



UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
Câmpus de Presidente Prudente

CARLOS SOUTO DOS SANTOS FILHO

**ANÁLISE DE FATORES DE RISCO AMBIENTAIS PARA RISCO RELATIVO DE
INTERNAÇÃO DE DOENÇAS CARDIOVASCULARES
VIA MODELO TEMPORAL DE CONTAGEM**

Revisado pelo Orientador

Assinatura do Orientador

Data ____ / ____ / 2025

PRESIDENTE PRUDENTE

2025

CARLOS SOUTO DOS SANTOS FILHO

**ANÁLISE DE FATORES DE RISCO AMBIENTAIS PARA RISCO RELATIVO DE
INTERNAÇÃO DE DOENÇAS CARDIOVASCULARES
VIA MODELO TEMPORAL DE CONTAGEM**

Relatório Final para Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Estatística da FCT/Unesp para aproveitamento na disciplina Trabalho de Conclusão de Curso.

Orientador: Prof. Dr. Guilherme Aparecido Santos Aguilar.

PRESIDENTE PRUDENTE

2025

S237a

Santos Filho, Carlos Souto dos

Análise de fatores de risco ambientais para risco relativo de internação de doenças cardiovasculares via modelo temporal de contagem / Carlos Souto dos Santos Filho. -- Presidente Prudente, 2025

54 p. : il., tabs.

Trabalho de conclusão de curso (Bacharelado - Estatística)

Universidade Estadual Paulista (UNESP), Faculdade de Ciências e Tecnologia, Presidente Prudente

Orientadora: Guilherme Aparecido Santos Aguiar


1. Série Temporal. 2. Modelo Aditivo Generalizado. 3. Risco relativo. 4. Doenças Cardiovasculares. 5. Poluentes do Ar. I. Título.

TERMO DE APROVAÇÃO

CARLOS SOUTO DOS SANTOS FILHO


ANÁLISE DE FATORES DE RISCO AMBIENTAIS PARA RISCO RELATIVO DE
INTERNAÇÃO DE DOENÇAS CARDIOVASCULARES
VIA MODELO TEMPORAL DE CONTAGEM

Relatório de Final de Trabalho de Conclusão de Curso aprovado como requisito para obtenção de créditos na disciplina Trabalho de Conclusão do curso de graduação em Estatística da Faculdade de Ciências e Tecnologia da Unesp, pela seguinte banca examinadora:

Documento assinado digitalmente
 GUILHERME APARECIDO SANTOS AGUILAR
Data: 02/12/2025 09:16:37-0300
Verifique em <https://validar.iti.gov.br>

Orientador: _____

Prof. Dr. Guilherme Aparecido Santos Aguilár
Departamento de Estatística

Documento assinado digitalmente
 MANOEL IVANILDO SILVESTRE BEZERRA
Data: 10/12/2025 11:32:09-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Manoel Ivanildo Silvestre Bezerra
Departamento de Estatística

Presidente Prudente, 02 de dezembro de 2025.

AGRADECIMENTOS

Primeiramente, gostaria de agradecer ao meu orientador, Professor Guilherme A. S. Aguiar. Sua condução neste trabalho foi exemplar, reconfortante e magistral. Agradeço sua atenção além das barreiras da Universidade e seus *insights* para sempre seguir adiante. Também agradeço ao Professor Manoel Silvestre por sempre exigir o rigor estatístico, mesclando momentos de generosa transmissão de conhecimentos em pontos cruciais desta monografia. À Professora Ana Júlia (UFES) pelo suporte científico. Minha gratidão eterna aos meus pais, que se foram fisicamente, mas cujo amor e sabedoria ecoam em cada passo da minha vida. A lembrança de vocês sempre será minha volição para buscar meus objetivos. Agradeço meus entes vivos, em particular, meu irmão, Gustavo, meus sobrinhos, Henry e Marina, meu primo, Paulo Roberto. A motivação de vocês, transmitida em doses homeopáticas, foi recebida com muito carinho. Ao todo Universo “Unesp”, estudantes, professores e funcionários por sempre primar o profissionalismo e a ética nas relações humanas. Por último, agradeço a natureza acolhedora do Campus Presidente Prudente, por *rapport* com papagaios, tucanos, gaviões, sabiás, ipês, palmeiras, flores e frutos. Encerrarei com o verso do poema "*Le Cimetière marin*", do poeta francês Paul Valéry, que expressa a adaptação e resiliência humana diante dos ventos aleatórios da vida:

“The wind rises, we must try to live”

RESUMO

A importância dos fatores de risco associados à doença cardiovascular (DCV) ganhou destaque pela publicação inédita em 1961 do Estudo Framingham dos preditores da doença coronária. Características do ambiente exercem influência importante no risco, evolução e gravidade das DCV e os modelos preditivos permitem identificar fatores intermediários que medeiam as relações entre fatores individuais e comportamentais. Objetivou-se explorar as associações entre concentrações diárias de poluentes atmosféricos e variável meteorológica (temperatura média) em admissões hospitalares por DCV em São Paulo, Brasil. Usando o DATASUS, foram extraídos dados de internações hospitalares (SIH) de 2010 a 2019. Dados de poluentes do ar e variável meteorológica foram obtidos da CETESB, ECMWF e INMET no mesmo período. Foi utilizado um modelo aditivo generalizado adicionado a um componente ARMA, capaz de lidar tanto com a estrutura de autocorrelação da série temporal quanto com a não linearidade existente nas covariáveis. Utilizou-se o software estatístico R (versão 4.4.0). Observou-se correlação moderada a forte entre diversos poluentes, evidenciando multicolinearidade relevante. Os poluentes com maior correlação com as internações foram NO_2 ($r \approx 0,32$), SO_2 ($r \approx 0,27$), $\text{MP}_{10\mu\text{m}}$ ($r \approx 0,19$) e $\text{MP}_{2,5\mu\text{m}}$ ($r \approx 0,13$). A decomposição da série temporal revelou sazonalidade anual bem marcada, com oscilação regular e repetitiva, e tendência de longo prazo com variações suaves. O modelo MAG-AR(1) apresentou melhor ajuste para $\text{MP}_{10\mu\text{m}}$ e $\text{MP}_{2,5\mu\text{m}}$. As estimativas de risco relativo foram significativas para $\text{MP}_{10\mu\text{m}}$ [RR = 1,224; IC95%: 1,223–1,236], indicando que incrementos neste poluente contribuíram significativamente para o aumento de internações por DCV. Para $\text{MP}_{2,5\mu\text{m}}$, não foram observadas associações estatisticamente significativas com o desfecho analisado. O modelo MAG-ARMA mostrou-se adequado para análise epidemiológica de séries temporais, pois considera simultaneamente a autocorrelação temporal e a não linearidade das relações entre poluentes e desfechos em saúde, permitindo estimativas mais precisas dos efeitos da exposição ambiental.

Palavras-chave: Doenças Cardiovasculares, Poluentes do Ar, Série Temporal, Modelo Aditivo Generalizado, Risco relativo.

ABSTRACT

The importance of risk factors associated with cardiovascular disease (CVD) gained prominence through the unprecedented publication in 1961 of the Framingham Study on coronary disease predictors. Environmental characteristics exert significant influence on CVD risk, progression, and severity, and predictive models can identify intermediate factors that mediate relationships between individual and behavioral factors. This study aimed to explore associations between daily concentrations of atmospheric pollutants and meteorological variables (mean temperature) in hospital admissions for CVD in São Paulo, Brazil. Using DATASUS, hospital admission data (SIH) were extracted from 2010 to 2019. Air pollutant and meteorological variable data were obtained from CETESB, ECMWF and INMET for the same period. A generalized additive model supplemented with an ARMA component was used, capable of handling both the autocorrelation structure of the time series and the non-linearity present in the covariates. Statistical software R (version 4.4.0) was utilized. Moderate to strong correlations were observed between several pollutants, highlighting significant multicollinearity. The pollutants with the highest correlation to hospitalizations were NO_2 ($r \approx 0.32$), SO_2 ($r \approx 0.27$), $\text{PM}_{10\mu\text{m}}$ ($r \approx 0.19$), and $\text{PM}_{2.5\mu\text{m}}$ ($r \approx 0.13$). Time series decomposition revealed a well-marked annual seasonality, with regular and repetitive oscillation, and a long-term trend with smooth variations. The MAG-AR model(1) showed the best fit for $\text{PM}_{10\mu\text{m}}$ and $\text{PM}_{2.5\mu\text{m}}$. Relative risk estimates were significant for $\text{PM}_{10\mu\text{m}}$ [RR = 1.224; 95% CI: 1.223–1.236], indicating that increases in this pollutant were significantly associated with increased CVD hospitalizations. For $\text{PM}_{2.5\mu\text{m}}$, no statistically significant associations were observed with the outcomes developed. The GAM-ARMA model proved suitable for epidemiological time series analysis because it simultaneously considers temporal autocorrelation and the non-linearity of the relationships between pollutants and health outcomes, allowing for more precise estimates of the effects of environmental exposure.

Keywords: Cardiovascular Diseases, Air Pollutants, Time Series, Generalized Additive Model, Relative Risk.

Lista de ilustrações

Figura 1 – Gráficos das Séries Temporais em contagem diária (A - Internação DCV; B - poluentes do ar; C – climáticas).....	31
Figura 2 – Gráficos das Séries Temporais em dados mensais (Internação DCV; poluentes do ar; Temperatura média).....	33
Figura 3 – Matriz de correlação entre poluentes atmosféricos, temperatura média e internações por DCV	34
Figura 4 - Decomposição da escala temporal (MP10 μ m e internação), 2010-2019	35
Figura 5 - Decomposição da escala temporal (MP2.5 μ m e temperatura média), 2010-2019	35
Figura 6 – Análise Gráfica da ACF e PACF dos resíduos	38
Figura 7 – Análise Gráfica da ACF e PACF dos resíduos (MAG -MP10 mensal)	39
Figura 8 - Análise Gráfica da ACF e PACF dos resíduos (MAG-AR(1) MP10 mensal).....	40
Figura 9 – Análise Gráfica da ACF e PACF dos resíduos (MAG e MAG-AR(1)	42

Lista de tabelas

Tabela 1 – Estatísticas Descritivas de Interações Cardiovasculares (2010–2019, SP)	28
Tabela 2 – Estatísticas Descritivas dos Poluentes Atmosféricos (2010–2019, SP)	29
Tabela 3 – Estatísticas Descritivas de Variáveis Meteorológicas (2010–2019, SP)	29
Tabela 4 - Estatísticas das Séries Temporais (Interação DCV)	30
Tabela 5 - Estatísticas das Séries Temporais de Poluentes do ar	30
Tabela 6 - Estatísticas das Séries Temporais Meteorológicas	30
Tabela 7- Resumo do modelo: covariável MP10 μ m (decomposição por escalas)	36
Tabela 8 - Resumo do modelo: covariável MP2,5 μ m (decomposição por escalas	36
Tabela 9 - Resultados do Modelo MAG para MP10	37
Tabela 10 - Resultados do Modelo MAG para MP10 μ m	39
Tabela 11 - Resultados do Modelo MAG-AR(1) para MP10 μ m	40
Tabela 12 - Estimativas de risco relativo e intervalos de confiança de 95% para a variação interquartil das concentrações de MP10 μ m entre os modelos MAG e MAG-AR(1)	41
Tabela 13 - Resultados dos Modelos MAG e MAG-AR(1) para MP2,5 μ m.....	41
Tabela 14 - Estimativas de RR e IC 95% para a variação interquartil das concentrações de MP2,5 μ m entre os modelos MAG e MAG-AR(1).....	42

Sumário

Lista de ilustrações	7
Lista de tabelas	8
1 INTRODUÇÃO	10
2 REVISÃO BIBLIOGRÁFICA / REFERENCIAL TEÓRICO	13
3 METODOLOGIA / MATERIAIS E MÉTODOS	15
3.1 Pergunta da Pesquisa e Revisão da Literatura	15
3.2 Área de estudo	15
3.3 Extração dos Dados	17
3.4 Análise Estatística Computacional	17
3.5 Técnica Estatística	18
3.5.1 Regressão em Séries Temporais em Epidemiologia Ambiental	18
3.5.2 Modelos de Séries Temporais Semiparamétricos	19
3.5.3 Modelo Aditivo Generalizado	19
3.5.4 Modelo MLG-ARMA	20
3.5.5 Modelo MAG-ARMA	21
3.5.6 Obtenção dos Parâmetros Estimados / Ajuste do Modelo	23
3.5.7 Graus de Liberdade e Seleção de Suavidade	25
3.5.8 Ajuste do Modelo para Cada Poluente Atmosférico	27
4 RESULTADOS E DISCUSSÃO	28
4.1 Análise Exploratória dos Dados	28
4.2 Decomposições por Escalas Temporais	34
4.3 Ajuste do Modelo para MP10 μ m e MP2,5 μ m	37
5 CONCLUSÕES	44
REFERÊNCIAS	45
APÊNDICE A	50
APÊNDICE B	51
APÊNDICE C	52

1 INTRODUÇÃO

Os efeitos da exposição a fatores ambientais em doenças crônico-degenerativas são os mais amplamente investigados na Toxicologia Ambiental. Diversos estudos epidemiológicos relataram que um acúmulo de fatores de risco ambiental aumenta a probabilidade de eventos de DCV e mortes (Erqou *et al.*, 2017; Caleyachetty *et al.*, 2014; Nandi *et al.*, 2012).

