

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

Juliana Bottoni de Souza

Análise de Componentes Principais e a Modelagem Linear Generalizada: uma associação entre o número de atendimentos hospitalares por causas respiratórias e a qualidade do ar, na Região da Grande Vitória, ES.

Vitória
2013

Juliana Bottoni de Souza

Análise de Componentes Principais e a Modelagem Linear Generalizada: uma associação entre o número de atendimentos hospitalares por causas respiratórias e a qualidade do ar, na Região da Grande Vitória, ES.

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 de poluição do ar.

Orientador: Prof. Dr. Valdério Anselmo Reisen.

Co-orientador: Prof. Dra. Jane Méri dos Santos

Vitória
2013

Dados Internacionais de Catalogação-na-publicação (CIP)
(Biblioteca Central da Universidade Federal do Espírito Santo, ES, Brasil)

S729a Souza, Juliana Bottoni de, 1988-
Análise de Componentes Principais e a Modelagem Linear
Generalizada: uma associação entre o número de atendimentos
hospitalares por causas respiratórias e a qualidade do ar,
na Região da Grande Vitória, ES / Juliana Bottoni de Souza. - 2013.
66 f.: il

Orientador: Valdério Anselmo Reisen.

Coorientadora: Jane Meri Santos.

Dissertação (Mestrado em Engenharia Ambiental)-
Universidade Federal do Espírito Santo, Centro Tecnológico.

1. Análise de componentes principais. 2. Ar - Qualidade
- Vitória, Região Metropolitana de (ES).3. Análise de
séries temporais. 4. Saúde pública. 5. Doenças respiratórias
em crianças.I. Reisen, Valdério Anselmo, 1961-.II. Santos,
Jane Meri.III. Universidade Federal do Espírito Santo.
Centro Tecnológico.IV. Título.

CDU: 628

Juliana Bottoni de Souza

Análise de Componentes Principais e a Modelagem Linear Generalizada: uma associação entre o número de atendimentos hospitalares por causas respiratórias e a qualidade do ar, na Região da Grande Vitória, ES.

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 de poluição do ar.

COMISSÃO EXAMINADORA

Prof. Dr. Valdério Anselmo Reisen - UFES

Prof. Dr. Antonio Ponce de Leon - UERJ

Prof. Dr. Neyval Costa Reis Júnior - UFES

A DEUS em primeiro lugar, a minha família, amigos, meu namorado e a todos que acreditaram na
minha vitória.

”Resista um pouco mais, mesmo que as feridas latejem e que sua coragem esteja cochilando. Resista mais um minuto e será fácil resistir aos demais. Resista mais um instante, mesmo que a derrota seja um ímã, mesmo que a desilusão caminhe em sua direção. Resista mais um pouco, mesmo que os invejosos digam para você parar, mesmo que sua esperança esteja no fim. Resista mais um momento, mesmo que você não possa avistar ainda a linha de chegada, mesmo que as inseguranças brinquem de roda à sua volta. Resista um pouco mais, mesmo que a sua vida esteja sendo pesada como a consciência dos insesatos e você se sinta como um pássaro de asas quebradas. Resista, porque o último instante da madrugada é sempre aquele que puxa a manhã pelo braço e essa manhã bonita, ensolarada, sem algemas, nascerá pra você em breve, desde que você resista. Resista, porque estamos sentados na arquibancada do tempo, torcendo ansiosos para que você vença e ganhe de Deus o troféu que você merece: a FELICIDADE!”. (Autor desconhecido)

Agradecimentos

- A DEUS sobre todas as coisas.
- A minha mãe Jussileni Bottoni de Souza e a meu pai João Gomes de Souza, aos meus irmãos pelos ensinamentos, paciência, cuidado, amor e dedicação ao longo dessa jornada.
- Ao meu namorado Rodrigo Lucio Mattos de Oliveira pelo amor, compreensão e apoio sempre me dedicado.
- Ao professor orientador Valdério Anselmo Reisen que durante esses anos de convivência me proporcionou ensinamentos, conselhos e principalmente sempre acreditou no meu potencial.
- A professora Co-orientadora Jane Méri dos Santos pelo apoio e conhecimento transmitido.
- Aos amigos conquistados durante o mestrado.
- Aos amigos do nunes que sempre me ajudaram com o conhecimento, conselhos e motivações para a conclusão desse trabalho.
- A todos os professores que contribuíram para a minha formação, com seus ensinamentos, conselhos e experiência.
- Aos funcionários da Universidade e do programa de Pós-graduação em Engenharia Ambiental em especial a Rose Mary Nunes Leão que sempre me ajudou.
- Àquelas pessoas que direta ou indiretamente colaboraram na realização e concretização dessa pesquisa.
- A Emersom Pedreira Matos por ter cedido o banco de dados;
- Ao Instituto Estadual de Meio Ambiente e Recursos Hídricos (IEMA), pelo fornecimento dos dados de poluição do ar;

Resumo

Esta dissertação faz uso do casamento entre duas ferramentas estatísticas, a Análise de Componentes Principais (ACP) e o Modelo Aditivo Generalizado (MAG), para estimar o efeito da associação entre a exposição atmosférica dos poluentes PM_{10} , SO_2 , NO_2 , O_3 e CO e o número de atendimentos por doenças respiratórias em crianças menores de 6 anos na região da Grande Vitória, ES, Brasil. Usualmente os poluentes atmosféricos são considerados as covariáveis explicativas no MAG; porém, por possuírem estrutura de autocorrelação, estes devem ser utilizados com cautela. A técnica de ACP fornece novas variáveis ortogonais, essas são combinações lineares das variáveis ambientais, nessa direção as componentes principais serão utilizadas no MAG, denotado aqui por MAG-ACP. Entretanto, as componentes principais obtidas por meio da matriz de variância/covariância aplicada a processos indexados no tempo também apresentam as propriedades de correlação temporal. Como procedimento alternativo, para atenuar a correlação temporal das componentes o método de filtragem é sugerido para transformar os dados atmosféricos num processo Ruído Branco. A matriz de resíduos é utilizada para obter as componentes e essas aplicadas ao modelo MAG - método aqui denominado VAR-MAG-ACP. Os resultados empíricos mostram que esse modelo remove as autocorrelações das componentes principais e indica estimativas mais significantes do Risco Relativo (RR) para cada poluente. Os resultados comprovam as hipóteses estabelecidas, as componentes principais selecionadas possuem correlação e autocorrelação nas defasagens. No ajuste final do modelo MAG-ACP foi necessária a inclusão de modelos do tipo ARMA(p,q) nos resíduos, já que esses carregavam as estruturas de autocorrelação presente nas componentes. O modelo VAR-MAG-ACP, além de produzir estimativas mais significantes no RR, gerou melhores ajustes residuais. Comparadas à modelagem MAG usual, as duas vertentes propostas apresentaram melhores resultados, tanto na estimativa do RR quanto na qualidade do ajuste.

Palavras Chave: Poluição do Ar Análise de Componentes Principais Doenças Respiratórias Modelo Aditivo Generalizado Modelo Vetorial Autoregressivo Séries Temporais.

Abstract

This dissertation uses two statistical tools, Principal Component Analysis (PCA) and Generalized Additive Model (GAM), jointly, to estimate the effect of the association between atmospheric exposure of PM₁₀, SO₂, NO₂, O₃ and CO and the number of admissions due respiratory diseases in children less than 6 years in the Região da Grande Vitória, Brazil. Usually the atmospheric pollutants are considered the explanatory covariables in MAG, but since they have an autocorrelation structure, they must be used with caution. The PCA technique provides a new set of orthogonal variables, these variables are linear combinations of environmental variables. Therefore, We use this approach in MAG, hereafter denoted by GAM-PCA. However, the principal components obtained through the matrix of variance / covariance applied to processes indexed by time also exhibit the properties of temporal correlation. A countermeasure to attenuate the temporal correlation of the components is use the filtering method to transform the data in an atmospheric white noise process. The residual matrix is used to obtain these components and applied to the model MAG - method here called VAR-GAM-PCA. The empirical results show that this model removes the autocorrelations of the main components and indicates significant estimates of relative risk (RR) for each pollutant. The results confirm the hypotheses established, the main components have selected correlation and the autocorrelation lags. To adjust the GAM-PCA model, an ARMA(p,q) model was used in the residual matrix since that structure carried autocorrelation from the original data. The VAR model-MAG-ACP, besides producing more significant in RR estimates, generated best fit residuals. Compared to the usual modeling MAG, the two strands proposals presented better results, both in estimating the RR and the quality of the fit.

Lista de Figuras

4.1	Concentrações dos poluentes atmosféricos em (a)-CO;(b)-NO ₂ ;(c)-SO ₂ ;(d)-PM ₁₀ ;(e)-O ₃ e o número de atendimentos diários (f)- Atendimentos	43
4.2	Função de Autocorrelação Amostral e Função Autocorrelação Amostral Parcial do poluente CO	43
4.3	Função de Autocorrelação Amostral e Função Autocorrelação Amostral Parcial do poluente O ₃	44
4.4	Função de Autocorrelação Amostral e Função Autocorrelação Amostral Parcial do poluente SO ₂	44
4.5	Função de Autocorrelação Amostral e Função Autocorrelação Amostral Parcial do poluente NO ₂	44
4.6	Função de Autocorrelação Amostral e Função Autocorrelação Amostral Parcial do poluente PM ₁₀	44
4.7	Correlação Cruzada(CCF) das componentes principais dos poluentes estudados.	46
4.8	Função de Autocorrelação amostral e Função de Autocorrelação Parcial amostral dos resíduos do modelo MAG-ACP.	47
4.9	Função de Autocorrelação amostral e Função de Autocorrelação Parcial amostral dos resíduos do modelo MAG-ACP.	48
4.10	Ajuste do modelo MAG-ACP-SAR	48
4.11	Correlação Cruzada(CCF) das componentes principais dos poluentes filtrados.	49
A.1	Série do número de atendimentos hospitalares por causas respiratórias diário de crianças de 0 a 6 anos da RGV.	62
A.2	(a)-Função de Autocorrelação Amostral; (b)-Função Autocorrelação Amostral Parcial	62
A.3	Box Plot das concentrações do poluente CO	63
A.4	Box Plot das concentrações do poluente NO ₂	63

A.5	Box Plot das concentrações do poluente O_3	63
A.6	Box Plot das concentrações do poluente PM_{10}	64
A.7	Box Plot das concentrações do poluente SO_2	64
A.8	Box Plot dos dados de atendimentos diários por doenças respiratórias em crinaças menores de 6 anos	64
A.9	Matriz de correlação dos poluentes	65
A.10	Gráfico de Scree plot das componentes principais	65
A.11	Gráfico do Biplot das componentes principais	65
A.12	(a)-Função de Autocorrelação Amostral; (b)-Função Autocorrelação Amostral Parcial	66
A.13	Resíduos do modelo final VAR-MAG-ACP.	66

Lista de Tabelas

3.1	Padrões Nacionais de qualidade do ar(Resolução CONAMA 003 de 28/06/1990).	21
3.2	OMS Diretrizes para Qualidade do Ar	22
4.1	Estatística descritiva dos atendimentos por doenças respiratórias nas áreas de abrangência de cada RAMQAr no período de jan/2005 a dez/2010	42
4.2	Correlação entre os poluentes, variáveis meteorológicas e atendimentos.	45
4.3	Resultado das cargas fatoriais e estatísticas da aplicação de análise de componentes principais.	45
4.4	Resultado da Regressão de Poisson para estimação dos efeitos dos poluentes nos atendimentos na RGV.	47
4.5	Resultado das cargas fatoriais e estatísticas da aplicação de análise de componentes principais(ACP) nos poluentes filtrados.	49
4.6	EQM estimado para os modelos MAG, MAG-ACP e VAR-MAG-ACP.	50
4.7	Risco Relativo(RR) e intervalo de confiança de 95% de atendimentos por doenças respiratórias em crianças menores de 6 anos para uma variação interquartilica dos poluentes PM ₁₀ , SO ₂ , NO ₂ e O ₃ e CO na RGV, jan-2005 a dez-2010.	51

Sumário

Abstract	8
Lista de Figuras	10
Lista de Tabelas	11
Lista de Abreviaturas	14
1 Introdução	15
2 Objetivo	18
2.1 Objetivo Geral	18
2.2 Objetivos Específicos	18
3 Revisão Bibliográfica	20
3.1 Padrões de Qualidade do Ar	20
3.2 Poluentes Ambientais	22
3.2.1 Material Particulado (PM_{10})	22
3.2.2 Dióxido de Nitrogênio (NO_2)	23
3.2.3 Ozônio (O_3)	23
3.2.4 Dióxido de Enxofre (SO_2)	24
3.2.5 Monóxido de Carbono (CO)	24
3.3 Revisão dos Estudos Envolvendo Poluição do Ar e Saúde Utilizando o Modelo Aditivo Generalizado (MAG) e a técnica de análise de componentes principais (ACP)	25

4 Artigo	31
4.1 Resumo	32
4.2 Introdução	33
4.3 Materiais e Métodos	35
4.3.1 Metodologia da Análise dos Dados	35
4.3.2 Risco Relativo	36
4.3.3 Modelo Aditivo Generalizado - MAG	38
4.3.4 Análise de Componentes Principais - ACP	39
4.3.5 MAG-ACP - Modelagem Aditiva Generalizada e Análise de Componentes Principais	40
4.3.6 VAR-MAG-ACP - Modelagem Vetorial autoregressiva e MAG-ACP	41
4.4 Resultados	42
4.5 Conclusão	51
4.6 Referências	53
5 Conclusão e Recomendações	56
Referências Bibliográficas	61
A	62

Lista de abreviaturas

ACP Análise de Componentes Principais

AIC Critério de Informação de Akaike

AR Auto regressivo

CID Classificação Internacional de Doenças

CO Monóxido de Carbono

CONAMA Conselho Nacional de Meio Ambiente

HINSG Hospital Infantil Nossa Senhora da Glória

IC Intervalo de Confiança

IEMA Instituto Estadual de Meio Ambiente e Recursos Hídricos

MAG Modelo Aditivo Generalizado

MAG-ACP Modelo Aditivo Generalizado com a Análise de Componentes Principais

MLG Modelo Linear Generalizado

NO₂ Dióxido de Nitrogênio

O₃ Ozônio

OMS Organização Mundial de Saúde

PM₁₀ Material Particulado

RAMQAr Rede Automática de Monitoramento da Qualidade do Ar

RGV Região da Grande Vitória

RR Risco Relativo

SO₂ Dióxido de Enxofre

VAR Modelo Vetorial Autoregressivo

VAR-MAG-ACP Modelo Vetorial autoregressivo com a modelagem aditiva generalizada e análise de componentes principais

Capítulo 1

Introdução

O impacto da poluição atmosférica sobre o bem-estar humano tem sido a principal motivação para o seu estudo e controle. A poluição atmosférica tem afetado a saúde da população, mesmo quando seus níveis encontram-se abaixo do que determina a legislação vigente (MARTINS et al 2001).

Os efeitos dos poluentes atmosféricos variam em função do tempo de exposição e de suas concentrações. De uma maneira geral, os efeitos podem ser classificados como: (a) agudos que estão relacionados a exposição a altas concentrações de poluentes e os efeitos são imediatos; e (b) crônicos que se relacionam a baixas concentrações de poluentes e os efeitos são a longo prazo.

Um nível elevado dos poluentes pode ocasionar desde irritação dos olhos, nariz e da garganta, bronquite e pneumonia até doenças respiratórias crônicas, câncer de pulmão, problemas cardíacos, etc. A exposição ao dióxido de nitrogênio (NO_2) pode levar ao decréscimo das funções respiratórias. As consequências da exposição ao dióxido de enxofre (SO_2) vão desde tosse e dor de garganta até danos permanentes ao sistema respiratório. Já a exposição ao monóxido de carbono (CO) pode levar a sérios problemas de coração e do sistema circulatório. A exposição ao material particulado (PM_{10}) pode ocasionar câncer de pulmão, redução da função pulmonar, etc. Um alto nível de exposição ao ozônio (O_3) pode ocasionar diversos problemas respiratórios como a asma e o câncer de pulmão (HOLGATE et al., 1999).

Esses efeitos devem ser considerados um problema de saúde pública, principalmente em grandes centros urbanos, onde suas concentrações são mais elevadas e seus efeitos são freqüentemente observados na

morbidade ¹ e mortalidade por doenças respiratórias e cardiovasculares, como o câncer de pulmão, diminuição da função respiratória, entre outros. Estudos relacionados ao tema sugerem que os grupos mais suscetíveis a exposição das concentrações dos poluentes, são as crianças e os idosos.

Essa exposição é considerada como um fator de risco para dor aguda e crônica de doenças respiratórias, e é um problema sério em cidades de todo o mundo, particularmente nos grandes centros dos países em desenvolvimento onde se estima que um quarto da população mundial esteja exposta à concentrações insalubres de poluentes atmosféricos. As crianças estão particularmente em risco devido à imaturidade dos seus sistemas de órgãos respiratórios.(WHO, 2012).

Os efeitos adversos dos poluentes atmosféricos na saúde humana têm sido uma preocupação dos órgãos regulamentadores de meio ambiente e de saúde pública ao longo dos últimos anos. A confirmação sobre esses efeitos na saúde pode ser obtida a partir de estudos populacionais através de investigação epidemiológica, orientando o desenvolvimento de práticas e legislação que visem o controle das emissões e da qualidade do ar.

Os estudos epidemiológicos que envolvem a análise de séries temporais objetivam verificar se os efeitos deletérios ou maléficos à saúde humana estão associados a um fator de risco, ou seja, busca-se verificar se os níveis de concentrações de determinados poluentes atmosféricos estão diretamente associados aos atendimentos hospitalares. Apesar dos diversos estudos sobre a associação entre mortalidade e poluição do ar nos últimos anos, as atenções se voltaram principalmente para as internações e admissões hospitalares por causas respiratórias ou cardiovasculares, como: Ostro et al. (1999), Saldiva et al. (2004), Chen et al. (2010).

As características de dados de séries temporais ambientais exigem métodos estatísticos complexos e suficientemente sensíveis, para detectar os efeitos combinados de outras covariáveis variantes no tempo. Isto é, no caso de estudos sobre a poluição do ar e saúde existem estruturas de tendências e flutuações sazonais, que interferem na dinâmica das variáveis observadas tais como fluxo de veículos, dias da semana, feriados, número de indústrias, entre outras. Além disso, influências de variáveis meteorológicas, como temperatura e umidade relativa do ar, também estão presentes na dispersão dos poluentes na atmosfera e na natureza de diversas doenças respiratórias (Schwartz et al 2000, Gouveia et al 2004).

