2016.

ESQUIVEL, R. M.; SENNA, V.; GOMES. G. S. S. Análise espectral singular: comparação de previsão em séries temporais. **Revista do Mestrado em Administração e Desenvolvimento Empresarial,** v. 16, n. 2, p. 87-101, 2012.

FELICIANO, R. A. **Uma proposta de gerenciamento integrado de demanda e distribuição, utilizando sistemas de apoio à decisão (SAD) com Bussiness Intelligence (BI).** 2009, 179f. Dissertação (Mestrado em Engenharia) Politécnica Universidade de São Paulo, São Paulo, 2009.

FIGUEREDO, C. J. **Previsão de Séries Temporais Utilizando a Metodologia Box & Jenkins e Redes Neurais para Inicialização de Planejamento e Controle de Produção.** 2008, 174f. Dissertação (Mestrado em Métodos Numéricos em Engenharia) Universidade Federal do Paraná (UFPR), Curitiba, 2008.

FEPAM. **Fundação Estadual de Proteção Ambiental Henrique Luiz Roessler.** RS. Disponível em http://www.fepam.rs.gov.br/. Acesso em: 15 nov. 2016.

GERHARDT, E. J. et al. Contribuição da análise multivariada na classificação de sítios em povoamentos de araucária angustifolia (bert.) o. ktze., baseada nos fatores físicos e morfológicos do solo e no conteúdo de nutrientes da serapilheira. **Ciência Florestal,** Santa Maria, v. 11, n. 2, p. 41-57, 2001.

GIL, A. C. **Como Elaborar Projetos de Pesquisa,** 4. ed. São Paulo: Atlas, p. 41-56, 2002.

GIODA, A.; GIODA, F. R. A influência da qualidade do ar nas doenças respiratórias, **Revista Ambiente e Saúde,** v. 7, n. 1, p. 15-23, 2006.

GONÇALVES, T. L. F. et al. Modelagem dos processos de remoção sulfato e dióxido de enxofre presente no particulado em diferentes localidades da região metropolitana de São Paulo. **Revista Brasileira de Geofísica,** v. 28, p. 109-119. 2010.

IEMA. Instituto de Energia e Meio Ambiente. **1° Diagnóstico da rede de monitoramento da qualidade do ar no Brasil.** 2014. Disponível em: <http://www.forumclima.pr.gov.br/ arquivos/File/Rosana/Diagnostico_Qualidade_do_Ar_Ver sao_Final_Std.pdf>. Acesso em: 02 mar. 2017.

INMET. **Instituto Nacional de Meteorologia.** Disponível em: http://www.inmet.gov.br/. Acesso em: 15 nov. 2016.

KUBRUSLY, L. S.; SABOIA, J. Uma análise multivariada da população ocupada nas regiões metropolitanas brasileiras. **Ensaios FEE,** Porto Alegre, v. 27, n. 2, p. 411-436, 2006.

LEITE, R. C. M. et al. Utilização de regressão logística simples na verificação da qualidade do ar atmosférico de Uberlândia. **Engenharia Sanitária e Ambiental,** v. 16, n. 01, 2011.

LÍRIO, G. S. W.; SOUZA, A. M. A satisfação dos serviços pós-venda de clientes de um veículo de comunicação escrita, sob a ótica da Análise Multivariada. **Revista Ciência e Natura,** v. 30, n. 2, p. 21-41, 2008.

LIZZI, E. A. S. **Predição do número de casos mensal dos casos de dengue por modelos de séries temporais.** 2012, 131f. Dissertação (Mestrado em Saúde na Comunidade) Universidade de São Paulo (USP), Ribeirão Preto, 2012.

LÚCIO, P. S. et al. Um modelo estocástico combinado de previsão sazonal para a precipitação no Brasil. **Revista Brasileira de Meteorologia,** v. 25, n. 1, p. 70-87, 2010.

MARCHESAN, T. M.; SOUZA, A. M.; MENEZES, R. Avaliação do processo de ensino: uma abordagem multivariada. **Revista Produção,** v. 21, n. 2, p. 271-283, 2011.

MARTÍNEZ, R. O.; ZAMPROGNO, B. Comparação de algumas técnicas de previsão em análise de séries temporais. **Revista Colombiana de Estatística,** v. 26, n. 2, p. 129-157, 2003.

MONTE, E. Z.; ALBUQUERQUE, T. T. A.; REISEN, V. A. Previsão da concentração de ozônio na Região da Grande Vitória, Espírito Santo, Brasil, utilizando o modelo ARMAX-GARCH. **Revista Brasileira de Meteorologia,** v. 30, n. 3, 285-294, 2015.

