

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO - UFES
CENTRO TECNOLÓGICO - CT
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA AMBIENTAL - PPGEA

WESLEY ROCHA GRIPA

Modelo sazonal fracionário auto-regressivo integrado média móvel com volatilidade: Estudo de previsão das concentrações de Material Particulado Inalável

Vitória
2008

WESLEY ROCHA GRIPA

Modelo sazonal fracionário auto-regressivo integrado média móvel com volatilidade: Estudo de previsão das concentrações de Material Particulado Inalável

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Ambiental do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Mestre em Engenharia Ambiental, na área de concentração Modelagem Matemática da Dispersão de Poluentes Atmosféricos em Ambientes Urbanos utilizando modelos estocásticos.

Orientador: Prof. Dr. Valdério Anselmo Reisen.

Co-orientador: Prof. Dr. Neyval Costa Reis Jr.

Vitória
2008

Gxxx Grippa, Wesley Rocha

Modelo sazonal fracionário auto-regressivo integrado média móvel com volatilidade: Estudo de previsão das concentrações de Material Particulado Inalável / Wesley Rocha Grippa – 2008.
62f.

Orientadores: Valdério A. Reisen, Neyval C. Reis Jr.
Dissertação (mestrado) - Universidade Federal do Espírito Santo,
Centro Tecnológico.

1. Séries Temporais. 2. Qualidade do Ar. 3. Volatilidade.
4. Sazonalidade. 5. Filtro de Kalman. 6. Longa Dependência

WESLEY ROCHA GRIPA

Modelo sazonal fracionário auto-regressivo integrado média móvel com volatilidade: Estudo de previsão das concentrações de Material Particulado Inalável

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Ambiental do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Mestre em Engenharia Ambiental, na área de concentração poluição do ar.

Aprovada em 20 de agosto de 2008.

COMISSÃO EXAMINADORA

Prof. Dr. Valdério Anselmo Reisen - UFES

Prof. Dr. Neyval Costa Reis Jr. - UFES

Prof. Dra Jane Meri Santos - UFES

Prof. Dr. Marcelo Cunha Medeiros - PUC-RJ

*Dedico este trabalho,
a DEUS, pela força constante.
Aos meus pais pelo apoio incondicional nesta minha caminhada.
E a minha querida namorada, pelo amor e apoio constante.*

*“É melhor tentar e falhar,
que preocupar-se e ver a vida passar;
é melhor tentar, ainda que em vão,
que sentar-se fazendo nada até o final.
Eu prefiro na chuva caminhar,
que em dias tristes em casa me esconder.
Prefiro ser feliz, embora louco,
que em conformidade viver ...”*

Martin Luther King

Agradecimentos

Agradeço essa conquista, primeiramente a DEUS, pois Ele é o meu MESTRE.

Aos meus pais, Pedro e Lizete, que por confiarem em mim, me ajudaram afetivamente, emocionalmente e financeiramente. Agradeço também aos meus irmãos, Weder e Karoliny, pelas brincadeiras que me ajudaram esquecer um pouco dos estudos e ao mesmo tempo, me fazia voltar mais disposto a caminhada.

Agradeço também as minhas avós pelo amor e a meu vovô, que me viu pegar o longo caminho junto a todos e hoje se encontra com DEUS.

A minha querida namorada Sandrélia, que ao meu lado, me apoiou e incentivou constantemente.

Ao meu orientador Prof. Valdério Anselmo Reisen pelos ensinamentos, cobranças e paciência.

Ao meu co-orientador Prof. Neyval Costa Reis Júnior, extremamente importante nesse trabalho.

Aos professores examinadores, Prof^a. Jane Meri Santos e Prof. Marcelo Cunha Medeiros pelos elogios e críticas que muito acrescentaram à dissertação.

Agradeço a todos os professores e funcionários que tive no Programa de Pós Graduação em Engenharia Ambiental (PPGEA).

Aos colegas de estudo, em especial aos amigos Geovane, Alexsander, Valdir, Kennedy e Alexandre.

Resumo

Estudos com séries temporais têm sido amplamente utilizados em análises de dados ambientais. Este trabalho avalia as concentrações diárias de Material Particulado Inalável (PM_{10}) para a Região da Grande Vitória, ES, Brasil, para os anos de 2003 e 2004, utilizando processos SARFIMA-GARCH, modelos que capturam as propriedades de longa dependência, sazonalidade e volatilidade condicional variável no tempo. Para verificar a eficácia da metodologia, realizou-se estudos de comparação de predição e previsão a modelos usuais da literatura, e os resultados evidenciaram que o modelo proposto, *SARFIMA-GARCH* captou melhor variabilidade dos dados em termos do erro quadrático médio.

Palavras Chave: Séries Temporais. Material particulado. PM_{10} . Partículas inaláveis. Volatilidade. Sazonalidade. Filtro de Kalman. Longa Dependência.

Abstract

Studies on time series have been widely used to analyze environmental data. This study evaluates the daily concentration of inhalable particulate matter (PM_{10}) for the Region of Grande Vitória, ES, Brazil, for the years 2003 and 2004, using SARFIMA-GARCH processes, models that capture the properties of long-term dependence, seasonality and volatility Conditional variable over time. To verify the effectiveness of the methodology was carried out studies for the comparison of prediction and forecasting models to the usual literature, and the results showed that the proposed model, SARFIMA-GARCH best captured variability of the data in terms of average square error.

Keywords: Time Series. Particulate matter. PM_{10} . Inhalable particulate. Volatility. Seasonality. Kalman Filter. Long Memory.

Lista de Figuras

| | | |
|------|--|----|
| 5.1 | Série Temporal das concentrações de PM_{10} na RGV; (a) série com valores faltantes; (b) série completa. | 41 |
| 5.2 | Correlograma da Série Temporal PM_{10}^* | 41 |
| 5.3 | Correlograma de \hat{U}_t ; (a) FAC e (b) FACP. | 43 |
| 5.4 | Gráfico de $\hat{\epsilon}_t^2$, $t = 1, \dots, n$ do modelo inicial <i>SARMA</i> e <i>SARFIMA</i> | 44 |
| 5.5 | Estrutura de autocorrelação de $\hat{\epsilon}_t^2$, $t = 1, \dots, n$ do modelo inicial <i>SARMA</i> . | 44 |
| 5.6 | Estrutura de autocorrelação de $\hat{\epsilon}_t^2$, $t = 1, \dots, n$ do modelo inicial <i>SARFIMA</i> | 45 |
| 5.7 | Correlograma residual do modelo I | 47 |
| 5.8 | Análise residual do modelo I | 47 |
| 5.9 | Correlograma residual do modelo II | 48 |
| 5.10 | Análise residual do modelo II | 48 |
| 5.11 | Valores Observador e Preditos; (a) Modelo I, (b) Modelo II | 49 |
| 5.12 | Previsão das concentrações de PM_{10} no período de 01 a 31 de Janeiro de 2005 para os modelos I e II; | 51 |
| 5.13 | Observação \times Previsão; (a) Modelo I e (b) Modelo II | 54 |

Lista de Tabelas

| | | |
|-----|---|----|
| 5.1 | Estimativas de d e D para diferentes <i>bandwidths</i> | 42 |
| 5.2 | Estatística do modelo I | 45 |
| 5.3 | Estatística do modelo II | 45 |
| 5.4 | Estatística descritiva dos resíduos dos modelos ajustados | 46 |
| 5.5 | Testes estatísticos de normalidade* e não correlação** dos resíduos | 46 |
| 5.6 | Avaliação de desempenho dos modelos | 48 |
| 5.7 | Análise de desempenho das previsões dos modelos | 50 |
| 5.8 | Combinação dos pares (O, P) para 01 a 31 de janeiro de 2006 com o índice de qualidade correspondente (1-excelente; 5-péssimo) | 53 |

Sumário

| | |
|---|-----------|
| Resumo | 7 |
| Abstract | 8 |
| Lista de Figuras | 9 |
| Lista de Tabelas | 10 |
| 1 Introdução | 13 |
| 2 Objetivo | 19 |
| 2.1 Objetivo Geral | 19 |
| 2.2 Objetivos Específicos | 19 |
| 3 Revisão Bibliográfica | 21 |
| 4 Modelos e Propriedades | 26 |
| 4.1 Modelo SARFIMA e propriedades | 26 |
| 4.2 Modelo de Volatilidade | 33 |
| 4.3 Modelos Espaço de Estados | 36 |

| | | |
|----------|-----------------------------------|-----------|
| 5 | Análises e Resultados | 39 |
| 5.1 | Modelos ajustados | 40 |
| 5.2 | Estudo de previsão | 50 |
| 6 | Conclusões e Recomendações | 55 |
| 6.1 | Conclusão | 55 |
| 6.2 | Recomendações | 56 |
| | Referências Bibliográficas | 58 |

Capítulo 1

Introdução

O ar é um dos mais importantes recursos naturais do planeta Terra. Com o crescimento da industrial, iniciado no período da Revolução Industrial, a qualidade do ar veio diminuindo, causando diversos episódios de altas concentrações de poluentes no início do século XX. Metodologias para análises de séries temporais vieram se difundindo na literatura, uma grande contribuição para o início do século, foi a idéia de séries temporais estocásticas, postulando que cada série temporal pode ser considerada como a realização de um processo estocástico (Yule, 1927). Baseando-se nesta idéia, um conjunto de métodos de análise para séries temporais foram desenvolvidos, como a primeira formulação do conceito de modelos auto-regressivo (AR) e médias móveis (MA), dados por diversos trabalhos de Slutsky, Walker, Yaglom e Yule. A publicação de *Time Series Analysis: Forecasting and Control* por Box e Jenkins, (1970), integrou o conhecimento já existente, causando um enorme impacto sobre a teoria e a prática da moderna análise e previsão de série

temporais, e com a chegada do computador eles popularizaram o uso dos modelos autorregressivos integrados médias móveis (ARIMA) e suas extensões em diversas áreas da ciência, (Gooijer & Hyndman, 2006). A partir daí, surgiram diversas publicações de trabalhos com análises de séries temporais de dados ambientais. No âmbito da poluição atmosférica, pode-se citar diversos autores que usaram desta metodologia para analisar séries de poluição do ar como McCollister e Wilson (1975), Finzi e Tebaldi (1982), Terry, Lee e Kumar (1986), Robeson e Steyn (1990) e Shively (1990).

Segundo Perez & Reyes (2002), fatores como densidade populacional, frota veicular, atividades industriais e condições meteorológicas, entre outros, contribuem para o aumento da poluição atmosférica. Um poluente muito comum às regiões urbanas é o Material Particulado Inalável (PM_{10}), isto é, partículas com um diâmetro aerodinâmico menor que ou igual a $10\mu m$, que possuem como fontes principais a combustão de combustíveis fósseis e processos industriais. Frota veicular é outra fonte significativa de poeira que contém um alto índice de partículas PM_{10} , que por serem minúsculas, mantêm-se suspensas por várias horas.

Um dos padrões de qualidade do ar especificados pelas legislações de diversos países é a média de 24h (média diária) das concentrações de PM_{10} . A preocupação com essas partículas é devido a seu minúsculo tamanho, pequenas o bastante para penetrarem no trato respiratório dos seres humanos e por isso são potencialmente prejudiciais ao funcionamento normal do organismo. De acordo com a Resolução CONAMA nº003 de 28/06/1990, $150\mu g/m^3$ é o limite máximo de concentração para o PM_{10} , que prevê o mínimo efeito

adverso sobre o bem-estar da população (IBAMA, 1992).

A exposição a concentrações elevadas de PM_{10} na atmosfera associa-se fortemente a diversos problemas de saúde. Estudos realizados por [Ostro et al. \(1996\)](#) em Santiago, Chile, mostraram através da análise de séries temporais uma associação significativa entre PM_{10} e problemas respiratórios em grupos de crianças. Especificamente, os autores detectaram um aumento de 7% no número de internações em grupos de crianças com idade de 2 a 15 anos para um aumento de $50\mu\text{g}/\text{m}^3$ na concentração de PM_{10} , com defasagem de 5 dias. Estudos similares conduzidos na Califórnia, EUA ([Ostro, Hurley & Lipsett, 1998b](#)) e Bancoque, Tailândia, ([Ostro et al., 1999](#)) indicaram uma associação estatística significativa entre PM_{10} e mortalidade, sendo que um incremento de $10\mu\text{g}/\text{m}^3$ na concentração diária de PM_{10} é associada com cerca de 1% no aumento de mortalidade total e cerca de 3% para doenças respiratórias e cardiovasculares. [Almeida \(2006\)](#) em estudo semelhante para Cariacica, ES, Brasil, detectou significativa correlação entre o PM_{10} e número de atendimentos hospitalares por causas respiratórias no grupo de crianças de 0 a 6 anos de idade. [Souza Jr. et al. \(2001\)](#), de 1995 a 1998, efetuaram um estudo de caracterização do PM_{10} da Região da Grande Vitória, ES, Brasil, realizados com filtros de monitoramento. As contribuições encontradas para as diversas fontes são, a contribuição industrial na qualidade do ar representa 34,6%, as atividades humanas (emissão veicular, queimadas, construção civil) 54,6%, e as emissões naturais cerca de 10,8%.

Várias estratégias de modelagens têm sido desenvolvidas e otimizadas, no intuito de gerar previsões satisfatórias das concentrações dos poluentes em diversos cenários. A utilização

de modelos de previsão é um artifício importante para conhecer os diversos comportamentos e características de determinados poluentes, podendo dessa forma, prever possíveis picos de concentração. Para isso, existem duas classes de modelos, os experimentais e os matemáticos. Nesse último, tem-se os modelos determinísticos e os modelos estocásticos. Este estudo se concentrará na classe de modelagem estocástica.