¹que corresponde ao comprometimento provocado por determinada doença à saúde do paciente

Numerosos estudos de séries temporais têm investigado a associação entre resultados adversos para a saúde diária da população e as concentrações de poluição no ambiente, nestes estudos normalmente se ajusta um modelo de Poisson log-linear para séries temporais simultâneas entre a mortalidade ou morbidade diária e a poluição atmosférica, e em muitos deles o efeito na saúde se concentra na estimativa de um único poluente, no entanto é conhecido que existe uma alta correlação entre os poluentes. Avaliar os efeitos adversos para a saúde de uma combinação de poluentes pode ser tanto mais interpretável e mais viável do que tentar isolar os efeitos de um poluente apenas. Um dos métodos propostos para se investigar os efeitos de vários poluentes correlacionados é Análise de Componentes Principais (ACP) (Burnett et al 2003, Cox 2000).

A Análise de Componentes Principais (ACP) é uma técnica estatística utilizada para reduzir a dimensionalidade de um conjunto de dados onde há um grande número de variáveis inter-relacionadas. Isto é feito de forma que o máximo de variabilidade presente nos dados seja mantido. A aplicação desta técnica no presente trabalho se deve ao fato de se querer selecionar os grupos (componentes) em análise para se inserir no modelo de regressão utilizado.

Nesse contexto, o presente estudo investigou a associação entre as concentrações dos poluentes atmosféricos e os atendimentos diários por causas respiratórias em crianças menores de 6 anos na Região da Grande Vitória (RGV), aplicando a técnica multivariada de análise de componentes principais (ACP) para avaliar os efeitos dos poluentes de uma forma conjunta, eliminando a correlação entre eles. Os modelos de regressão utilizados foram o Modelo Aditivo Generalizado (MAG), e suas extensões como o MAG-ACP e o Modelo Vetorial Autoregressivo VAR-MAG-ACP. Os efeitos da poluição sobre a saúde, foram quantificados através do Risco Relativo (RR), que expressa a probabilidade de um indivíduo desenvolver uma doença relativo à exposição a um fator de risco. A estimativa do RR também foi utilizada para comparação entre os modelos propostos.

Esta dissertação é apresentada como segue: O Capítulo 2 descreve os objetivos da pesquisa. O Capítulo 3 apresenta a revisão da literatura. O artigo central desta dissertação é o motivo do Capítulo 4. Finalmente no Capítulo 5 é apresentada as conclusões e recomendações para trabalhos futuros.

Capítulo 2

Objetivo

2.1 Objetivo Geral

Este trabalho tem como objetivo avaliar a aplicação da técnica de Análise de Componentes Principais (ACP) juntamente com o Modelo Aditivo Generalizado (MAG), na associação entre o número de atendimentos hospitalares por causas respiratórias e a concentração média dos poluentes atmosféricos.

2.2 Objetivos Específicos

- Analisar as séries temporais do número de atendimentos por causas respiratórias, os dados de concentração dos poluentes atmosféricos (PM_{10} , SO_2 , NO_2 , CO e O_3) e as variáveis meteorológicas (Temperatura - máxima e mínima, Umidade relativa do ar) na RGV;
- Aplicar a técnica de análise multivariada denominada Análise de Componentes Principais (ACP);
- Verificar se a técnica de ACP aplicada aos dados pode melhorar o ajuste e o risco relativo associado do modelo;
- Verificar os resíduos das componentes principais, se estas são autocorrelacionadas;
- Aplicar a metodologia de filtragem dos dados, denominada VAR-MAG-ACP;

- Comparar os resultados obtidos dos modelos propostos MAG-ACP e VAR-MAG-ACP com a modelagem MAG usual.

Capítulo 3

Revisão Bibliográfica

A exposição aos poluentes ambientais é reconhecida como um importante fator de risco para a ocorrência de determinadas doenças respiratórias e não respiratórias, como as cardiovasculares. Embora os efeitos dos poluentes possam estar relacionados à mortalidade, diversos estudos têm se concentrado nas admissões e internações hospitalares.

Para a investigação dessa associação e dos limites toleráveis de concentração dos poluentes na atmosfera, foram estabelecidos os padrões de qualidade do ar, que definem legalmente o limite máximo para a concentração de um poluente na atmosfera. Esses padrões são baseados em estudos científicos que assegurem a que níveis de concentração são toleráveis e proporcionem uma margem de segurança adequada.

3.1 Padrões de Qualidade do Ar

A resolução nº 03 do Conselho Nacional do Meio Ambiente - CONAMA, de 28 de junho de 1990 (que será aqui referida como Resolução CONAMA 03/1990) define como poluente atmosférico qualquer forma de matéria ou energia com intensidade e em quantidade, concentração, tempo ou características em desacordo com os níveis estabelecidos, e que tornem ou possam tornar o ar:

- I - impróprio, nocivo ou ofensivo à saúde;
- II - inconveniente ao bem-estar público;
- III - danoso aos materiais, à fauna e flora.
- IV - prejudicial à segurança, ao uso e gozo da propriedade e às atividades normais da comunidade.

Nesta mesma resolução são definidos os padrões de qualidade do ar como sendo as concentrações de poluentes atmosféricos que ultrapassadas, poderão afetar a saúde, a segurança e o bem estar da população, bem como ocasionar danos à flora e à fauna, aos materiais e ao meio ambiente em geral, conforme Tabela 3.1.

Tabela 3.1: Padrões Nacionais de qualidade do ar (Resolução CONAMA 003 de 28/06/1990).

Poluente	Tempo de Amostragem	Padrão Primário μgm^3	Padrão Secundário μgm^3	Método de Amostragem
PTS	24 horas ¹	240	150	separação/inercial filtração
	MGA ²	80	60	
PM ₁₀	24 horas ¹	150	150	separação/inercial filtração
	MMA ³	50	50	
SO ₂	24 horas	365	100	Infravermelho não dispersivo
	MAA ³	80	40	
NO ₂	1 hora	320	190	Quimiluminescência
	MAA	100	100	
CO	1 hora	40.000/35 ppm	40.000/35 ppm	Pararosanilina
	8 horas	10.000 (9ppm)	10.000 (9ppm)	
O ₃	1 hora ¹	160	160	Quimiluminescência

Fonte: CONAMA (1990)

(1) Não deve ser excedido mais que uma vez ao ano

(2) Média geométrica anual

(2) Média aritmética anual

Na Tabela 3.1 entende-se como padrão primário de qualidade do ar as concentrações dos poluentes que, ultrapassadas, poderão afetar a saúde da população. Já o padrão secundário de qualidade do ar representa as concentrações de poluentes abaixo das quais se prevê o mínimo efeito sobre o bem-estar da população, assim como o mínimo a fauna, à flora, e aos materiais e ao meio ambiente em geral.

Os padrões nacionais de qualidade do ar, fixados pela resolução CONAMA 03/90, estão defasados e seguiram os padrões europeus e americanos da época. Em 2005, a organização mundial de saúde (OMS) divulgou um relatório recomendando mudanças na medição da qualidade do ar, argumentando que 2 milhões de pessoas morrem por ano em todo o mundo em decorrência da poluição. Enquanto a OMS recomenda que a concentração diária de PM_{10} não pode ultrapassar $50\mu g/m^3$, o CONAMA 03/90 tem como parâmetro $150\mu g/m^3$; para o O_3 cujo índice recomendável passou a ser de $100\mu g/m^3$ a resolução CONAMA determina $160\mu g/m^3$ e para o SO_2 a OMS determina $20\mu g/m^3$ média diária o CONAMA 03/90 estabelece o máximo aceitável em $100\mu g/m^3$. As diretrizes da OMS para a qualidade do ar são mostradas na Tabela 3.2.

Tabela 3.2: OMS Diretrizes para Qualidade do Ar

Poluente	Diretriz	
PM_{25}	$10\mu g/m^3$ média anual	$25\mu g/m^3$ média 24 horas
PM_{10}	$20\mu g/m^3$ média anual	$50\mu g/m^3$ média 24 horas
O_3	$100\mu g/m^3$ média 8 horas	
SO_2	$20\mu g/m^3$ média 24 horas	$50\mu g/m^3$ média 10 minutos
NO_2	$40\mu g/m^3$ média anual	$200\mu g/m^3$ média 1 hora

Segundo a OMS (2005) estima-se que a redução dos níveis de PM_{10} de $50\mu g/m^3$ para $20\mu g/m^3$ poderia reduzir as mortes nas cidades poluídas em até 9% ao ano.

3.2 Poluentes Ambientais

3.2.1 Material Particulado (PM_{10})

Material particulado (MP) é o termo genérico utilizado para um tipo de poluente do ar, constituído por misturas complexas e variando em tamanho e composição, produzidos por uma grande variedade de recursos naturais e atividades antrópicas (Poschl, 2005). As principais fontes de poluição são indústrias, incineradores, veículos, atividades de construção civil, poeira, etc. O tamanho das partículas varia ($PM_{2.5}$ e PM_{10} para diâmetro aerodinâmico menor que $2.5\mu m$ e $10\mu m$, respectivamente). O tamanho das partículas está diretamente ligada ao seu potencial de causar doenças e local de deposição no corpo humano. As partículas de PM_{10} depositam-se principalmente no trato respiratório superior, enquanto que as partículas ultrafinas são capazes de atingir os alvéolos pulmonares.

Os efeitos gerais sobre a saúde humana: as partículas como o PM_{10} são as que causam maiores prejuízos à saúde, uma vez que não são retidas pelas defesas do organismo. Essas podem causar irritação nos olhos e na garganta, reduzindo a resistência às infecções e ainda provocando doenças crônicas. Além disso, atingem as partes mais profundas dos pulmões, no caso do $PM_{2.5}$, transportando para o interior do sistema respiratório substâncias tóxicas e cancerígenas.(WHO,2006)

Os efeitos gerais sobre o meio ambiente são alterações da visibilidade; alteração no balanço de nutrientes de lagos, rios e do solo; danificação da vegetação e alteração na diversidade do ecossistema. Além disso, podem causar danos estéticos (manchas e danificações de rochas e outros materiais).

3.2.2 Dióxido de Nitrogênio (NO_2)

O óxidos de nitrogênio são formados por dois poluentes: o monóxido de nitrogênio (NO) e o dióxido de nitrogênio (NO_2). Suas fontes principais são combustões em veículos automotores, indústrias, usinas térmicas que utilizam óleo ou gás e incineradores.

O NO_2 é altamente tóxico ao ser humano, sua exposição aumenta a susceptibilidade aos problemas respiratórios em geral. Além disso, é irritante às mucosas e pode nos pulmões ser transformado em nitrosaminas (algumas das quais são carcinogênicas).

Os efeitos gerais sobre o meio ambiente são: a formação da chuva ácida e conseqüentemente danos à vegetação e agricultura. Além disso, contribui para formação do ozônio na troposfera; para o aquecimento global; formação de compostos quimiotóxicos e alteração da visibilidade.

3.2.3 Ozônio (O_3)

O ozônio troposférico é um gás incolor e inodoro nas concentrações ambientais. Sua fonte principal é a formação, na troposfera, a partir da reação dos hidrocarbonetos e óxidos de nitrogênio na presença de luz solar.

O O_3 na saúde humana pode provocar danos na estrutura pulmonar, reduzindo sua capacidade e diminuindo a resistência às infecções. Causa, ainda, o agravamento de doenças respiratórias, aumen-

tando a incidência de tosse, asma, irritações no trato respiratório superior e nos olhos. O grau de efeitos adversos respiratórios produzido pelo ozônio depende de vários fatores, incluindo a concentração e duração da exposição, características climáticas, sensibilidade individual, doença respiratória preexistente, e status socioeconômico, características que foram mantidas constante na população, ou seja, que não foram incluídas no modelo. Quatro grupos de pessoas são particularmente sensíveis ao ozônio quando estão ativos ao ar livre: crianças, adultos saudáveis ao realizarem exercício físicos ao ar livre, pessoas com doenças respiratórias preexistentes e os idosos. E no meio ambiente o O_3 é agressivo às plantas, agindo como inibidor da fotossíntese e produzindo lesões características nas folhas.

3.2.4 Dióxido de Enxofre (SO_2)

O dióxido de enxofre é um gás incolor com forte odor (semelhante ao produzido na queima de palitos de fósforo). Suas fontes principais são os processos que utilizam queima de óleo combustível, refinaria de petróleo, veículos a diesel, polpa e papel.

A inalação do SO_2 pelo ser humano, mesmo em concentrações muito baixas, pode provocar espasmos passageiros dos músculos lisos dos bronquíolos pulmonares. Em concentrações progressivamente maiores, causam o aumento da secreção mucosa nas vias respiratórias superiores, inflamações graves da mucosa e redução do movimento ciliar do trato respiratório. Pode, ainda, aumentar a incidência de rinite, faringite e bronquite. No meio ambiente em certas condições, o SO_2 pode transformar-se em trióxido de enxofre (SO_3) e, com a umidade atmosférica, transformar-se em ácido sulfúrico, sendo assim um dos componentes da chuva ácida.

3.2.5 Monóxido de Carbono (CO)

O monóxido de carbono (CO) é um gás tóxico, incolor e inodoro. É um produto da combustão incompleta de substâncias que contêm carbono, ou seja, forma-se na combustão de carbono com uma quantidade limitada de ar. As principais fontes de CO são: queima de combustível fóssil, aquecedores a gás ou querosene, fogão, fumaça de cigarro. Em áreas urbanas, do total do monóxido de carbono emitido para a atmosfera, mais de 90% pode ser de origem veicular (EPA, 2008). Mesmo sendo gerados em maior quantidade em ambientes externos, esses poluentes estão presentes em ambientes internos, devido à entrada por sistemas de condicionamento de ar. O CO é tóxico porque forma um complexo

com a hemoglobina do sangue, carboxihemoglobina, que é mais estável que a oxihemoglobina, impedindo, assim, a hemoglobina das hemáceas de transportar oxigênio pelo corpo. A deficiência de oxigênio leva à inconsciência e posteriormente à morte.

Os poluentes possuem diferentes efeitos na população humana como mostrado acima, esse efeitos individuais são geralmente o motivo de estudo dos trabalhos científicos como (Ostro et al. 199, Saldiva et al. 2007, Chen et al. 2010 entre outros), no entanto é conhecido que existe uma alta correlação entre os poluentes e avaliar os efeitos adversos para a saúde de uma combinação de poluentes pode ser tanto mais interpretável e mais viável do que tentar isolar os efeitos de um poluente apenas. O que se faz nesta dissertação é incluir o efeito conjunto do grupo de poluentes estudados (PM_{10} , SO_2 , NO_x , CO e O_3) para se obter um valor mais real do efeito de exposição em relação as internações por doenças respiratórias.

3.3 Revisão dos Estudos Envolvendo Poluição do Ar e Saúde Utilizando o Modelo Aditivo Generalizado (MAG) e a técnica de análise de componentes principais (ACP)

Ostro et al. (1999) investigaram a associação entre os poluentes atmosféricos e as doenças respiratórias em crianças menores de 2 anos e das faixas etárias entre 3 a 15 anos de idade, em Santiago, Chile. Utilizando análise de séries temporais com o modelo de regressão de Poisson para o número de atendimentos médicos diários com sintomas de doenças respiratórias. Foram obtidas medições diárias de concentração de PM_{10} e de O_3 , temperatura e umidade relativa do ar. A análise de regressão múltipla indicou uma associação estatisticamente significativa entre PM_{10} e atendimentos médicos por sintomas respiratórios em crianças de ambas as faixas etárias avaliadas. Para as crianças menores de 2 anos, uma elevação de $50 \mu gm^3$ de PM_{10} foi associada a aumentos dos sintomas entre 4-12%. Para crianças entre 3-15 anos de idade o aumento dos sintomas variou entre 3-9%. E elevações de 5% nos registros de atendimentos foram explicados pela variação de concentração de 50 ppb de O_3 .

Saldiva et al. (2004) estudaram os efeitos de curto prazo da concentração de poluição atmosférica

na morbidade causada por doenças respiratórias em menores de 15 anos e na mortalidade de idosos na cidade de São Paulo. Foram analisadas as contagens diárias de admissões hospitalares, de menores de 15 anos e de mortes de idosos (>64 anos) no período de 1993 a 1997, em relação às médias diárias da concentração de poluentes atmosféricos (PM₁₀, CO e O₃). A regressão de Poisson em modelos aditivos generalizados (MAG) foi utilizada para análise. Os modelos foram ajustados para efeitos da tendência temporal, sazonalidade, dias da semana, fatores meteorológicos e autocorrelação. Os resultados foram analisados segundo os percentis, com variações do 10° ao 90° percentil da concentração dos poluentes onde foi detectada uma associação significativa com o aumento de admissões por doenças respiratórias em menores de 15 anos para PM₁₀ (%RR=10,0), CO (%RR=6,1) e O₃ (%RR=2,5). Associação similar foi encontrada para mortalidade em idosos e PM₁₀ (%RR=8,1) e CO (%RR=7,9). Os resultados encontrados são coerentes com os estudos anteriores que já apontavam a associação entre variações de curto prazo a exposição dos poluentes atmosféricos e o incremento na morbidade e mortalidade nos grandes centros urbanos.

Saldiva et al. (2007) avaliaram os efeitos agudos do material particulado inalável (PM₁₀) sobre os atendimentos em pronto-socorro por doenças respiratórias e cardiovasculares no Município de Itabira. Os dados foram diários de atendimento de pronto-socorro por doenças respiratórias (Classificação Internacional de Doenças, 10ª Revisão - CID-10: J00-J99) em crianças e adolescentes (menores de 19 anos) e idosos (com mais de 64 anos), e por doenças cardiovasculares (CID-10: I00-I99) em pacientes com mais de 44 anos foram obtidos do hospital do município, para o período entre 1° de janeiro de 2003 e 30 de junho de 2004. As análises foram feitas para todos os atendimentos e também estratificadas por faixa etária. Para estimar a associação entre o número de atendimentos diários e as concentrações de PM₁₀ foi utilizado o modelo aditivo generalizado (MAG). As variáveis de confusão foram controladas, como temperatura e sazonalidade. Os aumentos de 10 $\mu\text{g m}^{-3}$ no PM₁₀ foram associados com aumentos nos atendimentos de pronto-socorro por doenças respiratórias de 4% (IC95%: 2,2-5,8), no dia e no dia seguinte, para crianças menores de 13 anos, e de 12% (IC95%: 8,5-15,5), nos três dias subsequentes para os adolescentes entre 13 e 19 anos. Dentre os atendimentos por doenças cardiovasculares, o efeito foi agudo (4%; IC95%: 0,8-8,5), principalmente para os indivíduos com idade entre 45 e 64 anos. Esses resultados mostram que o PM₁₀ gerado principalmente através de mineração a céu aberto pode acarretar prejuízos à saúde da população exposta.