Evidências cumulativas relativas à poluição atmosférica publicadas pela Associação Americana do Coração comprovam a relação causal entre a exposição ao material particulado de dimensão inferior a 2,5µm (MP2,5µm) e a morbimortalidade cardiovascular (Sørensen; Pershagen, 2019). Outros estudos prévios referiram que a exposição a MP2,5µm ocasiona cerca de 70% a 80% de mortes prematuras de origem cardiovascular (Bhatnagar, 2006).

Poluentes ambientais, tais como produtos químicos orgânicos voláteis, Monóxido de Carbono (CO), Dióxido de Nitrogênio (NO₂), Ozônio (O₃) e sulfatos foram relatados por especialistas do Conselho de Epidemiologia e Prevenção da Associação Americana do Coração como fatores contribuintes para a carga de risco de DCV, pois estes poluentes constituem cerca de 98% da mistura gasosa de centros urbanos e podem otimizar os efeitos nocivos do material particulado (Brook *et al.*, 2010).

No cenário nacional, pesquisadores investigaram a relação entre mortalidade diária de idosos acima de 65 anos e poluentes ambientais na região metropolitana de São Paulo. Com a aplicação de modelos de séries temporais, foi apurado forte associação de mortalidade a material particulado < 10µm (MP10µm), óxidos de nitrogênio (NO_x), dióxido de enxofre (SO₂) e CO, sendo que a associação com MP10µm foi a mais importante, pois teve um aumento na mortalidade geral de 13% (Saldiva *et al.*, 1995).

A partir da década de 2000, pesquisa sobre exposição ambiental a poluentes do ar e internações por DCV de idosos na cidade de São Paulo foi publicada com ênfase nos modelos aditivos generalizados de regressão de Poisson ajustados para efeitos defasados até 20 dias após a exposição. Os poluentes MP10µm e SO₂ foram associados a um aumento de 3,17% na insuficiência cardíaca congestiva e 0,89% no total de doenças cardiovasculares. Autores comentaram sobre a forte associação

encontrada entre exposição a poluentes do ar e internações cardiovasculares (Martins *et al.*, 2006).

Apesar da importância de MP10 μ m, ainda há escassez de evidências sobre a exposição a MP2,5 μ m e risco relativo (RR) de internação por DCV na cidade de São Paulo. Somado a isso, não existem informações para faixa etária inferior a 65 anos.

Partindo desta lacuna, o presente estudo levanta o questionamento se existe diferença no tipo de poluente (MP2,5 μ m e MP10 μ m) na associação com RR de internações por DCV na cidade de São Paulo. Esta pergunta foi elaborada segundo as estratégias facilitadoras usadas na área da Saúde e que foram descritas na Metodologia.

O objetivo principal deste estudo foi quantificar o RR de internação hospitalar por DCV e identificar os fatores associados, principalmente os preditores ambientais MP (2,5 μ m e 10 μ m) com a aplicação de um modelo estatístico de estimação na cidade de São Paulo, Brasil.

A delimitação de objetivos secundários (específicos) desta pesquisa foram:

- Explorar a influência da associação de diferentes MP (2,5 μ m e 10 μ m) e RR de internações por DCV;
- Avaliar a associação entre exposição diária a poluentes atmosféricos (CO, O₃, NO₂, SO₂, MP) e variáveis climáticas com as hospitalizações por DCV;
- Propor um Modelo Aditivo Generalizado (MAG) para dados de contagem (internações por DCV) capaz de lidar tanto com a estrutura de autocorrelação da série temporal quanto com a não linearidade existente nas covariáveis, descrito na literatura como MAG adicionado a um componente ARMA (*autoregressive–moving-average*);
- Ajustar e verificar a performance de MAG-ARMA para avaliar o impacto de poluentes atmosféricos e variáveis meteorológicas no número de internações de DCV.

Partindo dos estudos descritos anteriormente, a evidência empírica do impacto das exposições ambientais de agentes tóxicos atmosféricos e condições meteorológicas na saúde humana é assunto de relevância mundial (Kelsall *et al.*, 1997). Tais estudos são um alerta sobre a importância de controlar e reduzir as emissões de poluentes atmosféricos e dar suporte aos gestores de saúde na alocação de recursos.

O interesse pelo assunto tem crescido exponencialmente com publicações focadas também na quantificação e monetização dos efeitos de covariáveis (poluentes do ar e climáticas) em diversos setores econômicos, recursos ambientais, consumo

de energia, qualidade do ar e saúde humana. De investigações sobre o impacto econômico destas covariáveis na Saúde, pesquisa de meta-análise associou as temperaturas médias com taxas de mortalidade por doenças respiratórias e cardiovasculares (Martens,1998). Já Patz *et al.* (2005) realizaram estudos preditivos dos incrementos nos riscos para a saúde sob as projeções futuras de mudança climática. Com os resultados obtidos, os autores também apontaram aumentos de morbidade e mortalidade em escala global.

Diante do exposto, e levando-se em consideração da responsabilidade social da Universidade, esta pesquisa atende alguns Objetivos de Desenvolvimento Sustentável (ODS) da Agenda 2030, adotada pelos membros das Nações Unidas na Assembleia Geral em 2015. Dos 17 ODS, este estudo trata do ODS 3 (“Garantir vidas saudáveis e promover o bem-estar para todos, em todas as idades”) e ODS 13 (“Tomar medidas urgentes para combater as alterações climáticas e os seus impactos”) (Guia [...], 2024).

2 REVISÃO BIBLIOGRÁFICA / REFERENCIAL TEÓRICO

Nos últimos anos, as pesquisas alertaram para a desaceleração do declínio das mortes por DCV, mas estas ainda lideram as causas de mortes no mundo (Virani *et al.*,2021). A carga crescente de fatores de risco cardiovascular em adultos jovens poderia explicar parcialmente este achado (Jacobs *et al.*,2022). No Brasil, o comportamento epidemiológico segue a mesma tendência mundial, sendo a principal causa de morte, especialmente em idosos (Mansur; Favarato,2012).

Por muito tempo, a importância dos fatores de risco associados à DCV foi negligenciada. O interesse pelo assunto ganhou destaque pela publicação inédita em 1961 do Estudo Framingham dos preditores da doença coronária (Kannel *et al.*,1961). O interesse pelo assunto reascendeu com a descoberta de que características do ambiente exercem influência importante no risco, evolução e gravidade das DCV (Bhatnagar, 2017). As modificações no ambiente concomitante ao crescimento populacional e aumento da longevidade acarretaram as aglomerações e crescimento das cidades, modificações no estilo de vida, mudanças nas políticas sociais e nos aspectos culturais com consequente impacto no risco de DCV, mesmo na ausência de alterações genéticas (Bhatnagar, 2017).

A construção dos modelos preditivos identificou os ambientes físicos e sociais como fatores intermediários que medeiam as relações entre fatores individuais e fatores comportamentais (estilo de vida, aspectos nutricionais) que, em última análise, influenciam os indicadores antropomórficos e fisiológicos de risco cardiovascular (Mansur; Favarato,2012). Tendências recentes em pesquisas epidemiológicas em doenças crônicas levaram a uma proliferação de estudos que mostram o acúmulo de evidências ao longo das últimas décadas dos fatores de risco associados a DCV. Dentre estes, os mais citados na literatura são índice elevado de urbanização (Alizadeh *et al.*,2022), baixa escolaridade (Alizadeh *et al.*,2022), taxa de violência (Alizadeh *et al.*,2022; Powell-Wiley *et al.*,2022), aspectos organizacionais do trabalho (Alizadeh *et al.*,2022), desigualdades sociais (Alizadeh *et al.*,2022; Bhatnagar, 2017; Münzel *et al.*,2022), privação socioeconômica (Bhatnagar, 2017; Alizadeh *et al.*,2022; Powell-Wiley *et al.*,2022), desorganização social (Alizadeh *et al.*,2022; Münzel *et al.*,2022; Powell-Wiley *et al.*,2022), moradia em regiões periféricas (Alizadeh *et al.*,2022), desigualdade de renda (Alizadeh *et al.*,2022;

Münzel *et al.*,2022; Powell-Wiley *et al.*,2022), poluição atmosférica por MP_{2,5}µm e MP₁₀µm (Bhatnagar, 2017; Schulz *et al.*,2005; Alizadeh *et al.*,2022; Münzel *et al.*,2022), nível de O₃ (Münzel *et al.*,2022), intensidade da atividade de petróleo e gás natural (Alizadeh *et al.*,2022), tráfego (Alizadeh *et al.*,2022; Münzel *et al.*,2022), ruído ambiental (Alizadeh *et al.*,2022; Münzel *et al.*,2022), desastre natural (Alizadeh *et al.*,2022), altitude (Bhatnagar, 2017), latitude (Bhatnagar, 2017), espaços verdes (Bhatnagar, 2017; Powell-Wiley *et al.*,2022), acesso ao apoio social (Schulz *et al.*,2005; Sabzmakan *et al.*, 2014) e de saúde (Powell-Wiley *et al.*,2022; Sabzmakan *et al.*, 2014), variáveis climáticas (temperatura, insolação) (Münzel *et al.*,2022), poluição luminosa artificial (Münzel *et al.*,2022), queimadas (Münzel *et al.*,2022), índice de caminhabilidade (Powell-Wiley *et al.*,2022), SO₂ (Jingesi *et al.*,2023) e NO₂ (Jingesi *et al.*,2023) e COVID-19 (Tateishi *et al.*,2023).

Ao avaliar o impacto cumulativo do ambiente no risco de DCV, as pesquisas ainda deixam lacunas nos mecanismos fisiopatológicos de exposição ambiental. Isso se deve a variabilidade e nível de complexidade das covariáveis representadas nos domínios naturais, sociais e pessoais (Bhatnagar, 2017).

Estudo recente de regressão de Poisson multivariada realizado com o objetivo de explorar o efeito incremental de fatores ambientais, além dos fatores sociais, na prevalência de DCV mostrou que pessoas que vivem em áreas com adversidades sociais e ambientais podem ter um risco até duas vezes maior de desenvolver doenças cardíacas e acidente vascular encefálico (Tateishi *et al.*,2023).

Com relação a exposição de longo prazo a poluentes ambientais, existem publicações que mostraram as associações entre risco social cumulativo, dois preditores atmosféricos urbanos (MP_{2,5}µm e carbono negro) e o risco de DCV incidente e mortalidade (Canterbury *et al.*,2020).

Uma compreensão de como diferentes domínios afetam o risco de DCV, nomeadamente a exposição a agentes nocivos ambientais, poderia levar a uma melhor avaliação de predição e ajudar no desenvolvimento de novas estratégias preventivas e terapêuticas.

3 METODOLOGIA / MATERIAIS E MÉTODOS

3.1 Pergunta da Pesquisa e Revisão da Literatura

Para a construção da pergunta norteadora da pesquisa, foram usadas as estratégias PICOS (Methley *et al.*, 2014) e SPIDER (Cooke *et al.*, 2012), que são boas ferramentas para pesquisas na área da Saúde. Optou-se em realizar revisão rápida com a finalidade de unir estudos sobre diferentes tópicos para reinterpretação ou interconexão, a fim de fornecer uma descrição histórica do desenvolvimento da teoria e da pesquisa sobre o tema. Para responder à pergunta da pesquisa, levantaram-se descritores e potenciais sinônimos em língua inglesa usando a ferramenta “Mesh” (Pubmed). Na sequência, foi realizada uma busca utilizando-se as seguintes bases de dados científicas: *Medline (PubMed)*, *Web of Science*, *Scopus* e *Embase*.

3.2 Área de estudo

A cidade de São Paulo pertence a uma divisão política e socioeconômica regional do estado de São Paulo, na região sudeste do Brasil. O município possui uma área territorial de 1.521,202 km² e área urbanizada de 914,56 km² (em 2019). Com uma densidade demográfica de 7.528,26 habitantes por quilômetro quadrado e cerca de 11.451.999 habitantes (dados de 2022). O setor industrial e econômico do município é robusto com PIB per capita de R\$ 66.872,84 (em 2021). Em 2023, ocupou o primeiro lugar no total de receitas realizadas e no total de despesas empenhadas. A emissão veicular é uma das fontes predominante de poluição, o município em agosto de 2024 tinha uma frota total de 9.721.123 veículos, sendo 6.353.930 de automóveis e 146.701 de caminhões (Ministério dos Transportes [...], 2024). Sua delimitação é dada pelo Trópico de Capricórnio e está localizada entre as latitudes 23°20' e 24°00' S e as longitudes 46°20' e 46°50' W, a 760 metros acima do nível do mar (Oliver; Ribeiro, 2016). Segundo a classificação de Köppen-Geiger (Alvares *et al.*, 2013), o clima em São Paulo é subtropical Cfa. Suas características territoriais permitem uma variação climática de transição entre os climas tropicais úmidos de altitude e os climas subtropicais permanentemente úmidos, demarcando uma estação quente e úmida no verão, e uma fria e seca no inverno.