MOREIRA, D. M.; TIRABASSI, T.; MORAES, M. R.; Meteorologia e Poluição Atmosférica. **Ambiente & Sociedade.** v. 11, n 1, 2008.

MORETTIN, P. A. **Econometria Financeira-Um curso em séries temporais financeiras.** Editora São Paulo: Edgard Blucher LTDA., 2008, 319 p.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de Séries Temporais.** São Paulo: Edgard Blucher LTDA., 2004, 535 p.

_____. **Previsão de Séries Temporais.** 2. ed. São Paulo: Editora Atual Editora LTDA., 1987, 436p.

NASCIMENTO, H. H. S. **Planejamento de mercado de energia elétrica no novo modelo do setor elétrico brasileiro pela aplicação de análise de componente principal como métodos de combinação de previsão.** 2013, 100f, Dissertação (Mestre em Engenharia Elétrica) Universidade Federal do Mato Grosso do Sul (UFMS), Campo Grande, 2013.

NETO, M. M. J. Estatística multivariada: Uma visão didática-metodológica. **Revista de Filosofia e Ensino,** maio, p. 1-13, 2004.

OZAKI, V. A.; ADMI, A. C. O. Modelagem estatística dos prêmios do seguro rural. **Revista Política Agrícola,** v. 21, n. 1, p. 60-75, 2012.

REISEN, V. A.; MOLINARES, F. A. A.; TEIXEIRA, E. C. Modelagem de Séries Temporais Sazonais na Presença de Outliers Estudo de Caso da Vazão Máxima Mensal do Rio Jucu, ES, Brasil. **Revista Brasileira de Recursos Hídricos,** v. 13, n. 2, 2008.

RINALDI, J. G. S.; MORABITO, S.; TACHIBANA, V. M. A importância da rapidez de atendimento em supermercados: um estudo de caso. **Revista Gestão e Produção,** v. 16, n. 1, p. 1-14, 2009.

SAAD, D. S. **Aplicação de técnicas estatísticas multivariadas em dados de cerâmica vermelha produzidas na região central do Rio Grande do Sul.** Dissertação 166f. Programa de Pós-Graduação em Engenharia Civil, Universidade Federal de Santa Maria (UFSM), Santa Maria, 2009.

SILVA, L. C. **Utilizando o ARMAX para estimar a influência do PIB, SELIC e Inflação no faturamento de empresas do setor siderúrgico com ações na bolsa.** Monografia 82f. Bacharelando (Administração), Universidade de Brasília (UNB), Brasília 2011.

SOUSA, W. R. N. **Estimativa de tendência de séries temporais de imagens de sensoriamento remoto para evolução de processos de desertificação.** 2012, 85f. Dissertação (Mestrado em Ciência da Computação) Universidade do Estado do Rio Grande do Norte (UERN)/Universidade Federal do Semiárido (UFERSA), Mossoró, 2012.

SOUTO, E. et al. Uma abordagem simples para a previsão do consumo de energia em redes de sensores sem fio. **25º Simpósio Brasileiro de Redes de Computadores e Sistemas Sem Fio,** Belém do Pará, de 28 maio a 01 junho de 2007.

SOUZA, A. M.; SILVA, F. M. Perfil dos alunos e dos Colégios Militares: um enfoque multivariado. **Revista Ciência e Natura,** v. 31, n. 2, p. 7-24, 2009.

SOUZA, A. M.; SCHIMIDT, A. S.; TURCATO, C. R. S. O aprendizado em organizações certificadas pela NBR ISO 9001:2000. **Revista Produção Online,** v. 11, n. 2, p. 289-318, 2011.

SOUZA, R. C.; CAMARGO, M. E. **Análise de Séries Temporais: Os Modelos ARIMA**, Ed. SEDIGRAF, 1996, 242p.

TIDRE, P. V.; BIASE, N. G.; SILVA, M. I. S. Utilização dos modelos de séries temporais na previsão do consumo mensal de energia elétrica da região norte do Brasil. **Revista Eletrônica Matemática e Estatística em Foco,** v. 1, n. 1, p. 57-63, 2013.

VERÍSSIMO, A. J. et al. Métodos estatísticos de suavização exponencial Holt-Winters para a previsão da demanda em uma empresa do setor metal mecânico. **Revista Gestão Industrial,** v. 8, n. 4, p. 154-171, 2012.

VICINI, L. **Análise multivariada da teoria à prática.** 215 p. Monografia (Especialização Estatística e Métodos Quantitativos) - Universidade Federal de Santa Maria, Santa Maria, 2005.

WERNER, L.; RIBEIRO, J. L. D. Previsão de Demanda: Uma Aplicação dos Modelos Box Jenkins na Área de Assistência Técnica de Computadores Pessoais. **Revista Gestão & Produção,** v. 10, n. 1, p. 47-67, 2003.