Chen et al. (2010) realizaram um estudo de séries temporais utilizando o MAG com regressão de Poisson para examinar a associação entre os poluentes PM₁₀, SO₂ e NO₂ e o número de internações

por causas específicas, em Xangai, China, usando três anos de dados de janeiro de 2005 a dezembro de 2007. A área de estudo incluiu nove distritos urbanos, a população alvo foi todos os residentes permanentes que vivem nesta área urbana. O número diário de admissões hospitalares dos residentes nos nove distritos urbanos de 1 de janeiro de 2005 a 31 de dezembro de 2007 (1.095 dias) foram coletados do banco de dados de Xangai Seguro de Saúde Bureau (SHIB). As causas de internação foram codificadas de acordo com a Classificação Internacional de Doenças, o CID 10, todas as causas não-acidental, as doenças cardiovasculares e doenças respiratórias, os dados diários dos poluentes PM_{10} , SO_2 e NO_2 foram obtidos a partir da base de dados do Centro de Monitoramento Ambiental de Xangai. Para permitir o ajuste do efeito do clima sobre a internação hospitalar, as variáveis meteorológicas como média diária de temperatura e umidade relativa do ar foram obtidas do banco de dados de meteorologia de Xangai, essas variáveis de confusão foram controladas por funções splines na modelagem do MAG. Para controlar a sazonalidade de curto prazo (dias da semana) foram utilizadas variáveis dummy e para controlar a sazonalidade de longo prazo (dias transcorridos) foram utilizadas funções loess. O estudo concluiu que a poluição do ar estava associada com maior risco a internação cardiovascular em Xangai. Para uma defasagem de 5 dias, um aumento de $10 \mu g m^{-3}$ na concentração de PM_{10} , SO_2 e NO_2 , correspondiam a 0,18%, 0,63% e 0,99% para aumento de internação hospitalar total. Para um aumento de 0,23%, 0,65% e 0,80% em internações por doença cardiovascular, respectivamente. Em todos os resultados, observou-se um intervalo de confiança de 95%. Este estudo forneceu as primeiras evidências de que na China, especificamente em Xangai, sobre o efeito da poluição do ar na admissão hospitalar.

Estudo realizado por Almeida (2006) para verificar, através do MAG, a relação existente entre os poluentes atmosféricos PM_{10} , O_3 , SO_2 e NO_2 e o número de atendimentos diários por causas respiratórias da população infantil da Região da Grande Vitória (RGV), E.S., tem evidenciado para o grupo investigado no período de 2001 a 2004 que os níveis da poluição atmosférica são significativos, com destaque para os danos causados por PM_{10} e NO_2 . Dessa forma, pode-se definir como objetivo deste estudo a apresentação da modelagem, para dados referentes ao número de atendimentos hospitalares por causas respiratórias, através do MAG, que é uma das diversas técnicas de modelagem para dados discretos. Foram encontrados no modelo de regressão final coeficientes das defasagens simples e acumuladas significativas dos poluentes e para essas defasagens foram calculados os riscos relativos, os aumentos percentuais para um aumento de $10 \mu g m^{-3}$ nos níveis dos poluentes e os números de atendimentos diários esperados no ano, foi considerado o nível de significância 5% em todas as análises. Os resultados evidenciaram que para o grupo estudado, os efeitos agudos da poluição atmosférica são

significativos, o principal poluente associado com a morbidade em crianças de 0 a 6 anos, por causas respiratórias foi o PM_{10} , seguido de NO_x , SO_2 , e O_3 . Nesse mesmo estudo os seguintes poluentes PM_{10} e SO_2 apresentaram efeitos protetores nas estações Enseada do Suá e Vila Velha - IBES, uma análise mais cuidadosa tem que ser investigada nessas estações para esses resultados que não estão em acordo com a hipótese descrita.

Barbosa (2009) analisou a associação entre as concentrações dos poluentes atmosféricos, material particulado (PM_{10}), dióxido de nitrogênio (NO_2), ozônio (O_3) e o número de atendimentos hospitalares por doenças respiratórias (Classificação Internacional de Doenças, 10^a Revisão CID-10) em crianças de 0 a 6 anos residente no Município de Serra, no período de janeiro de 2001 a dezembro de 2004. Foi utilizado o Modelo Aditivo Generalizado (MAG) e a técnica de Bootstrap para a construção de intervalos de confiança para os parâmetros do modelo, tendo assim uma replicação das amostras. Os atendimentos por doenças aéreas respiratórias, de acordo com o CID-10, em crianças de 0 a 6 anos de idade foi considerada como variável resposta, esses dados foram coletados no banco de dados do Hospital Infantil Nossa Senhora da Glória (HINSG), localizado no município de Vitória, porém apenas as crianças do município da serra foram consideradas. As variáveis explicativas foram às concentrações médias diárias dos poluentes NO_2 , O_3 e PM_{10} e as variáveis meteorológicas consideradas como fatores de confusão: como as médias diárias da temperatura e umidade relativa do ar obtidos junto ao Instituto Estadual de Meio Ambiente e Recursos Hídricos (IEMA), sendo a estação de Laranjeiras/Serra a região de abrangência do estudo. Funções suavizadoras e variáveis indicadoras de dias da semana e feriados foram utilizadas para controlar a sazonalidade, tanto de curta, como de longa duração. A conclusão do trabalho foi que os níveis de concentrações dos poluentes em ambientes urbanos afeta diretamente os atendimentos hospitalares em crianças menores de 6 anos por causas respiratórias e que intervalos de confiança com amplitude menores ou próximos a 95% de confiança, podem ser utilizados para medir o impacto das concentrações dos poluentes atmosféricos na saúde da população.

Muitos destes estudos que investigam os efeitos adversos à saúde em associação com a concentração dos poluentes atmosféricos, relevam o fato de que os múltiplos poluentes envolvidos na modelagem são altamente correlacionados e que, portanto, as abordagens tradicionais lidam com o problema da multicolinearidade, neste contexto a Análise de Componentes Principais (ACP), traz na construção dos modelos as correlações existentes entre os poluentes.

Análise de Componentes Principais é um dos métodos multivariados mais simples. O objetivo da análise é tomar p variáveis X_1, X_2, \dots, X_p e encontrar combinações destas para produzir índices Z_1, Z_2, \dots, Z_p que sejam não correlacionados na ordem de sua importância, e que descreva a variação nos dados; esse procedimento implica que toda a variância é comum ou compartilhada.

Roberts e Martin (2006) estudaram a associação dos diversos poluentes como (PM_{10} , O_3 , SO_2 , NO_2 e CO) e seus efeitos à saúde. Os dados utilizados consistiam em séries temporais diárias de mortalidade, das variáveis meteorológicas e das concentrações dos poluentes, em nove cidades nos Estados Unidos, no período de 1987 a 2000. Os dados foram retirados do banco de dados do estudo nacional da morbidade, mortalidade e poluição do ar. Neste trabalho é observado um problema comum no contexto dos modelos de regressão, chamado de multicolinearidade, a análise de componentes principais (ACP) é utilizada como medida corretiva deste problema. Como esta técnica permite agrupar os poluentes em um grupo principal e assim elimina as altas correlações existentes entre eles. Neste artigo também é comparado e descrito o uso Supervisionado de Análise de Componentes Principais (SACP), onde se considera a relação entre as covariáveis (poluentes) e os efeitos adversos à saúde. O estudo demonstrou que ambas as técnicas são eficazes para controlar a multicolinearidade e determinam melhor o modelo de regressão, porém o método de SACP foi mais bem sucedido em identificar os componentes principais dos poluentes associados com a mortalidade, ao permitir a exclusão de poluentes que não estão associados diretamente aos resultados adversos à saúde.

Wang e Pham (2011) propuseram a aplicação dos modelos aditivos generalizados (MAG), baseado no método de análise de componentes principais (PCA), para quantificar a associação entre a mortalidade diária e as concentrações de poluentes atmosféricos, principalmente o PM_{10} , CO , NO_2 , SO_2 e O_3 . Neste estudo analisaram os efeitos das variáveis de confusão dentro do modelo, tais qual a tendência temporal de longo prazo, a sazonalidade, o dia da semana e os fatores meteorológicos. O estudo foi realizado na capital da Coreia, Seul, no período de 1º de junho de 1999 a 31 de dezembro de 2002, os dados foram fornecidos pelo Instituto Nacional de Estatística Coreano e pelo Ministério do Meio Ambiente. Utilizaram a análise de componentes principais, a fim de se controlar a multicolinearidade das variáveis em estudo, primeiramente construíram o modelo utilizando o GAM, que consistia nos fatores de confusão da tendência de longo prazo, fatores meteorológicos, a sazonalidade e os dias da semana. Depois com base neste modelo determinaram os efeitos dos "lags" para cada poluente, e num terceiro e último momento aplicaram a técnica de PCA para eliminar a multicolinearidade entre as covariáveis dos poluentes. Os resultados mostraram que o SO_2 e o CO tiveram uma grande contri-

buição para a mortalidade diária nos dados de Seul. O valor de RR SO_2 foi de 1,017692 (95% IC, 1,010368-1,025069), enquanto o valor do RR do CO foi de 1,014076 (95% IC, 1,010396-1,017769). O efeito de O_3 foi o maior com RR de 1,010238 (95%IC 1,003202-1,017323). O efeito do PM_{10} para a morte diária foi o menor comparado com os outros poluentes analisados, com RR de 1,001165 (95% IC, 997356-1,004989). Os autores sugerem que novas pesquisas podem ser realizadas com um foco ainda maior na análise de sensibilidade e análise de desempenho de predição para o modelo aditivo generalizado com análise de componentes principais.

Os trabalhos pesquisados mostraram a relevância de estudar o comportamento das séries temporais dos poluentes atmosféricos, para que sejam comprovados os seus efeitos adversos na saúde. Nos trabalhos revisados percebe-se a aplicação da metodologia dos modelos lineares generalizados e dos modelos aditivos generalizados o MLG e o MAG, respectivamente. A proposta deste trabalho é casar a metodologia multivariada de ACP e a modelagem MAG para obter as estimativas do efeito da associação entre a exposição atmosférica e o número de atendimentos hospitalares por doenças respiratórias, em crianças menores de 6 anos, na região da Grande Vitória, ES.

Capítulo 4

Artigo

¹ **Análise de Componentes Principais e a Modelagem Linear Generalizada: uma associação entre o número de atendimentos hospitalares por causas respiratórias e a qualidade do ar, na Região da Grande Vitória, ES.**

Juliana Bottoni de Souza

Valdério A. Reisen

Jane Meri Santos

Glaura C. Franco

Programa de Pós-Graduação em Engenharia Ambiental, CT-UFES

Departamento de Estatística, CCE-UFES

Departamento Engenharia Ambiental, CT-UFES

Departamento de Estatística, UFMG

¹Normalização segundo a Revista Jornal de Epidemiologia

4.1 Resumo

Este artigo propõe o casamento entre duas ferramentas estatísticas, a Análise de Componentes Principais (ACP) e o Modelo Aditivo Generalizado (MAG), para estimar o efeito da associação entre a exposição atmosférica dos poluentes PM_{10} , SO_2 , NO_2 , O_3 e CO e o número de atendimentos por doenças respiratórias em crianças menores de 6 anos na região da Grande Vitória, ES, Brasil. Usualmente, os poluentes atmosféricos são considerados como as covariáveis explicativas no MAG; porém, por possuírem estrutura de autocorrelação, esses devem ser utilizados com certa cautela. A técnica de ACP fornece novas variáveis ortogonais, essas são combinações lineares das variáveis ambientais. Nessa direção as componentes principais serão utilizadas no MAG, denotado aqui por MAG-ACP. Entretanto, as componentes principais obtidas por meio da matriz de variância/covariância aplicada a processos indexados no tempo também apresentam as propriedades de correlação temporal. No ajuste final do modelo MAG-ACP foi necessária a inclusão de modelos do tipo SARMA(p,q)(P,Q) nos resíduos, com o objetivo de eliminar as estruturas de autocorrelação presente nas componentes. Para atenuar a correlação temporal das componentes, o método de filtragem, por meio do modelo Vetorial Autoregressivo (VAR), é sugerido como procedimento alternativo para transformar os dados atmosféricos num processo Ruído Branco. A matriz de resíduos é utilizada para obter as componentes e essas aplicadas ao modelo MAG - método aqui denominado VAR-MAG-ACP. Os resultados empíricos mostram que esse modelo remove as autocorrelações das componentes principais e indica estimativas mais significantes do Risco Relativo (RR) para cada poluente, além de gerar melhores ajustes residuais. Comparadas à modelagem MAG usual, em geral, as duas vertentes propostas (MAG-ACP e VAR-MAG-ACP) apresentaram melhores resultados, tanto na estimativa do RR quanto na qualidade do ajuste. Por exemplo, um aumento de $10.49 \mu g/m^3$ (Intervalo interquartil) nos níveis de PM_{10} resultou num aumento de pelo menos 5% do valor do RR estimado, por meio do modelo VAR-MAG-ACP, comparado ao valor obtido no modelo MAG usual.

Palavras Chave: Séries Temporais. Qualidade do Ar. Modelo Aditivo Generalizado. ACP. Modelo Vetorial Autoregressivo. Efeitos na saúde.

4.2 Introdução

O impacto da poluição atmosférica no bem-estar humano tem sido a principal motivação para o seu estudo e controle. A poluição atmosférica afeta a saúde da população mesmo quando seus níveis encontram-se abaixo do que determina a legislação vigente (Martins et al 2001). Inúmeros estudos têm encontrado associações significativas entre os níveis diários de concentração de poluentes e atendimentos por causas respiratórias ou cardiovasculares (Ostro 1999), (Schwartz 2000), (Saldiva 2004), (Gouveia 2004), (Saldiva 2007), (Chen et al 2010), entre outros, pois sabe-se que os pulmões constituem o principal alvo de ataque dos poluentes do ar, entre os quais os principais são material particulado(PM_{10}), dióxido de enxofre(SO_2) e ozônio (O_3).

Os efeitos adversos dos poluentes atmosféricos na saúde humana têm sido uma preocupação dos órgãos regulamentadores de meio ambiente e de saúde pública. A confirmação desses efeitos na saúde pode ser obtida a partir de estudos populacionais, através de investigações epidemiológicas, orientando o desenvolvimento de práticas e legislação que visem o controle das emissões e da qualidade do ar.

Pelas características da variável de desfecho em saúde (atendimentos), em geral, o modelo de regressão não linear aditivo generalizado com distribuição marginal de Poisson têm sido a ferramenta estatística para medir e quantificar a associação entre os poluentes atmosféricos e os efeitos adversos à saúde. Nessa metodologia, a inclusão das covariáveis (poluentes) no modelo de regressão ocorre de forma individual, pois os poluentes são correlacionados. Uma forma de contornar essa questão é utilizar a análise de componentes principais da matriz de covariância dos poluentes.

Avaliar os efeitos adversos para a saúde de uma combinação de poluentes pode ser tanto mais interpretável e mais viável do que tentar isolar os efeitos de um poluente apenas. Essa vertente de pesquisa é bem atual e de relevância já explorada pelos autores Roberts e Martin (2006), Wang e Pham (2010). O artigo de Roberts e Martin (2006), por exemplo, avaliou a associação dos diversos poluentes como (PM_{10} , O_3 , SO_2 , NO_2 e CO) e seus efeitos à saúde. O problema da multicolinearidade (correlação entre os poluentes) foi resolvida por meio de ACP, esta técnica permitiu agrupar os poluentes no modelo de regressão. Os autores sugerem ainda um método descrito como "supervisionado de ACP", onde a relação entre as covariáveis (poluentes) e os efeitos deletérios à saúde são observados antes da inserção destes no modelo de regressão. Mas recentemente, no artigo de Wang e Pham (2010) em Seul, Coréia, os autores também verificaram a associação conjunta dos poluentes na mortalidade diária, utilizando a ACP e um método Robusto. Os resultados encontrados foram mais significativos quando aplicada

a técnica multivariada de ACP, produzindo melhores estimativas do RR.

A ACP é uma técnica estatística multivariada utilizada para reduzir a dimensionalidade de um conjunto de dados com a preservação do máximo de variabilidade presente nas covariáveis Jhonson (2007). A aplicação da técnica de ACP, requer em geral, que os dados sejam obtidos através de replicações independentes. Entretanto, o uso de ACP para realizar inferências estatísticas necessita devida atenção em covariáveis que são correlacionadas no tempo, como é o caso dos poluentes atmosféricos.

Primeiramente, a propriedade de estacionariedade deve ser satisfeita. Como as componentes principais são combinações lineares das covariáveis, as propriedades dessas serão transferidas de forma linear para as componentes. O efeito das correlações no contexto inferencial é motivo da tese de Zamprognio (2013). Nesse trabalho os autores mostram teórica e empiricamente o efeito da correlação temporal das covariáveis, quando essa é negligenciada no procedimento. Na Proposição 1 os autores evidenciam que as componentes são correlacionadas temporalmente. Como as ACPs carregam a estrutura temporal das covariáveis, o seu uso em regressão deve ser feito de forma cautelosa. Essa é uma das vertentes científicas propostas no presente artigo.

Como forma de contornar o problema da correlação temporal das ACPs na aplicação do MAG, este artigo também propõe o uso do procedimento de filtros de modelos multivariados em séries temporais aplicados nas covariáveis, antes da obtenção das componentes principais. A utilização do casamento entre a técnica de ACP e dos modelos de regressão evidenciaram, para o estudo de caso neste artigo, que o uso de MAG-ACP e VAR-MAG-ACP produziu melhores estimativas no Risco Relativo (RR) comparadas a utilização da modelagem MAG usual.