3.3 Extração dos Dados

As fontes de informação foram dados de saúde digitalizados e anonimizados do município de São Paulo do Departamento de Informática do Sistema Único de Saúde (DATASUS), extraídos do Sistema de Internação Hospitalar (SIH) de 2010 a 2019 para a variável dependente (VD) (Doenças isquêmicas do coração e Doenças cerebrovasculares com respectivos Códigos Internacionais de Doenças (CID-10) I21, I24, I50, I60, I61, I62, I63, I64, I65, I66, I67, I68, I69 do Capítulo IX Doenças do aparelho circulatório) estratificada por faixa etária (abaixo 65 anos, 65 a 75 anos e acima de 75 anos) e gênero (masculino e feminino). Critérios raciais, de escolaridade e ocupação não foram estudados. Dados de poluentes do ar (MP10 μ m, NO₂, SO₂) foram obtidos da Companhia Ambiental do Estado de São Paulo (CETESB) das seguintes estações de monitoramento Santana, Parque Dom Pedro II, Congonhas, São Miguel Paulista, Capão Redondo, Ibirapuera, Mooca, Cerqueira Cesar, Cidade Universitária USP-Ipen, Nossa Senhora do Ó, Itaquera, Interlagos, Itaim Paulista, Marginal Tiete-Pte Remédios e Perus. Adicionalmente, os dados de MP2,5 μ m, CO e O₃ foram extraídos da *API CAMS Global Reanalysis* do *Copernicus Atmosphere Monitoring Service*, operado pelo *European Centre for Medium-Range Weather Forecasts* (ECMWF), devido à presença de muitos dados faltantes na base CETESB. O *CAMS Global Reanalysis* combina dados de modelos atmosféricos com observações de satélites através de assimilação de dados, produzindo campos tridimensionais consistentes temporalmente de composição atmosférica (Inness et al., 2019). A validação dos dados CAMS demonstrou adequada qualidade para aplicação em estudos epidemiológicos, com desempenho satisfatório quando comparado com dados de estações de superfície na América do Sul (Requia et al., 2021). Apresenta-se, no APÊNDICE A, imagem de satélite, na qual estão indicadas as localizações das estações de monitoramento operadas pela CETESB no município de São Paulo e na Região Metropolitana.

As covariáveis climáticas (de confusão) (insolação total, precipitação total, temperatura máxima, temperatura média compensada, temperatura mínima e umidade relativa do ar) foram extraídas da base de dados meteorológicos do Instituto Nacional de Meteorologia (INMET). As covariáveis (poluentes do ar e climáticas) foram extraídas no mesmo período de tempo da variável dependente.

3.4 Análise Estatística Computacional

Como ponto de partida, foi feita análise exploratória dos dados. Isso incluiu pesquisa de valores anômalos e dados faltantes e possível imputação de dados. Na sequência, as covariáveis foram selecionadas para compor a análise final. A estratificação da idade em faixa etária em abaixo 65 anos, 65 a 75 anos e acima de 75 anos foi adotada segundo critérios estabelecidos em outro estudo (Peng; Dominici; Louis, 2006) e devido alterações fisiológicas relacionadas ao envelhecimento. A matriz de correlação entre as covariáveis e a variável resposta foi elaborada para identificar possíveis relações lineares entre os preditores e o desfecho de interesse, bem como para detectar multicolinearidade entre as covariáveis. Além disso, a série temporal foi decomposta em tendências de longo prazo, tendências sazonais e tendências de curto prazo de maior frequência (decomposição clássica de séries temporais) para VD e covariáveis. Foi utilizado o software estatístico R (versão 4.4.0) (APÊNDICE C) e os seguintes pacotes (*dyn*, *broom*, *stargazer*, *quantmod*, *mgcv*, *dlnm*, *stats*, *tsModel*, *dplyr*, *splines*, *gam*, *glarma*, *rempsyc*, *dyn*, *lattice*, *mda*).

A decomposição por escalas de tempo é conceptualizada conforme a referência teórica utilizada, onde uma série temporal Y_t é representada como:

$$Y_t = \text{tendência}_t + \text{sazonalidade}_t + \text{curto_prazo}_t. \quad (1)$$

Esta decomposição permite separar os componentes de variação de diferente extensão temporal (longo prazo, sazonal e curto prazo) e avaliá-los isoladamente para a variável de interesse e para as exposições (por exemplo, MP10 μm ou MP2,5 μm). A descrição e a implementação prática desta abordagem seguem os conceitos apresentados na bibliografia sobre decomposições por escalas (ex.: métodos de médias móveis, decomposição sazonal-tendência por suavização local por regressão polinomial e decomposição por transformada de Fourier).

Para avaliar a associação entre as covariáveis e a variável desfecho em cada escala temporal foi ajustado um modelo de regressão linear múltipla onde cada componente de tempo da exposição entra como covariável separada. A forma genérica do modelo de regressão ajustado foi:

$$Y_t = \beta_0 + \beta_1 X_t^{(\text{anual})} + \beta_2 X_t^{(\text{sazonal})} + \beta_3 X_t^{(\text{semanal})} + \beta_4 X_t^{(\text{residual})} + \varepsilon_t, \quad (2)$$

em que $X_t^{(\text{anual})}$ (*componente anual; representa a variação de longuíssimo prazo ou baixa frequência, geralmente associada a padrões estruturais do ano*),

$X_t^{(sazonal)}$ (componente sazonal; captura ciclos intraanuais de médio prazo, como padrões trimestrais), $X_t^{(semanal)}$ (componente semanal; descreve oscilações de curto prazo ao longo dos dias da semana), $X_t^{(residual)}$ (componente de alta frequência; representa a variação subsemanal ou ruído de alta frequência, contendo oscilações muito rápidas e flutuações não explicadas pelas escalas anteriores) e ε_t é o termo de erro. Essa parametrização permite interpretar cada coeficiente β_j como a associação específica entre a componente temporal correspondente da exposição e o desfecho, preservando as características próprias de cada escala temporal.

3.5 Técnica Estatística

3.5.1 Regressão em Séries Temporais em Epidemiologia Ambiental

Tradicionalmente, modelo de regressão de séries temporais é uma abordagem padrão para estudar associações de curto prazo entre exposições diárias (por exemplo, poluição do ar, temperatura) e desfechos de saúde (por exemplo, mortalidade ou internações hospitalares cardiovasculares) (Bhaskaran *et al.*, 2013). A contagem dos dados geralmente é feita em intervalos regulares (diariamente) e os modelos visam estimar as relações exposição-resposta, ajustando-as para tendências temporais e fatores de confusão (Bhaskaran *et al.*, 2013; Dominici, 2004).

A maioria dos estudos sobre Epidemiologia Ambiental e Séries Temporais são baseados em modelos lineares generalizados (MLG) (Nelder e Wederburn, 1972) com ligação logarítmica e erro de Poisson, regredindo contagens diárias de níveis de poluentes de ar e covariáveis. Termos de dia da semana e meteorológicos (por exemplo, temperatura e umidade) são incluídos como covariáveis, e funções suaves de tempo de calendário são usadas para controlar tendências de longo prazo e sazonalidade. Esta estrutura de regressão semiparamétrica de Poisson produz estimativas da taxa logarítmica relativa (risco) por unidade de mudança de exposição (Dominici, 2004).

Segundo Tomov *et al.* (2023), diferentemente dos modelos de previsão, que têm como objetivo projetar resultados futuros, os modelos de séries temporais aplicados ao contexto ambiental concentram-se em estimar os efeitos da exposição, buscando explicar os dados observados, e não prever eventos futuros.

De forma geral, em Epidemiologia Ambiental, o modelo básico é a regressão de Poisson (variância igual à média), mas as contagens reais de saúde

frequentemente exibem superdispersão (variância > média). Para acomodar isso, os estatísticos usam modelos quase-Poisson (introduzindo um parâmetro de dispersão) ou regressão binomial negativa. Ignorar a superdispersão pode subestimar os erros-padrão e superestimar a significância. A abordagem de quase verossimilhança ou as alternativas binomiais negativas oferecem flexibilidade: ambas têm um parâmetro extra (φ) de modo que $Var(Y) = \varphi * E(Y)$ (Samoli et al., 2005).

3.5.2 Modelos de Séries Temporais Semiparamétricos

Em modelos de series temporais de contagem, o padrão de modelo é Log-Linear de Poisson. Neste presente estudo, foi considerado a variável resposta Y_t tendo distribuição de Poisson com média μ_t e o logaritmo de μ_t é o preditor linear. Considera-se como preditores, duas classes de exposição: de interesse e fatores de confusão. A fórmula do modelo é dada por:

$$Y_t \sim \text{Poisson}(\mu_t) \quad (3)$$

$$\log \mu_t = \alpha + \beta x_{t-\ell} + \eta z_t + s(t, \lambda)_t + \varepsilon_t$$

Nesta pesquisa, pretende-se observar Y_t , uma série temporal de contagens diárias de internação hospitalar por doença cardiovascular ($t = 1, \dots, n$), x_t , uma série temporal de níveis diários de agentes poluidores do ar (MP2,5 μm e MP10 μm), z_t , uma série temporal de temperatura média diária, $s(t, \lambda)_t$, uma série de outros fatores de confusão não mensurados que variam suavemente com o tempo (variáveis proxy podem representar estes fatores) e ε_t é erro aleatório, ou seja, a parte não explicada do modelo (com média zero e variância constante). O controle da suavidade de $s(t, \lambda)_t$ é dada pelo parâmetro escalar de suavização (λ) (ou grau de liberdade). Para valores altos de (λ) tem-se menos suavidade de $s(t, \lambda)_t$ e valores menores representam mais suavidade da função.

3.5.3 Modelo Aditivo Generalizado

Hastie e Tibshirani (1990) pioneiramente apresentaram MAG como uma extensão flexível do MLG para fins de modelagem e análise exploratória. O MAG mantém a estrutura do MLG, mas substitue parte dos termos paramétricos por funções suaves das covariáveis, permitindo modelar relações não lineares de forma flexível. A estrutura de MAG é expressa como:

$$g(\mu_i) = \mathbf{X}_i^* \boldsymbol{\theta} + f_1(x_{1i}) + f_2(x_{2i}) + f_3(x_{3i}, x_{4i}) + \dots \quad (4)$$

onde $\mu_i \equiv \mathbb{E}(Y_i)$ e $Y_i \sim$ distribuição da família exponencial de interesse. O vetor \mathbf{X}_i^* representa os componentes estritamente paramétricos do modelo, $\boldsymbol{\theta}$ é o vetor de coeficientes associado, e f_i são funções suaves das covariáveis x_k . Essa abordagem permite capturar a complexidade da relação entre a variável resposta e as covariáveis sem exigir a especificação de formas funcionais detalhadas, evitando modelos excessivamente rígidos e difíceis de interpretar. No entanto, essa flexibilidade introduz dois desafios teóricos a saber: (i) a necessidade de representar as funções suaves de maneira computacionalmente viável e (ii) a escolha apropriada do grau de suavização dessas funções (Hastie; Tibshirani, 1990).

A aplicação de MAG na toxicologia atmosférica e epidemiologia foi proposta inicialmente por Schwartz (1994), e desde então tornou-se uma abordagem padrão nessa área. Nesses modelos, o log da média da resposta é especificado como a soma de um termo linear para o poluente atmosférico e de funções suaves para variáveis de confusão, como tempo e temperatura. Splines de suavização — como splines cúbicos naturais ou *LOESS* — são utilizados para ajustar padrões sazonais e tendências de longo prazo, ao mesmo tempo em que se estima o efeito do poluente. Assim, funções suaves não paramétricas operam como proxies para fatores de confusão temporais não mensurados.

Contudo, a implementação de MAG também apresenta limitações. Dominici *et al.* (2002) demonstraram que softwares estatísticos padrão para MAG, como o *gam* no S-Plus, podem apresentar falhas de convergência ou produzir estimativas viesadas dos coeficientes quando os efeitos dos poluentes são pequenos e muitos termos suaves são incluídos no modelo. Para contornar essas limitações, é comum empregar técnicas mais robustas, como splines penalizados — por exemplo, aquelas implementadas no pacote *mgcv* do software R — que promovem maior estabilidade e confiabilidade na estimação dos parâmetros.

3.5.4 Modelo MLG-ARMA

O foco da modelagem epidemiológica explicativa é a relação entre a ocorrência de um desfecho e as covariáveis usando procedimentos que não são capazes de capturar a dependência inerente às observações, como o MGL e o MAG. Seguindo uma linha do tempo, Mckenzie (1985) e Al-Osh e Alzaid (1987) introduziram o modelo autorregressivo de valores inteiros. O interesse por séries temporais de contagem reascendeu a partir de novas metodologias introduzidas por Shephard (1995) que

publicou o Modelo Linear Generalizado Autorregressivo de Médias Móveis (GLARMA), posteriormente generalizado por Davis et al. (2003) que adicionaram uma estrutura ARMA, introduzido por Box e Jenkins (1976), ao MLG (com modelo de Poisson). Os autores aplicaram MLG-ARMA a dados de contagens diárias de asma, investigando os efeitos dos níveis de poluição do ar e de variáveis meteorológicas sobre essas ocorrências, demonstrando a eficiência do modelo em capturar dependências temporais. Em outra linha de investigação, Benjamin et al. (2003) propuseram o modelo ARMA generalizado. O modelo de Poisson condicional autorregressivo para contagem de dados com dependência temporal e superdispersão foi publicado por Heinen (2003). Mais recentemente, Gamerman et al. (2013) apresentaram uma família de modelos de espaço de estados não gaussianos que possibilita a mensuração exata da verossimilhança marginal.

Os modelos MLG-ARMA assumem que a relação entre a variável resposta e as covariáveis é linear, o que limita sua capacidade de capturar padrões complexos presentes em séries temporais de contagem. Essa suposição linear pode levar a inferências imprecisas ou à subestimação de efeitos não lineares importantes.

3.5.5 Modelo MAG-ARMA

Para superar essas limitações, Câmara et al. (2021) propuseram maior flexibilidade em contextos epidemiológicos ao incorporar uma estrutura ARMA no MAG. O modelo MAG-ARMA resultante é capaz de lidar tanto com a autocorrelação das séries temporais quanto com a não linearidade das covariáveis, permitindo ajustar modelos semiparamétricos e não paramétricos aos dados, capturando simultaneamente relações lineares e não lineares, o que resulta em estimativas mais precisas e informativas.