Sazonalidade, longa dependência e volatilidade são características e geralmente encontradas em séries de dados ambientais, características estas presentes pelos inúmeros ciclos que encontramos no meio ambiente, portanto essas propriedades devem ser consideradas para que modelos possam prever satisfatoriamente as concentrações esperadas dos poluentes em estudo. A sazonalidade é um fenômeno geralmente intra-anual que se repete com certa regularidade em determinado período. Longa dependência é caracterizada por autocorrelações significantes entre observações distantes. Essas propriedades têm sido bem exploradas de maneira teórica e prática em [Doukhan, Oppenheim & Taqqu, \(2003\)](#). A propriedade de volatilidade é caracterizada pela não-constância da variância, ou seja, a variância do processo é função do tempo. No âmbito da poluição do ar, as características de sazonalidade e longa dependência presentes nessas séries ocorrem geralmente para o período de 7 dias (sazonalidade semanal), pois durante os dias úteis existe uma grande movimentação de veículos e intensa atividade industrial, havendo uma acumulação contínua de poluentes, fato observado em menor volume para os finais de semana. A propriedade de volatilidade caracteriza a aleatoriedade nas séries de dados da poluição do ar. Essas propriedades são exploradas em [Reisen, Rodrigues & Palma \(2006a,b\)](#), [Engle \(1982\)](#), [Jor-](#)

quera et al. (1998), Chelani & Devotta (2005), Zolghadri & Cazaurang (2005), Goyal, Chan & Jaiswal (2006), Teyssière & Kirman (2006) entre outros.

Recentemente, análises de séries temporais com longa dependência têm sido alvo de muitos estudos, e diversos métodos de estimação para o parâmetro de memória longa foram propostos. Processos com essa propriedade tiveram estudos iniciados por Hurst (1951) em análises a série de níveis mínimos de água no Rio Nilo. Posteriormente, Granger e Joyeux (1980) e Hosking (1981) propuseram uma generalização do ARIMA para lidar com longa dependência. As propriedades desses estimadores de longa dependência vêm sendo ostensivamente avaliadas em diversas situações, como por exemplo, na presença de erros não Gaussianos e *outliers*, ver por exemplo Sena, Reisen & Lopes (2006), Jensen (1999) e Whitcher (2004). Em séries temporais com longa dependência e sazonalidade, trabalhos de Porter-Hudak (1990) e Hassler (1994) propõem o método Geweke & Porter-Hudak (1983) para estimação dos processos *ARFIMA* sazonal. Uma generalização desses modelos sazonais de longa memória são os modelos *ARUMA* e o *GARMA*, respectivamente, propostos por Giraitis & Leipus (1995) e Woodward, Cheng & Gray (1998). Reisen, Rodrigues & Palma (2006a,b) apresentaram estudos do modelo *ARFIMA* sazonal que é um caso particular do modelos *ARUMA/GARMA* e sugeriram estimadores ao parâmetro de memória. Estudos empíricos, considerados pelos autores, evidenciaram a eficácia dos estimadores quando comparados aos métodos já existentes na literatura. Reisen et al. (2008) considera modelos fracionários com sazonalidade múltipla, mostrando propriedades teóricas e assintóticas dos estimadores dos parâmetros de memória do processo.

Séries temporais provenientes de monitoramentos da qualidade do ar possuem frequentemente a presença de observações faltantes. Essas observações faltantes ocorrem geralmente por falha na medição ou defeito nos equipamentos de medida. A série das concentrações diárias de PM_{10} usada neste estudo possui essas características. Para estimação das observações faltantes, foi usada uma modificação do Filtro de Kalman proposta por [Palma & Chan \(1997\)](#). [Palma & Del Pino \(1999\)](#) e [Iglesias, Jorqueira & Palma \(2005\)](#) são exemplos de pesquisas relacionadas à estimação dessas observações faltantes.

A sazonalidade é uma importante propriedade que deve ser explorada em estudos de séries ambientais. Trabalhos envolvendo previsões em séries de concentração de PM_{10} como [Stadlober, Hörmann & Pfeiler \(2008\)](#) incorpora a componente sazonal através do uso de variáveis binárias ou dicotômicas (*dummy*) enquanto [Goyal, Chan & Jaiswal \(2006\)](#) não incorpora ao modelo de previsão a influência sazonal. Neste trabalho, uma inovação na análise temporal, é inclusão de uma componente sazonal multiplicativa fracionária, como explorado de forma teórica em [Reisen et al. \(2008\)](#).

O trabalho está dividido como se segue. A próxima Seção, mostra os objetivos do trabalho. A Seção 3 apresenta uma revisão bibliográfica do estudo, a Seção 4 descreve a metodologia de modelagem e propriedades dos modelos. A Seção 5 apresenta as análises e os resultados de modelagem e de estimação. As conclusões da pesquisa e recomendações para trabalhos futuros estão na Seção 6.

Objetivo

2.1 Objetivo Geral

O objetivo deste trabalho é utilizar a modelagem de séries temporais, com dois parâmetros de longa dependência e erros heterocedásticos na análise de séries temporais das concentrações médias diárias de PM_{10} , através do processo Sazonal Auto-regressivo Integrado Fracionário Média Móvel (*SARFIMA*), aplicado-o às concentrações de material particulado inalável (PM_{10}) do município de Cariacica.

2.2 Objetivos Específicos

- Estudar a estrutura de correlação da série de PM_{10} ;

- Estimar as observações faltantes presentes na série através do Filtro de Kalman;
- Avaliar as propriedades de sazonalidade, longa dependência e volatilidade, características presentes na série das concentrações de material particulado inalável;
- Comparar os valores preditos e previstos do modelo *SARFIMA* aos do modelo *SARMA* (Box & Jenkins), ambos com erros heterocedásticos e ajustados a série de material particulado inalável.

Revisão Bibliográfica

Diversos estudos têm utilizado ferramentas estatísticas para explicar a variabilidade e influência à saúde dos principais poluentes atmosféricos. Importantes estudos envolvendo essas questões estão listados abaixo.

[Palma & Del Pino \(1999\)](#) discutem questões teóricas e metodológicas para predição em modelos de longa dependência com dados incompletos. Os autores investigaram a influência das observações faltantes nos erros de predição para processos com curta e longa dependência e a técnica é ilustrada com uma análise do nível mínimo de água do Rio Nilo, série com forte dependência. Os resultados da avaliação da influência dos valores faltantes na predição, indicam um contraste nítido entre os modelos ARMA e ARFIMA, essa influência desaparece no erro quadrático médio predito a uma taxa exponencial para o ARMA e a uma taxa hiperbólica para o ARFIMA, persistindo neste ao longo do tempo.

A variância do erro de previsão cresce mais lentamente para o modelo ARFIMA do que para o ARMA, dando ao processo de longa dependência uma clara vantagem diante dos processos de curta dependência. Os dados do nível anual mínimo de água do rio Nilo, foram usados para o período de 622 à 1921, série explorado por Hosking. O impacto dos valores faltantes foi avaliado para três combinações. Primeiro para a série de 622 à 1281, com 7% de valores faltantes, a segunda com a série de 1282 à 1921, com 46% de valores faltantes e a terceira, com toda a série e 27% de dados faltantes. Os resultados da avaliação mostraram que a estimação do parâmetro de longa dependência (d) por máxima verossimilhança foi similar para as três séries.

A concentração de CO₂ na atmosfera é uma variável importante para os principais modelos de simulação ambiental e é considerado um fundamental fator para a predição da temperatura e mudanças climáticas. No entanto, a variância condicional dos níveis de CO₂ não tem sido previamente examinado. [McAleer & Chan \(2005\)](#) analisaram a tendência e a volatilidade das concentrações de CO₂ usando dados de janeiro de 1965 a dezembro de 2002 do conhecido conjunto de dados Mauno Lao. A série apresenta um padrão sazonal, tendência positiva ao longo do tempo e uma estrutura de autocorrelação persistente. Foi estimado diversos modelos da classe GARCH e GJR, este último usado geralmente para conjunto de dados financeiros. Os autores verificaram que a modelagem da variância condicional das concentrações de CO₂ permitiu previsões um passo a frente com intervalos de confiança mais exatos. Para um estudo de previsão, os autores reestimaram os parâmetros dos modelos para a série de janeiro de 1965 a dezembro de 2001 e verifica-

ram o desempenho de previsão dos modelos para janeiro de 2002 a dezembro de 2002. Os critérios avaliados foram a Raiz do Erro Quadrático Médio (RMSE), o Erro Médio Absoluto (MAE) e o Erro Médio Absoluto Percentual (MAPE), que mostraram o modelo EGARCH(1) como melhores resultados.

Iglesias, Jorquera & Palma (2005) propõem uma metodologia estatística para manipular regressões com longa dependência nos erros e valores faltantes. A estratégia de estimação é desenvolvida com uma abordagem clássica e Bayesiana e é ilustrada com uma aplicação em um conjunto de dados de poluição do ar da cidade de Santiago, Chile, para período de 01 de janeiro de 1989 a 31 de dezembro de 1996 (*Ambient Air Quality Monitoring Network*), com um número recorde de 531 dias sem observação. Para correção das observações faltantes, foi utilizado o Filtro de Kalman. Para explicar a variação nas concentrações de PM_{10} , os autores utilizaram como variáveis explanatórias da regressão, a velocidade do vento, a precipitação, as concentrações de CO_2 e SO_2 , variáveis estas que mostraram-se correlacionadas com a concentração do PM_{10} . A aplicação da metodologia aos dados reais, com a abordagem clássica, mostrou que a inferência pode ser distorcida se a longa dependência nos erros não for considerada.

Changa et al. (2005) realizaram um estudo na cidade de Tapey (Taiwan) no período compreendido entre 1997 a 2001 para verificar a associação entre os níveis dos poluentes (PM_{10} , SO_2 , NO_2 , CO e O_3) para variações interquartílicas e os atendimentos por doenças cardiovascular na população de Tapey. Os resultados encontrados relatam que para os períodos mais quentes, ou seja, temperatura acima de $20^{\circ}C$, associações estatisticamente

significativas foram encontradas entre os níveis de concentração dos poluentes PM_{10} (24.51 $\mu g/m^3$), NO_2 (9.95 ppb), CO (0.49 ppm) e O_3 (9.95 ppb), evidenciando um aumento nas admissões hospitalares por patologia cardiovascular na ordem de 8.5%, 17.7%, 9% e 18.9% respectivamente. Nos dias com temperatura abaixo de $20^\circ C$ todos os poluentes se mostraram associados aos aumentos no número de atendimentos por doenças cardiovascular, exceto O_3 e SO_2 , sendo verificados aumentos de 14.20%, 11.2% e 7.3% para um variação interquartílica dos poluentes PM_{10} , NO_2 e O_3 respectivamente, evidenciando que elevados índices dos poluentes atmosféricos, aumentam o risco de admissões hospitalares.

Goyal, Chan & Jaiswal (2006) aplicaram três modelos estatísticos à média diária da concentração de PM_{10} nas cidades de Delhi e Hong Kong. O modelo 1 é uma regressão linear múltipla dos parâmetros meteorológicos. O modelo 2 é um modelos de série temporal ARIMA e o modelo 3 é uma combinação dos dois primeiros modelos. O objetivo dos autores era desenvolver um modelo estatístico de previsão, confiável, da concentrações de PM_{10} para as cidades de Delhi e Hong Kong e promover um estudo comparativo através do desempenho dos modelos. Os dados utilizados na modelagem foram material particulado inalável ($\mu g/m^3$), velocidade do vento (km/h), temperatura ($^\circ C$), radiação solar (MJ/m^2h) e umidade relativa do ar (%) medidos no período de julho de 2000 a junho de 2001, para a duas cidades. Na comparação de desempenho dos modelos, as medidas de erro mostraram que o modelo 3 captou melhor a variabilidade dos conjuntos de dados e por isso foi utilizado no estudo de previsão. O estudo de previsão ocorreu apenas para a cidade de Delhi, para o período de julho de 2001 a junho de 2002 e o resultado da pre-

visão mostrou-se satisfatório com um erro absoluto médio de apenas 25% entre os valores observados e a previstos.

Stadlober, Hormann & Pfeiler (2008) realizaram um estudo de previsão para as cidades de Bolzano, Klagenfurt e Graz em períodos compreendidos entre 2001 e 2006. Este trabalho aplicou um modelo de regressão linear múltipla, prevendo as concentrações de PM_{10} através da velocidade do vento, variação da temperatura, precipitação e variáveis *dummy* para controle da sazonalidade. Para refletir a qualidade das previsões, os autores definiram uma função qualidade baseada no erro de previsão que classifica o valor previsto como “excelente”, “bom”, “satisfatório”, “ruim” e “péssimo”. Os resultados mostraram que a função de qualidade para previsão proporcionou resultados adequados, com bom desempenho. As previsões para as cidades de Bolzano e Graz, indicaram, respectivamente, 86% e 97% dos valores previstos classificados na três primeiras categorias.

Os trabalhos citados mostram a importância de se conhecer e modelar o comportamento da série da concentração de PM_{10} , para que sejam evidenciados seus efeitos na saúde. O presente trabalho utiliza técnicas de séries temporais para modelar o comportamento e prever a concentração de PM_{10} em ambientes urbanos, contribuindo na elaboração de políticas ambientais e de saúde.

Modelos e Propriedades

4.1 Modelo SARFIMA e propriedades

O processo $X_t \equiv \{X_t\}_{t \in \mathbb{Z}}$ é definido como $SARFIMA(p, d, q) \times (P, D, Q)_s$ (Auto-regressivo Integrado Fracionário Média Móvel Sazonal), se satisfaz a equação de diferença

$$X_t = \nabla^{-d} U_t, \quad (4.1)$$

onde $U_t = \frac{\Theta(B^s)\theta(B)}{\Phi(B^s)\phi(B)}\epsilon_t$, $\Phi(z^s)\phi(z) \neq 0$, $z \in \mathbb{C}$, é um processo $SARMA(p, q) \times (P, Q)_s$, $\{\epsilon_t\}$ é um processo ruído branco com $E(\epsilon_t) = 0$ e $Var(\epsilon_t) = \sigma_\epsilon^2$, B é o operador de retardo ($B^k X_t = X_{t-k}$, $k \in \mathbb{N}$), $\Phi(z^s)$, $\Theta(z^s)$, $\phi(z)$ e $\theta(z)$ são polinômios de ordem P , Q , p , $q \in \mathbb{N}$, respectivamente, com raízes fora do círculo unitário, $\{z : |z| \leq 1\}$, e sem raízes comuns,

representados nas formas polinomiais

$$\Phi(z^s) = 1 - \Phi_1 z^s - \Phi_2 z^{2s} - \dots - \Phi_P z^{Ps},$$

$$\Theta(z^s) = 1 - \Theta_1 z^s - \Theta_2 z^{2s} - \dots - \Theta_Q z^{Qs},$$

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p,$$

$$\theta(z) = 1 - \theta_1 z - \theta_2 z^2 - \dots - \theta_q z^q,$$

com as constantes reais Φ_i ($i = 1, 2, \dots, P$), Θ_l , ($l = 1, 2, \dots, Q$), ϕ_j ($j = 1, 2, \dots, p$) e θ_g ($g = 1, 2, \dots, q$). $\mathbf{d} = (d, D)$ é o vetor dos parâmetros $d, D \in \mathbb{R}$ que são, respectivamente, os graus de diferenciação fracionária e fracionária sazonal de período s .