Os resultados desta pesquisa estão apresentados neste artigo da seguinte forma: a Seção 2 descreve a área de estudos e explora as características dos dados que incluem as variáveis de saúde e de poluição do ar e, também, detalha algebricamente a modelagem estatística utilizada como a ACP, MAG e o VAR. A Seção 3 apresenta os resultados obtidos e as estatísticas descritivas dos poluentes atmosféricos, das variáveis meteorológicas e dos dados de saúde, além dos ajustes do modelos de regressão de Poisson para o estudo proposto. Conclusões e recomendações para trabalhos futuros são as motivações da Seção 4.

4.3 Materiais e Métodos

Para os estudo proposto, foram utilizados dados referentes aos atendimentos hospitalares, obtidos junto ao sistema de registro de atendimentos do hospital Infantil Nossa Senhora da Glória (HINSG), onde se encontra o principal pronto-socorro infantil da cidade de Vitória, e que são atendidas crianças da RGV como um todo. As doenças aéreas respiratórias são classificadas de acordo com a Classificação Internacional de Doenças (CID-10) e o grupo estudado foi de crianças menores de 6 anos. O período do estudo foi de 1º de janeiro de 2005 a 31 de dezembro de 2010.

As informações sobre os níveis diários dos poluentes atmosféricos, Material Particulado (PM_{10}), Dióxido de Enxofre (SO_2), Monóxido de Nitrogênio (NO_2), Ozônio (O_3) e Monóxido de Carbono (CO), bem como as variáveis meteorológicas foram obtidas junto ao Instituto Estadual de Meio Ambiente e Recursos Hídricos (IEMA), medidas coletadas nas 8 estações de monitoramento (RAMQAr), localizadas na RGV.

Para todos os poluentes a coleta de dados compreendeu um período de 24 horas , que se iniciava na primeira meia hora do dia. Para cada estação, foi considerada a média de 24 horas para os poluentes PM_{10} e o SO_2 , a média móvel da máxima de 8 horas para o CO e o O_3 , e a máxima de 24 horas para o NO_2 . A média diária dessas variáveis entre as estações foram as covariáveis utilizadas no modelo de regressão MAG.

4.3.1 Metodologia da Análise dos Dados

As variáveis de interesse descritas acima foram modeladas utilizando técnicas de séries temporais, modelos de regressão e análise multivariada. Em geral, o MAG e suas vertentes, permitem funções paramétricas e não- paramétricas para o ajuste da curva média dos dados. As vertentes do MAG proposta neste artigo são o modelo MAG com erro correlacionado, tipo ARMA, o modelo MAG-ACP e esse após a filtragem dos dados originais, por meio do modelo VAR, o VAR-MAG-ACP. Cada desfecho foi modelado tomando inicialmente como pressuposto básico de distribuição que as contagens de eventos de saúde (admissões hospitalares) seguem uma distribuição de Poisson. O número diário de atendimentos por doenças respiratórias foi considerada como variável dependente, e os níveis de concentrações diárias dos poluentes atmosféricos como variáveis independentes.

Uma característica comum das variáveis envolvidas no estudo é a falta de observações (*missings*), seja por medições realizadas de forma incorreta, por falhas nos equipamentos ou medições inválidas, essas foram ajustadas via imputação. O procedimento de imputação seguiu a metodologia descrita por Junger (2008). Nesse método as estimativas são obtidas pela correlação espacial entre os níveis do próprio poluente e pela autocorrelação dos níveis desse poluente, ao longo do tempo.

A análise e ajuste dos modelos consistiram em vários procedimentos implementados em etapas. Inicialmente, tratou-se a sazonalidade de curta duração com variáveis indicadores para os dias da semana e os feriados. Para a sazonalidade de longa duração foi utilizada a função suavizadora "loess" Friedman (1991), que permite controlar uma dependência não linear entre a variável de interesse (atendimentos) e a sazonalidade. As covariáveis de confusão (Temperatura e Umidade relativa) também foram modelados através de curvas suavizadoras "splines" Friedman (1991), Wahba (2000). Após essa etapa, as componentes principais foram calculadas por meio da matriz de covariância dos poluentes considerados. O modelo ajustado MAG-ACP foi obtido por meio de análises e de testes usuais de resíduos e da qualidade do ajuste (AIC - critério de Akaike (parcimônia do modelo) entre outros (Akaike 1973).

No caso do modelo VAR-MAG-ACP, as séries de poluentes foram filtrados pelo modelo Vetorial Autoregressivo (VAR(p)), onde p é ordem do modelo. A matriz de resíduos foi utilizada para a obtenção das componentes principais, por meio da aplicação de ACP. As etapas seguintes de ajuste do modelo VAR-MAG-ACP são equivalentes as descritas anteriormente para o modelo MAG-ACP.

4.3.2 Risco Relativo

Uma medida muito utilizada em estudos epidemiológicos para mensurar o impacto da concentração de poluentes atmosféricos na saúde da população exposta é o risco relativo (RR). Pode-se definir o RR como uma relação de probabilidade de um evento ocorrer após certa exposição a um fator de risco, que no estudo se refere à exposição de níveis de concentração de poluentes. O risco relativo é utilizado na análise estatística de dados com desfechos binários (0 ou 1). É particularmente relevante devido a sua utilização em análises de regressão de contagens, tipicamente quando a distribuição dos dados segue uma distribuição de probabilidade Poisson.

Em análise de variáveis discretas, o risco relativo de eventos A e B é definido de acordo com Bishop (2007) como

$$RR(A, B) = \frac{P(A/B)}{P(A/B^c)} \quad (4.1)$$

onde $P[\cdot]$ corresponde a probabilidade definida em um espaço de probabilidade e B^c é o complementar de B . Por analogia, a função risco relativo de uma variável Y em um nível x da variável independente X , denotado por $RR(x)$, é definida segundo Baxter (1997) como:

$$RR(x) = \frac{E(Y/X = x)}{E(Y/X = 0)} \quad (4.2)$$

O $RR(x)$ é a razão entre o valor esperado de Y em um nível x da variável independente X e valor esperado de Y para $X=0$ (grupo controle). No caso do modelo MAG, com distribuição marginal de Poisson, $RR(x)$ é estimado da mesma forma encontrada em Zou (2004),

$$\hat{RR}(x = \xi) = e^{(\xi \hat{\beta})} \quad (4.3)$$

onde ξ é a variação da concentração do poluente que pode, por exemplo, assumir o valor de $10\mu g/m^3$, da variação interquartílica, entre outros e $\hat{\beta}$ é o coeficiente estimado associado ao poluente estudado.

Para um nível de significância α , a hipótese a ser testada é definida com $H_0 : RR(x) = 1$ contra $H_1 : RR(x) > 1$. A não rejeição de H_0 implica estatisticamente que o poluente estudado não causa efeito adverso à saúde.

De acordo com o Teorema Central do Limite (T.C.L), os intervalos de confiança (IC) para o risco relativo $RR(x)$ é dado por:

$$e^{\hat{\beta}\xi \pm z_{\alpha/2} s_p \xi} \quad (4.4)$$

onde $\hat{\beta}$ é o coeficiente estimado de β , s_p é o desvio padrão de $\hat{\beta}$, ξ é a variação da concentração do poluente (variação interquartílica) e $z_{\alpha/2}$ é o quantil $\alpha/2$ da distribuição normal padrão.

4.3.3 Modelo Aditivo Generalizado - MAG

O modelo aditivo generalizado(MAG)(Hastie e Tibshirani 1990) com distribuição marginal de Poisson é usualmente referido nas análises de associação entre a variável desfecho e as covariáveis explicativas. Sua ampla utilização se deve ao fato da utilização de modelagem não-linear descrevendo a relação entre as variáveis de interesse, por exemplo, Ostro (1999), Schwartz (2000), Saldiva (2004), Gouveia (2004), Chen (2010).

Seja $\{Y_t\}$, $t = 1, \dots, N$, uma série temporal de contagem formada por números inteiros não negativos. A densidade condicional de $\{Y_t\}$ dado o passado \mathfrak{S}_{t-1} , denotada por $\{Y_t/\mathfrak{S}_{t-1}\}$, possui uma distribuição de Poisson, com média μ_t , se satisfazer

$$f(y_t; \mu_t/\mathfrak{S}_{t-1}) = \frac{e^{-\mu_t} \mu_t^{y_t}}{y_t!}, \quad t = 1, \dots, N. \quad (4.5)$$

O modelo de Poisson descrito na Equação 4.5, isto é, $Y_t/\mu_t \sim Poisson(\mu_t)$ tem a seguinte propriedade:

$$E[Y_t/\mathfrak{S}_{t-1}] = Var[Y_t/\mathfrak{S}_{t-1}] = \mu_t, \quad t = 1, \dots, N \quad (4.6)$$

onde $E[Y_t/\mathfrak{S}_{t-1}]$ e $Var[Y_t/\mathfrak{S}_{t-1}]$ representam o valor esperado e a variância condicionada de $\{Y_t\}$, respectivamente (Benjamin e Stasinopoulos 2003).

Seja

$$X = [x_1 \dots x_p]' \quad (4.7)$$

o vetor de dimensão p de covariáveis que podem incluir os valores passados de Y_t e outras informações auxiliares tais como, os poluentes, as variáveis de confusão (tendência, sazonalidade, variáveis meteorológicas entre outras) descritas em Kedem e Fokianos (2002).

A curva que descreve a relação entre Y_t e o vetor de covariáveis X é obtida por meio da transformação logarítmica de μ_t :

$$\log(\mu_t) = \sum_{j=0}^q \beta_j X_j + \sum_{j=q+1}^p f(x_j), \quad q \leq p \quad (4.8)$$

onde β_j é o vetor dos coeficientes a serem estimados (covariáveis) e $f(x_j)$ são funções suavizadoras

para as variáveis de confusão (temperatura e umidade) e a sazonalidade de longa duração presente nos dados. β_0 corresponde ao intercepto da curva associado ao vetor de valores unitários. O Modelo 4.10 é usualmente referido como semi-paramétrico, pois envolve adição de relação paramétrica e não paramétrica. Para estimação e inferência dos parâmetros do modelo utiliza-se, em geral, métodos de estimação por máxima ou quase-verosimilhança e suas propriedades assintóticas são dadas em Kedem e Fokianos (2002). As funções não paramétricas utilizadas são avaliadas por meio das funções "splines", "loess", média móvel entre outras (Friedman 1991) e (Wahba 2000).

4.3.4 Análise de Componentes Principais - ACP

A Análise de Componentes Principais (ACP) é uma técnica multivariada estatística que tem por finalidade a redução da dimensionalidade do espaço matricial dos dados, por meio de transformações lineares das variáveis originais (Hotelling 1933). A correlação existente entre as variáveis implica na ocorrência de multicolinearidade nos modelos de regressão. A aplicação da técnica de ACP no presente trabalho se deve ao fato de solucionar o problema da correlação entre os poluentes. Em geral, a explicação de toda a variabilidade do sistema determinado por k variáveis só pode ser efetuada por k componentes principais. No entanto, uma grande parte dessa variabilidade pode ser explicada por um número r menor de componentes, $r \leq k$ (Johnson 2007).

Sejam $(\lambda_1, \mathbf{a}_1), (\lambda_2, \mathbf{a}_2), \dots, (\lambda_k, \mathbf{a}_k)$ os pares de autovalores-autovetores, respectivamente, da matriz Σ de covariância do vetor X . A i -ésima componente principal de Σ é dada por

$$CP_i = \mathbf{a}'_i \mathbf{x} = a_{i1}x_1 + a_{i2}x_2 + \dots + a_{ik}x_k, \quad i = 1, 2, \dots, k \quad (4.9)$$

onde

$$\begin{aligned} Var(CP_i) &= \mathbf{a}'_i \Sigma \mathbf{a}_i \quad i = 1, 2, \dots, k. \\ Cov(CP_i, CP_s) &= \mathbf{a}'_i \Sigma \mathbf{a}_s \quad i, s = 1, 2, \dots, k. \end{aligned}$$

Como já mencionado, a técnica de ACP é usualmente aplicada em amostras independentes. Entretanto, quando as covariáveis são correlacionadas temporalmente o seu uso requer atenção. As componentes principais são combinações lineares dessas covariáveis e, portanto as propriedades dessas serão transferidas de forma linear para as componentes obtidas. Logo, o seu uso em modelos de regressão deve ser feito de forma cautelosa. Um método alternativo de controlar a presença de autocorrelação nas componentes principais é umas das propostas deste artigo e está descrita na Seção 4.6.

4.3.5 MAG-ACP - Modelagem Aditiva Generalizada e Análise de Componentes Principais

Uma das vertentes apresentadas neste artigo é a utilização conjunta entre a técnica de ACP e a modelagem MAG, denotado por MAG-ACP. As covariáveis geradas pela aplicação de ACP, definidas aqui por CP, são combinações lineares das variáveis ambientais originais. A inclusão dessas novas covariáveis no modelo MAG é definida da seguinte forma:

$$\log(\mu_t) = \sum_{i=0}^q v_i CP_i + \sum_{i=q+1}^p f(x_i), \quad q \leq p \quad (4.10)$$

onde v_i é o vetor estimado das componentes principais(CP) e $f(x_i)$ são as funções suavizadoras para as variáveis de confusão (temperatura e umidade).

Nesse contexto, a estimativa do RR para o modelo 4.10 é da forma,

$$\hat{RR}^*(x = \xi) = e^{\left(\xi \hat{v}_i^*\right)} \quad (4.11)$$

onde ξ é a variação da concentração do poluente dada pela variação interquartílica. \hat{v}_i^* é representado pela expressão:

$$\hat{v}_i^* = \sum_{i=1}^k a_{ij} \hat{v}_i, \quad j = 1, \dots, k. \quad (4.12)$$

onde a_{ij} corresponde aos autovetores associados das covariáveis j (poluentes); \hat{v}_i o coeficiente esti-

mado da i -ésima componente principal. \hat{v}_i^* é facilmente obtido por meio das equações 4.9 e 4.10.

4.3.6 VAR-MAG-ACP - Modelagem Vetorial autoregressiva e MAG-ACP

Como já especificado o uso da técnica de ACP em processos indexados no tempo, causa correlação temporal nas componentes principais. A proposta deste artigo é atenuar a autocorrelação e correlação cruzada dessas componentes, através do filtro VAR(1) nas observações originais, a fim de torná-los em processo ruído branco (McGrevy et al 2012). A escolha do modelo VAR(1) é baseada nas características das covariáveis consideradas no estudo, essas descritas na Seção 3.

Seja $W_t = (W_{1t}, W_{2t}, \dots, W_{kt})'$ um processo linear vetorial autoregressivo estacionário (VARMA), definido como a solução do sistema (Hamilton 1994):

$$\Phi(B)\Delta^d(B)(W_t - \gamma) = \Theta(B)\epsilon_t \quad (4.13)$$

onde B é o operador de atraso, γ é o vetor de médias e ϵ_t é um ruído branco com $E(\epsilon_t) = 0$ e $\text{Var}(\epsilon_t) = \Sigma_{\epsilon_t}$. O operador $\Phi(B) = I - \sum_{i=1}^p \Phi_i B^i$ e $\Theta(B) = I - \sum_{i=1}^q \Theta_i B^i$ são matrizes polinomiais de ordem p, q respectivamente, $d \in \mathbb{N}$ e I é a matriz identidade de dimensão $(k \times k)$ e Φ_i, Θ_i são matrizes de dimensão $(k \times k)$ de constantes.

O VAR(1) é um caso particular do modelo VARMA(p,q), com $p = 1, q = 0$.

Se W_t é um processo estacionário VAR(1), então $W_t = \Phi W_{t-1} + \epsilon_t$

Procedimentos de identificação e de estimação do Modelo 4.13 podem ser vistos em Hamilton (1994).

O Risco Relativo do modelo VAR-MAG-ACP é denotado por **.

4.4 Resultados

A amostra do número de atendimentos hospitalares correspondeu a um total de $N=2191$, no período de janeiro de 2005 a dezembro de 2010. A Tabela 4.1 apresenta as estatísticas descritivas (médias, desvio padrão, quartil entre outras) das variáveis estudadas. Observa-se que o número médio de atendimentos diários foi igual a 27.1 com desvio padrão de 18.15. Os resultados evidenciaram que os níveis de concentração dos poluentes estudados não ultrapassam o padrão primário de qualidade do ar recomendados pelo Conselho Nacional de Meio Ambiente (CONAMA), e nem os limites estabelecidos pela diretriz da Organização Mundial da Saúde (OMS). Porém, além desta pesquisa outros estudos também têm mostrado que os poluentes, mesmo em níveis de exposição abaixo dos padrões aceitáveis, causam efeitos deletérios para a saúde humana (Bakonyi et al 2004).

Tabela 4.1: Estatística descritiva dos atendimentos por doenças respiratórias nas áreas de abrangência de cada RAMQAR no período de jan/2005 a dez/2010

	média	desvio	mínimo	25	Percentis		
					50	75	máximo
PM ₁₀ ($\mu\text{g}/\text{m}^3$)	33.45	8.83	8.98	27.90	32.75	38.39	86.74
SO ₂ ($\mu\text{g}/\text{m}^3$)	12.44	3.11	4.89	10.06	12.16	14.57	26.48
O ₃ ($\mu\text{g}/\text{m}^3$)	31.86	8.36	12.10	25.97	30.73	36.58	72.34
NO ₂ ($\mu\text{g}/\text{m}^3$)	24.82	6.93	9.03	19.59	24.13	29.37	62.59
CO ($\mu\text{g}/\text{m}^3$)	885.79	231.28	295	724.82	866.60	1031.09	2141.50
Temperatura mínima ($^{\circ}\text{C}$)	20.86	2.47	13.10	19.08	21.15	22.80	25.98
Temperatura média ($^{\circ}\text{C}$)	24.43	2.45	17.00	22.62	24.40	26.35	30.80
Temperatura máxima ($^{\circ}\text{C}$)	29.35	3.28	19.40	27.20	29.41	31.60	39.70
Umidade relativa do ar (%)	77.43	6.03	61.60	73.24	77.19	81.14	97.28
Atendimentos	27.09	18.15	1.00	13.00	24.00	37.00	121.00

As variáveis meteorológicas (Temperatura e Umidade Relativa) utilizadas neste trabalho são da estação de monitoramento de Carapina. A média da temperatura máxima utilizada no modelo foi de 29.35°C com desvio padrão de 3.28°C e a umidade relativa do ar obteve uma média de 77.43% com desvio padrão de 6.03%.