No MAG-ARMA são utilizadas distribuições condicionais das observações y_t às informações passadas $\mathcal{F}_{t-1}^y = \sigma\{y_s, s \leq t-1\}$, onde \mathcal{F}_{t-1}^y representa a filtragem gerada pelas observações passadas até o tempo $t-1$ sobre o Processo y_t , $\sigma\{y_s, s \leq t-1\}$ é a σ -álgebra gerada pelos valores observados y_1, y_2, \dots, y_{t-1} . Supondo que o modelo segue uma distribuição de Poisson:

$$y_t \mid \mathcal{F}_{t-1}^y \sim \text{Poi}(\mu_t), \quad (5)$$

onde $\mu_t = E(y_t \mid \mathcal{F}_{t-1}^y)$. O modelo dos preditores $\eta_t = \ln(\mu_t)$ é dado por:

$$\eta_t = \beta_0 + \sum_{j=1}^k \beta_j x_{t,j} + \sum_{j=1}^l s_j(w_{t,j}) + Z_t, \quad (6)$$

onde $(x_{t,1}, \dots, x_{t,k})$ representa covariáveis relacionadas de forma linear com η_t , $(w_{t,1}, \dots, w_{t,l})$ são covariáveis vinculadas a η_t por intermédio de funções suaves s_1, \dots, s_l e Z_t representa a correlação temporal. Adotando-se os pressupostos de Davis et al. (2003),

$$Z_t = \sum_{i=1}^{\infty} \tau_i \varepsilon_{t-i} \quad (7)$$

onde para algum $\lambda \in (0,1]$,

$$\varepsilon_t = (y_t - \mu_t) \mu_t^{-\lambda} = (y_t - e^{\eta_t}) e^{-\lambda \eta_t} \quad (8)$$

os parâmetros τ_i correspondem aos coeficientes da expansão em série de potências, assim representada:

$$\sum_{i=1}^{\infty} \tau_i z^i = \left(1 - \sum_{i=1}^p \phi_i z^i\right)^{-1} \left(1 + \sum_{i=1}^q \theta_i z^i\right) - 1, |z| \leq 1, \quad (9)$$

onde os polinômios $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ e $\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$ não possuem zeros em comum e todos os seus zeros estão fora do círculo unitário. A partir de (7) e (9), segue-se que Z_t pode ser calculado de forma recursiva por meio da equação de diferenças.

$$Z_t = \phi_1(Z_{t-1} + \varepsilon_{t-1}) + \dots + \phi_p(Z_{t-p} + \varepsilon_{t-p}) + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad (10)$$

de acordo com (8), $E(\varepsilon_t | \mathcal{F}_{t-1}^y) = \mu_t^{-\lambda} (E(y_t | \mathcal{F}_{t-1}^y) - \mu_t) = 0$. Seja, $\mathcal{F}_{t-1}^\varepsilon = \sigma\{\varepsilon_s, s \leq t-1\}$, A relação (8) implica que $\mathcal{F}_{t-1}^\varepsilon \subset \mathcal{F}_{t-1}^y$. Portanto, tem-se que,

$$E(\varepsilon_t | \mathcal{F}_{t-1}^\varepsilon) = E[E(\varepsilon_t | \mathcal{F}_{t-1}^y) | \mathcal{F}_{t-1}^\varepsilon] = 0, \quad (11)$$

o que mostra que (ε_t) forma uma sequência de diferenças de martingala, ou seja, que cada termo tem expectativa condicional nula dado o passado. Desta forma, $\text{cov}(\varepsilon_s, \varepsilon_t) = 0$ para $s \neq t$, e a variância de ε_t é dada por:

$$\text{var}(\varepsilon_t) = E(\varepsilon_t^2) = E[E(\varepsilon_t^2 | \mathcal{F}_{t-1}^y)] = E(\mu_t^{-2\lambda} E[(y_t - \mu_t)^2 | \mathcal{F}_{t-1}^y]) = E(\mu_t^{1-2\lambda}). \quad (12)$$

De (6), (10) e (12) acarreta que,

$$\begin{aligned} E(\eta_t) &= \beta_0 + \sum_{j=1}^k \beta_j x_{t,j} + \sum_{j=1}^l s_j(w_{t,j}) \\ \text{var}(\eta_t) &= \sum_{i=1}^{\infty} \tau_i^2 E(\mu_{t-i}^{1-2\lambda}) \end{aligned} \quad (13)$$

$$\text{cov}(\eta_t, \eta_{t+h}) = \begin{cases} \sum_{i=1}^{\infty} \tau_i \tau_{i+h} E(\mu_{t-i}^{1-2\lambda}), & \text{if } h \geq 0, \\ \sum_{i=1}^{\infty} \tau_i \tau_{i-h} E(\mu_{t+h-i}^{1-2\lambda}), & \text{if } h < 0, \end{cases}$$

quando $\lambda = 0.5$, (ε_t) correspondem aos resíduos de Pearson e as covariâncias de (η_t) não dependem de t , mesmo que (η_t) não seja estritamente estacionária.

3.5.6 Obtenção dos parâmetros estimados / Ajuste do Modelo

Em relação aos tipos de funções suaves, são mencionadas algumas representações principais a saber, os *splines* naturais (cúbicos), *splines* penalizados (por expansões de base, *B-splines*) e por suavizadores *kernel/LOESS* (Peng; Dominici; Louis, 2006). Cada suavizador possui um parâmetro de suavização ou graus de liberdade (gl) que controlam a ondulação. Por exemplo, uma spline de suavização com parâmetro de suavização λ possui um "gl efetivo" equivalente que pode ser interpretado como controlando a flexibilidade do modelo. Independentemente da base funcional empregada, o elemento essencial consiste na capacidade das funções de suavização em capturar padrões de confundimento temporal sem incorrer em sobreajuste aos componentes de ruído dos dados (Dominici, 2004).

As curvas *B-splines* foram utilizadas neste estudo devido à sua facilidade para obter suavização flexível. Conceitualmente, *B-spline* $B_{i,d}$ depende dos nós $t_i \leq \dots \leq t_{i+d+1}$, onde d é a ordem do polinômio. Seja o vetor de nós dado por $(t_1, t_2, \dots, t_{m+d+1})$, para algum número inteiro positivo m . A partir dele, podem ser construídas *B-splines* $B_{1,d}, \dots, B_{m,d}$, de grau d . Uma função *spline* s_j pode então ser expressa como uma combinação linear dessas *B-splines*, isto é,

$$s_j = \sum_{i=1}^m \alpha_{i,j} B_{i,d}, \quad (14)$$

os valores reais $\alpha_{1,j}, \dots, \alpha_{m,j}$, são denominados coeficientes *B-spline* da função s_j . Foi adotado $d = 3$, correspondente ao uso de splines cúbicas naturais. Nesse caso, os segmentos polinomiais situados antes do primeiro nó e após o último nó são representados por funções lineares, o que implica que a segunda derivada da *spline* é nula nos extremos do intervalo.

A determinação do número ótimo de nós segue a abordagem proposta por Harrell (2004) e está relacionada ao tamanho da amostra n . Em amostras pequenas $n \leq 100$, o uso de três a quatro nós tende a proporcionar um ajuste adequado, equilibrando flexibilidade e precisão. Para amostras maiores, a escolha de cinco nós costuma ser apropriada como ponto inicial. Além disso, o Critério de Informação de Akaike (AIC), pode ser empregado como ferramenta de seleção do número de nós mais adequado ao modelo. Esta metodologia foi detalhada nos próximos tópicos.

Manipulando (6) e (14), e eliminando $d=3$ na notação, o modelo do preditor é:

$$\eta_t = \beta_0 + \sum_{j=1}^k \beta_j x_{t,j} + \sum_{j=1}^l \sum_{i=1}^m \alpha_{i,j} B_i(w_{t,j}) + Z_t, \quad (15)$$

o vetor Z_1 é definido conforme a expressão (8). Assim, para um inteiro fixo m e um conjunto de nós fixos $(t_1, t_2, \dots, t_{m+4})$, o vetor de parâmetros do modelo MAG-ARMA é dado por

$$\delta = (\beta_0, \dots, \beta_k, \alpha_{1,1}, \dots, \alpha_{m,l}, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q). \quad (16)$$

De acordo com a equação (5), a função de log-verossimilhança condicional é expressa como

$$L_n(\delta) = \sum_{t=1}^n (y_t \eta_t(\delta) - e^{\eta_t(\delta)}) \quad (17)$$

onde $\eta_t(\delta)$ é definida em (15) e $\eta_t(\delta)$ é obtida a partir da equação (10). A maximização de $L_n(\delta)$ pode ser realizada por meio do método de Newton, utilizando valores iniciais nulos para todos os parâmetros.

As medidas de qualidade de ajuste do modelo proposto podem ser avaliadas pelos critérios AIC, BIC (Bayesian Information Criterion), o Erro Quadrático Médio (EQM) e a *Deviance*, os quais permitem comparar diferentes especificações de modelos, equilibrando qualidade de ajuste e complexidade. BIC é definido por

$$\text{BIC} = -2 \ln \left(L_n(\hat{\delta}_n) \right) + r \ln(n) \quad (18)$$

onde $\hat{\delta}_n$ representa o vetor de estimativas que maximiza $L_n(\delta)$ e r corresponde ao número de parâmetros ajustados no modelo. O critério para comparar a qualidade do ajuste neste estudo foi o BIC. Adicionalmente, é fundamental realizar diagnósticos dos resíduos, como a inspeção gráfica dos resíduos padronizados e a análise das funções de autocorrelação (ACF) e autocorrelação parcial (PACF), a fim de verificar a presença de estrutura temporal não capturada pelo modelo. Para fins confirmatórios, o teste de Ljung-Box é uma excelente opção da presença ou não de ruído branco. A ausência de autocorrelação significativa nos resíduos indica que o modelo ajustado é capaz de captar adequadamente a dependência temporal dos dados.

O RR é amplamente empregado para quantificar o impacto da poluição atmosférica sobre a saúde humana. No contexto do modelo, o RR associado à covariável poluente $x_j = (x_{t,j})$ em (15) representa a variação relativa na contagem esperada de desfechos decorrente de uma mudança unitária ξ em x_j , mantendo as demais covariáveis constantes, e é definido por

$$\widehat{RR}_{x_j}(\xi) = \exp(\hat{\beta}_j \xi). \quad (19)$$

O RR e seu intervalo de confiança (IC) ao nível de confiança $1 - \alpha$ são estimados conforme

$$\begin{aligned} \widehat{RR}_{x_j}(\xi) &= \exp(\hat{\beta}_j \xi) \\ \widehat{CI}\{\widehat{RR}_{x_j}(\xi)\} &= \exp(\hat{\beta}_j \xi \pm z_{\alpha/2} \text{se}(\hat{\beta}_j) \xi) \end{aligned} \quad (20)$$

onde $\hat{\beta}_j$ é o estimador de máxima verossimilhança condicional $\hat{\beta}_{j,n}$ de β_j , $\text{se}(\hat{\beta}_j)$ é o erro padrão estimado de $\hat{\beta}_j$, e $z_{\alpha/2}$ denota o quantil $(1 - \alpha/2)$ da distribuição normal padrão.

3.5.7 Graus de Liberdade e Seleção de Suavidade

Ao avaliar modelos MAG em séries temporais é fundamental levar em conta o efeito do gl (Dominici, 2004). A alocação adequada de gl para os *splines* constitui aspecto crítico no controle de confusão e na obtenção de estimativas de efeito confiáveis. A definição inadequada desses parâmetros pode comprometer significativamente os resultados: graus de liberdade excessivamente baixos (suavização excessiva) podem resultar em ajuste inadequado dos fatores de confusão, enviesando o coeficiente de poluente; por outro lado, graus de liberdade

excessivamente altos (suavização insuficiente) podem ocasionar superajuste e inflacionar a variância.

Dominici (2004) destacou que as escolhas de graus de liberdade podem ter um grande impacto na magnitude e na incerteza estatística das taxas relativas estimadas, evidenciando a relevância dessa decisão metodológica. Apesar de não existir uma padronização, atualmente, a seleção automatizada por meio de critérios como AIC, escore de validação cruzada generalizada ou critérios personalizados tornou-se prática comum para ajustar termos de suavização nesses modelos, proporcionando maior objetividade e reprodutibilidade na definição dos parâmetros de suavização.

Neste estudo, a modelagem da tendência temporal foi conduzida por meio de funções suaves (tipo natural spline), com o objetivo de ajustar potenciais fatores de confusão sazonais e de longo prazo nas estimativas do RR associado à exposição aos poluentes atmosféricos (MP10 μm e MP2,5 μm). Nesse contexto, considerou-se também que a temperatura média, incluída como variável confundidora em todos os modelos, apresentava uma relação não linear com as internações por DCV.

Dado que diferentes quantidades de suavização podem alterar substancialmente as estimativas, a definição dos gl desse termo suavizado foi realizada por meio de uma abordagem híbrida. Inicialmente, obteve-se uma estimativa objetiva do gl ótimo por meio de validação cruzada generalizada (VCG), conforme o método proposto por Gu (2002), utilizando-se a função `bruto`, disponível no pacote `mda` do R.

Na segunda etapa da abordagem híbrida, o número inicial de gl obtido pelo método VCG foi refinado por meio da minimização do AIC. Para isso, construiu-se um vetor contendo valores de gl próximos ao estimado inicialmente — tanto menores quanto maiores — e ajustaram-se modelos sucessivos variando o número de nós da função natural *spline* da temperatura média. Para cada ajuste, calculou-se o AIC a partir da log-verossimilhança do modelo. A expressão utilizada segue a formulação:

$$AIC = -2\ln(L) + 2p, \quad (21)$$

em que L denota a função de verossimilhança do modelo ajustado e p o número total de parâmetros. O gl final foi selecionado como aquele que apresentou o menor AIC, representando o melhor compromisso entre ajuste adequado e parcimônia para a suavização adequada da temperatura média.