A definição do modelo (4.1) é uma generalização de modelos de Box & Jenkins e foi apresentada por Peiris & Sings (1996). O filtro linear $\nabla^{-\mathbf{d}}$ (Eq. (4.1)) pode ser escrito como

$$\nabla^{\mathbf{d}} \equiv (1 - B)^d (1 - B^s)^D = \prod_{j=0}^{\lfloor \frac{s}{2} \rfloor} [(1 - Be^{iw_j})(1 - Be^{-iw_j})]^{d_j}, \quad (4.2)$$

onde $w_j = \frac{2\pi j}{s}$, $j = 1, 2, \dots, \lfloor \frac{s}{2} \rfloor$,

$$d_j = \begin{cases} \frac{d+D}{2}, & j = 0 \ (w_0 = 0) \\ D, & j = 1, \dots, \lfloor \frac{s}{2} \rfloor - 1 \ (0 < w_j < \pi) \\ \frac{D}{2}, & j = \lfloor \frac{s}{2} \rfloor \ (w_{\lfloor \frac{s}{2} \rfloor} = \pi, \text{ se } s \text{ for par}) \end{cases},$$

sendo $\lfloor \cdot \rfloor$ a função maior inteiro.

Baseado em (4.2), Reisen et al. (2008) mostrou que o modelo (4.1) é estacionário e invertível, se e somente se, $|d + D| < \frac{1}{2}$ e $|d|, |D| < \frac{1}{2}$ o processo possui características de longa dependência quando $d + D > 0$.

O filtro linear $\nabla^{\mathbf{d}}$ (Eq. (4.1)), pode ser representado pela expansão infinita

$$\nabla^{\mathbf{d}} \equiv (1 - B)^d (1 - B^s)^D = \sum_{j=0}^{\infty} \psi_j^* B^j, \quad (4.3)$$

onde

$$\psi_j^* = \pi_j + \sum_{i=1}^{\infty} \pi_i^{(s)} \pi_{j-is}, \quad (4.4)$$

$$\pi_l = \frac{\Gamma(l - d)}{\Gamma(l + 1)\Gamma(-d)}, \quad l = 0, 1, \dots \quad (4.5)$$

e

$$\pi_k^{(s)} = \frac{\Gamma(k - D)}{\Gamma(k + 1)\Gamma(-D)}, \quad k = 0, 1, \dots \quad (4.6)$$

A prova da Equação 4.3 é consequência imediata do produto das expansões binomiais $(1 - B)^d$ e $(1 - B^s)^D$.

Corolário 1 (Reisen et al. (2008)). *Seja X_t o processo definido pela Eq. (4.1) onde $|d + D| < \frac{1}{2}$ e $|d|, |D| < \frac{1}{2}$. Se $\Phi(z^s)$, $\phi(z) \neq 0$, $|z| = 1$, então o modelo (4.1) possui uma solução única estacionária, representada pela expansão $MA(\infty)$*

$$X_t = \sum_{j \in \mathbf{Z}} \psi_j \nabla^{-\mathbf{d}} \epsilon_{t-j}, \quad (4.7)$$

onde ψ_j são determinados pela expansão de Laurent

$$\sum_{j \in \mathbb{Z}} \psi_j z^j = \frac{\Theta(z^s) \theta(z)}{\Phi(z^s) \phi(z)}, \quad (4.8)$$

para algum $|z| = 1$.

Para maiores detalhes sobre o Corolário 1, ver [Reisen et al. \(2008\)](#).

Uma série temporal com parâmetros de diferenciação fracionários sazonais e não sazonais possui função densidade espectral dada por

$$f(w) = f^*(w) |w|^{-v} \prod_{i=1}^k \prod_{j=1}^{\varphi_i} |w - w_{ij}|^{-v_i}, \quad (4.9)$$

onde $w \in (-\pi, \pi]$, $f^*(\cdot)$ é uma função contínua, limitada acima e longe de zero, $0 \leq v$, $v_i < 1$ e $w_{ij} \neq 0$ para $j = 1, \dots, \varphi_i$, $i = 1, \dots, k$ ([Reisen et al., 2008](#)). Processos com densidade espectral dada pela Eq. (4.9) têm sido discutidos em [Palma & Chan \(2005\)](#), [Giraitis & Leipus \(1995\)](#), entre outros. Especificamente, a densidade espectral do processo X_t , satisfazendo o modelo (4.1) com d , $D \in (-\frac{1}{2}, \frac{1}{2})$ pode ser escrita

$$f_X(\omega) = f_U(\omega) \left[2 \operatorname{sen} \left(\frac{\omega s}{2} \right) \right]^{-2D} \left[2 \operatorname{sen} \left(\frac{\omega}{2} \right) \right]^{-2d}, \quad \omega \in (-\pi, \pi], \quad (4.10)$$

onde $f_U(\cdot)$ é a densidade espectral do processo *ARMA* estacionário U_t . Para maiores detalhes, veja [Reisen, Rodrigues & Palma \(2006a, b\)](#).

Granger & Joyeux (1980) e Hosking (1981) propuseram o modelo $ARFIMA(p,d,q)$, caso particular do modelo $SARFIMA$ (Eq.(4.1)) quando $P = Q = D = 0$. Os modelos $ARFIMA$ são amplamente utilizados para modelar séries temporais com características de memória longa e possuem as propriedades sumarizadas a seguir. O processo $ARFIMA$ é estacionário e invertível quando $|d| < 0.5$; $d > 0$ caracteriza a propriedade de longa dependência, $d = 0$ e $d < 0$ indicam que o processo é de dependência curta e intermediária, respectivamente. A função densidade espectral para o modelo $ARFIMA$ comporta-se da forma $f(w) \sim C|w|^{-2d}$ para $w \rightarrow 0$ em que C é uma constante não negativa. A autocorrelação entre X_t e X_{t+k} satisfaz $\rho(k) \sim k^{2d-1}$ quando $k \rightarrow \infty$. Desde o trabalho de Geweke & Porter-Hudak (1983), muitos estimadores para o parâmetro d foram propostos, veja Reisen (1994) e Arteché & Robinson (2000), entre outros.

Seja $\{X_1, \dots, X_n\}$ uma amostra gerada do processo X_t (Eq. (4.1)). Para estimação dos parâmetros d e D , Reisen, Rodrigues & Palma (2006a, b) sugerem uma modificação ao método de Geweke & Porter-Hudak (1983) em processos sazonais. O método consiste em obter o estimador $\hat{\mathbf{d}} = (\hat{d}, \hat{D})$ do logaritmo aplicado na Eq. (4.10), que resulta na equação linear

$$\log f_X(\omega) = \log f_U(\omega) - D \log \left[2 \operatorname{sen} \left(\frac{\omega S}{2} \right) \right]^2 - d \log \left[2 \operatorname{sen} \left(\frac{\omega}{2} \right) \right]^2, \quad (4.11)$$

onde $-\pi \leq \omega \leq \pi$. A estimativa de \mathbf{d} pode ser obtida utilizando a função periodograma amostral dada por

$$I_X(w) = \frac{1}{n} \left| \sum_{t=1}^n X_t e^{iwt} \right|^2,$$

como estimador de $f_X(\cdot)$. A equação linear dada em (4.11) pode ser aproximada pela

equação de regressão linear múltipla

$$\log I(\omega) \cong a_o - D \log \left[2 \operatorname{sen} \left(\frac{\omega s}{2} \right) \right]^2 - d \log \left[2 \operatorname{sen} \left(\frac{\omega}{2} \right) \right]^2 + u, \quad (4.12)$$

onde a_o é uma constante e $u = \ln \frac{I(w)}{f(w)} - E \left[\ln \frac{I(w)}{f(w)} \right]$.

Segundo [Reisen et al. \(2008\)](#), o estimador $\hat{\mathbf{d}}$ possui as seguintes propriedades assintóticas:

(a) Vício:

$$E(\hat{\mathbf{d}}) - \mathbf{d} = \frac{1}{\sqrt{M}} W^{-1} b_n (1 + o(1)), \quad (4.13)$$

(b) Variância:

$$\operatorname{Var}(\hat{\mathbf{d}}) = \frac{\pi^2}{6M} W^{-1} (1 + o(1)), \quad (4.14)$$

onde M é o “*bandwidth*” usado na equação de regressão (4.12),

$$b_n = -2 \begin{pmatrix} \sum_{k=0}^{\lfloor \frac{s}{2} \rfloor} b_k \delta_k (2\pi)^{v_k} \frac{v_k}{(v_k+1)^2} \left(\frac{M}{n} \right)^{v_k} \\ \sum_{k \in I_k} b_k \delta_k (2\pi)^{v_k} \frac{v_k}{(v_k+1)^2} \left(\frac{M}{n} \right)^{v_k} \end{pmatrix},$$

em que b_k é função de $f_A(\cdot)$, Q é uma matriz 2×2 dada por

$$W = 4 \begin{pmatrix} \sum_{k=0}^{\lfloor \frac{s}{2} \rfloor} \delta_k & \sum_{k \in I_k} \delta_k \\ \sum_{k \in I_k} \delta_k & \sum_{k \in I_k} \delta_k \end{pmatrix},$$

$I_k = \{0, k \text{ quando } ks_2 \text{ é múltiplo de } s\}$ e

$$\delta_k = \begin{cases} 1, & k = 0, \lfloor \frac{s}{2} \rfloor \\ 2, & k = 1, 2, \dots, \lfloor \frac{s}{2} \rfloor - 1. \end{cases}$$

(c) Distribuição:

$$\sqrt{M}(\hat{\mathbf{d}} - \mathbf{d}) \xrightarrow{d} N\left(W^{-1}b, \frac{\pi^2}{6}W^{-1}\right), \quad (4.15)$$

para

$$b = -2(2\pi)^{v^*} \frac{v^*}{(v^* + 1)^2} K \begin{pmatrix} \sum_{k \in J_k} b_k \delta_k \\ \sum_{k \in I_k \cap J_k} b_k \delta_k \end{pmatrix},$$

onde $J_k = \{k \text{ para } v_k = v^*\}$.

Os resultados dados nas Eq. (4.13), (4.14) e (4.15) foram obtidos por [Reisen et al. \(2008\)](#), supondo $\epsilon_t \sim N(0, 1)$, na Eq. (4.1). Entretanto, no caso mais simples dos processos de longa memória, isto é, o modelo $ARFIMA(p, d, q)$, [Sena, Reisen & Lopes \(2006\)](#) mostraram, através de ensaios de Monte Carlo, que o estimador \hat{d} é robusto com a suposição teórica de erros não Gaussianos, com ênfase em processos heterocedásticos. Os autores apresentam justificativas teóricas e empíricas. Em processos $SARFIMA$, com parâmetros sazonais fracionários, como estudado neste trabalho, não foi encontrada literatura referente à essa questão. Portanto, a estimação de modelos $SARFIMA$, com erros não Gaussianos, é uma vertente de pesquisa que necessita de estudos. Entretanto, os resultados teóricos e empíricos discutidos em [Sena, Reisen & Lopes \(2006\)](#) dão suporte para o estudo presente neste trabalho, isto é, estrutura do modelo $SARFIMA$, com erros não Gaussianos.

4.2 Modelo de Volatilidade

Nesse trabalho, a modelagem da volatilidade será efetuada nos resíduos do modelo 4.1, modelo com dois parâmetros de memória, ajustado previamente ao conjunto de dados em questão. O primeiro modelo que fornece uma estrutura sistemática para a modelagem da volatilidade é o modelo $ARCH(m)$ (Auto-regressivo Condicional Heterocedástico), proposto por Engle (1982), sumarizado a seguir.

Seja $Y_t \equiv \{Y_t\}_{t \in \mathbb{Z}}$ um processo dado pela Eq. 4.1, escrito na forma

$$Y_t = E(Y_t | \mathfrak{S}_{t-1}) + \xi_t, \quad (4.16)$$

onde \mathfrak{S}_t denota o conjunto de informações avaliadas até o instante t e ξ_t é dada por

$$\xi_t = \eta_t \sqrt{h_t}, \quad \eta_t \sim iid(0, 1), \quad (4.17)$$

onde $E(\xi_t | \mathfrak{S}_{t-1}) = 0, \forall t$. O modelo $ARCH(m)$ assume que a variância de ξ_t é função do tempo passado, e é definido por

$$h_t = \alpha_0 + \sum_{j=1}^m \alpha_j \xi_{t-j}^2, \quad (4.18)$$

onde $m \in \mathbb{N}$ é a ordem do modelo, $\alpha_0 > 0, \alpha_i \geq 0$ para $i = 1, \dots, m$. O modelo $ARCH(m)$,

Eq. (4.18), pode ser reescrito como uma auto-regressão da forma

$$\xi_t^2 = \alpha_0 + \sum_{j=1}^m \alpha_j \xi_{t-j}^2 + u_t, \quad (4.19)$$

onde $u_t = \xi_t^2 - h_t$ é imprevisível e considerado uma inovação da auto-regressão de ξ_t^2 . Se o processo definido pela Eq. (4.18) é estacionário de segunda ordem, a variância de ξ_t é

$$Var(\xi_t) = \frac{\alpha_0}{1 - \sum_{j=1}^m \alpha_j}, \quad (4.20)$$

onde $\sum_{i=1}^m \alpha_i < 1$, condição de estacionariedade.

Processos $ARCH(m)$ possuem excesso de curtose (κ), que ocorre devido às pesadas caudas presentes nessas distribuições. Essa propriedade pode ser verificada pelo cálculo de κ para o processo $ARCH(1)$ que é dada por

$$\kappa = \frac{E(\xi_t^4)}{(Var(\xi_t))^2} = \frac{3(1 - \alpha_1^2)}{1 - 3\alpha_1^2}, \quad (4.21)$$

onde $0 < \alpha_1^2 \leq \frac{1}{3}$ e $\kappa > 3$. Para maiores detalhes, veja [Gourieroux \(1997\)](#), [Hamilton \(1994\)](#) e [Tsay \(2002\)](#).