Os gráficos da Figura 4.1 mostram que as séries dos poluentes ambientais e dos atendimentos por doenças respiratórias são sazonais, propriedade esperada para esses fenômenos. Outra característica observada das séries é a estacionariedade aparentemente fraca. Essa confirmação é feita pelos gráficos das funções amostrais de autocorrelação e correlação parcial de cada poluente apresentadas nas Figuras 4.2 a 4.6.

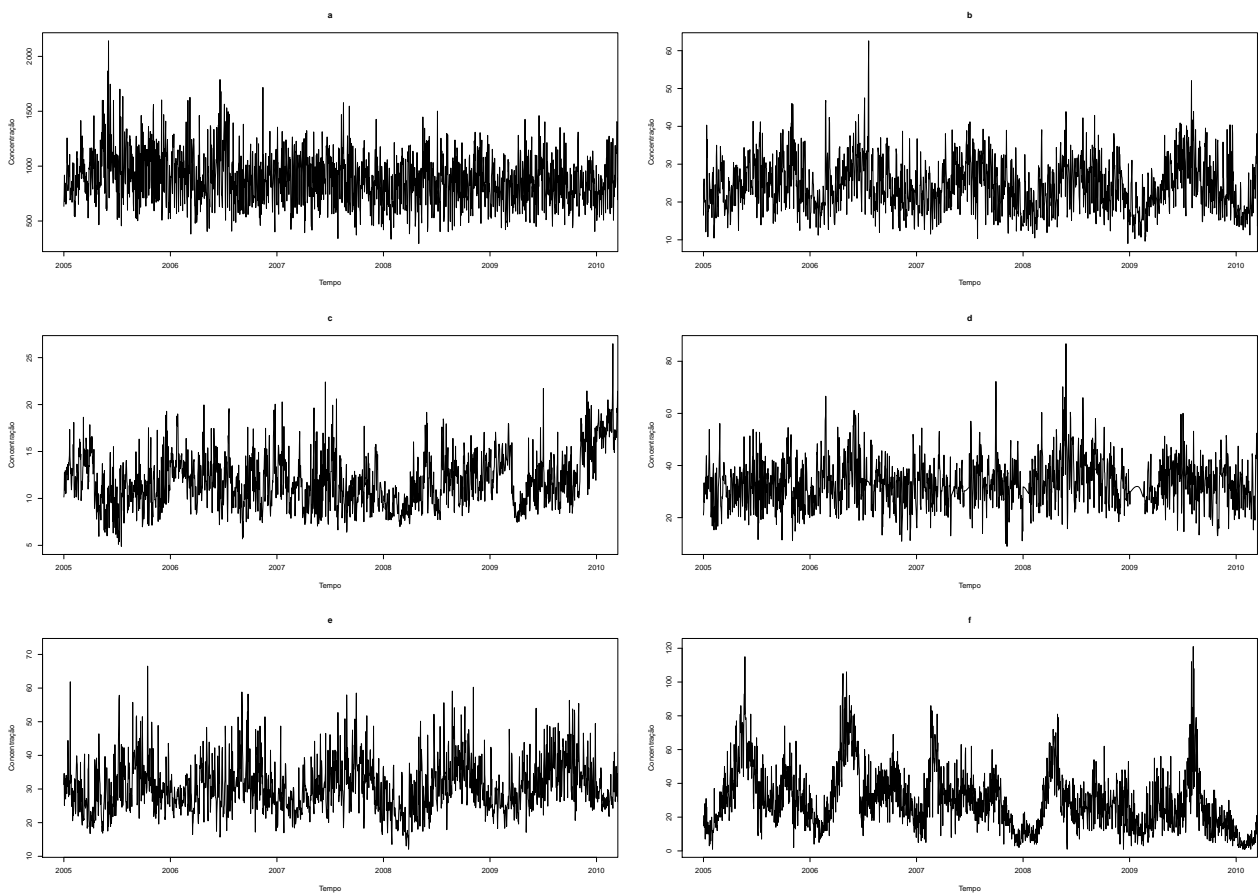


Figura 4.1: Concentrações dos poluentes atmosféricos em (a)-CO;(b)-NO₂;(c)-SO₂;(d)-PM₁₀;(e)- O₃ e o número de atendimentos diários (f)- Atendimentos

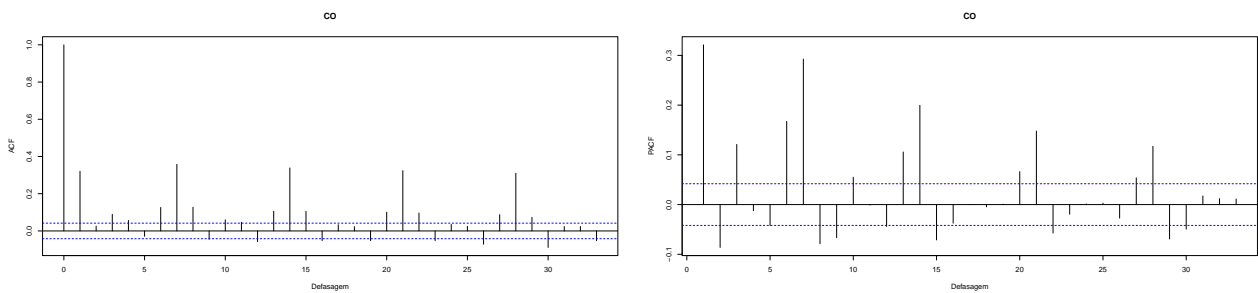


Figura 4.2: Função de Autocorrelação Amostral e Função Autocorrelação Amostral Parcial do poluente CO

A Tabela 4.2 exibe as correlações entre os poluentes atmosféricos, as variáveis meteorológicas e os atendimentos. Pode-se notar que há uma indicação de correlação entre os poluentes atmosféricos, aparentemente, em geral, não significativa. Com base nos estudos teóricos e empíricos apresentados em Zamprognio (2013), a técnica de ACP pode ser aplicada sem causar interpretações e testes espúrios quando o processo possui uma estrutura de autocorrelação fraca.

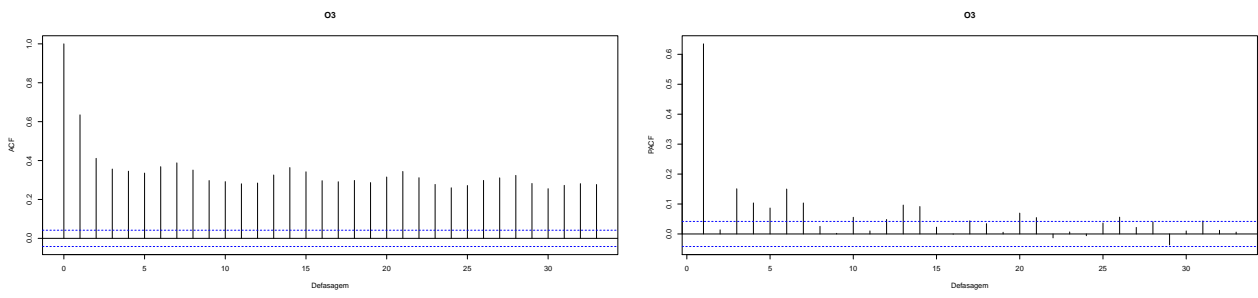


Figura 4.3: Função de Autocorrelação Amostral e Função Autocorrelação Amostral Parcial do poluente O_3

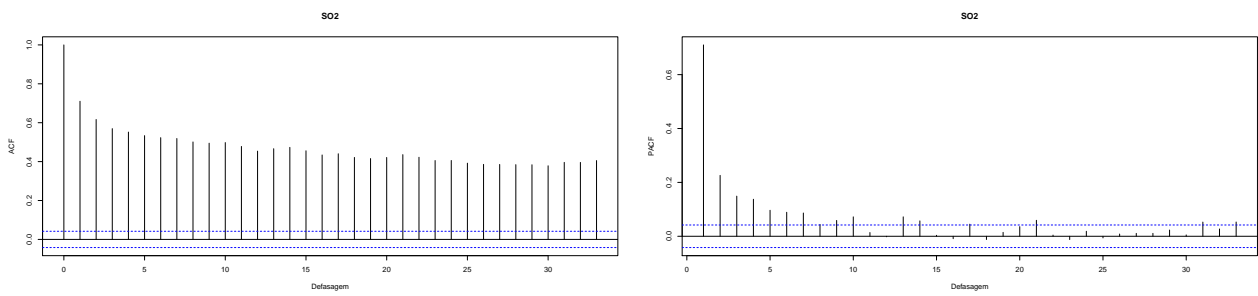


Figura 4.4: Função de Autocorrelação Amostral e Função Autocorrelação Amostral Parcial do poluente SO_2

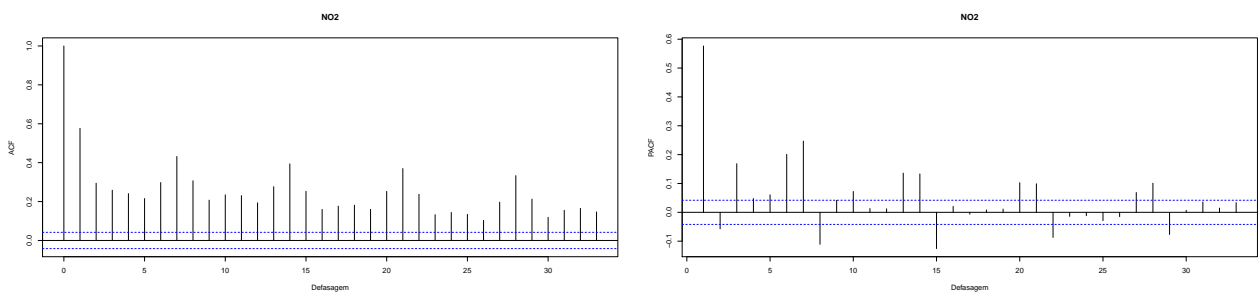


Figura 4.5: Função de Autocorrelação Amostral e Função Autocorrelação Amostral Parcial do poluente NO_2

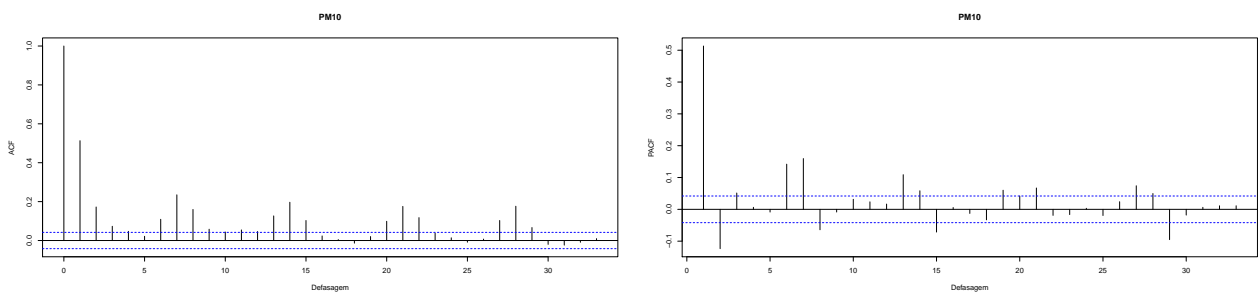


Figura 4.6: Função de Autocorrelação Amostral e Função Autocorrelação Amostral Parcial do poluente PM_{10}

As temperaturas mínima e máxima possuem correlação negativa com os poluentes O_3 e NO_2 e positiva com o PM_{10} , respectivamente. A associação positiva entre a temperatura máxima e mínima com o poluente PM_{10} é explicada pelo fato que nos períodos mais quentes a dispersão dos poluentes se torna mais acelerada e a baixa temperatura leva ao acúmulo de poluição no ar, o que impede que as partículas se dispersem, mantendo-as ao nível da atmosfera.

Tabela 4.2: Correlação entre os poluentes, variáveis meteorológicas e atendimentos.

	PM10	SO2	NO2	CO	O3	T(max)	T(min)	UR	Atendimentos
PM10	1.00								
SO2	0.31	1.00							
NO2	0.34	0.04	1.00						
CO	0.35	0.22	0.61	1.00					
O3	-0.04	-0.08	0.04	-0.40	1.00				
T(max)	0.20	0.44	-0.43	-0.06	-0.23	1.00			
T(min)	-0.10	0.16	-0.48	-0.10	-0.16	0.62	1.00		
UR	-0.28	-0.29	0.23	0.26	-0.22	-0.44	-0.03	1.00	
Atendimentos	0.05	-0.33	0.09	0.09	-0.08	-0.15	-0.19	0.14	1.00

T= Temperatura ($^{\circ}C$); UR= Umidade relativa do ar (%)

As análises descritivas e gráficas acima motivaram o uso da técnica de ACP nos dados de poluentes atmosféricos, com o objetivo de atenuar a correlação existente entre eles.

A Tabela 4.3 apresenta resultados da técnica de ACP aplicada aos poluentes PM_{10} , SO_2 , NO_2 , O_3 e CO . Observa-se que as 3 primeiras componentes explicam 83,2% da variabilidade total das variáveis. Como critério de escolhas das componentes utilizadas na modelagem utilizou-se a proporção da variância acumulada. Assim foram escolhidas as 3 primeiras componentes (destacadas em negrito).

Tabela 4.3: Resultado das cargas fatoriais e estatísticas da aplicação de análise de componentes principais.

	CP1	CP2	CP3	CP4	CP5
Desvio Padrão	1.4315	1.0431	1.0115	0.7741	0.4904
Proporção da variância	0.4098	0.2176	0.2046	0.1198	0.0481
Proporção variância acumulada	0.4098	0.6274	0.8320	0.9519	1.0000
CO	-0.6074*	-0.1999	-0.2311	-0.2146	-0.7012
NO_2	-0.5058*	0.3316	-0.4786	-0.2599	0.5810
O_3	0.2523	0.8615*	-0.0363	-0.1995	-0.3911
PM_{10}	-0.4680	0.3213	0.2784	0.7746	-0.0151
SO_2	-0.3041	0.0680	0.7992*	-0.4966	0.1327

Os maiores coeficientes (autovetores) das componentes 1, 2 e 3 são, respectivamente, dos poluentes CO , O_3 e SO_2 . Como complementação de análise da Tabela 4.3, observa-se que em cada componente pode-se fazer uma divisão de "clusters" que agrupe, por exemplo, os poluentes que possuem cargas fatoriais maiores que 0.5. Na Tabela 4.3 sugestões de "clusters", para cada componente principal, estão indicadas por (*).

Na Figura 4.7 são apresentados os comportamentos temporais das componentes principais obtidas das séries de concentrações dos poluentes, isto é, dos dados originais. Verifica-se que a componente 1 é autocorrelacionada e as correlações cruzadas são não nulas, o que corrobora o resultado da Proposição 1 de Zamprogno (2013). O fenômeno sazonal das variáveis de poluição também é evidentes nas componentes que é um resultado esperado. Os gráficos evidenciam que a estrutura de autocorrelação dos poluentes persiste nas componentes. Portanto, a aplicação da técnica de ACP deve ser de forma cautelosa, mesmo para processos com estrutura de autocorrelação aparentemente fraca, caso deste artigo.

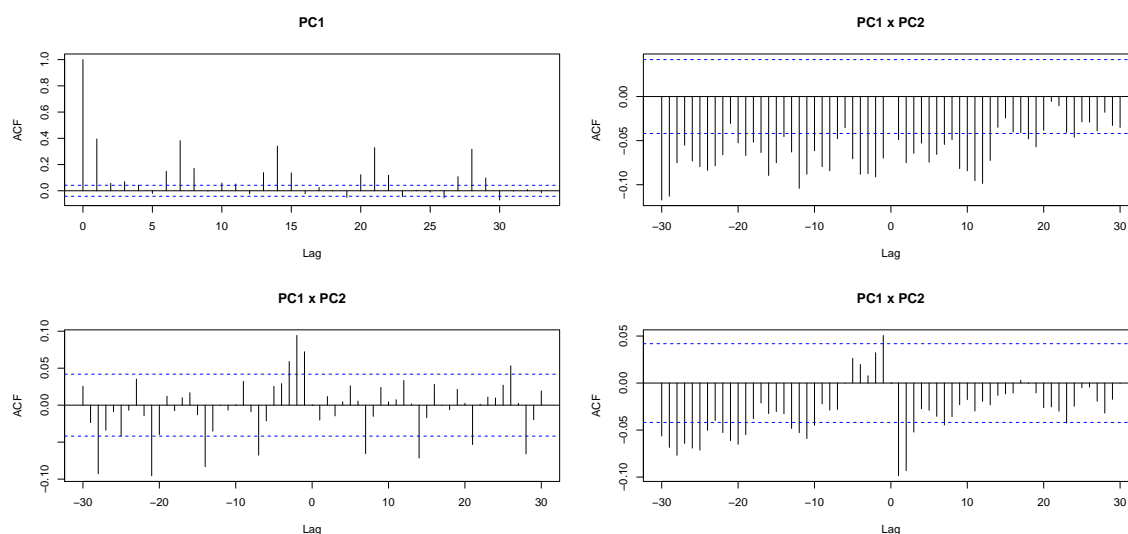


Figura 4.7: Correlação Cruzada(CCF) das componentes principais dos poluentes estudados.

Devido a limitação de espaço, a análise da qualidade do ajuste dos modelos de regressão para explicar a associação entre os atendimentos diários e os níveis de concentrações dos poluentes apresentadas neste artigo é restrita ao modelo MAG-ACP. As referentes aos modelos MAG e VAR-MAG-ACP satisfazem as condições de bom ajuste na modelagem e encontram-se disponíveis com os autores. A comparação da qualidade do ajuste entre os modelos propostos é aqui apresentado por meio do Erro Quadrático Médio estimado (\widehat{EQM}), e os resultados estão discutidos a seguir (Ver Tabela 4.6).

Como já mencionado, as componentes principais do processo sem filtragem possuem autocorrelação. Conseqüentemente, essa propriedade também é transferida para os resíduos do modelo ajustado MAG-ACP, como mostra a Figura 4.8. Diante dessa problemática, para os resíduos do modelo MAG-ACP foi ajustado um modelo $SARMA(1, 0, 0)(1, 0, 0)_7$, o que resultou no modelo final MAG-ACP-SAR. As estimativas dos parâmetros desse modelo estão na Tabela 4.4.