3.5.8 Ajuste do Modelo para Cada Poluente Atmosférico

Para cada poluente atmosférico (MP10 μm ou MP2,5 μm), foi ajustado um modelo MAG-ARMA com distribuição de Poisson, seguindo um algoritmo sistemático de modelagem. O procedimento inicia-se pela preparação dos dados, etapa em que se constrói a base analítica contendo o número diário de internações por DCV (variável resposta), a concentração diária do poluente de interesse e a temperatura média diária. Após a remoção de observações com valores ausentes, define-se o vetor de tempo $t = 1, \dots, n$, representando o total de dias ou meses observados.

Em seguida, constroem-se os termos temporais destinados a captar padrões de longo prazo e sazonalidade, incluindo um termo linear em t e harmônicos anuais e ou semestrais e ou semanais, a depender da análise gráfica de decomposição temporal da variável resposta. Posteriormente, procede-se à seleção do grau de suavização para a temperatura média. Testam-se diferentes gl para o *spline* natural pela abordagem híbrida descrita anteriormente, e ajusta-se um MAG inicial para cada configuração, selecionando-se o valor de gl que minimiza o AIC.

Com o grau de suavização definido, ajusta-se o modelo MAG base, utilizando distribuição de Poisson, suavizador para temperatura, tendência e harmônicos pré-estabelecidos empiricamente. Caso apresentem ruído branco, o MAG é considerado adequado, prosseguindo-se com ajuste de um modelo MAG-ARMA, que incorpora estrutura de dependência temporal por meio de termos autorregressivos e/ou de médias móveis, conforme sugerido pela análise gráfica das funções ACF e PACF dos resíduos. O BIC é calculado para fins de comparação entre os modelos. O diagnóstico é conduzido por meio da análise gráfica da ACF e PACF dos resíduos de Pearson e teste de Ljung-Box para todos os modelos.

Após a seleção do modelo adequado, estimam-se o RR e os intervalos de confiança de 95% para a variação interquartil do poluente escolhido. O algoritmo detalhado do modelo MAG-ARMA está apresentado no APÊNDICE B. Esse procedimento é então repetido integralmente para o segundo poluente, permitindo comparar o efeito estimado e o desempenho dos modelos.

A modelagem começou em escala diária, testando tendências, harmônicos e suavização da temperatura média, além de análises de sensibilidade usando distribuições alternativas e controle de dependência temporal. Devido à persistência da autocorrelação residual, a análise passou para frequência mensal, revisando tendência e harmônicos com base na inspeção das séries.

4 RESULTADOS E DISCUSSÃO

4.1 Análise Exploratória dos Dados (AED)

A AED foi conduzida com foco na contagem diária de internações, complementada por visualizações de frequência diária e mensal, em conformidade com as especificações do modelo estatístico adotado. Durante o período de 2010 a 2019 em São Paulo, foram registradas 362.474 hospitalizações por doenças isquêmicas do coração e cerebrovasculares, representando uma média diária de 99,25 admissões hospitalares. A distribuição por gênero demonstrou predominância masculina, com aproximadamente 55,83% das internações ($n = 202.377$). A análise das estatísticas descritivas das internações (Tabela 1) revelou variabilidade considerável nos dados, com valores oscilando entre 1 e 457 internações diárias. O primeiro quartil (Q1) situou-se em 80 internações diárias, enquanto o terceiro quartil (Q3) alcançou 116 internações, indicando concentração dos dados próxima à média. A mediana de 100 internações diárias, próxima à média de 99,25, sugere distribuição relativamente simétrica dos dados centrais.

A estratificação por gênero e idade evidenciou padrões distintos de internação. Entre os homens, a faixa etária com menor número de internações foi a superior a 75 anos (média de 9,59 internações diárias), contrastando com o grupo masculino menor que 65 anos, que apresentou a maior média (31,72 internações diárias). Para as mulheres, observou-se distribuição mais equilibrada entre as faixas etárias, com destaque para o grupo acima de 75 anos (12,66 internações diárias), superando numericamente os homens da mesma faixa etária.

Tabela 1 – Estatísticas Descritivas de Internações Cardiovasculares (2010–2019, SP).

Variável	Mínimo	Q1	Mediana	Média	Q3	Máximo
Número internação	1,00	80,00	100,00	99,25	116,00	457,00
homem	0,00	44,00	55,00	55,42	65,00	334,00
mulher	0,00	36,00	44,00	43,84	51,00	185,00
homem <65a	0,00	24,00	31,00	31,72	38,00	177,00
homem 65-75a	0,00	10,00	13,00	14,11	17,00	98,00
homem >75a	0,00	7,00	9,00	9,59	12,00	73,00
mulher <65a	0,00	16,00	20,00	20,49	25,00	93,00
mulher 65-75a	0,00	8,00	10,00	10,69	13,00	57,00
mulher >75a	0,00	10,00	12,00	12,66	15,00	109,00

Fonte: Elaborado pelo autor

O monitoramento dos poluentes atmosféricos (Tabela 2) revelou concentrações médias de MP10 de 26,02 $\mu\text{g}/\text{m}^3$, MP2,5 de 28,33 $\mu\text{g}/\text{m}^3$, NO_2 de 38,39 $\mu\text{g}/\text{m}^3$, SO_2 de 2,62 $\mu\text{g}/\text{m}^3$, O_3 de 44,99 $\mu\text{g}/\text{m}^3$ e CO de 0,22 ppm. A variabilidade do MP10 foi substancial, com valores máximos atingindo 110,12 $\mu\text{g}/\text{m}^3$, muito superiores à média. O MP2,5 apresentou comportamento semelhante, com concentração máxima de 142,51 $\mu\text{g}/\text{m}^3$ e mediana de 23,86 $\mu\text{g}/\text{m}^3$, evidenciando episódios de poluição intensa no período analisado. O NO_2 apresentou distribuição mais ampla, com concentrações máximas de 126,92 $\mu\text{g}/\text{m}^3$, enquanto o SO_2 manteve-se em níveis relativamente baixos, com máximo de 18,79 $\mu\text{g}/\text{m}^3$. Ozônio teve valores máximos de 112,72 $\mu\text{g}/\text{m}^3$ e mínimos de 9,93 $\mu\text{g}/\text{m}^3$, demonstrando variação temporal característica deste poluente secundário. O monóxido de carbono apresentou máximo de 0,93 ppm, refletindo a influência das emissões veiculares na área urbana estudada.

Tabela 2 – Estatísticas Descritivas dos Poluentes Atmosféricos (2010–2019, SP).

Variável	Mínimo	Q1	Mediana	Média	Q3	Máximo
MP10 μm	0,00	15,75	24,25	26,02	34,08	110,12
MP2,5 μm	5,69	15,60	23,86	28,33	36,14	142,51
NO_2	0,00	25,50	37,38	38,39	50,19	126,92
SO_2	0,00	1,08	2,08	2,62	3,67	18,79
O_3	9,93	35,93	43,48	44,99	52,53	112,72
CO	0,05	0,14	0,19	0,22	0,28	0,93

Fonte: Elaborado pelo autor

As condições meteorológicas (Tabela 3) caracterizaram-se por temperatura média de 20,72°C, com amplitude térmica considerável entre mínimas (média de 16,71°C) e máximas (média de 26,41°C). A umidade relativa do ar apresentou média de 73,10%, indicando condições de umidade moderada a alta. A precipitação total média foi de 4,45 mm, com alta variabilidade representada pelo desvio padrão elevado.

Tabela 3 – Estatísticas Descritivas de Variáveis Meteorológicas (2010–2019, SP).

Variável	Mínimo	Q1	Mediana	Média	Q3	Máximo
Insolação total (h)	0,00	2,20	5,50	5,42	8,40	12,60
Precipitação total (mm)	0,00	0,00	0,00	4,45	2,20	123,60
Temperatura máxima (°C)	8,70	23,40	26,70	26,41	29,80	37,80
Temperatura média (°C)	7,60	18,40	20,90	20,72	23,23	29,20
Temperatura mínima (°C)	3,50	14,40	17,00	16,71	19,20	25,40
Umidade relativa do ar (%)	28,00	66,30	74,50	73,10	81,30	98,30

Fonte: Elaborado pelo autor

A análise detalhada das séries temporais (Tabelas 4, 5 e 6) revelou características interessantes dos dados. Para as interações por DCV, o coeficiente de variação de 0,26 indicou variabilidade moderada. A assimetria positiva (1,63) e curtose elevada (21,49) sugeriram distribuição com cauda direita mais pesada que a normal e maior concentração de valores extremos.

Os poluentes atmosféricos apresentaram variabilidade distinta, com SO₂ exibindo o maior coeficiente de variação (0,82), seguido por MP10µm (0,59), MP2,5µm (0,59), NO₂ (0,51) e CO (0,52). O ozônio apresentou coeficiente de variação de 0,30, indicando menor variabilidade relativa em comparação aos demais poluentes. Todas as variáveis meteorológicas demonstraram distribuições não-normais, com precipitação apresentando a maior variabilidade (coeficiente de variação de 2,58).

Tabela 4 - Estatísticas das Séries Temporais (Interação DCV).

Variável	DP	Assimetria	Curtose	Variância	Coef. Variação
Interação DCV	26,07	1,63	21,49	679,61	0,26
Homem	17,30	2,99	42,50	299,09	0,31
Mulher	11,55	1,06	14,64	133,39	0,26
Homem <65a	10,45	1,73	21,66	109,14	0,33
Homem 65-75a	6,72	4,34	44,85	45,19	0,48
Homem >75a	4,11	2,43	29,20	16,92	0,43
Mulher <65a	6,72	0,80	8,23	45,20	0,33
Mulher 65-75a	4,04	0,78	8,00	16,36	0,38
Mulher >75a	4,81	3,39	55,13	23,10	0,38

Fonte: Elaborado pelo autor. Legenda: DP = Desvio padrão. Variância = Covariância.

Tabela 5 - Estatísticas das Séries Temporais de Poluentes do ar.

Variável	DP	Assimetria	Curtose	Variância	Coef. Variação
MP10	15,32	0,79	4,40	234,70	0,59
MP2,5	16,65	1,28	1,78	277,26	0,59
NO2	19,75	0,38	3,53	389,99	0,51
SO2	2,15	1,45	6,64	4,63	0,82
O ₃	13,34	0,69	1,18	178,05	0,30
CO	0,12	1,27	2,02	0,01	0,52

Fonte: Elaborado pelo autor. Legenda: DP = Desvio padrão. Variância = Covariância.

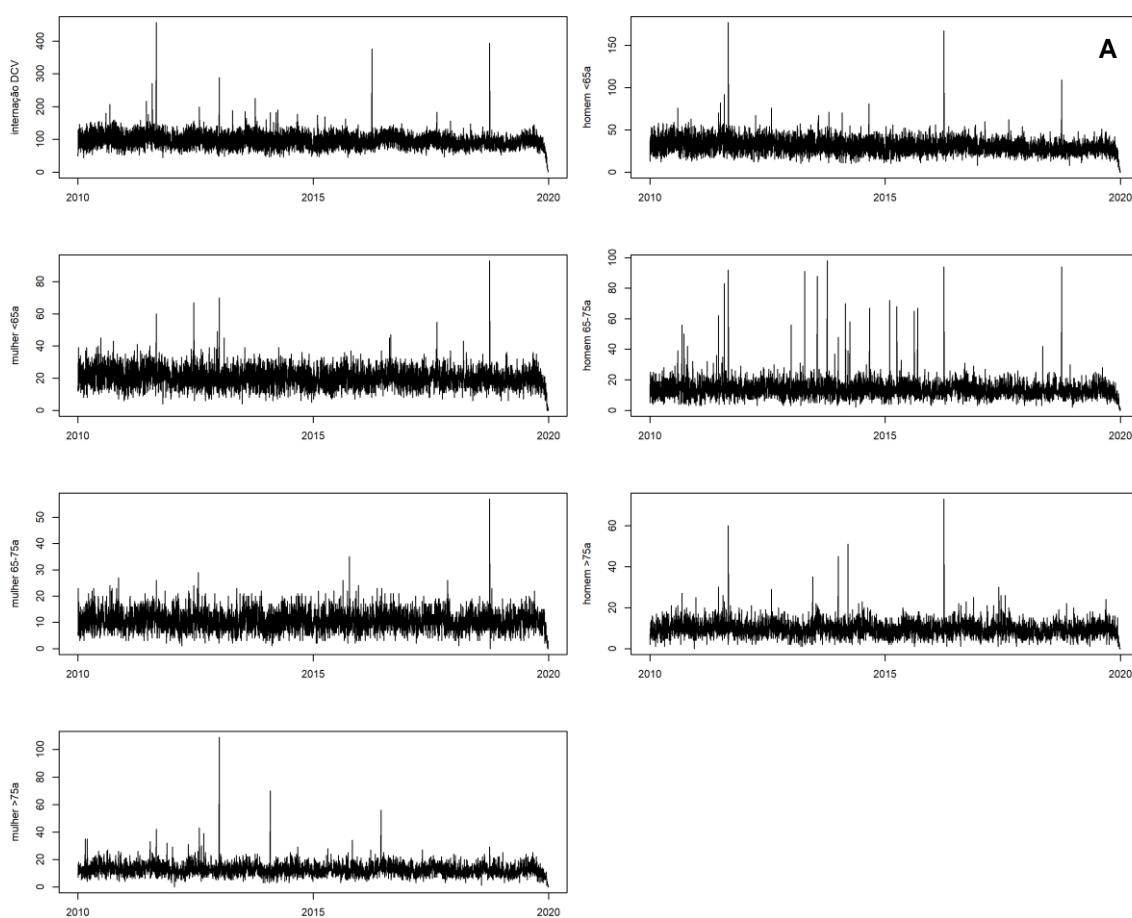
Tabela 6 - Estatísticas das Séries Temporais Meteorológicas.