[Bollerslev \(1986\)](#) propôs uma generalização ao modelo de [Engle \(1982\)](#) (Eq. (4.18)) em que a variância do processo dado pela Eq. (4.17) é da forma

$$h_t = \alpha_0 + \sum_{i=1}^m \alpha_i \xi_{t-i}^2 + \sum_{j=1}^r \beta_j h_{t-j}, \quad (4.22)$$

onde $m, r \in \mathbb{N}$ são as ordens do modelo, $\alpha_0 > 0$ e $\alpha_i, \beta_j \geq 0$, para $i = 1, 2, \dots, m$ e $j = 1, 2, \dots, r$. O modelo (4.22) é definido por $GARCH(m, r)$, (Auto-regressivo Condicional Heterocedástico Generalizado) e pode ser mais parcimonioso que o modelo $ARCH(m)$ (Gourieroux, 1997).

O modelo $GARCH(m, r)$, definido pela Eq. (4.22), pode ser reescrito da forma

$$\xi_t^2 = \alpha_0 + \sum_{i=1}^{\max(m, r)} (\alpha_i + \beta_i) \xi_{t-i}^2 + \zeta_t - \sum_{j=1}^r \beta_j \zeta_{t-j}, \quad (4.23)$$

onde $\zeta_{t-j} = \xi_{t-j}^2 - h_{t-j}$ com $E(\zeta_t) = 0$ e $Cov(\zeta_t, \zeta_{t-j}) = 0$, $j = 1, \dots, r$. Bollerslev (1986), mostrou que o processo $GARCH(m, r)$, é estacionário se $\sum_{i=1}^m \alpha_i + \sum_{j=1}^r \beta_j < 1$, sempre que $E(\xi_t) = 0$ e

$$Var(\xi_t) = \frac{\alpha_0}{1 - \left(\sum_{i=1}^m \alpha_i + \sum_{j=1}^r \beta_j \right)}. \quad (4.24)$$

Como observado nos processos $ARCH(m)$, os processos $GARCH(m, r)$ também apresentam caudas pesadas em suas distribuições, causada pelo excesso de curtose. Para o caso mais simples onde $m = r = 1$, a curtose (κ) é dada por

$$\kappa = \frac{E(\xi_t^4)}{(Var(\xi_t))^2} = \frac{3[1 - (\alpha_1 + \beta_1)^2]}{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2}, \quad (4.25)$$

onde $1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2 > 0$ e $\kappa > 3$.

4.3 Modelos Espaço de Estados

Para representar o modelo $SARFIMA(p, d, q) \times (P, D, Q)_s$ (Eq. (4.1)) na representação espaço de estado é preciso utilizar a representação $MA(\infty)$

$$X_t = \sum_{j=0}^{\infty} \Psi_j \epsilon_{t-j}, \quad (4.26)$$

onde Ψ_j são coeficientes reais de $\Psi(z) = \sum_{j=0}^{\infty} \psi_j \nabla^{-d} z^j = \frac{\Theta(z^s)\theta(z)}{\Phi(z^s)\phi(z)}(1 - z^s)^{-D}(1 - z)^{-d}$ (Corolário 1). A partir da Eq. (4.26), a representação espaço de estado do processo

$\{X_t\}_{t=1, \dots, n}$ é dada por

$$\begin{cases} \Lambda_{t+1} = F\Lambda_t + H\epsilon_t \\ X_t = G\Lambda_t + \epsilon_t \end{cases}, \quad (4.27)$$

para $\Lambda_t = [X(t|t-1) \quad X(t+1|t-1) \quad X(t+2|t-1) \cdots]'$, $X(t|j) = E(X_t | X_j, X_{j-1}, \dots)$,

$$F = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ 0 & 0 & 0 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad H = \begin{bmatrix} \Psi_1 \\ \Psi_2 \\ \Psi_3 \\ \vdots \end{bmatrix} \quad G = [1 \quad 0 \quad 0 \quad \cdots]. \quad (4.28)$$

A expansão (4.7) truncada para ϑ componentes ($\vartheta \in \mathbb{N}$), pode ser escrita por $X_t =$

$\sum_{j=0}^{\vartheta} \Psi_j \epsilon_{t-j}$, e a representação na forma espaço de estado é dada por

$$\Lambda_{t+1} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \Lambda_t + \begin{bmatrix} \Psi_1 \\ \Psi_2 \\ \vdots \\ \Psi_m \end{bmatrix} \epsilon_t, \quad (4.29)$$

$$X_t = [1 \ 0 \ 0 \ \cdots \ 0] \Lambda_t + \epsilon_t, \quad (4.30)$$

onde Λ_t , F , H e G possuem, respectivamente, as ordens $[\vartheta \times 1]$, $[\vartheta \times \vartheta]$, $[\vartheta \times 1]$ e $[1 \times \vartheta]$.

Para o modelo espaço de estado (4.27), a predição um passo à frente do estado é $\hat{\Lambda}_t = P_{t-1} \Lambda_t$, onde $P_{t-1} \Lambda_t$ é a projeção de Λ_t no espaço linear $K_t = \{X_{t-1}, \dots, X_1\}$ e $\Omega_t = E(\Lambda_t \Lambda_t' - \hat{\Lambda}_t \hat{\Lambda}_t')$ é a matriz de covariância do erro de predição. O procedimento recursivo do filtro de Kalman considerando como condições iniciais $\hat{\Lambda}_1 = E(\Lambda_1)$ e $\Omega_1 = E(\Lambda_1 \Lambda_1') - E(\hat{\Lambda}_1 \hat{\Lambda}_1')$ é dado pelas equações

$$\Delta_t = G \Omega_t G' + R, \quad (4.31)$$

$$\Xi_t = F \Omega_t G' + S, \quad (4.32)$$

$$\Omega_{t+1} = F \Omega_t F' + T - \Xi_t \Delta_t^{-1} \Xi_t', \quad (4.33)$$

$$\hat{\Lambda}_{t+1} = F \hat{\Lambda}_t + \Xi_t \Delta_t^{-1} (X_t - G \hat{\Lambda}_t), \quad (4.34)$$

$$\hat{X}_t = G \hat{\Lambda}_t, \quad (4.35)$$

onde $R = Var(\epsilon_t) = \sigma_\epsilon^2$, $T = Var(H\epsilon_t) = \sigma_\epsilon^2 HH'$ e $S = Cov(\epsilon_t, H\epsilon_t) = \sigma_\epsilon^2 H$.

Para análise de séries temporais com observações faltantes, [Palma & Chan \(1997\)](#) propuseram uma modificação ao Filtro de Kalman para estimação dessas observações. A modificação ocorre nas Eq. (4.33) e (4.34), reescritas, respectivamente, por

$$\Omega_{t+1} = \begin{cases} F\Omega_t F' + T - \Xi_t \Delta_t^{-1} \Xi_t', & \text{se } X_t \text{ é observado} \\ F\Omega_t F' + T, & \text{se } X_t \text{ é faltante} \end{cases}, \quad (4.36)$$

$$\hat{\Lambda}_{t+1} = \begin{cases} F\hat{\Lambda}_t + \Xi_t \Delta_t^{-1} (X_t - G\hat{\Lambda}_t), & \text{se } X_t \text{ é observado} \\ F\hat{\Lambda}_t, & \text{se } X_t \text{ é faltante.} \end{cases}, \quad (4.37)$$

As modificações nas equações da matriz de covariância do erro de predição e da predição do estado são obtidas, observando que, se X_t é uma observação faltante, denotada por X_t^* , então $K_t = K_{t-1}$. A estimativa de X_t^* é dada por

$$\hat{X}_t^* = E(G\Lambda_t + \epsilon_t | K_{t-1}) = G\hat{\Lambda}_t + E(\epsilon_t | K_{t-1}) = G\hat{\Lambda}_t. \quad (4.38)$$

Portanto, pelo resultado da Eq. (4.38), a estimativa de X_t^* corresponde à estimativa de X_t , onde X_t é observado.

Análises e Resultados

Neste capítulo são apresentadas e discutidas as análises de modelagem e previsão das concentrações médias diárias de PM_{10} para a cidade de Cariacica, Região da Grande Vitória (RGV), Brasil, utilizando a metodologia descrita na Seção 4. Para implementação do Filtro de Kalman foi utilizado o *software MATLAB 7.0* e a estimação dos parâmetros de longa dependência e análises dos resultados foram computadas nos *softwares OX* e *R*.

Este trabalho foi realizado no município de Cariacica localizado na Região da Grande Vitória (RGV), constituída de 1.337.187 habitantes (IBGE/2000), a região de Cariacica corresponde a 43% da população total do Estado do Espírito Santo, vinculada à região da Grande Vitória (RGV). Abrange uma área de 1.461 km², sendo um dos principais pólos de desenvolvimento urbano e industrial do Estado. Cerca de 55% a 65% das atividades industriais potencialmente poluídoras do Espírito Santo tais como siderurgia, pelotização,

mineração (pedreiras), cimenteiras, indústria alimentícia, usina de asfalto, entre outras, estão instaladas nessa região, influenciando na elevação dos atendimentos por patologias respiratórias (Martins et al., 2004). Foram consideradas 731 observações compreendidas entre 01 de janeiro de 2003 e 31 de dezembro de 2004, de médias diárias das concentrações de PM_{10} , medidas em $\mu\text{g}/\text{m}^3$. Nesse período, Cariacica apresentou as maiores médias de concentração em PM_{10} de RGV.

As Figuras 5.1(a) e 5.1(b) mostram, respectivamente, a série de PM_{10} com observações faltantes (*missing data*) e a série completa, denota por $PM_{10}^* \equiv X_t$, essa obtida utilizando a metodologia descrita na Subseção 4.3. A série das concentrações de PM_{10} apresentam características de sazonalidade, possíveis observações atípicas e volatilidade. Todas as análises de estimação, modelagem e previsão dos modelos em estudo foram realizados para a série de PM_{10}^* (Figura 5.1(b)).

5.1 Modelos ajustados

A Função de Autocorrelação Amostral (FAC) de PM_{10}^* (Figura 5.2(a)), apresenta correlações significativas com um decaimento lento para os *lags* múltiplos de 7, que indica sazonalidade $s = 7$ (confirmando a variação entre as concentrações medidas nos dias úteis e finais de semana) e possível presença da propriedade de longa dependência na componente sazonal. O decaimento lento das correlações também é observado para os *lags* entre os períodos sazonais, que sugere longa dependência entre os períodos sazonais.

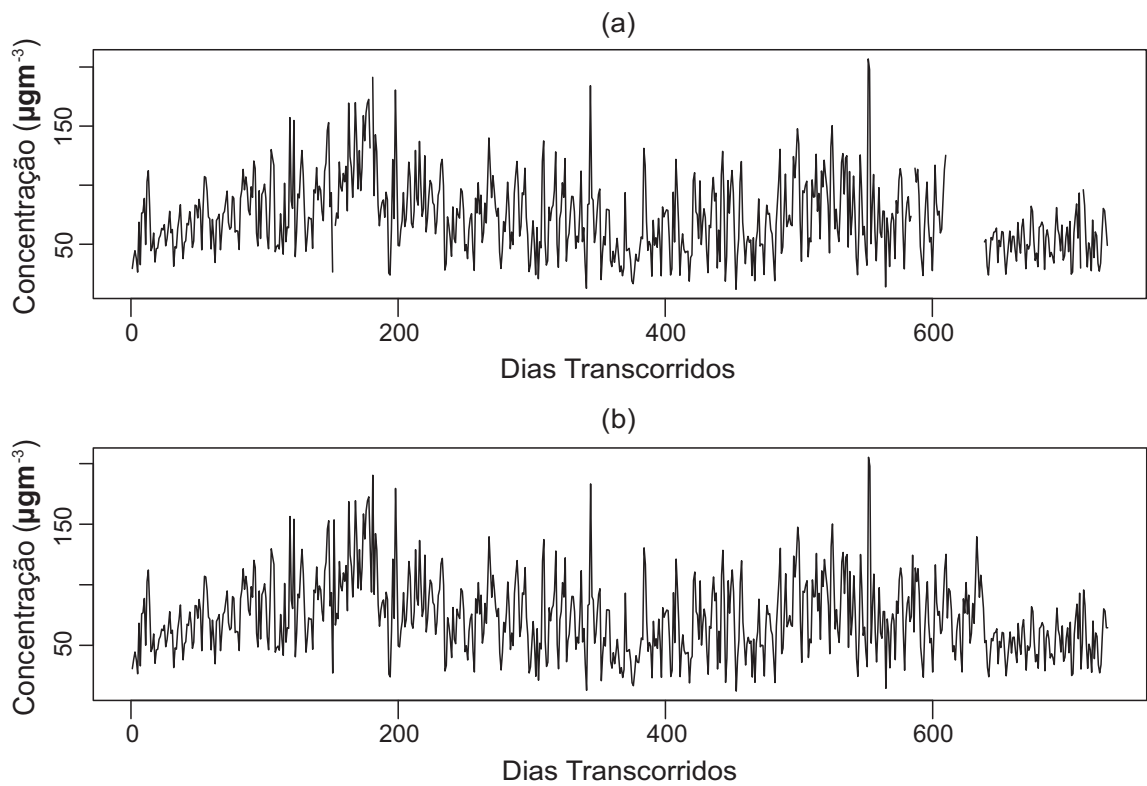


Figura 5.1: Série Temporal das concentrações de PM_{10} na RGV; (a) série com valores faltantes; (b) série completa.