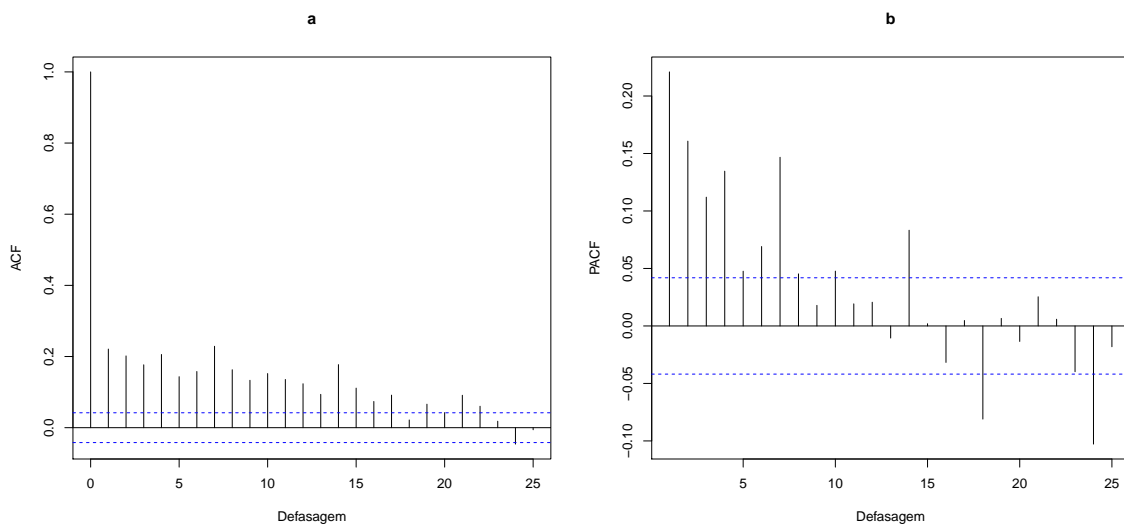


Figura 4.8: Função de Autocorrelação amostral e Função de Autocorrelação Parcial amostral dos resíduos do modelo MAG-ACP.

Tabela 4.4: Resultado da Regressão de Poisson para estimação dos efeitos dos poluentes nos atendimentos na RGV.

Variáveis	Estimativas	Erro Padrão	z valor	valor-p
(Intercepto)	4.4871	0.0901	49.82	0.0000***
Terça-feira	-0.1596	0.0152	-10.50	0.0000***
Quarta-feira	-0.2176	0.0154	-14.14	0.0000***
Quinta-feira	-0.1321	0.0151	-8.76	0.0000***
Sexta-feira	-0.1571	0.0154	-10.22	0.0000***
Sabado	-0.1204	0.0150	-8.04	0.0000***
Domingo	-0.0860	0.0154	-5.59	0.0000***
feriado2	0.1886	0.0440	4.29	0.0000***
feriado3	0.3189	0.0384	8.30	0.0000***
PC1	-0.0244	0.0040	-6.16	0.0000***
PC2	0.0163	0.0055	2.99	0.0028 **
PC3	-0.0157	0.0056	-2.79	0.0052***
AR(1)	0.1772	0.0216	45.14	<2e-16***
SAR(1)	0.1893	0.0217	83.16	<2e-16***

Significância: **** 0.001 *** 0.01 ** 0.05 * 0.1 ' ' 1

feriado2 = Corpus+Penha, feriado3 = Cinzas+Tirad+Indep

Como observado na Figura 4.9, a estrutura de autocorrelação do modelo MAG-ACP-SAR é eliminada, com correlações grandes não significativas.

A Figura 4.10 mostra o ajuste (os valores preditos) do modelo MAG-ACP-SAR. Com base nesse gráfico nota-se que esse modelo se ajustou bem aos dados da variável de interesse o número de atendimentos diários em crianças menores de 6 anos na RGV.

Na sequência da modelagem do vetor de observações, o modelo multivariado sazonal com período de

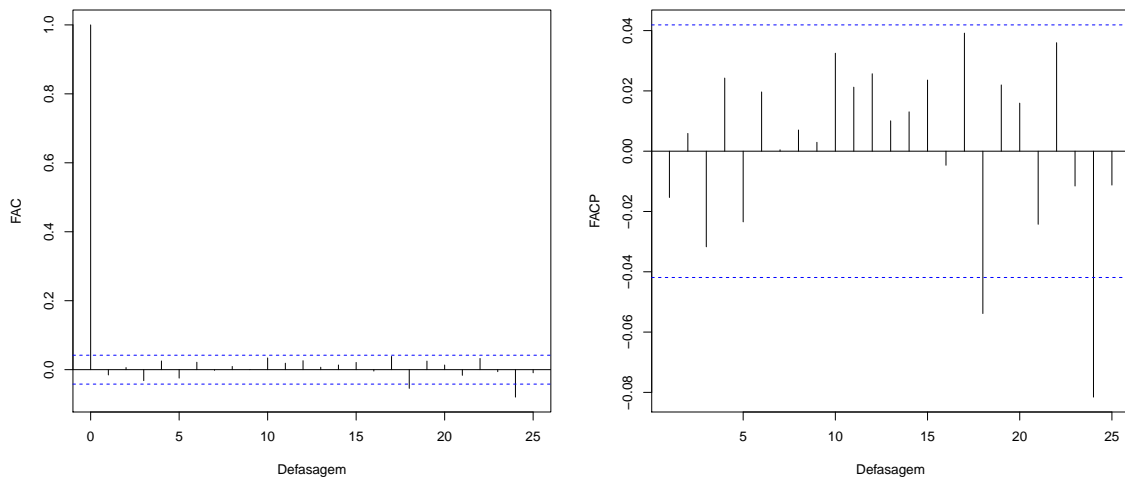


Figura 4.9: Função de Autocorrelação amostral e Função de Autocorrelação Parcial amostral dos resíduos do modelo MAG-ACP.

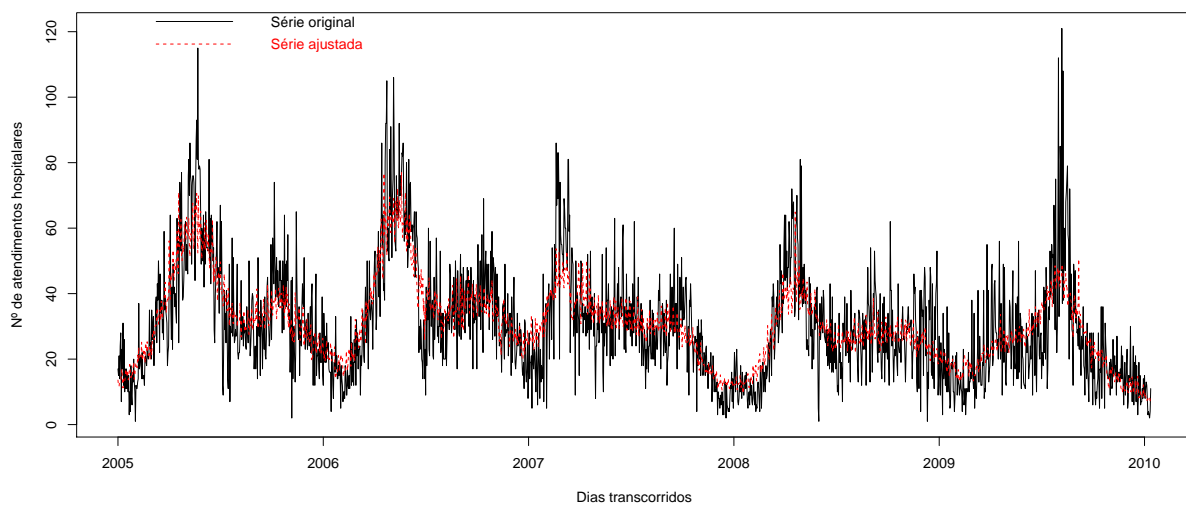


Figura 4.10: Ajuste do modelo MAG-ACP-SAR

7 dias, VAR(1) foi indicado para o ajuste do vetor de poluentes. A escolha desse modelo seguiu os procedimentos de estimação, de identificação e de qualidade de ajuste em modelos de séries temporais multivariados. As estimativas do modelo assim como as análises do ajuste do modelo estão disponíveis com os autores.

A Tabela 4.5 mostra o resultado da técnica de ACP aplicada a matriz de resíduos do modelo sazonal VAR(1). Os resultados indicam que a estrutura temporal dos poluentes não altera a proporção da variância acumulada, (comparar as Tabelas 3 e 4), isto é, as 3 primeiras componentes explicam 83% da variabilidade dos dados filtrados, resultado equivalente ao encontrado na Tabela 3.

Entretanto, a análise de "clusters" gerou uma interpretação diferente do caso anterior, mais coerente com os fenômenos das variáveis estudadas. Nessa metodologia os "clusters" estão indicados por (**). Os resultados evidenciaram uma associação entre os poluentes NO₂, O₃, não observada no caso anterior. Esse dois poluentes estão fisicamente associados pois a formação do O₃ depende de liberação de partícula do NO₂.

Para o processo Ruído Branco dos dados filtrados por meio do modelo VAR(1), a Figura 4.11 mostra que a autocorrelação da componente 1 e as correlações cruzadas são praticamente eliminadas, resultado esperado em função das discussões anteriores.

Tabela 4.5: Resultado das cargas fatoriais e estatísticas da aplicação de análise de componentes principais(ACP) nos poluentes filtrados.

	CP1	CP2	CP3	CP4	CP5
Desvio Padrão	1.4774	1.0223	0.9628	0.7228	0.5680
Proporção da variância	0.4366	0.2090	0.1854	0.1045	0.0645
Proporção variância acumulada	0.4366	0.6456	0.8310	0.9355	1.0000
CO	0.5711**	-0.1431	0.2918	-0.1469	0.7393
NO ₂	0.4205	-0.6527**	0.2543	-0.0905	-0.5695
O ₃	-0.3693	-0.5801**	-0.4685	-0.4896	0.2606
PM ₁₀	0.4012	-0.1409	-0.7040**	0.5663	0.0532
SO ₂	0.4468	0.4441	-0.3675	-0.6402	-0.2414

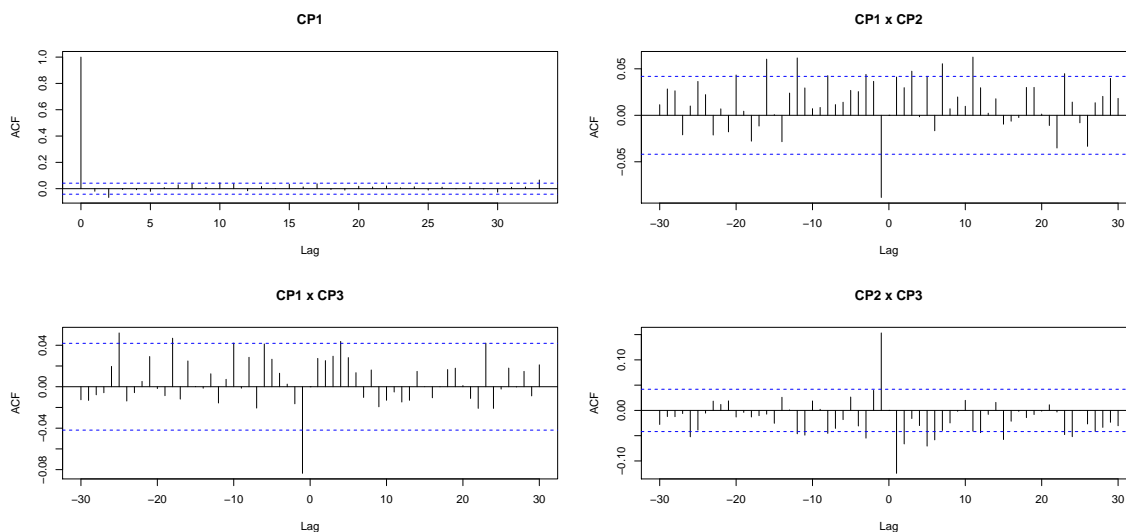


Figura 4.11: Correlação Cruzada(CCF) das componentes principais dos poluentes filtrados.

O estudo comparativo da qualidade dos ajustes dos três modelos propostos foi feito por meio do \widehat{EQM} , definido da forma:

$$\widehat{EQM} = \sum_{i=1}^n e_i^2 / N \quad (4.14)$$

onde $e_i = Y_i - \hat{Y}_i$, com \hat{Y}_i é o valor predito de Y_i , o número de atendimentos hospitalares .

Os resultados do \widehat{EQM} estão Tabela 4.6. Observa-se que os modelos propostos MAG-ACP-SAR e VAR-MAG-ACP apresentam similares resultados (valores em negrito). Entretanto, o \widehat{EQM} do modelo MAG é aproximadamente 35% superior aos obtidos pelos outros dois modelos.

Tabela 4.6: EQM estimado para os modelos MAG, MAG-ACP e VAR-MAG-ACP.

	EQM
MAG	1.480
MAG-ACP-SAR	1.143
VAR-MAG-ACP	1.144

Para finalizar, os valores do RR estimado, para cada poluente e modelo, foram calculados para a comparação do desempenho dos modelos ajustados MAG, MAG-ACP-SAR e VAR-MAG-ACP nas variáveis em estudo. Os resultados estão apresentados na Tabela 4.7 e esses estão expressos por incremento da variação interquartilica. Essa escolha é baseada pelo fato da análise do RR ser feita para poluentes de diferentes escalas. Os resultados do RR estimado foram, em geral, significativos para os todos os modelos considerados, isto é, os valores indicam que, para cada poluente, houve uma contribuição significativa no aumento do número de atendimentos por doenças respiratórias. Na maioria dos casos, os valores mais expressivos do RR estimado foram obtidos por meio do modelo proposto VAR-MAG-ACP.

Como exemplo de uma análise específica e comparativa dos valores do RR, observa-se que, para o poluente PM_{10} os Riscos Relativos estimados aumentaram aproximadamente de 2% (\hat{RR}) para 3% (\hat{RR}^*) e 7% (\hat{RR}^{**}). Aumentos expressivos nos valores estimados do Risco Relativo também são observados para o poluente CO. Nesse caso, $\hat{RR} = 1,020$, $\hat{RR}^* = 1.048$ e $\hat{RR}^{**} = 1.077$.

Portanto, os modelos propostos MAG-ACP-SAR e VAR-MAG-ACP apresentaram, em geral, resultados mais significativos no aumento esperado do número de atendimentos por causas respiratórias, comparados ao modelo MAG usual.

Tabela 4.7: Risco Relativo(RR) e intervalo de confiança de 95% de atendimentos por doenças respiratórias em crianças menores de 6 anos para uma variação interquartilica dos poluentes PM_{10} , SO_2 , NO_2 e O_3 e CO na RGV, jan-2005 a dez-2010.

	\hat{RR}	\hat{RR}^*	\hat{RR}^{**}
PM_{10}	1.020(1.010,1.039)	1.029(1.001,1.090)	1.075(1.001,1.092)
SO_2	1.040(1.010,1.080)	0.982(0.972,1.001)	1.027(1.010,1.040)
CO	1.020(1.010,1.030)	1.048(1.002,1.071)	1.077(1.002,1.100)
NO_2	1.000(0.990,1.020)	1.028(1.010,1.040)	1.012(1.010,1.030)
O_3	0.980(0.972,1.001)	1.081(1.003,1.093)	0.992(0.992,1.020)

4.5 Conclusão

Este trabalho propõe o casamento híbrido entre duas ferramentas estatísticas a Análise de Componentes Principais (ACP) e o Modelo Aditivo Generalizado (MAG) para estimar o efeito da associação entre a exposição atmosférica dos poluentes PM_{10} , SO_2 , NO_2 , O_3 e CO e o número de atendimentos por doenças respiratórias, em crianças menores de 6 anos na RGV, no período entre 2005 e 2010.

Os modelos propostos são denotados por MAG-ACP-SAR e VAR-MAG-ACP. O primeiro modelo utiliza as componentes principais dos dados originais como covariáveis no modelo MAG. Como as componentes principais do processo original dos dados possuem autocorrelação, essa propriedade consequentemente, também é transferida para os resíduos do modelo ajustado MAG-ACP. Os resíduos desse modelo foram ajustados por meio do modelo $SARMA(1, 0, 0)(1, 0, 0)_7$. O modelo final é definido por MAG-ACP-SAR.

As propriedades das componentes principais obtidas por meio do modelo MAG-ACP sugeriu a utilização do filtro VAR nos dados originais e esse procedimento resultou no modelo híbrido denotado por VAR-MAG-ACP. Nesse modelo as componentes são não autocorrelacionadas e a correlação cruzada é nula.

A qualidade do ajuste dos modelos supracitados, foi calculada por meio do EQM estimado. Os resultados indicam que o \widehat{EQM} do modelo MAG usual foi 35% superior aos dos outros dois modelos propostos.

Os efeitos deletérios provocados pela exposição as concentrações dos poluentes na saúde da população infantil da RGV, foram obtidos por meio da estimativa do Risco Relativo nos modelos de regressão propostos MAG, MAG-ACP-SAR e VAR-MAG-ACP. Todos os poluentes mostraram uma associação

com os atendimentos por doenças respiratórias.

Os resultados do RR estimado foram, em geral, significativos para os todos os modelos considerados, isto é, os valores indicaram que, para cada poluente, houve uma contribuição significativa no aumento do número de atendimentos por doenças respiratórias. Na maioria dos casos, os valores mais significativos do RR foram obtidos por meio do modelo proposto VAR-MAG-ACP. Por exemplo, para o poluente PM_{10} os Riscos Relativos estimados aumentaram aproximadamente de 2% (\hat{RR}) para 3% ($R\hat{R}^*$) e 7% ($R\hat{R}^{**}$).

No modelo MAG-ACP-SAR, um aumento de $10.49 \mu g/m^3$ (intervalo interquartilico) nos níveis de material particulado (PM_{10}) acarretou num $R\hat{R}^*$ para 1.029(IC 95% 0.991,1.09). Similares interpretações podem ser observadas para os outros poluentes e modelos propostos.

Os resultados encontrados neste artigo por meio dos modelos MAG e MAG-ACP, estão coerentes com os estudos realizados por (Wang e Pham 2011). Nesse trabalho, os autores consideraram a relação entre a morbidade e as concentrações de poluentes atmosféricos, para dados registrados na Coréia. O artigo comprova que a utilização da técnica de ACP melhora as estimativas finais do risco relativo.

Em resumo, os resultados encontrados neste artigo evidenciaram a relação significativa entre os níveis de concentração dos poluentes e o número de atendimentos hospitalares em crianças menores de 6 anos, mesmo em um ambiente com níveis abaixo dos padrões recomendados pelo CONAMA e OMS.