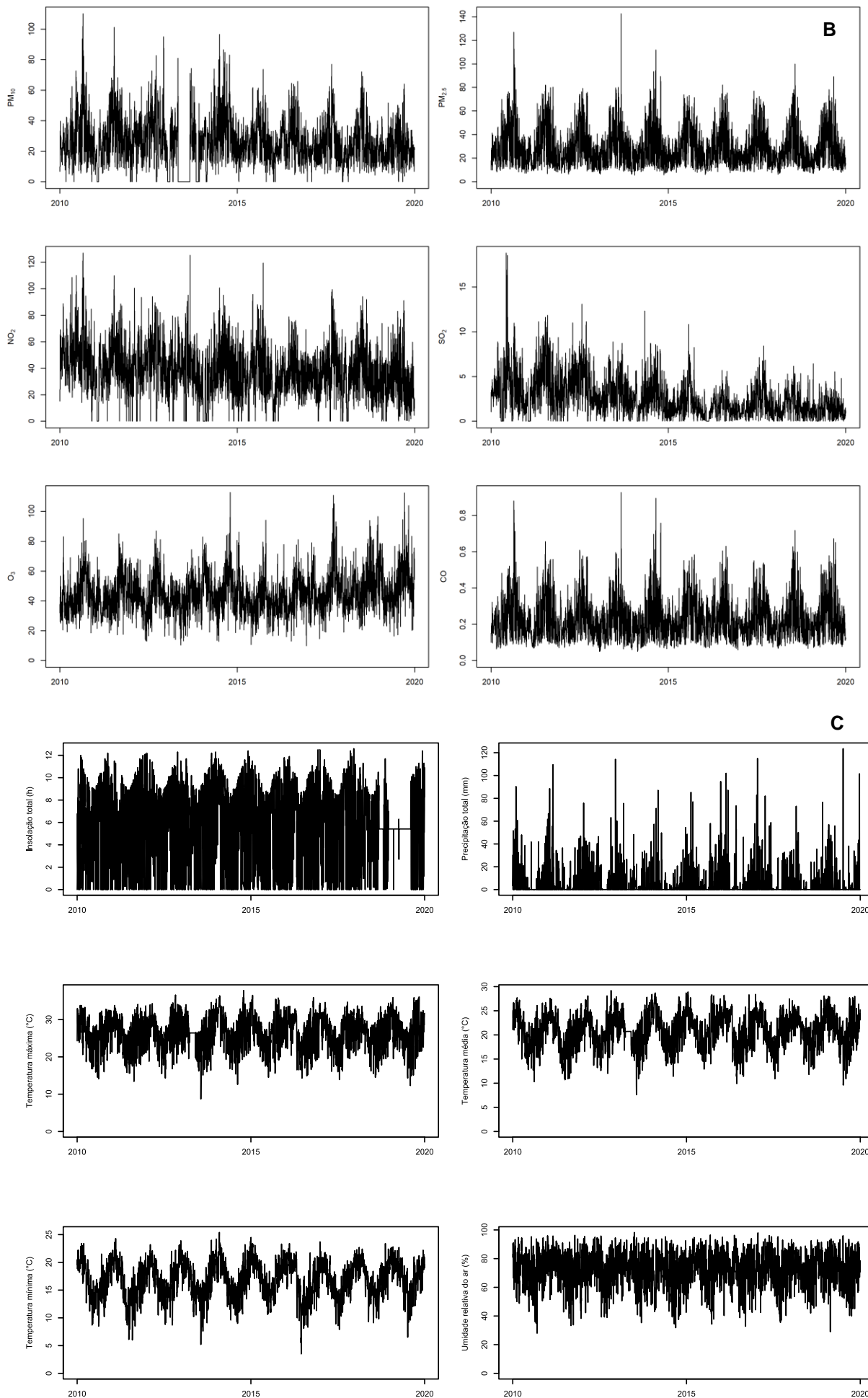
Variável	DP	Assimetria	Curtose	Variância	Coef. Variação
Insolação Total	3,49	-0,20	1,84	12,18	0,64
Precipitação Total	11,50	4,28	27,46	132,17	2,58
Temperatura Máxima	4,38	-0,32	2,66	19,24	0,17
Temperatura Média	3,43	-0,28	2,76	11,78	0,16
Temperatura Mínima	3,21	-0,39	2,78	10,33	0,19
Umidade Relativa Ar	11,75	-0,71	3,40	138,11	0,16

Fonte: Elaborado pelo autor. Legenda: DP = Desvio padrão. Variância = Covariância.

A seguir, tem-se a análise gráfica de todas as séries temporais (contagem diária) desta pesquisa (Figura 1). Na análise da variabilidade temporal do poluente MP10 μ m, tem-se evidência de uma redução na variabilidade das concentrações ao longo do período de 10 anos. Observa-se que, após aproximadamente 2015, houve uma diminuição na frequência de valores extremamente elevados de MP10 μ m, contrastando com o padrão observado no período anterior. Esta mudança temporal sugere uma possível alteração nas condições ambientais ou nas fontes de emissão durante o período estudado. Cabe destacar que a representação gráfica dos dados de MP10 μ m incorpora uma tendência temporal, permitindo uma visualização mais clara dos padrões de longo prazo e da evolução da variabilidade ao longo da série temporal.

Figura 1 – Gráficos das Séries Temporais em contagem diária (A - Internação DCV; B - poluentes do ar; C – climáticas).

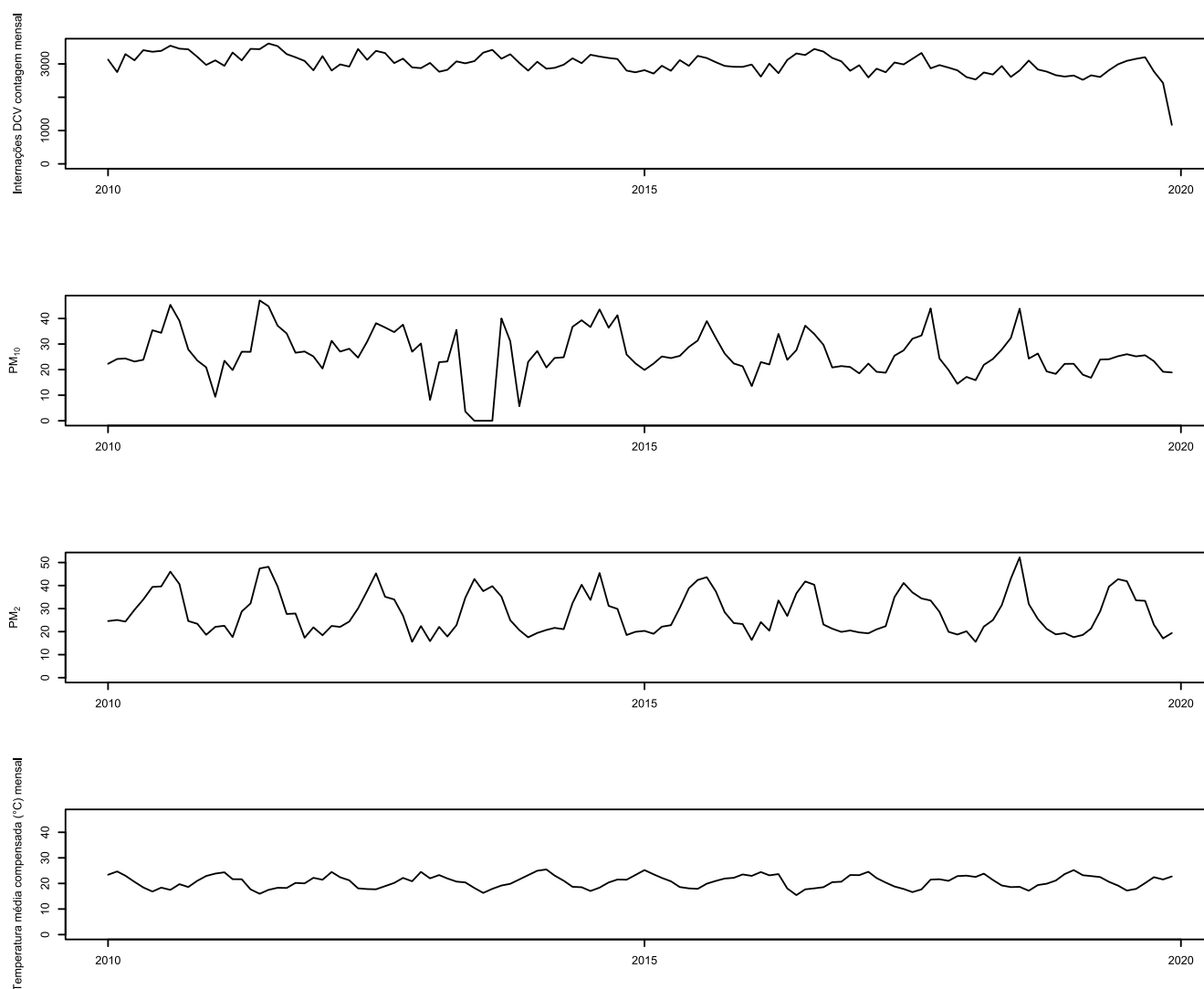




Fonte: Elaborado pelo autor

A visualização da frequência mensal dos casos de hospitalização por DCV ao longo dos 120 meses analisados, bem como das médias mensais das covariáveis consideradas, poluentes atmosféricos PM_{10} , $PM_{2,5}$ e temperatura média ($^{\circ}C$), permitiu identificar padrões sazonais e tendências de longo prazo relevantes para a especificação do modelo. Esses gráficos fornecem uma compreensão inicial da variabilidade temporal das séries e evidenciam comportamentos distintos entre as variáveis ambientais, que posteriormente são incorporados ao ajuste do modelo MAG-ARMA. A distribuição mensal apresentada na Figura 2 ilustra claramente essas dinâmicas, contribuindo para a interpretação preliminar das associações investigadas.

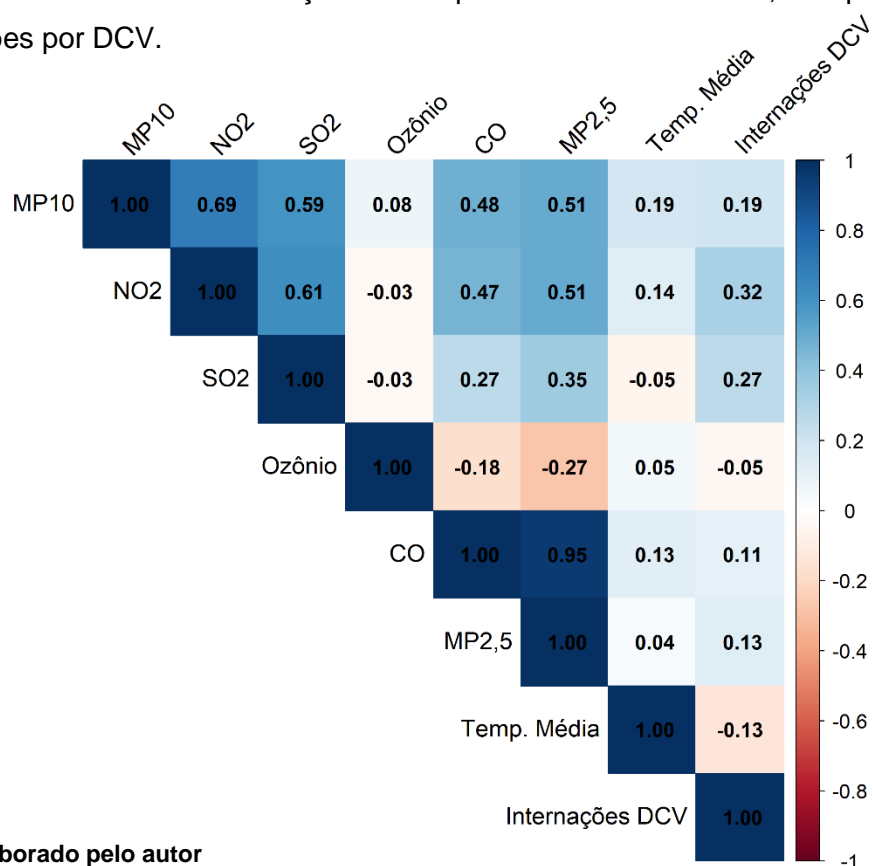
Figura 2 – Gráficos das Séries Temporais em dados mensais (Internação DCV; poluentes do ar; Temperatura média).



Fonte: Elaborado pelo autor

Na matriz de correlação (Figura 3) entre poluentes atmosféricos, temperatura média e o número de internações por DCV observa-se uma correlação moderada a forte entre diversos poluentes atmosféricos, especialmente entre CO e MP2,5µm, assim como entre MP10µm, NO₂ e SO₂. As correlações entre poluentes e temperatura foram baixas, indicando pouca relação direta entre essas variáveis. Os poluentes que apresentaram maior correlação com variável resposta foram NO₂ ($r \approx 0,32$), SO₂ ($r \approx 0,27$), MP10µm ($r \approx 0,19$) e MP2,5µm ($r \approx 0,13$). A matriz evidencia presença de multicolinearidade relevante entre alguns poluentes.