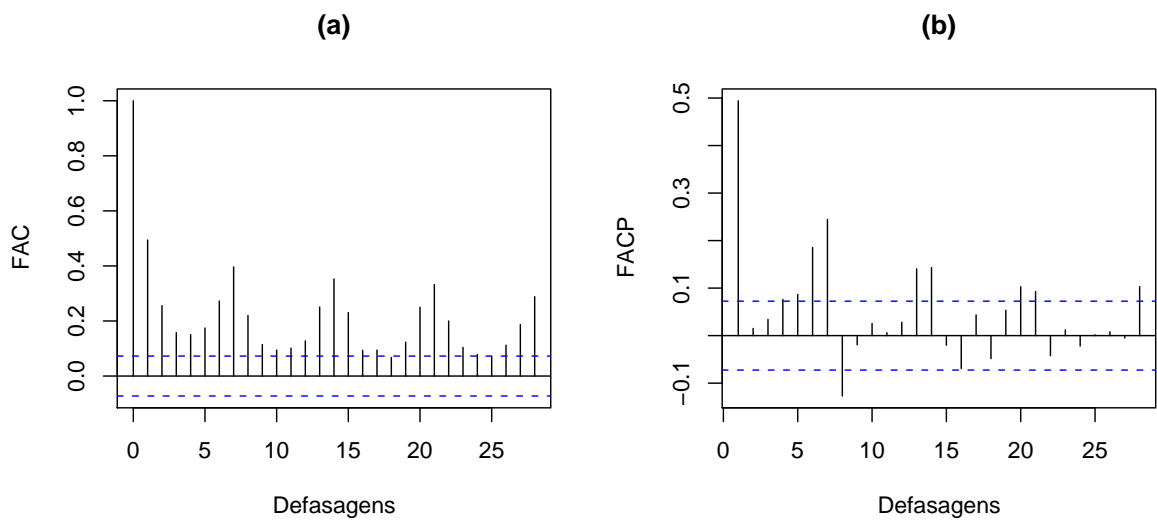


Figura 5.2: Correlograma da Série Temporal PM_{10}^*

As evidências empíricas da FAC sugerem modelos das classes $SARMA(p, q) \times (P, Q)_7$ e $SARFIMA(p, d, q) \times (P, D, Q)_7$. Para o modelo $SARFIMA$, foi utilizada a metodologia discutida na seção 4.1 para estimação do vetor $\mathbf{d} = (d, D)$. A Tabela 5.1 mostra as estimativas dos parâmetros d e D , obtidas para diversos valores do *bandwidth* $M = n^\alpha$, $0 < \alpha < 1$, em que M satisfaz a condição $\left(\frac{M}{n}\right) \log M + \frac{1}{M} \rightarrow 0$ para $M \rightarrow \infty$ e $n \rightarrow \infty$ (Reisen et al. 2008).

Tabela 5.1: Estimativas de d e D para diferentes *bandwidths*.

| α | 0.5 | 0.6 | 0.7 | 0.8 | Máx |
|----------|-----------------|-----------------|-----------------|-----------------|-----------------|
| d | 0.3300 (0.0048) | 0.3218 (0.0007) | 0.3207 (0.0019) | 0.3274 (0.0007) | 0.3378 (0.0009) |
| D | 0.3696 (0.0006) | 0.1459 (0.0030) | 0.1116 (0.0007) | 0.1164 (0.0013) | 0.1195 (0.0007) |

Pela Tabela, 5.1 observa-se que as estimativas dos parâmetros são estáveis para $\alpha > 0.7$. Diante desses valores empíricos, foi fixado $\alpha = 0.7$. O vetor $\hat{\mathbf{d}} = (0.3207, 0.1116)$ mostra que o modelo $SARFIMA$ apresenta as propriedades de estacionariedade, invertibilidade e memória longa. Note que, mesmo utilizando todas as frequências, $w_j = \frac{2\pi j}{n}$, $j = 1, 2, \dots, \frac{n-1}{2}$, as estimativas de \mathbf{d} não apresentam mudanças significativas. Isso é uma indicação que se existe componentes $ARMA$ no modelo, a contribuição dos coeficientes do mesmo na estrutura de dependência das observações não é de grande impacto. Isso será melhor explorado a seguir.

Dando seqüência na modelagem da série PM_{10}^* , U_t na Eq. (4.1) é estimado por $\hat{U}_t = \nabla^{\hat{\mathbf{d}}} X_t$, onde \hat{U}_t tem representação AR finita dada por $\hat{U}_t = \sum_{j=0}^n \hat{\psi}_j^* X_{t-j}$, $n = 731$. $\hat{\psi}_j^*$, $j = 1, 2, \dots, 731$ são estimativas dos coeficientes ψ_j^* obtidos de acordo com o Corolário ???. Por exemplo, para $j = 731$, $\hat{\psi}_j^*$ é de ordem 10^{-6} . Isso mostra que a contribuição das

observações, na expansão ∇^d , para distantes *lags* é de impacto praticamente nulo. As Figuras 5.3(a) e 5.3(b) mostram, respectivamente, a função de autocorrelação amostral e a função de autocorrelação parcial amostral de \hat{U}_t , que evidenciam uma possível estrutura auto-regressiva de ordem 1.

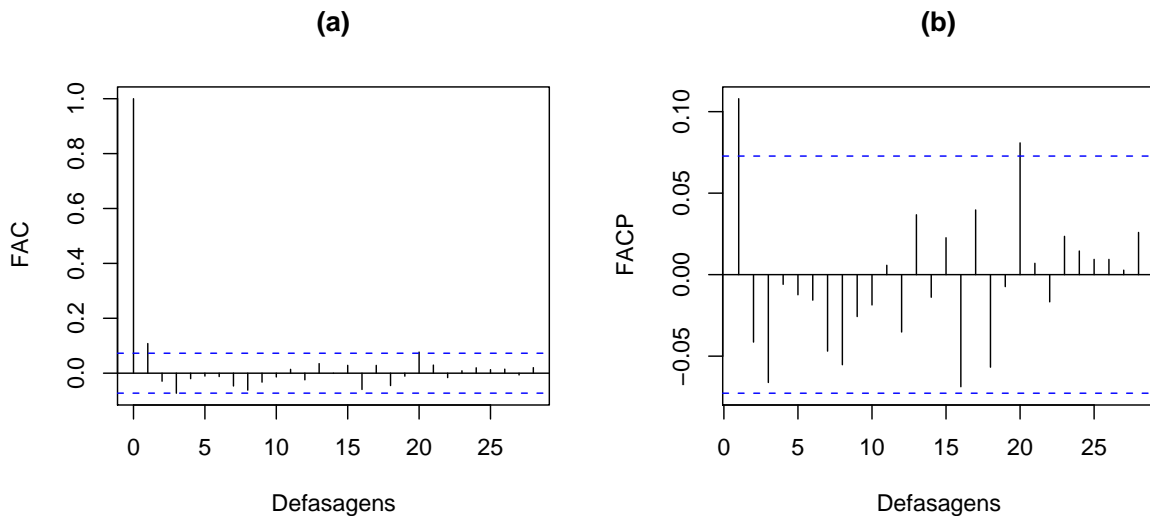


Figura 5.3: Correlograma de \hat{U}_t ; (a) FAC e (b) FACP.

Na estimação da parte Auto-regressiva Média Móvel do modelo (4.1), a série \hat{U}_t foi usada, desconsiderando as cinco primeiras observações para evitar os efeitos das aproximações iniciais. Para identificação da ordem do modelo utilizou o Critério de Akaike (AIC) e as funções FAC e FACP, que sugerem como modelo inicial o $SARFIMA(1, \hat{d}, 0) \times (0, \hat{D}, 0)_7$. Através de metodologias análogas, o modelo inicial $SARMA(1, 0) \times (1, 1)_7$ foi sugerido ao conjunto PM_{10} .

As Figuras 5.4, 5.5 e 5.6 correspondem aos gráficos de $\hat{\epsilon}_t^2$, $t = 1, \dots, n$ e suas funções de autocorrelação amostral e autocorrelação parcial amostral para os modelos ajustados

SARMA e *SARFIMA*, respectivamente. As figuras indicam variância não constante, isto sugere uma componente que modele a heterocedasticidade. Para confirmar a presença de heterocedasticidade nos resíduos dos respectivos modelos, foi efetuado o teste de multiplicadores de Lagrange, onde a hipótese nula de homoscedasticidade residual foi rejeitada para ambos os modelos, com valores p 's menores que 0.001.

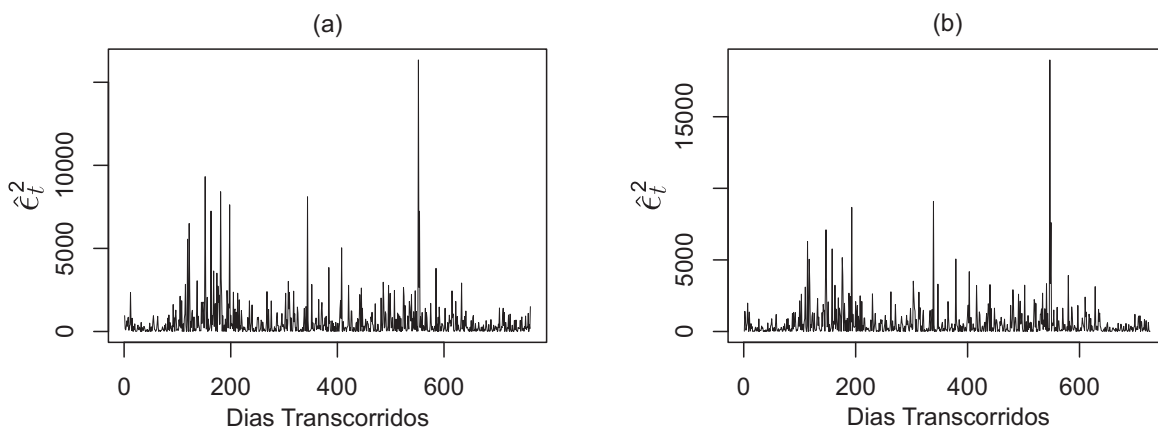


Figura 5.4: Gráfico de $\hat{\epsilon}_t^2$, $t = 1, \dots, n$ do modelo inicial *SARMA* e *SARFIMA*

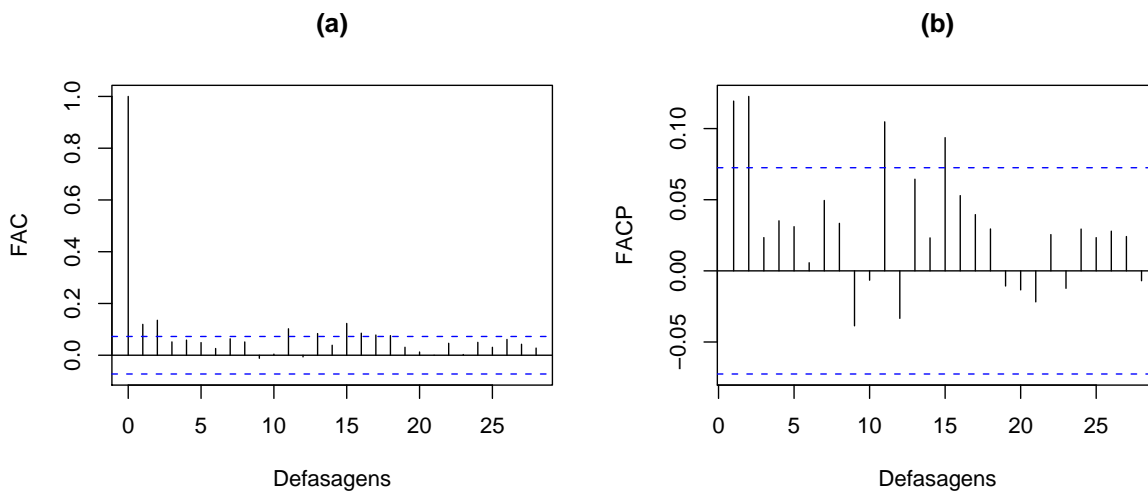


Figura 5.5: Estrutura de autocorrelação de $\hat{\epsilon}_t^2$, $t = 1, \dots, n$ do modelo inicial *SARMA*

Após devidos procedimentos de modelagem, foram ajustados os modelos heterocedásticos

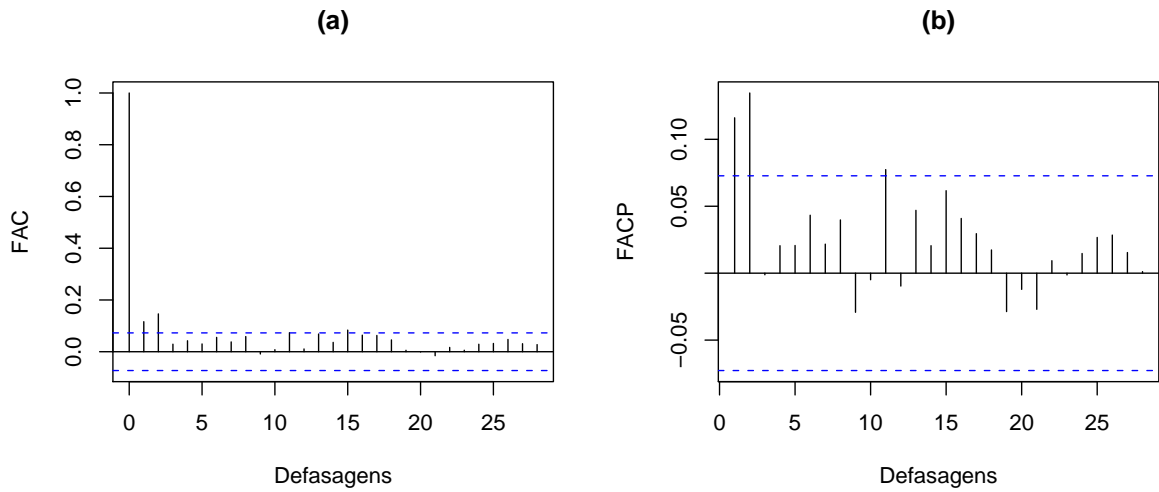


Figura 5.6: Estrutura de autocorrelação de $\hat{\epsilon}_t^2$, $t = 1, \dots, n$ do modelo inicial *SARFIMA* *GARCH* e *ARCH* nos resíduos dos modelos *SARMA* e *SARFIMA*, respectivamente. Os modelos completos $SARMA(1, 0) \times (1, 1)_7 - GARCH(1, 1)$ e $SARFIMA(1, \hat{d}, 0) \times (0, \hat{D}, 0)_7 - ARCH(2)$ são denominados, respectivamente, de modelos I e II e os parâmetros estimados estão apresentados na Tabela 5.2 (modelo I) e Tabela 5.3 (modelo II).