Para trabalhos futuros, recomenda-se a utilização da técnica de ACP na análise de dados de frequência. Outros grupos suscetíveis da população devem ser estudados, bem como outras doenças, por exemplo, as cardiovasculares. Quanto a metodologia pode-se utilizar outras técnicas como a do "bootstrap" para estimar os intervalos de confiança com mais precisão e a modelagem GLARMA, para resolver o problema da autocorrelação serial dos dados. Essas vertentes de pesquisas e suas aplicações são temas atuais do segundo autor deste artigo.

Os autores agradecem ao apoio dos órgãos:

- CNPq
- CAPES
- FAPES

- FAPEMIG

4.6 Referências

- MARTINS, L.; et. al. Relação entre poluição atmosférica e atendimentos por infecções de vias aéreas superiores no município de São Paulo: avaliação do rodízio de veículos. *Revista Brasileira de Epidemiologia*. 2001.
- SCHWARTZ, J. Harvesting and long term exposure effects in the relationship between air pollution and mortality. *American Journal Epidemiologic* 151, 440- 448. 2000.
- SALDIVA, P.H.N. FREITAS, C. et al. Internações e óbitos e sua relação com a poluição atmosférica em São Paulo, 1993 a 1997. *Revista Saúde Pública*. 2004.
- OSTRO, B. D., ESKELAND, G. S., SÁNCHEZ, J. M. e FEYZIOGLU T. Air pollution and health effects: A study of medical visits among children in Santiago, Chile. *Environmental Health Perspect* 107(1), 69-73. 1999.
- SALDIVA, P.H.N. FREITAS, C. et al. Associação entre poluição atmosférica e doenças respiratórias e cardiovasculares na cidade de Itabira, Minas Gerais, Brasil. *Caderno Saúde Pública*, Rio de Janeiro, 23. 2007.
- GOUVEIA, N.; BREMNER, S. A.; NOVAES, H. M. Association between ambient air pollution and birth weight in São Paulo, Brazil. *Journal Epidemiologic Community Health* 58, 11-17. 2004.
- CHEN R. J., CHU C., TAN, J. G., et al. Ambient air pollution and hospital admission in Shanghai, China. *Journal of Hazardous Materials*. 2010.
- RAMSEY TO, BURNETT RT, KREWSKI D. 2003. The effect of concurvity in generalized additive models linking mortality to ambient particulate matter. *Epidemiology*, Vol. 14, p. 18-23. 2003.
- ROBERTS S, MARTIN M. Using Supervised Principal Components Analysis to Assess Multiple Pollutant Effects. *Environmental Health Perspectives*, Vol. 116, No. 12. 2006.

- WANG, Yaping and PHAM, Hoang. Analyzing the effects of air pollution and mortality by generalized additive models with robust principal components. *International Journal of System Assurance Engineering and Management*, 2, p. 253- 259. 2011.
- BURNETT RT, Brook J, Dann T, Delocla C, Philips O, Cakmak, et al. Association between particulate and gas phase components of urban air pollution and daily mortality in eight Canadian cities. *Inhal Toxicol*, 12, p. 15-39. 2003.
- COX LH. Statistical issues in the study of air pollution involving airborne particulate matter. *Environmetrics*, 11, p. 611-626. 2000.
- FRIEDMAN, J. Multivariate adaptive regression splines. *The annals of statistics*, p. 1- 67. 1991.
- WAHBA, G. Splines in nonparametric regression. *Encyclopedia of Environmetrics*. 2000.
- AKAIKE, H. Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csàki F, editors. *Second International Simposium on Information Theory. Budapest*, p. 267-281. 1973.
- JUNGER, W.L. Análise, Imputação de dados e interfaces computacionais em estudos de séries temporais epidemiológicas; Analisis, data input and computer interfaces in epidemiological seies studies, Ph.D. Universidade Estadual do Rio de Janeiro. Instituto de Medicina Social. 2008.
- BISHOP, Y.; FIENBERG, S.; HOLLAND, P. *Discrete multivariate analysis: theory and practive*. Cambridge:MIT, 575 p.. 2007.
- BAXTER, L, S. Finch, F. Lipfert, Q. Yu. *Risk analysis* p. 273-278. 1997.
- ZOU, Guangyong. A Modified Poisson Regression Approach to Prospective Studies with Binary Data. *American Journal of Epidemiology*, Vol. 159, No. 7. 2004.
- HASTIE T, J, TIBSHIRANI R, J. Generalized additive models. *Chapman & Hall*, Londres. 1990.
- MCCULLAGH, P.; NELDER, J.A. Generalized Linear Models. *Chapman & Hall*, Londres, 2nd. 1989.
- BENJAMIN, M. A., RIGBY, R. A., AND STASINOPOULOS, D. M. Generalized Autoregressive Moving Average Models. *Journal of the American Statistical Association*, Vol. 98, No. 461. 2003.
- KEDEM, B. e FOKIANOS, K. *Regression Models for Time Series Analysis*. Wiley, USA ,2th edition. 2002.

- R Development Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Viena, Austria. ISBN 3-900051-07-0. 2011.
- HOTELLING, H. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology* 24, 417-441. 1933.
- JOHNSON, R.A.; WICHERN, D.W. *Applied multivariate statistical analysis*. 6th edition. Prentice Hall, New Jersey, 800 p. 2007.
- BAKONYI SMC, DANNI-OLIVEIRA IM, MARTINS LC, BRAGA ALF. Air pollution and respiratory diseases among children in the city of Curitiba, Brazil. *Revista de Saúde Pública* 38(5), p. 675-700. 2004.
- ZAMPROGNO, B. Análise de componentes principais no domínio do tempo e suas implicações em dados autocorrelacionados. Tese (Doutorado) Programa de Pós-Graduação em Engenharia Ambiental do Centro Tecnológico, UFES, Vitória, 2013.
- MCGREY, G.R, HAN, CHIROK, SUL, D. Estimating the number of common factors in serially dependent approximate factor models. *Economics Letters* 116, p. 531-534. 2012.
- HAMILTON, J.D. *Time Series Analysis*. Princeton, NJ: Princeton University Press. 1994.

Capítulo 5

Conclusão e Recomendações

Esta dissertação avaliou o efeito da associação entre as concentrações atmosféricas dos poluentes PM_{10} , SO_2 , NO_2 , O_3 e CO e o número diário de atendimentos por doenças respiratórias em crianças menores de 6 anos, na RGV, no período entre janeiro de 2005 a dezembro de 2010. Para estimar essa associação, utilizou-se o Modelo Aditivo Generalizado (MAG), com duas diferentes vertentes. O trabalho propôs o casamento entre a técnica de Análise de Componentes Principais (ACP) e o MAG e também a utilização de um modelo de filtragem dos dados ambientais, descrito pelo Modelo Vetorial Autoregressivo (VAR). Para as análises foi utilizada a média geral dos poluentes entre as estações de monitoramento (RAMQAr). Estatísticas descritivas foram utilizadas para a análise dos dados, através de gráficos e cálculos de medidas de tendência central e variabilidade.

Um comportamento sazonal foi verificado nos atendimentos e nas concentrações dos poluentes, uma propriedade esperada para esses fenômenos ambientais, conforme gráficos do Box plot no Apêndice. Foi verificada a correlação entre os poluentes e as variáveis meteorológicas. Essas correlações são significativas estatisticamente, sendo observada mais fortemente entre os poluentes NO_2 e CO , seguidos por PM_{10} e NO_2 e PM_{10} e SO_2 .

O modelo central utilizado foi construído com base nas variáveis explicativas dadas pelas 3 componentes principais selecionadas, as variáveis meteorológicas (funções não paramétricas de alisamento) e as de controle como: dias da semana e feriados (variáveis indicadoras) e o número de dias transcorridos (função não paramétrica de alisamento). Após a modelagem inicial do modelo MAG-ACP, verificou-se a necessidade de inserção do processo AR sazonal, para estimativas finais mais precisas, já que o modelo apresentou resíduos autocorrelacionados. Com o modelo final ajus-

tado, as estimativas se mostraram mais adequadas, para o ajuste dos atendimentos e os poluentes estudados. O modelo MAG-ACP-SAR foi uma das propostas desta dissertação.

Na modelagem MAG-ACP-SAR, o risco relativo estimado para um aumento de $10.49 \mu\text{g}/\text{m}^3$ (intervalo interquartilico) nos níveis de material particulado (PM_{10}) é de 1.03(IC 95%:0.99,1.09), enquanto para um aumento de $4.51 \mu\text{g}/\text{m}^3$ nos níveis de SO_2 , o RR é de 0.98(IC 95%:0.97,1.00). O RR do CO foi o maior observado com 1.05(IC 95%:1.00,1.07), para um aumento de $306.27 \mu\text{g}/\text{m}^3$. Já para um aumento de $9.78 \mu\text{g}/\text{m}^3$ nos níveis de NO_2 , o RR foi de 1.03(IC 95%:1.01,1.04). O RR para um aumento de $10.61 \mu\text{g}/\text{m}^3$ nos níveis do O_3 , foi de 1.08(IC 95 %:1.00,1.09).

Como procedimento alternativo o modelo VAR-MAG-ACP foi ajustado aos dados. Esse método de filtragem dos dados atmosféricos foi realizado para eliminar a autocorrelação. Os resultados encontrados comprovaram que o modelo VAR-MAG-ACP produz melhores estimativas e como esperado, elimina quase por completo a autocorrelação entre as componentes.

Na modelagem VAR-MAG-ACP, o RR apresenta valores ainda mais significativos que nas duas outras modelagens, exceto para o SO_2 . O efeito relativo ao CO corresponde a um aumento percentual de 8% nos atendimentos. O PM_{10} apresenta um aumento de 7%, o SO_2 um aumento de 3% e o NO_2 2%. O efeito do poluente O_3 não apresentou aumento no números de atendimentos por essa modelagem.

Os resultados encontrados comprovaram que os modelos propostos MAG-ACP-SAR e VAR-MAG-ACP produziram melhores estimativas no RR e na qualidade do ajuste. O modelo VAR-MAG-ACP se mostrou ainda mais significativo nas estimativas do RR, comprovando que a presença de autocorrelação nos dados ambientais gera estimativas diferentes, o que corrobora as hipóteses descritas nesta pesquisa. Esses resultados evidenciaram a relação entre os níveis de concentração dos poluentes e o número de atendimentos hospitalares em crianças, por causas respiratórias, comprovando que mesmo em um ambiente com níveis abaixo dos padrões recomendados pelo CONAMA e OMS as concentrações dos poluentes atmosféricos causam efeitos adversos à saúde.

Para trabalhos futuros, recomenda-se a utilização da técnica de ACP na análise de dados de frequência. Outros grupos suscetíveis da população, por exemplo, os idosos. Outras doenças podem ser associadas aos níveis de concentração dos poluentes, as cardiovasculares, por exemplo. Quanto a metodologia pode-se utilizar outras técnicas como a de bootstrap para estimar os in-

tervalos de confiança com mais precisão e a modelagem GLARMA, para resolver o problema da autocorrelação serial dos dados.

Referências Bibliográficas

- [1] ALMEIDA, M. A. I. Modelo Aditivo Generalizado (MAG) no estudo da relação entre o número de atendimentos hospitalares por causas respiratórias e a qualidade do ar. Dissertação (Mestrado) Programa de Pós-Graduação em Engenharia Ambiental do Centro Tecnológico, UFES, Vitória, 2006.
- [2] BARBOSA, G.C. O modelo aditivo generalizado e a técnica de Bootstrap: um estudo entre o número de atendimento hospitalar por causas respiratórias e a qualidade do ar. 60 f. Dissertação (Mestrado) Programa de Pós-Graduação em Engenharia Ambiental do Centro Tecnológico, UFES, Vitória. 2009.
- [3] BENJAMIN, M. A., RIGBY, R. A., AND STASINOPOULOS, D. M. Generalized Autoregressive Moving Average Models. *Journal of the American Statistical Association*, Vol. 98, No. 461. 2003.
- [4] CHEN R. J., CHU C., TAN, J. G., et al. Ambient air pollution and hospital admission in Shanghai, China. *Journal of Hazardous Materials*. 2010.
- [5] DAVIS, R.A.; DUNSMUIR, W.T.M.; WANG, Y. Modelling time series of count data. *Asymptotics, Nonparametrics & Time Series*, 63-114, New York. 1999.
- [6] DAVIS, R.A.; DUNSMUIR, W.T.M.; STREETT, S.B. Observation-driven models for Poisson counts. *Biometrika*, 90(4), 777-790. 2003.
- [7] GOUVEIA, N.; BREMNER, S. A.; NOVAES, H. M. Association between ambient air pollution and birth weight in São Paulo, Brazil. *Journal Epidemiologic Community Health* 58, 11-17. 2004.
- [8] HOLGATE, S. T. et al. *Air Pollution and Health*. 1999.

- [9] HOTELLING, H. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology* 24, 417-441. 1933.
- [10] IBGE. Dados do Censo 2010. Primeiros resultados. Disponível em: <http://www.ibge.gov.br/home/estatistica/populacao/censo2010/>. Acesso dia 05 de dezembro de 2011.
- [11] EPA (ENVIRONMENTAL PROTECTION AGENCY). How nitrogen oxides affect the way we live and breathe. Disponível em: <http://www.epa.gov/air/urbanair/nox/noxfldr.pdf>. Acesso em: 11/10/2012.
- [12] JOHNSON, R.A.; WICHERN, D.W. *Applied multivariate statistical analysis*. 6th edition. Prentice Hall, New Jersey, 800p. 2007.
- [13] KEDEM, B. e FOKIANOS, K. *Regression Models for Time Series Analysis*. Wiley, USA, 2th edition. 2002.
- [14] Martins, L.; et. al. Relação entre poluição atmosférica e atendimentos por infecções de vias aéreas superiores no município de São Paulo: avaliação do rodízio de veículos. *Revista Brasileira de Epidemiologia*. 2001.
- [15] MCCULLAGH, P.; NELDER, J.A. *Generalized Linear Models*. Chapman and Hall, Londres, 2nd. 1989.
- [16] NELDER, J. A.; WEDDERBURN, R. W. M. *Generalized Linear Models*. *Journal of the Royal Statistical Society A* 135, 370-384. 1972.
- [17] OSTRO, B. D., ESKELAND, G. S., SÁNCHEZ, J. M. e FEYZIOGLU T. Air pollution and health effects: A study of medical visits among children in Santiago, Chile. *Environmental Health Perspect* 107(1), 69-73. 1999.
- [18] POSCHL, U., 2005. Atmospheric aerosols: composition, transformation, climate and health effects. *Angew. Chem. Int. Ed. Engl.* 44, 7520.
- [19] ROBERTS S, MARTIN M. Using Supervised Principal Components Analysis to Assess Multiple Pollutant Effects. *Environmental Health Perspectives*, Vol. 116, No. 12. 2006.
- [20] SALDIVA, P.H.N. FREITAS, C. et al. Internações e óbitos e sua relação com a poluição atmosférica em São Paulo, 1993 a 1997. *Revista Saúde Pública*. 2004.

- [21] SALDIVA, P.H.N. FREITAS, C. et al. Associação entre poluição atmosférica e doenças respiratórias e cardiovasculares na cidade de Itabira, Minas Gerais, Brasil. *Caderno Saúde Pública*, Rio de Janeiro, 23. 2007.
- [22] SCHWARTZ, J. Harvesting and long term exposure effects in the relationship between air pollution and mortality. *American Journal Epidemiologic* 151, 440- 448. 2000.
- [23] ZOU, Guangyong. A Modified Poisson Regression Approach to Prospective Studies with Binary Data. *American Journal of Epidemiology*, Vol. 159, No. 7. 2004.
- [24] WANG, Yaping and PHAM, Hoang. Analyzing the effects of air pollution and mortality by generalized additive models with robust principal components. *International Journal of System Assurance Engineering and Management*, 2, p. 253- 259. 2011.
- [25] WHO. World Health Organization. 2005.
- [26] WHO. World Health Organization. 2006.
- [27] WHO. World Health Organization. 2012.

Apêndice A

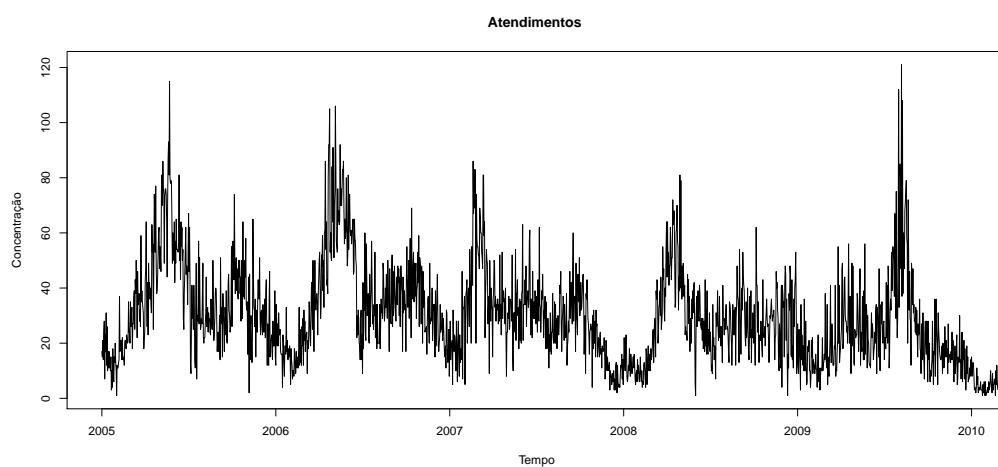


Figura A.1: Série do número de atendimentos hospitalares por causas respiratórias diário de crianças de 0 a 6 anos da RGV.