Figura 3 – Matriz de correlação entre poluentes atmosféricos, temperatura média e internações por DCV.

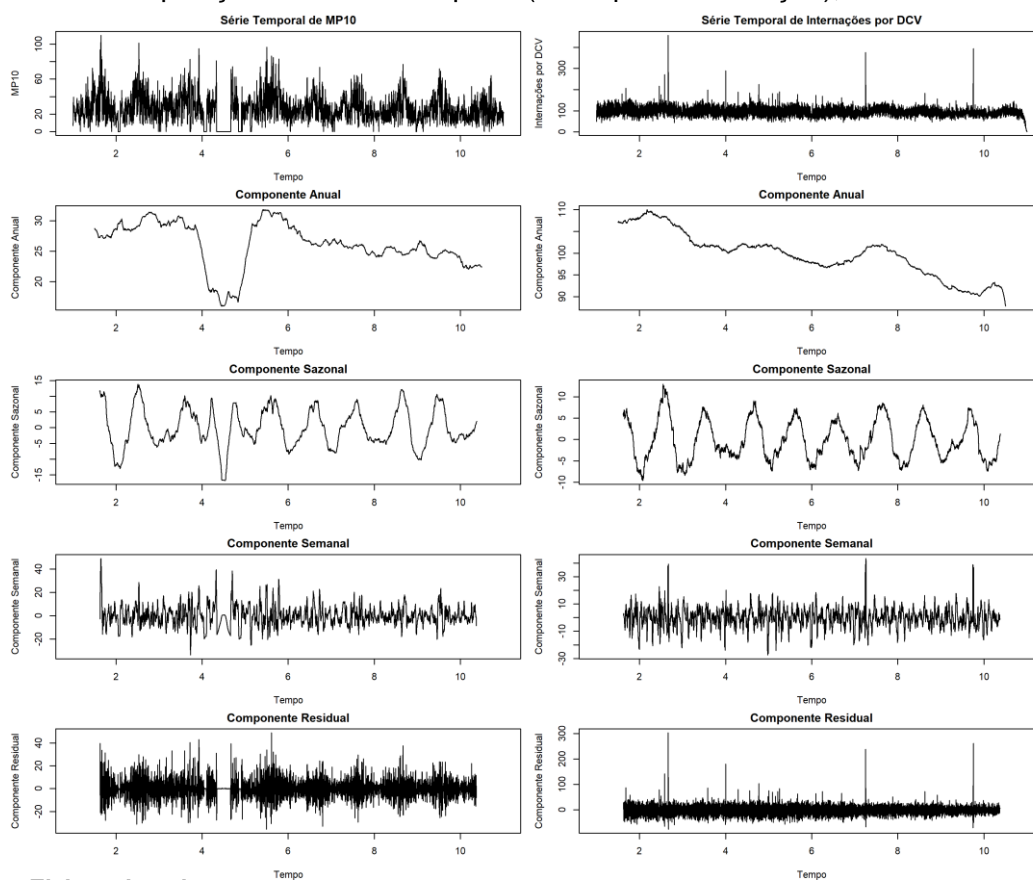


Fonte: Elaborado pelo autor

4.2 Decomposições por Escalas Temporais

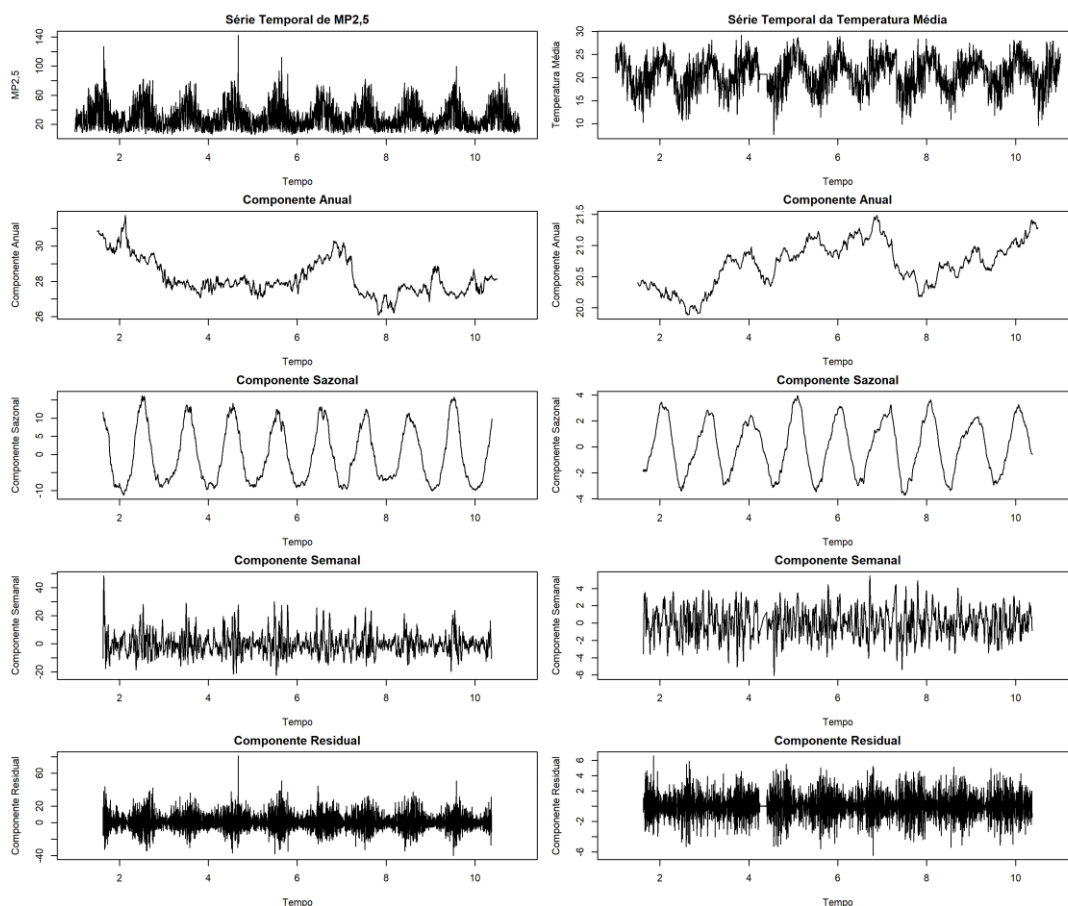
Foi realizado a decomposição completa das principais séries temporais (MP10µm, MP2,5µm, temperatura média e Internação DCV) em três escalas de tempo distintas pelo método de médias móveis para obter uma imagem mais detalhada desta associação. Previamente, foi verificado a presença de dados faltantes e presença de outliers para aplicar o tratamento adequado (Figura 4 e 5).

Figura 4 - Decomposição da escala temporal (MP10 μ m e internação), 2010-2019.



Fonte: Elaborado pelo autor

Figura 5 - Decomposição da escala temporal (MP2.5 μ m e temperatura média), 2010-2019.



Fonte: Elaborado pelo autor

A decomposição da série temporal de internações por DCV em três escalas (longo prazo, sazonalidade anual e curto prazo) evidenciou a presença de uma sazonalidade anual bem marcada, com oscilação regular e repetitiva ao longo dos anos. A tendência de longo prazo mostrou variações suaves, caracterizando alterações estruturais ao longo do período analisado. Por outro lado, não foi identificada sazonalidade semestral, já que nenhum dos componentes da decomposição apresentou um ciclo compatível com um período de seis meses. O componente semanal apresentou oscilações de alta frequência típicas de dados diários, sem periodicidade mais longa.

Foram ajustados dois modelos de regressão linear múltipla contendo a decomposição temporal das exposições: um para MP10 μ m (tabela 7) e outro para MP2,5 μ m (tabela 8). A seguir apresentam-se os resultados sumarizados dos ajustamentos (coeficientes, erro padrão, estatística t e valores-p).

Tabela 7 — Resumo do modelo: covariável MP10 μ m (decomposição por escalas).

Termo	Estimativa	Erro padrão	t valor	Pr(> t)
Intercepto	90,75	2,92	31,09	< 2e-16 (***)
MP10 anual	0,32	0,11	2,89	0.00384 (**)
MP10 sazonal	0,51	0,07	7,70	1.82e-14 (***)
MP10 semanal	0,02	0,05	0,49	0.62438
MP10 resíduos	0,42	0,04	9,61	< 2e-16 (***)

Fonte: Elaborado pelo autor. R² = 0,051; R² ajustado = 0,050. Estatística F = 42,76 (gl = 4; 3188); p < 0,001.

Tabela 8 — Resumo do modelo: covariável MP2,5 μ m (decomposição por escalas).

Termo	Estimativa	Erro padrão	t valor	Pr(> t)
Intercepto	31,20	11,59	2,69	0,00715 (**)
MP2,5 anual	2,41	0,41	5,87	4,77e-09 (***)
MP2,5 sazonal	0,55	0,05	10,47	< 2e-16 (***)
MP2,5 semanal	0,01	0,05	0,18	0,85373
MP2,5 resíduos	0,07	0,04	1,90	0,05802 (.)

Fonte: Elaborado pelo autor. R² = 0,043; R² ajustado = 0,042. Estatística F = 35,68 (gl = 4; 3188); p < 0,001.

Os coeficientes das diferentes escalas devem ser interpretados com cautela devido à presença de autocorrelação temporal potencial e à não-normalidade dos resíduos (teste de Shapiro–Wilk rejeitou normalidade). Ainda assim, os sinais e magnitudes dos coeficientes fornecem informação útil sobre em quais escalas temporais as exposições estão mais fortemente associadas com internações por DCV.

4.3 Ajuste do Modelo para MP10 μm e MP2,5 μm

Da análise descritiva das séries temporal, especificou-se o preditor linear dos modelos MAG-ARMA com um único poluente atmosférico com a incorporação do termo de tendência de longo prazo (trend_t), dos componentes harmônicos anuais de variação sazonal ($\sin\left(\frac{2\pi t}{365}\right), \cos\left(\frac{2\pi t}{365}\right)$) e do efeito não linear da temperatura média diária por meio da função spline natural (Temp_t, k).

Inicialmente para MP10 μm ($MP10_t$), o termo linear do modelo MAG-ARMA Poisson ajustado foi definido como:

$$\eta_t = \beta_1 MP10_t + \beta_2 \sin\left(\frac{2\pi t}{365}\right) + \beta_3 \cos\left(\frac{2\pi t}{365}\right) + \beta_4 \text{trend}_t + ns(\text{Temp}_t, k) + \sum_{i=1}^{\infty} \tau_i \varepsilon_{t-i}. \quad (22)$$

O modelo MAG segue a mesma estrutura apresentada em (22), excluindo-se o componente autorregressivo ($\sum_{i=1}^{\infty} \tau_i \varepsilon_{t-i}$). Após a aplicação do método híbrido para seleção do número de gl, foram testados ajustes variando de 12 a 16 nós e, mediante comparação dos valores de AIC obtidos, constatou-se que o melhor ajuste foi alcançado com 16 nós. Apresentam-se inicialmente as estimativas do modelo MAG ajustado (Tabela 9), nas quais todos os coeficientes mostraram-se estatisticamente significativos ($p < 0,05$). A qualidade do ajuste, avaliada pelo critério BIC, resultou em 37561,05.

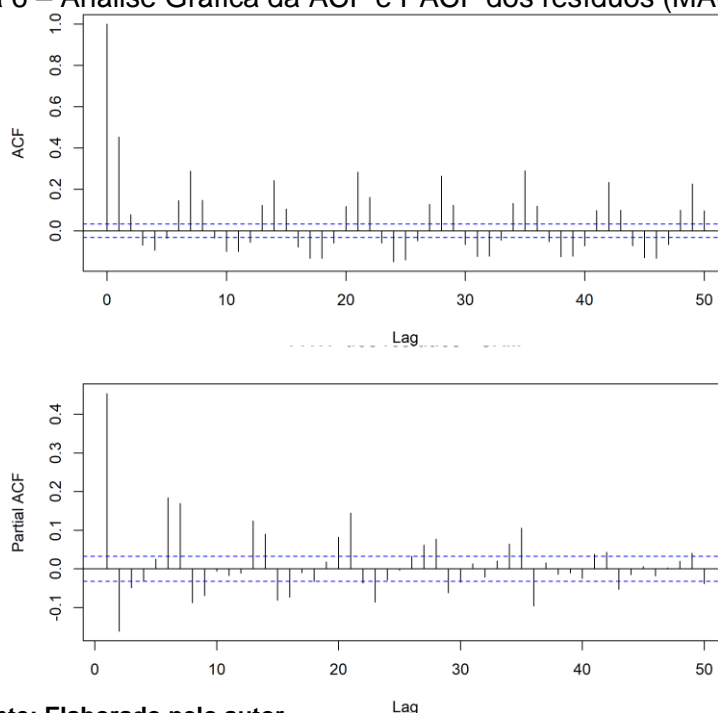
Tabela 9 - Resultados do Modelo MAG para MP10 μm .

Variável	Estimativa	Erro Padrão	p-valor
MP10	0,0023	0,0001	<0,0001
tendência	-0,000046	0,0000016	<0,0001
sen12	-0,0522	0,0054	<0,0001
cos12	-0,0443	0,0036	<0,0001

Fonte: Elaborado pelo autor.

Da análise dos gráficos de ACF e PACF para os resíduos do modelo MAG – MP10 μm (Figura 6), não foi obtido ruído branco, uma vez que há picos fora das bandas de confiança nas defasagens iniciais. O teste de Ljung-Box corroborou essa conclusão, apresentando p-valor $< 0,05$. Esse comportamento indica a necessidade de um ajuste considerando a estrutura temporal da série. Com base nos padrões observados, um modelo AR(1) é o mais simples e provavelmente adequado, dado que a ACF apresenta decaimento após lag 1 e a PACF mostra corte abrupto após lag 1.