Tabela 5.2: Estatística do modelo I

| SARMA | | | | GARCH | | | |
|------------|------------|-------------|---------|------------|------------|-------------|---------|
| Parâmetro | Estimativa | Erro Padrão | Teste t | Parâmetro | Estimativa | Erro Padrão | Teste t |
| ϕ | 0.4405 | 0.0334 | 13,188 | α_0 | 9.6552 | 5.8572 | 1.6484 |
| Φ_1 | 0.9503 | 0.0180 | 52.794 | α_1 | 0.0633 | 0.0168 | 3.7569 |
| Θ_1 | -0.7972 | 0.0384 | -20.760 | β_1 | 0.9216 | 0.0205 | 44.960 |

Tabela 5.3: Estatística do modelo II

| SARFIMA | | | | ARCH | | | |
|-----------|------------|-------------|---------|------------|------------|-------------|---------|
| Parâmetro | Estimativa | Erro Padrão | Teste t | Parâmetro | Estimativa | Erro Padrão | Teste t |
| d | 0.3207 | 0.0019 | 163.65 | α_0 | 491,17 | 24.813 | 19.794 |
| D | 0.1116 | 0.0007 | 144.53 | α_1 | 0.0917 | 0.0402 | 2.2804 |
| ϕ | 0.1078 | 0.0369 | 2.9214 | α_2 | 0.0842 | 0.0364 | 2.3095 |

A Tabela 5.4 apresenta a estatística descritiva dos resíduos dos modelos ajustados. Os correlogramas residuais dos modelos I e II são, respectivamente, mostrados nas Figuras

5.7 e 5.9. As Figuras 5.8 e 5.10 mostram a análise residual para os modelos I e II, respectivamente. Os testes de normalidade de [Shapiro-Wilk \(1965\)](#) e [Jarque Bera \(1981\)](#) rejeitaram, para ambos os modelos, a hipótese de que as distribuições de probabilidade dos resíduos são normais. No entanto, essa rejeição ocorre pela presença de caudas pesadas nessas distribuições empíricas. A propriedade de assimetria pode ser observada para ambas distribuições empíricas dos resíduos, causada pela presença de valores extremos na série. Entretanto, para ambos os modelos, a hipótese nula de não correlação ($H_0 : \rho_{res}=0$) nos resíduos não foi rejeitada pelos testes [Box-Pierce \(1970\)](#) e [Ljung-Box \(1978\)](#) como mostrado na Tabela 5.5.

Tabela 5.4: Estatística descritiva dos resíduos dos modelos ajustados

| | Modelo I | Modelo II |
|---------------|----------|-----------|
| Média | 0.008 | 0.006 |
| Desvio Padrão | 0.997 | 1.000 |
| Assimetria | 1.763 | 1.937 |
| Curtose | 4.088 | 5.296 |

Tabela 5.5: Testes estatísticos de normalidade* e não correlação** dos resíduos

| Teste | Modelo I | Modelo II |
|---------------|-----------|-----------|
| | valor p | valor p |
| Shapiro-Wilk* | < 0.0001 | < 0.0001 |
| Jarque-Bera* | < 0.0001 | < 0.0001 |
| Box-Pierce** | 0.9462 | 0.5689 |
| Box-Ljung** | 0.9460 | 0.5681 |

Dando continuidade à modelagem da série de PM_{10}^* , é apresentada a seguir a análise comparativa da qualidade dos modelos ajustados I e II. As Figuras 5.11(a) e 5.11(b) mostram, respectivamente, uma análise visual dos modelos ajustados I e II, plotados com a série PM_{10}^* . Essas evidenciam um bom desempenho para ambos os modelos, uma vez que

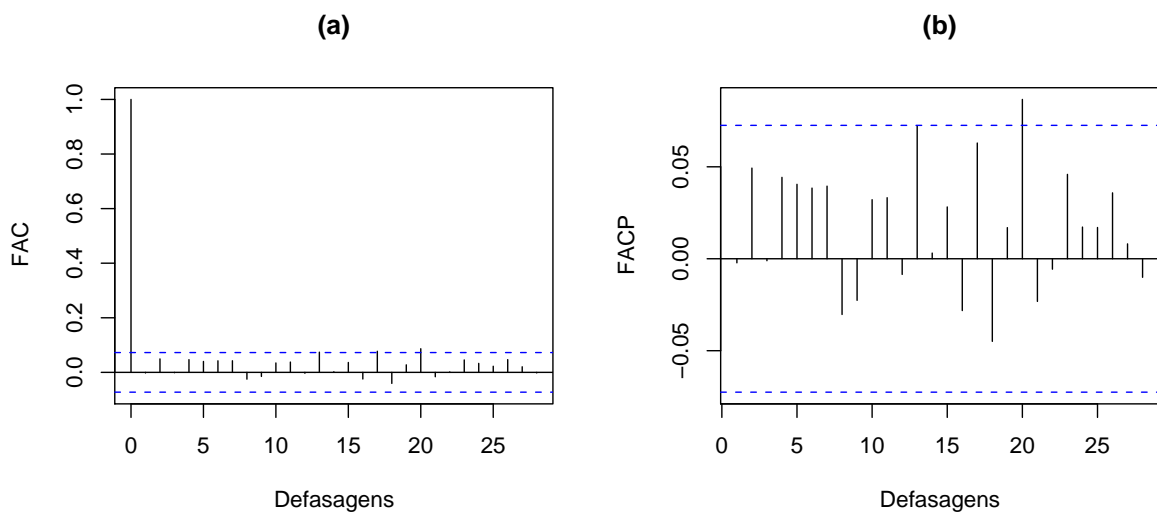


Figura 5.7: Correlograma residual do modelo I

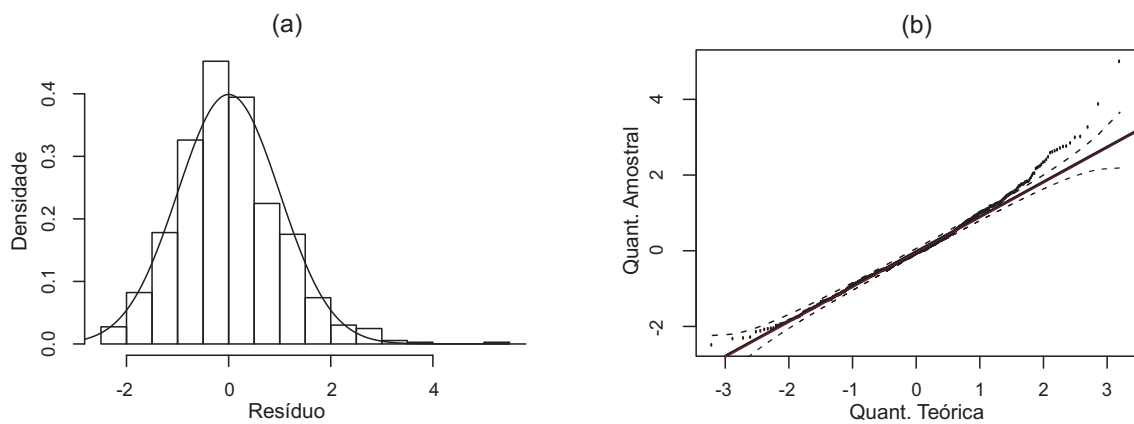


Figura 5.8: Análise residual do modelo I

esses mostraram captar a variabilidade, tendência e sazonalidade da série. No entanto, para quantificar os erros no ajuste, medidas dos erros de predição são apresentadas na Tabela 5.6. As medidas de erro são o Erro Médio Percentual (EMP), Erro Quadrático Médio (EQM), Erro Absoluto Médio (EAM), Erro Absoluto Médio Percentual (EAMP) e a Raiz do Erro Quadrático Médio (REQM).

Pela Tabela 5.6, verifica-se que para todas as medidas, o modelo II é superior, com

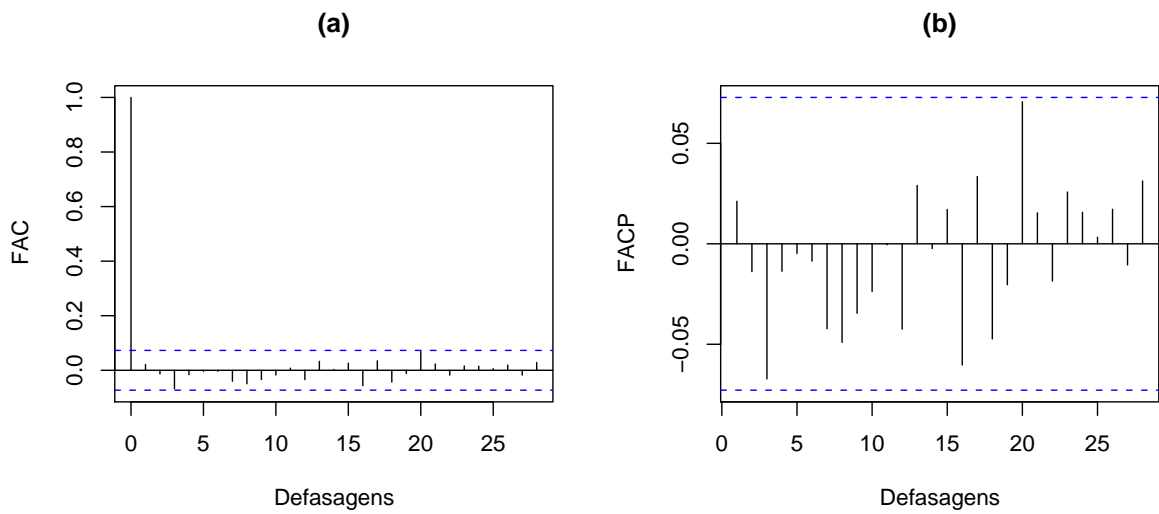


Figura 5.9: Correlograma residual do modelo II

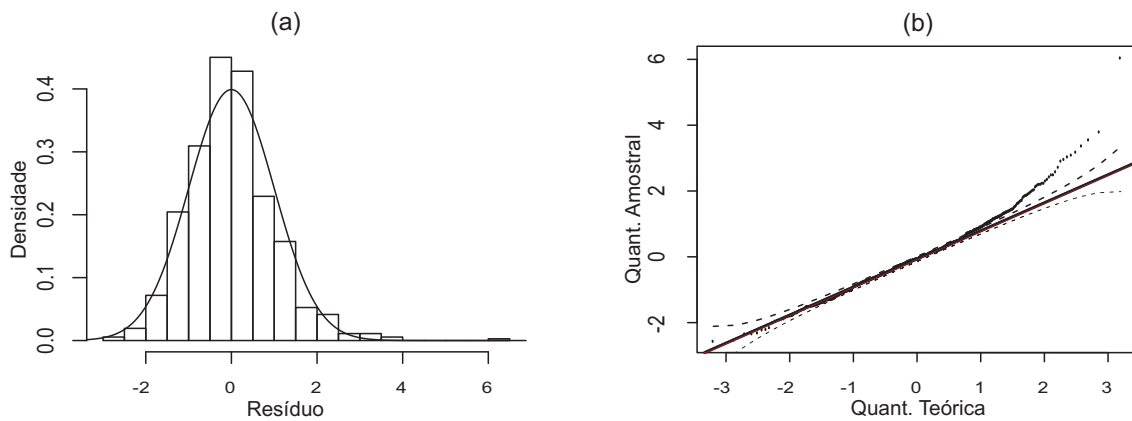


Figura 5.10: Análise residual do modelo II

Tabela 5.6: Avaliação de desempenho dos modelos

| Medidas de erro | Modelo I | Modelo II |
|-----------------|----------|-----------|
| EMP | -0.12005 | -0.11214 |
| EQM | 394.5024 | 183.2088 |
| EAM | 15.6387 | 10.6056 |
| EAMP | 0.28275 | 0.20240 |
| REQM | 19.8620 | 13.5354 |

destaque à diferença no erro quadrático médio entre os modelos.

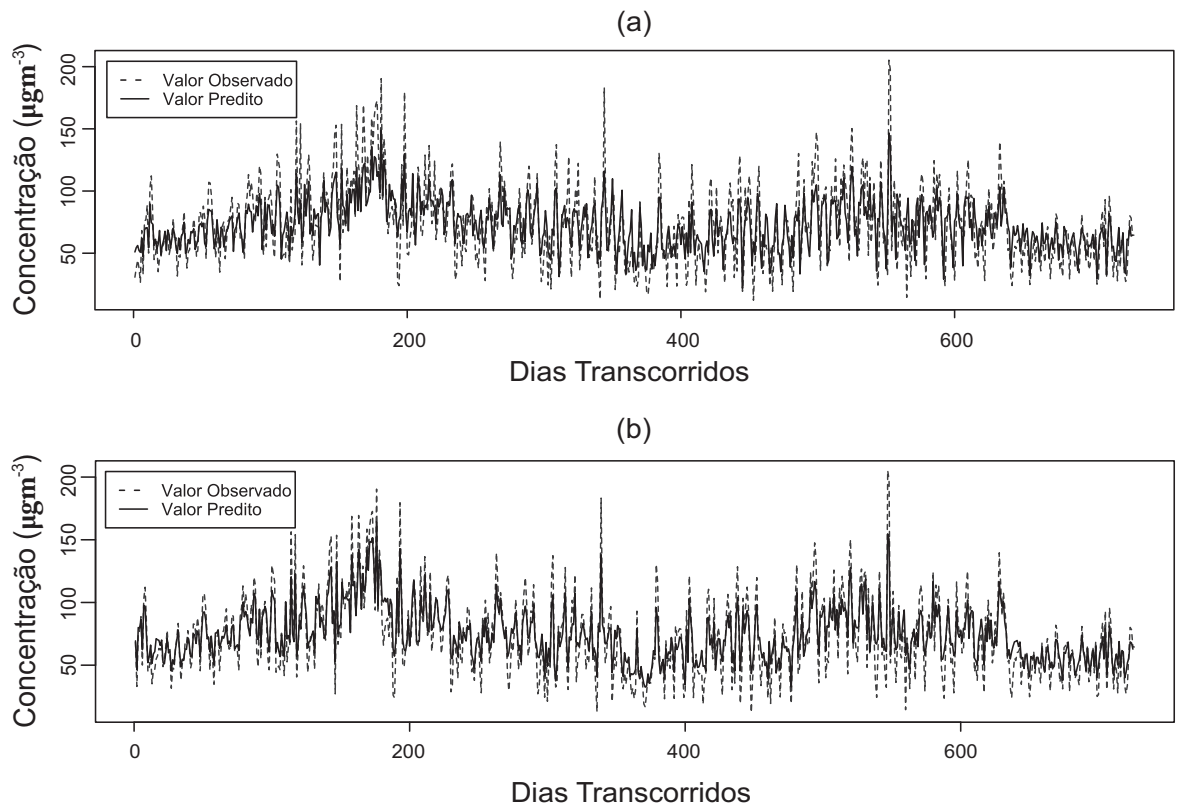


Figura 5.11: Valores Observador e Preditos; (a) Modelo I, (b) Modelo II

Como já mencionado anteriormente, a pequena variabilidade nas estimativas de \mathbf{d} em relação a *bandwidth* indica que, se existem componentes *ARMA* na série PM_{10}^* , essas não causam grande impacto. Essa observação empírica é reforçada, observando a estimativa de ϕ para o modelo II, onde $\hat{\phi} = 0.1078$. Em função dessas evidências, o Erro Quadrático Médio (EQM) do ajuste do modelo $SARFIMA(0, d, 0) \times (0, D, 0)_7 - GARCH(m, r)$ foi calculado e forneceu um valor 135,5% maior que o EQM do modelo II. Portanto, mesmo $\hat{\phi}$ não ser uma estimativa significativamente tão expressiva, a inclusão da componente *AR*(1) no modelo estimado fornece uma redução substancial no EQM do processos $SARFIMA(0, \hat{d}, 0) \times (0, \hat{D}, 0)_7 - ARCH(1)$.