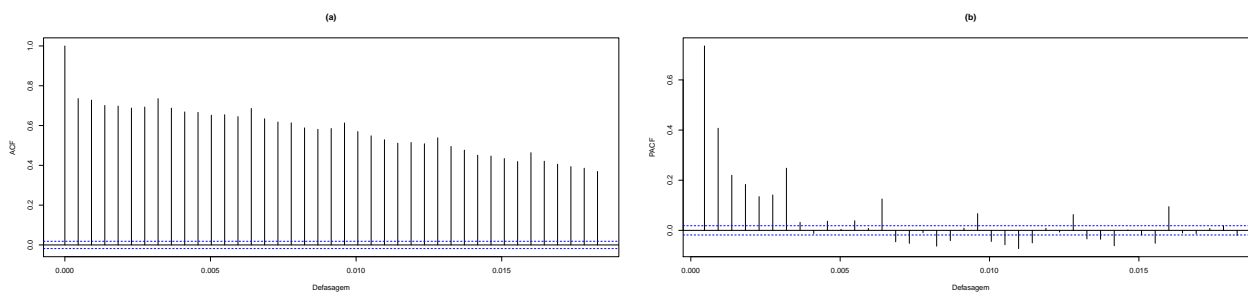


Figura A.2: (a)-Função de Autocorrelação Amostral; (b)-Função Autocorrelação Amostral Parcial

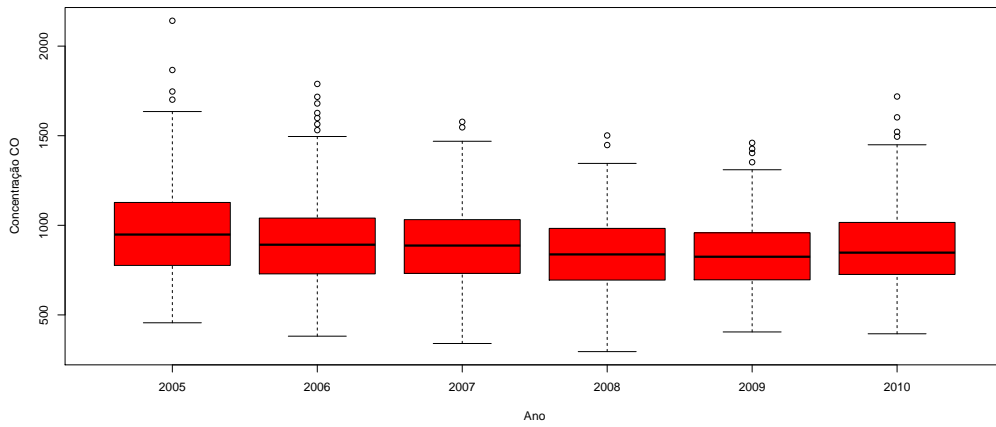


Figura A.3: Box Plot das concentrações do poluente CO

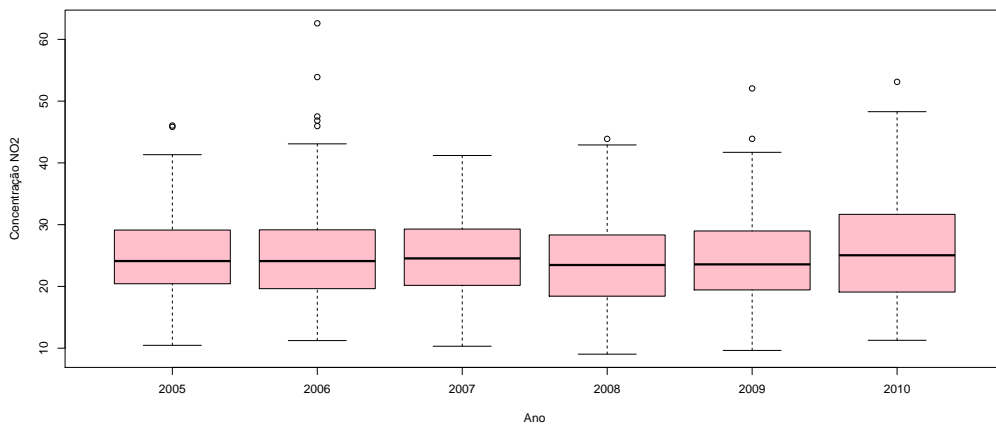


Figura A.4: Box Plot das concentrações do poluente NO₂

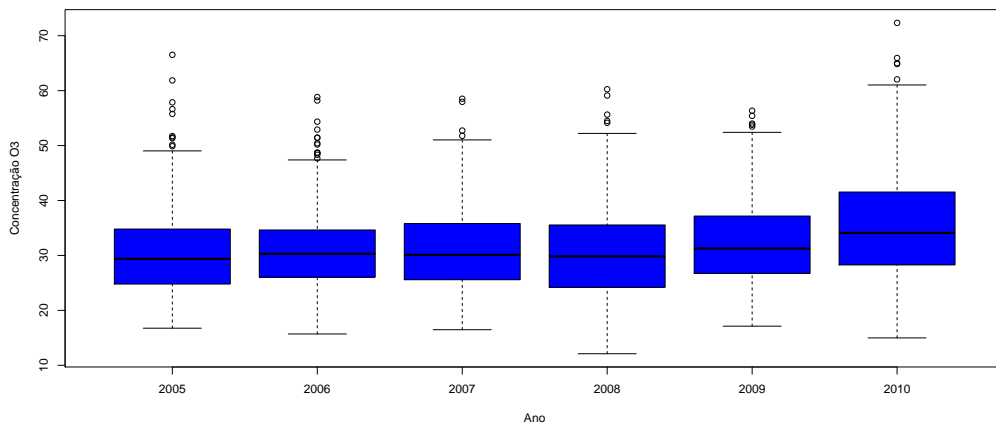


Figura A.5: Box Plot das concentrações do poluente O₃

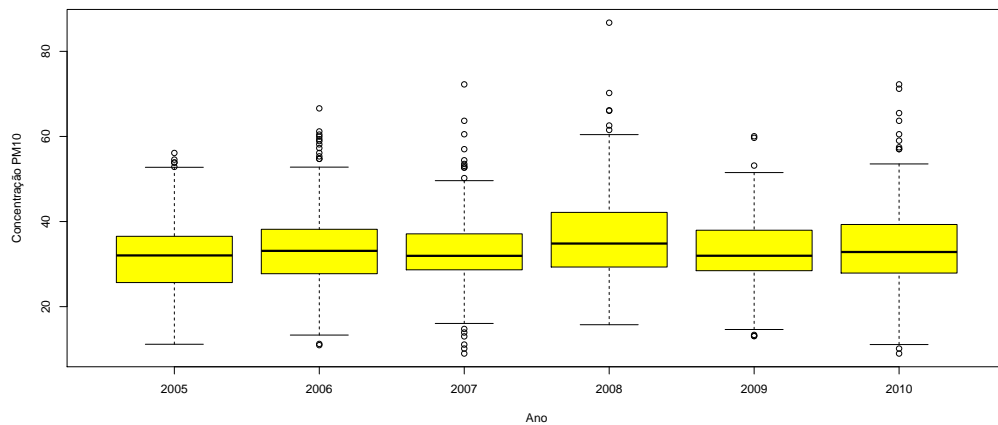


Figura A.6: Box Plot das concentrações do poluente PM₁₀

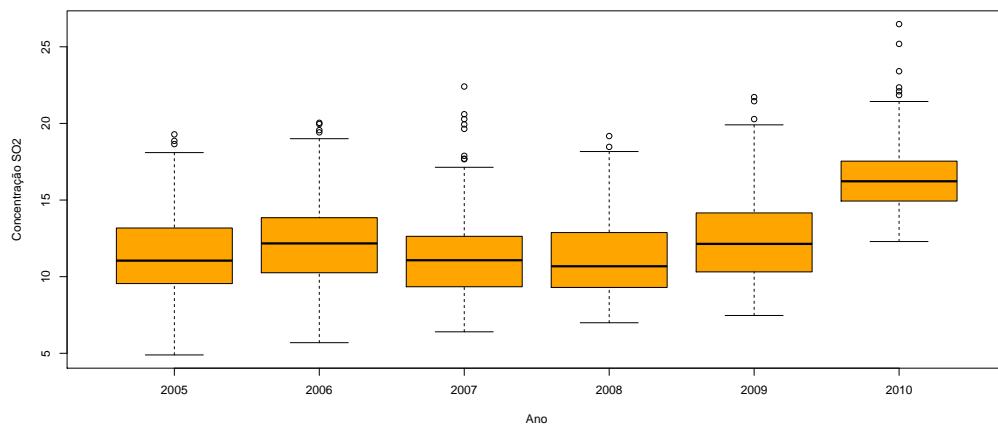


Figura A.7: Box Plot das concentrações do poluente SO₂

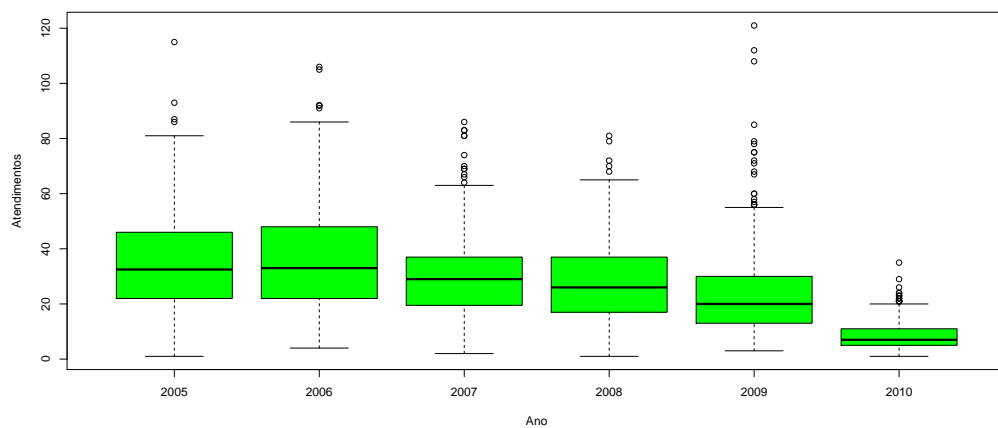


Figura A.8: Box Plot dos dados de atendimentos diários por doenças respiratórias em crianças menores de 6 anos

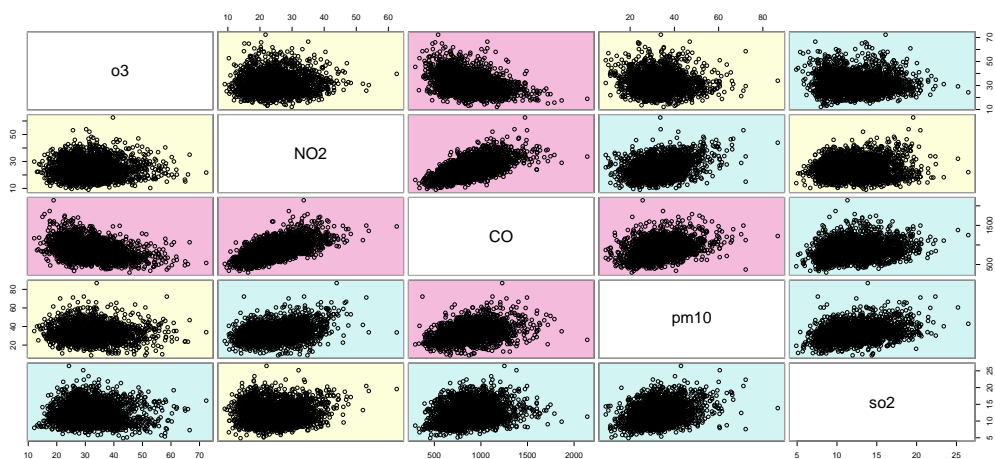


Figura A.9: Matriz de correlação dos poluentes

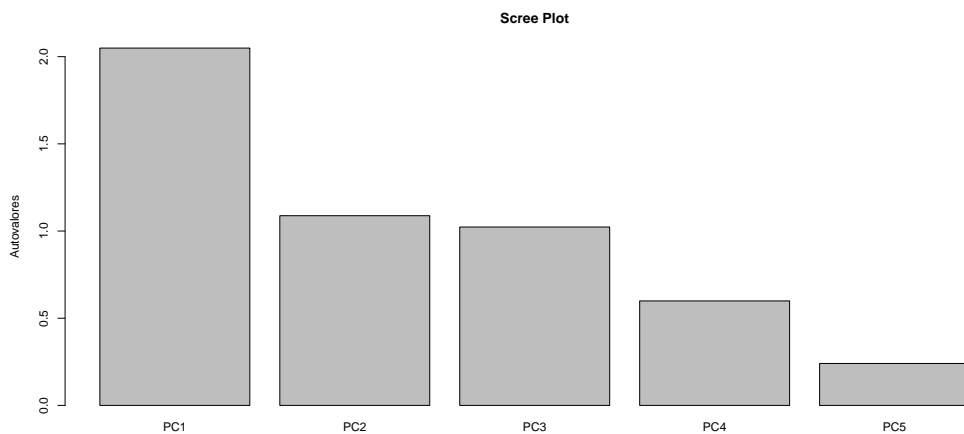


Figura A.10: Gráfico de Scree plot das componentes principais

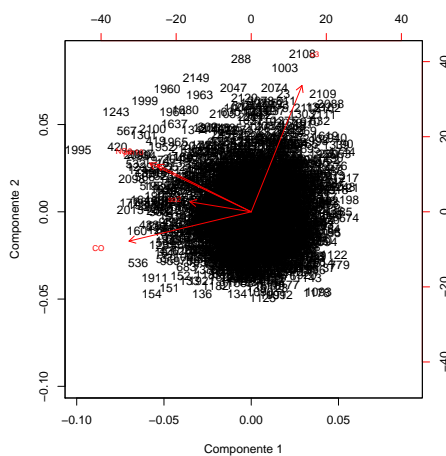


Figura A.11: Gráfico do Biplot das componentes principais

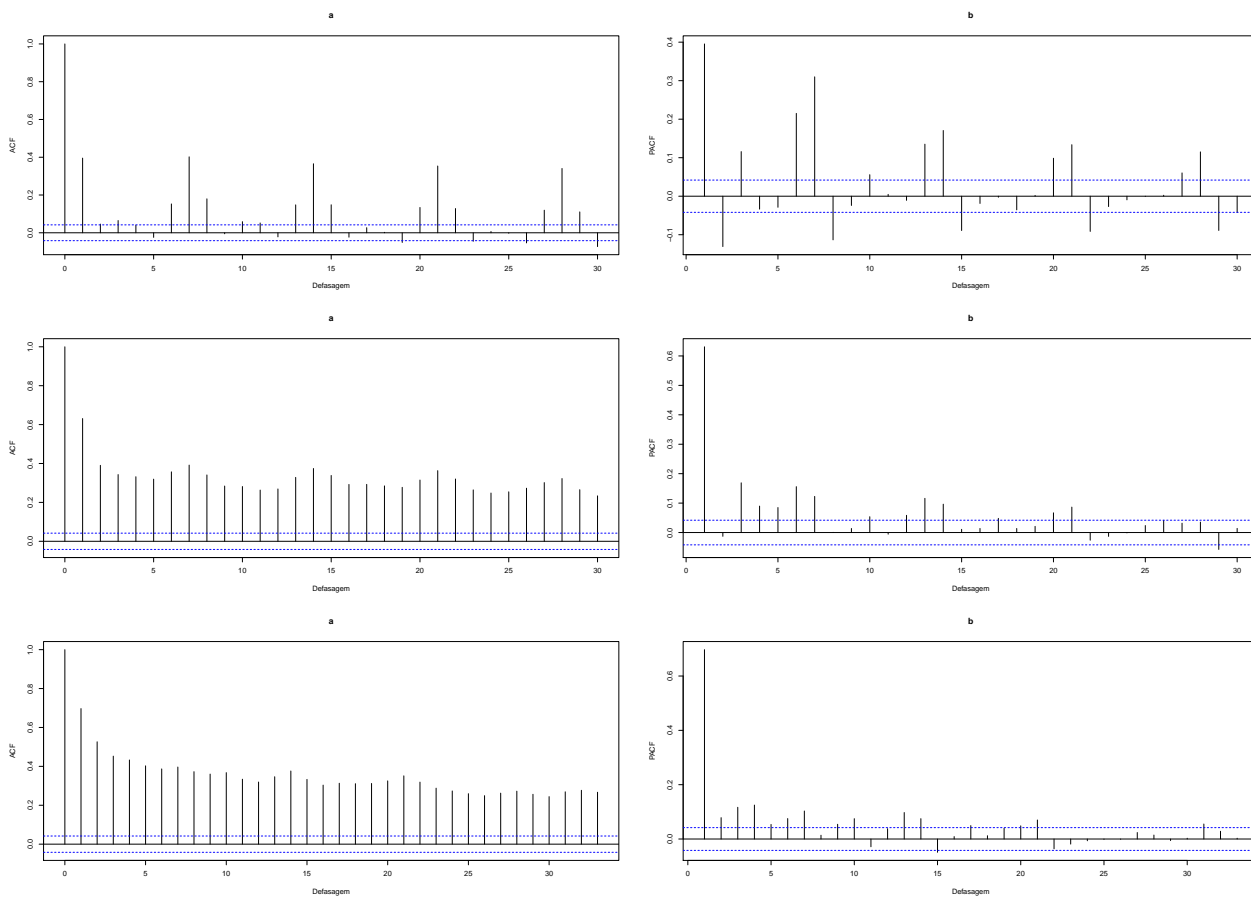


Figura A.12: (a)-Função de Autocorrelação Amostral; (b)-Função Autocorrelação Amostral Parcial

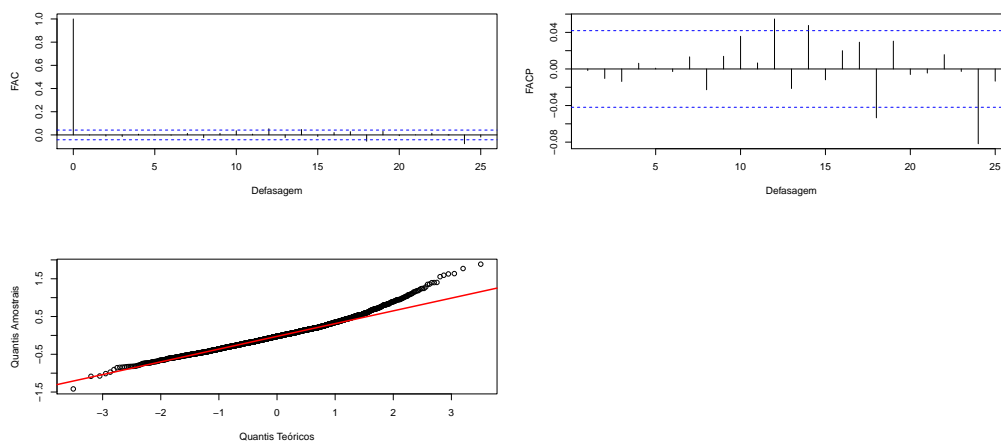


Figura A.13: Resíduos do modelo final VAR-MAG-ACP.