5.2 Estudo de previsão

Como já mencionado anteriormente, foi utilizado para ajuste dos modelos o conjunto de dados das concentrações diárias de PM_{10} para o período de 01 de janeiro de 2003 a 31 dezembro de 2004. Nessa seção, é apresentado o estudo de previsão um passo à frente (24 horas à frente), que avaliou o desempenho dos modelos ajustados para o período de 01 a 31 de janeiro de 2005. Os resultados da comparação de medida dos erros de previsão estão na Tabela 5.7.

Tabela 5.7: Análise de desempenho das previsões dos modelos

| Medidas de erro | Modelo I | Modelo II |
|-----------------|----------|-----------|
| EMP | -0.17052 | -0.04054 |
| EQM | 290.1885 | 112.0082 |
| EAM | 1.54116 | 8.625393 |
| EAMP | 0.321753 | 0.197565 |
| REQM | 17.03492 | 10.58339 |

O modelo II obteve melhor desempenho nas previsões, fato semelhante já observado na Tabela 5.6 para os valores preditos. Para uma análise visual, a Figura 5.12 mostra as previsões dos modelos I e II, respectivamente, com os valores observados para a concentração diária de PM_{10} . As figuras trazem para cada previsão os respectivos intervalos de confiança para os níveis de 95% de confiabilidade. Estes intervalos de confiança, reafirmam o melhor desempenho do modelo $SARFIMA(1, \hat{d}, 0) \times (0, \hat{D}, 0)_7 - ARCH(2)$ diante do conjunto de dados analisados.

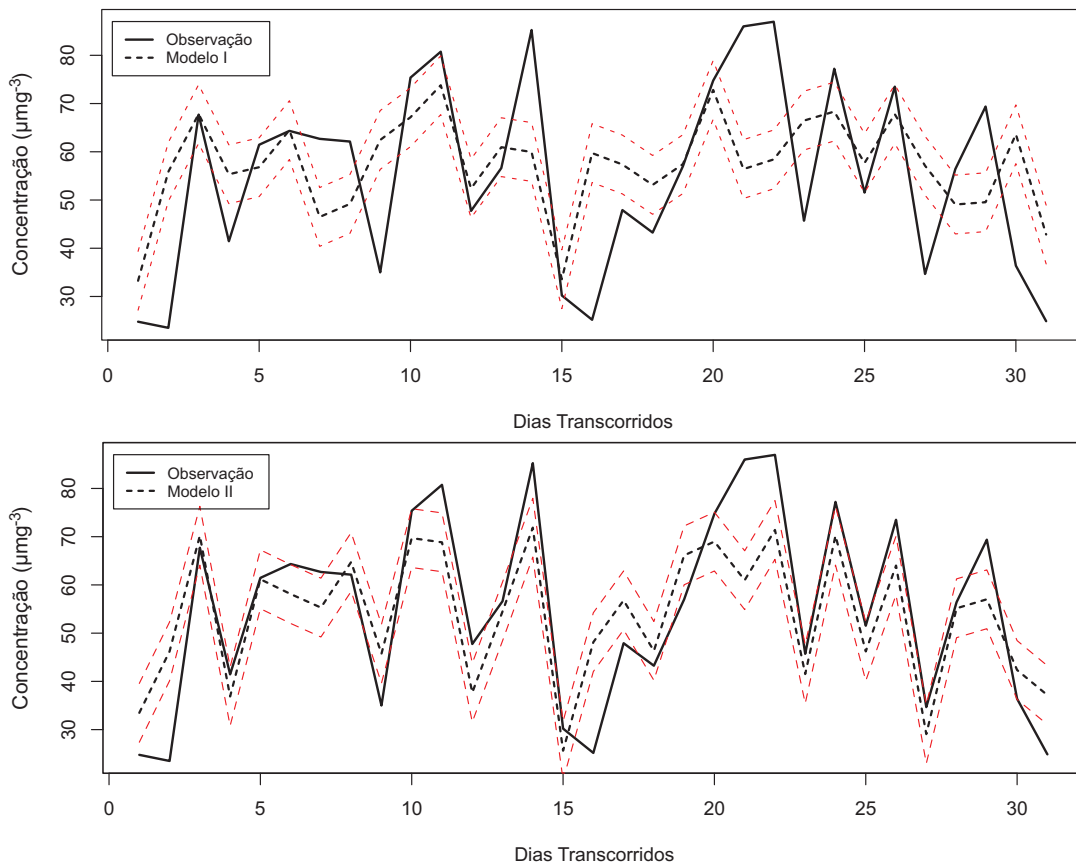


Figura 5.12: Previsão das concentrações de PM_{10} no período de 01 a 31 de Janeiro de 2005 para os modelos I e II;

Para medir a qualidade da previsão do modelos, usaremos a função de qualidade $Q : \mathbb{R}^2 \rightarrow [0, 1]$ proposta por [Stadlober, Hormann & Pfeiler \(2008\)](#) e que atribui um valor em $[0, 1]$ para cada par ordenado ($O =$ Observação, $P =$ Previsão). Os autores, aplicaram esta metodologia em um estudo de previsão para as cidades de Bolzano, Klagenfurt e Graz, Austria e os valores próximos de 1 significam boas previsões enquanto que próximos de 0 indicam baixa qualidade.

A função Q é dada pela relação

$$Q(O, P) = 1 - \min \left\{ \frac{a \cdot |O - P|}{D}, 1 \right\}, \quad (5.1)$$

onde

$$D = 1 + \frac{1}{2} \sqrt{|O - 50| + |P - 50| + b \cdot I_{\{O \leq 50, P \leq 50\}} + c \cdot I_{\{O \geq 150, P \geq 150\}}}$$

que inclui a função indicadora I_E definida por: $I_E = 1$ se a condição E for satisfeita, e $I_E = 0$ se não. Os parâmetros $a = 10^{-1}$, $b = 10^2$ e $c = 10^3$ foram utilizados com base em avaliações feitas por [Stadlober, Hormann & Pfeiler \(2008\)](#). A função indicadora I_E , mede a capacidade do modelo em capturar elevados picos (valores de concentração para qualidade do ar dita “inadequada”) e baixas concentrações.

Baseado no resultado de $Q(O, P)$ foi atribuído a seguinte classificação: $Q(O, P) \geq 0.8 \Rightarrow$ “excelente”, $Q(O, P) \in [0.6, 0.8) \Rightarrow$ “bom”, $Q(O, P) \in [0.4, 0.6) \Rightarrow$ “satisfatório”, $Q(O, P) \in [0.2, 0.4) \Rightarrow$ “ruim” e $Q(O, P) \in [0, 0.2) \Rightarrow$ “péssimo”.

A Tabela 5.8 apresenta a classificação de cada previsão para os modelos I e II, respectivamente, confirmando o melhor desempenho do modelo II. Esse modelo apresentou 100% das previsões classificadas nas três primeiras categorias (“excelente”(61.3%), “bom”(35.48%) e “satisfatório”(3.22%)) contra 70.96% do modelo I. O índice de qualidade média das previsões para o modelo I foi 0.627 (“bom”) enquanto para o modelo II ficou 0.814 (“excelente”). Resultados semelhantes foram encontrados por [Stadlober, Hormann & Pfeiler \(2008\)](#), onde a função de qualidade de previsão Q obteve 86% e 97% dos valores classifi-

Tabela 5.8: Combinação dos pares (O, P) para 01 a 31 de janeiro de 2006 com o índice de qualidade correspondente (1-excelente; 5-péssimo)

| MODELO I | | | MODELO II | | |
|-----------------|--------------|---------------|-----------------|--------------|---------------|
| Observação (O) | Previsão (P) | Qualidade (Q) | Observação (O) | Previsão (P) | Qualidade (Q) |
| 24.77 | 33.26 | 1 | 24.77 | 33.48 | 1 |
| 24.00 | 55.85 | 5 | 24.00 | 46.21 | 2 |
| 67.70 | 67.82 | 1 | 67.70 | 70.20 | 1 |
| 41.45 | 55.33 | 3 | 41.45 | 36.87 | 1 |
| 61.45 | 56.80 | 1 | 61.45 | 61.16 | 1 |
| 64.33 | 64.48 | 1 | 64.33 | 58.10 | 1 |
| 62.68 | 46.49 | 3 | 62.68 | 55.30 | 2 |
| 62.12 | 49.18 | 3 | 62.12 | 64.75 | 1 |
| 35.00 | 62.45 | 4 | 35.00 | 45.71 | 1 |
| 75.37 | 67.15 | 1 | 75.37 | 69.70 | 1 |
| 81.00 | 73.79 | 1 | 81.00 | 68.83 | 2 |
| 48.00 | 52.40 | 2 | 48.00 | 37.66 | 1 |
| 56.62 | 60.95 | 1 | 56.62 | 54.50 | 1 |
| 85.00 | 59.94 | 3 | 85.00 | 71.87 | 2 |
| 30.20 | 33.57 | 1 | 30.20 | 25.62 | 1 |
| 25.16 | 59.72 | 5 | 25.16 | 48.13 | 2 |
| 47.91 | 57.36 | 2 | 47.91 | 56.79 | 2 |
| 43.00 | 53.13 | 2 | 43.00 | 46.32 | 1 |
| 57.00 | 57.44 | 1 | 57.00 | 66.12 | 2 |
| 75.00 | 72.81 | 1 | 75.00 | 68.98 | 1 |
| 86.00 | 56.41 | 4 | 86.00 | 60.98 | 3 |
| 86.95 | 58.41 | 4 | 86.95 | 71.39 | 2 |
| 45.70 | 66.46 | 4 | 45.70 | 41.51 | 1 |
| 77.21 | 68.33 | 2 | 77.21 | 70.20 | 1 |
| 51.54 | 57.67 | 2 | 51.54 | 46.21 | 2 |
| 74.00 | 67.78 | 1 | 74.00 | 64.01 | 2 |
| 34.66 | 57.10 | 4 | 34.66 | 29.06 | 1 |
| 56.41 | 49.06 | 2 | 56.41 | 55.15 | 1 |
| 69.37 | 49.54 | 4 | 69.37 | 57.01 | 2 |
| 36.37 | 63.62 | 4 | 36.37 | 42.35 | 1 |
| 24.87 | 42.84 | 2 | 24.87 | 37.15 | 1 |
| Qualidade Média | | 2 | Qualidade Média | | 1 |

cados nas três primeiras classes para as cidades de Bolzano e Graz (Austria), respectivamente.

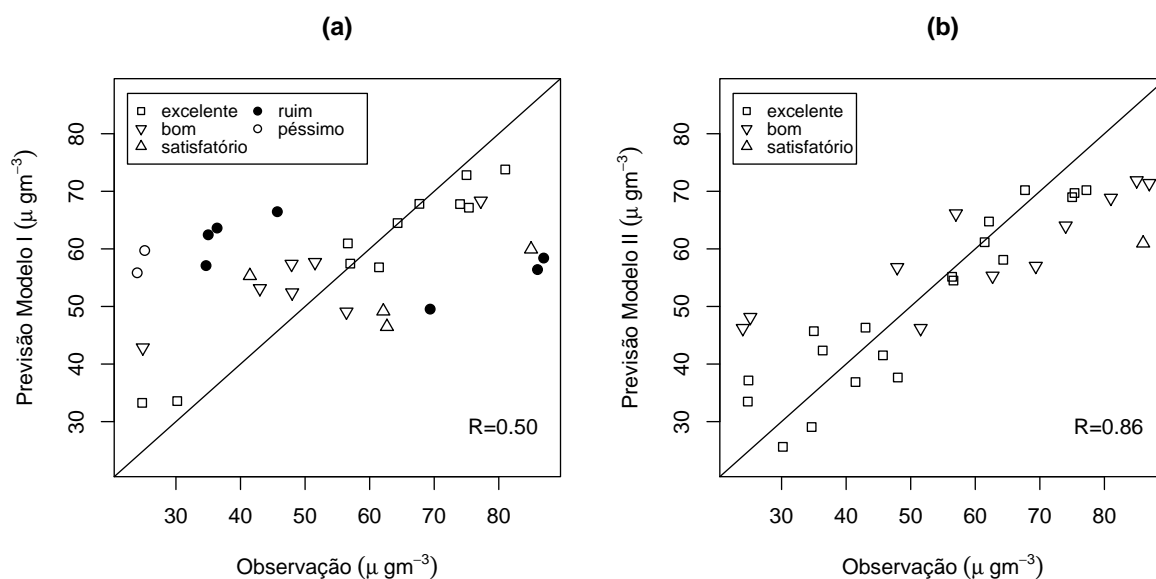


Figura 5.13: Observação \times Previsão; (a) Modelo I e (b) Modelo II

As Figuras 5.13(a) e (b) mostram um *scatter plot* dos valores observados *versus* os valores previstos pelos modelos I e II, respectivamente, classificados de acordo com a função de qualidade de previsão expressa pela Equação 5.1. A figura mostra uma clara diferença no desempenho dos modelos, o modelo I captura a média, ou seja, valores baixos são superestimados e picos subestimados, enquanto o modelo II consegue ser mais preciso, isto é, o modelo II prever as observações com menores erros, no entanto os picos são ainda subestimados.

Conclusões e Recomendações

6.1 Conclusão

A presente dissertação apresentou um estudo de previsão em uma série de concentração média diária de Material Particulado Inalável (PM_{10}) utilizando técnicas de modelagens de séries temporais. Para estimação do modelo selecionado, modelo $SARFIMA(p, d, q) \times (P, D, Q)_s$ com dois parâmetros de memória na presença de erros heterocedásticos e com propriedades de longa dependência, sazonalidade e volatilidade, utilizou-se a metodologia apresentada na Seção 4.1, proposta por [Reisen et al. \(2008\)](#).

A estrutura de correlação da série de PM_{10} , analisada nesse trabalho, apresentou as propriedades de sazonalidade, longa dependência nos períodos sazonais e entre esses períodos, e heterocedasticidade residual. Propriedades incorporadas a modelagem desta série.

A série das concentrações diárias de PM_{10} usada neste estudo, apresentava diversas observações faltante, que foram estimadas através do Filtro de Kalman, como proposto por [Palma & Chan \(1997\)](#).

Para verificar a acurácia do modelo proposto, o modelo $SARFIMA(p, d, q) \times (P, D, Q)$ - $ARCH(2)$, a metodologia foi aplicada a um conjunto de dados de poluição do ar e comparado ao modelo $SARMA-GARCH$. Os resultados dessa aplicação mostrou que o modelo $SARFIMA(1, \hat{d}, 0) \times (0, \hat{D}, 0)$ - $ARCH(2)$ conseguiu uma melhor captura da variabilidade dos dados, obtendo um ajuste com melhores medidas de erro. Para verificar a qualidade das previsões dadas pelos modelos I e II, foi implementada uma função de qualidade de previsão proposta por [Stadlober, Hörmann & Pfeiler \(2008\)](#), em que esta função mostrou que 96,78% das previsões dadas pelo modelo II apresenta qualidade classificadas como “excelente” ou “bom”, contra 54,8% do modelo I.

6.2 Recomendações

- Utilizar técnicas de modelagens que efetuem a estimação dos parâmetros dos modelos, assumindo a presença de observações atípicas (*outliers*). [Molinales, F. A. F., \(2007\)](#) é exemplo de estudo teórico para a estimação de modelos com um parâmetro de memória na presença de *outliers*;
- Efetuar um estudo de comparação das diversas técnicas de imputação de dados presente na literatura. Neste contexto, podemos citar [Junninen et al. \(2004\)](#).

- Estimar a Função de Autocorrelação da série considerando a presença das observações faltantes. [Yajima et al. \(1999\)](#) efetuou um estudo teórico com simulações para a estimação da FAC quando a série de dados apresenta valores faltantes.

Referências Bibliográficas

- [1] ALMEIDA, M. A. I., 2006. Modelo Aditivo Generalizado (MAG) no estudo da relação entre o número de atendimentos hospitalares por causas respiratórias e a qualidade do ar. Dissertação de Mestrado, Vitória: Programa de Pós-Graduação em Engenharia Ambiental: Universidade Federal do Espírito Santo.
- [2] ARTECHE, J. e ROBINSON, P.M., 2000. Semiparametric inference in seasonal and cyclical long memory processes. *Journal Time Series Analysis* 21(1), 1-25.
- [3] BERA, A. K. e JARQUE C. M., 1981. Efficient tests for normality, homoscedasticity and serial independence of regression residuals: Monte Carlo evidence. *Economics Letters* 7, 313-318.
- [4] BOLLERSLEV, T., 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307-327.
- [5] BOX, G. E. P. e PIERCE, D. A., 1970, Distribution of residual correlations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association* 65, 1509-1526.
- [6] BOX, G. E. P. e JENKINS, G. M., 1970. Time Series Analysis, Forecasting and Control. Holden-Day.
- [7] MARTINS, L. C, LATORRE, M.R.D.O, CARDOSO, M. R. A, et al., 2004. Poluição atmosférica e atendimentos por pneumonia e gripe em São Paulo, Brasil, *Revista de Saúde Pública* 36(1), 88-94.
- [8] CHELANI, A. B. e DEVOTTA, S., 2005. Air quality forecasting using a hybrid autoregressive and nonlinear model. *Atmospheric Environment* 40, 1774-1780.
- [9] CHANGA, C. C., TSAIB, S. S., HOA, S. C. e YANG, C. Y., 2005. Air pollution and hospital admissions for cardiovascular disease in Taipei, Taiwan. *Environmental Research* 98, 114-119.
- [10] DOUKHAN, P., OPPENHEIM, G. e TAQQU, M. S., 2003. Theory and Applications of Long-Range Dependence. Birkhäuser.

- [11] ENGLE, R.F., 1982. Autoregressive conditional heterocedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987-1007.
- [12] FINZI, G. e TEBALDI G., 1982. A mathematical model for air pollution forecast and alarm in an urban area, *Atmospheric Environment* 16, 2055-2059.
- [13] GEWEKE, J. E PORTER-HUDAK, S., 1983. The estimation and application of long memory times series model. *Journal of the Time Series Analysis* 4(4), 221-238.
- [14] GIRAITIS, L. e LEIPUS, R., 1995. A generalized fractionally diferencing approach in long memory modelling. *Lithuanian Mathematical Journal* 35, 53-65.
- [15] GOYAL, P., CHAN, A. T. e JAISWAL, N., 2006. Statistical models for the prediction of respirable suspended particulate matter in urban cities. *Atmospheric Environment* 40, 2068-2077.
- [16] GRANGER, C. W.J. e JOYUEX, R., 1980. An introduction to long-memory times series models and fractional differencing. *Journal of the Time Series Analysis* 1, 15-29.
- [17] GOOIJER, J. G. e HYNDMAN, R. J., 2006. 25 years of time series forecasting. *International Journal of Forecasting* 22, 443-473.
- [18] GOURIÉROUX, C., 1997. ARCH Models and Financial Applications. Springer.
- [19] HAMILTON, J. D., 1994. Time Series Analysis. Princeton University Press.
- [20] HASSLER, U, 1994. (Mis)specification of long memory and seasonal time series. *Journal Time Series Analysis* 15(1), 19-30.
- [21] HOSKING, J., 1981. Fractional Differencing. *Biometrika* 68(1), 165-176.
- [22] HURST, H.E., 1951. Long-term storage capacity of reservoirs. *Transaction of the American Society of Civil Engineers*, 116, 770-799.
- [23] IBAMA. Resoluções CONAMA, 1984/91, 1992. 4a. ed. rev. e aum, Brasília.
- [24] IGLESIAS, P., JORQUERA, H. e PALMA, W., 2006. Data analysis using regression models with missing observations and long memory: an application study. *Computational statistics e Data Analysis* 50, 2028-2043.
- [25] JENSEN, M. J., 1999. Using wavelets to obtain a consistent ordinary least squares estimator of the long-memory parameter. *Journal of Forecasting* 18, 17-32.
- [26] JORQUERA, H., PÉREZ, R., CIPRIANO, A., ESPEJO, A. S., LETELIER, M. V. e ACUÑA G., 1998. Forecasting Ozone Daily Maximum Levels at Santiago, Chile. *Atmospheric Environment* Vol. 32(20), 3415-3424.
- [27] JUNNINEN, H., NISKAA, H., TUPPURAINENC, K., RUUSKANENA, J. e KOLEHMAINEN, M., 2004. Methods for imputation of missing values in air quality data sets. *Atmospheric Environment* 38, 2895-2907.
- [28] LJUNG, G. M. e BOX, G. E. P., 1978. On a measure of lack of fit in time series models. *Biometrika* 65, 553-564.

- [29] McALEER, M., CHAN, F., 2005. Modeling trend and volatility in atmospheric carbon dioxide concentrations. *Environmental Modelling and Software* 21, 1273-1279.
- [30] McCOLLISTER, G. M. e WILSON, K R., 1975. Linear stochastic models for forecasting daily maxima and hourly concentrations of air pollutants. *Atmospheric Environment* 9, 417-423.
- [31] MOLINARES, F. A. F., 2007. Estimação robusta em processos de memória longa na presença de *outliers* aditivos. Dissertação de Mestrado, Recife: Programa de Pós-Graduação em Estatística: Universidade Federal de Pernambuco.
- [32] OSTRO, B., SANCHES, J. M., ARANDA, C. e ESKELAND, G. S., 1996. Air pollution and mortality: results from a study of Santiago, Chile. *Journal of Exposure Analysis and Environmental Epidemiology* 6, 97-114.
- [33] OSTRO, B., HURLEY, S. e LIPSETT, M. J., 1998b. Air Pollution and Daily Mortality in the CoachellaValley, California: A Study of PM₁₀ Dominated by Coarse Particles. *Environmental Research* A81, 231-238.
- [34] OSTRO, B. D., ESKELAND, G. S., SÁNCHEZ, J. M. e FEYZIOGLU T., 1999. Air pollution and health effects: A study of medical visits among children in Santiago, Chile. *Environ Health Perspect* 107(1), 69-73.
- [35] PALMA, W., CHAN, N.H., 1997. Estimation and forecasting of long-memory processes with missing values. *Journal of Forecasting* 16, 395-410.
- [36] PALMA, W., DEL PINO, G., 1999. Statistical analysis of incomplete long-range dependent data. *Biometrika* 86, 165-172.
- [37] PALMA, W., CHAN, N.H., 2005. Efficient Estimation of Seasonal Long-Range-Dependent Processes. *Journal of Time Series Analysis*, 26(6), 863-892.
- [38] PEIRIS, M. S. e SINGH, N., 1996. Predictors for Seasonal and Nonseasonal Fractionally Integrated ARIMA Models. *Biometrics*, 38(6), 741-752.
- [39] PEREZ, P. e REYES, J., 2002. Prediction of maximum of 24-h average of PM₁₀ concentrations 30 h in advance in Santiago, Chile. *Atmospheric Environment* 36, 4555-4561.
- [40] PORTER-HUDAK, S, 1990. An application of the seasonal fractionally differenced model to the monetary aggregates. *Journal American Statistical Assoc.* 85, 338-344.
- [41] REISEN, V. A., 1994. Estimation of the fractional difference parameter in the ARIMA(p,d,q) model using the smoothed periodogram. *Journal of the Time Series Analysis* 15(3), 335-350.
- [42] REISEN, V. A. e LOPES, S., 1998. Some simulations and applications of forecasting long-memory time-series models. *Journal of Statistical Planning and Inference* 80, 269-287.
- [43] REISEN, V. A., RODRIGUES, A. L. e PALMA, W., 2006a. Estimation of seasonal fractionally integrated processes. *Computational Statistics e Data Analysis* 50, 568-582.

- [44] REISEN, V. A., RODRIGUES, A. L. e PALMA, W., 2006b. Estimating seasonal long-memory processes: a Monte Carlo study. *Journal of Statistical Computation and Simulation* 76(4), 305-316.
- [45] REISEN, V. A., ZAMPROGNO, B., PALMA, W and ARTECHE, J., 2008. Estimating multiple fractional seasonal long-memory parameters. Preprint.
- [46] ROBESON S.M. e STEYN, D.G., 1990. Evaluation and comparison of statistical forecast models for daily maximum ozone concentrations. *Atmospheric Environment. Part B.* 24(2), 303-312.
- [47] SENA Jr., M. R., REISEN, V. A. e LOPES, S. R., 2006. Correlated Error in the Parameters Estimation of the ARFIMA Model: A simulated Study. *Communications in Statistics - Simulation and Computation* 35, 789-802.
- [48] SHAPIRO, S. S e WILK, M. B, 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52, 591-611.
- [49] SHIVELY, T. S., 1990. An analysis of the long-term trend in ozone data from two Houston, Texas monitoring sites *Atmospheric Environment. Part B. Urban Atmosphere* 24(2), 293-301.
- [50] SOUZA Jr., P. A., KLINGELHÖFER, G., GUIMARÃES, A. F., MORIMOTO, T., QUEIROZ, R. S. de, 2001. Intelligent Receptor Modelling: A Receptor Modelling Based on Adaptive Techniques. AeWMA Conference.
- [51] STADLOBER, E., HÖRMANN, S e PFEILER, B., 2008. Quality and performance of a PM10 daily forecasting model. *Atmospheric Environment* 42, 1098-1109.
- [52] TERRY, W.R., LEE, J.B. e KUMAR, A., 1986. Time series analysis in acid rain modeling: Evaluation of filling missing values by linear interpolation. *Atmospheric Environment (1967)* 20(10), 1941-1943.
- [53] TEYSSIÈRE, G. e KIRMAN, A. P., 2006. Long Memory in Economics. Springer.
- [54] TSAY, R. S., 2002. Analysis of Financial Time Series. John Wiley e Sons, Inc.
- [55] WHITCHER, B., 2004. Wavelet-based estimation for seasonal long-memory processes. *Technometrics* 46, 225-238.
- [56] WOODWARD, W. A., CHENG, Q. C. and GRAY, H.L., 1998. A k -factor GARMA long-memory model. *Journal of Time Series Analysis* 19, 485-504.
- [57] YAJIMA, Y., NISHINO, H., 1999. Estimation of the autocorrelation function of a stationary time series with missing observations. *The Indian Journal of Statistics* 61, 189-207.
- [58] YULE, G. U., 1927. On the method of investigating periodicities in disturbed series, with special reference to Wölfer's sunspot numbers. *Philosophical Transactions of the Royal Society London A* 226, 267-298.
- [59] ZOLGHADRI, A. e CAZAURANG, F., 2005. Adaptive nonlinear state-space modelling for the prediction of daily mean PM₁₀ concentrations. *Environmental Modelling and Software* 21, 885-894.