Figura 6 – Análise Gráfica da ACF e PACF dos resíduos (MAG diário).



Fonte: Elaborado pelo autor.

A suposição de independência serial não foi atendida também para modelo MAG-AR(1). Diversas configurações de dependência temporal para contagem diária foram avaliadas com o pacote glarma, explorando diferentes ordens autoregressivas e de média móvel. Foram testadas, entre outras, as seguintes estruturas: AR(2), AR(7), MA(1), MA(3), ARMA(1,1), ARMA(1,2), ARMA(1,3), ARMA(2,1), ARMA(2,3). Adicionalmente, avaliou-se a sensibilidade do modelo a diferentes distribuições da resposta, incluindo Poisson e Binomial Negativa.

Apesar da ampla varredura de estruturas ARMA, observou-se de forma consistente que os modelos ajustados aos dados diários apresentavam autocorrelação residual persistente, com autocorrelações significativas em múltiplos lags nos gráficos de ACF e PACF, bem como forte rejeição do teste de Ljung–Box para diversos *lags* (ex.: lag 7, 14, 21, 28, 35, 42 e 50; todos com $p\text{-valor} < 0.001$). Estes resultados indicaram que, mesmo com estruturas AR, MA ou ARMA relativamente complexas, os resíduos permaneciam longe de um comportamento de ruído branco, sugerindo que o modelo diário não capturava adequadamente a dependência temporal intrínseca à série.

Diante desse cenário, optou-se pela conversão da série diária para frequência mensal, agregando a variável resposta por meio da soma mensal das internações e calculando as médias mensais das covariáveis ambientais. Além disso, todos os

componentes de tendência e sazonalidade foram novamente reconstruídos considerando a nova escala temporal, com especial atenção aos harmônicos anual ($\sin\left(\frac{2\pi t}{12}\right)$ e $\cos\left(\frac{2\pi t}{12}\right)$) e semi-anual ($\sin\left(\frac{4\pi t}{12}\right)$ e $\cos\left(\frac{4\pi t}{12}\right)$), sendo o modelo MAG Poisson definido como:

$$\eta_t = \beta_1 MP10_t + \sum_j \gamma_j sTemp_{j,t} + \beta_2 \sin\left(\frac{2\pi t}{12}\right) + \beta_3 \cos\left(\frac{2\pi t}{12}\right) + \beta_4 \sin\left(\frac{4\pi t}{12}\right) + \beta_5 \cos\left(\frac{4\pi t}{12}\right) + \beta_6 \text{trend}_t \quad (23)$$

Foram adotados 4 gl para o ajuste da spline natural. As estimativas do modelo MAG ajustado (Tabela 10) indicaram que todos os coeficientes foram estatisticamente significativos ($p < 0,05$). O critério de informação bayesiano obtido foi de 195929,5.

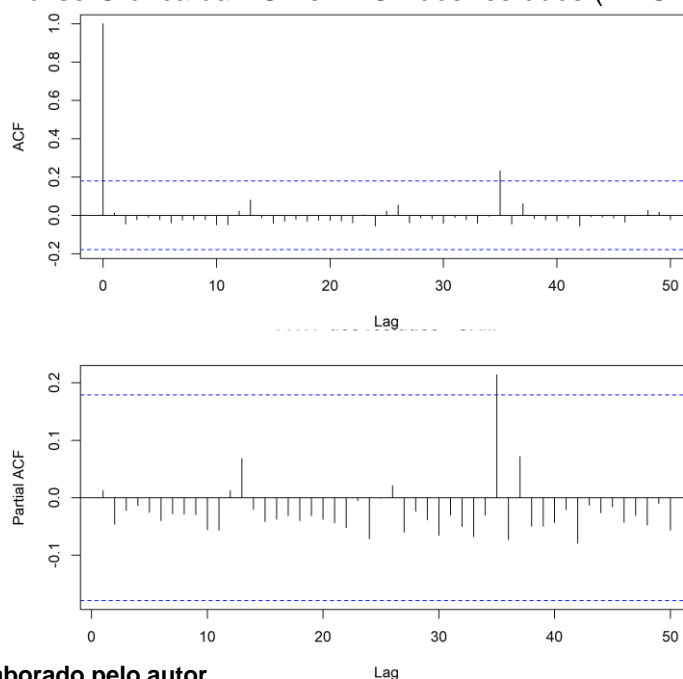
Tabela 10 - Resultados do Modelo MAG para MP10 μm .

Variável	Estimativa	Erro Padrão	p-valor
MP10	0,02069	0,00028	<0,0001
tendência	-0,00386	0,00054	<0,0001
sen12	-0,08056	0,00376	<0,0001
cos12	-0,46020	0,00707	<0,0001
sen6	0,15040	0,03025	<0,0001
cos6	0,25350	0,03164	<0,0001

Fonte: Elaborado pelo autor.

A análise dos gráficos de ACF e PACF para os resíduos do modelo MAG – MP10 μm (Figura 7) não evidenciou autocorrelação temporal significativa, com os valores das autocorrelações situando-se dentro das bandas de confiança, caracterizando comportamento de ruído branco. O teste de Ljung-Box corroborou essa constatação ($X^2 = 3,2753$; gl = 20; p-valor = 1,00).

Figura 7 – Análise Gráfica da ACF e PACF dos resíduos (MAG -MP10 mensal).



Fonte: Elaborado pelo autor.

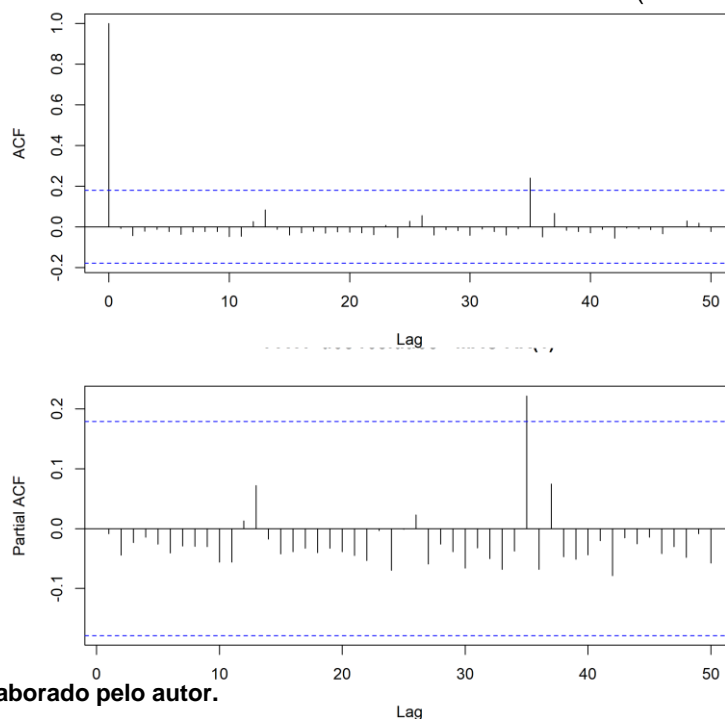
O modelo MAG-AR(1) Poisson para MP10 μm foi ajustado com acréscimo do termo $\sum_{i=1}^{\infty} \tau_i \varepsilon_{t-i}$ em (23). Teve valor menor para BIC (193762,7), indicando melhor ajuste em comparação ao modelo anterior. A análise dos resíduos por meio do teste de Ljung-Box ($X^2 = 3,2811$; gl = 20; p-valor = 1,00) e por análise gráfica (Figura 8), confirmando que os resíduos se comportam como ruído branco. A seguir os dados do modelo ajustado (Tabela 11).

Tabela 11 - Resultados do Modelo MAG-AR(1) para MP10 μm .

Variável	Estimativa	Erro Padrão	p-valor
MP10	0,02112	0,00029	<0,0001
tendência	-0,02453	0,00561	<0,0001
sen12	-0,06362	0,00384	<0,0001
cos12	-0,43340	0,00718	<0,0001
sen6	0,07821	0,03083	0,8000
cos6	0,24460	0,00323	<0,0001
ϕ_1	0,00043	0,00001	<0,0001

Fonte: Elaborado pelo autor.

Figura 8 – Análise Gráfica da ACF e PACF dos resíduos (MAG-AR(1) MP10 mensal).



Fonte: Elaborado pelo autor.

A Tabela 12 apresenta as estimativas de RR e seus respectivos IC de 95% para a variação interquartil das concentrações de MP10 μm . O modelo MAG estimou um RR de 1,224 (IC 95%: [1,218–1,231]), indicando que o aumento na concentração de MP10 μm correspondente à amplitude interquartil está associado a um acréscimo de aproximadamente 22,4% no risco de internações por doenças cardiovasculares. O modelo MAG-AR(1), apresentou estimativa ligeiramente superior, com RR de 1,229 (IC 95%: [1,223–1,236]), correspondendo a um aumento de 22,9% no risco. Ambas

as estimativas foram estatisticamente significativas, com IC que não incluem o valor nulo (RR = 1), confirmando a associação positiva entre exposição ao MP10 μ m e interações por DCV. A pequena variação entre os modelos reforça a robustez dos resultados, embora o modelo MAG-AR(1) apresente estimativas mais precisas ao considerar a estrutura de autocorrelação temporal dos dados.

Tabela 12 - Estimativas de risco relativo e intervalos de confiança de 95% para a variação interquartil das concentrações de MP10 μ m entre os modelos MAG e MAG-AR(1).

MP10 μ m	RR	IC 95%
MAG	1,22419	[1,21763 ; 1,23077]
MAG-AR(1)	1,22934	[1,22252 ; 1,23619]

Fonte: Elaborado pelo autor.

A expressão conceitual do modelo MAG para MP2,5 μ m foi:

$$\eta_t = \beta_1 PM_{2.5,t} + \sum_j \gamma_j s_{Temp_{j,t}} + \beta_2 \sin\left(\frac{2\pi t}{12}\right) + \beta_3 \cos\left(\frac{2\pi t}{12}\right) + \beta_4 \sin\left(\frac{4\pi t}{12}\right) + \beta_5 \cos\left(\frac{4\pi t}{12}\right) + \beta_6 trend_t, \quad (24)$$

onde a estrutura MAG-AR(1) tem o acréscimo do termo $\sum_{i=1}^{\infty} \tau_i \varepsilon_{t-i}$ em (24).

Na Tabela 13, os ajustes dos modelos MAG e MAG-AR(1) foram apresentados.

Tabela 13 - Resultados dos Modelos MAG e MAG-AR(1) para MP2,5 μ m.

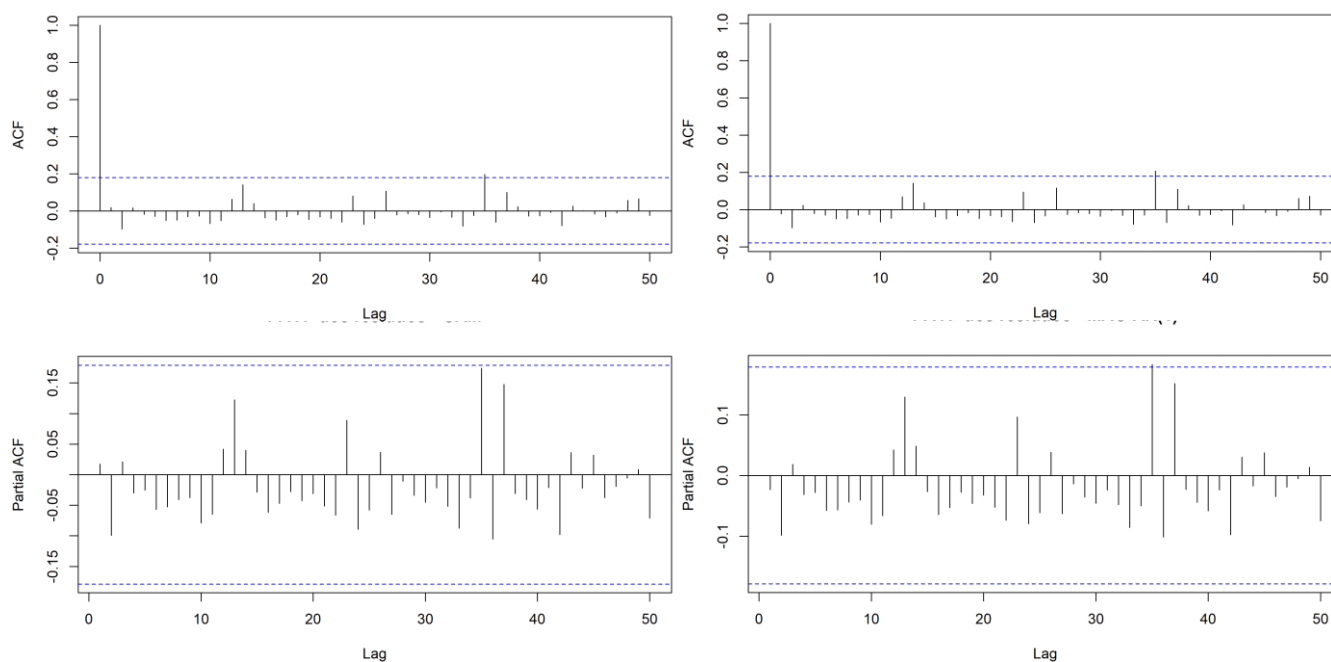
MAG			
Variável	Estimativa	Erro Padrão	p-valor
MP2,5	0,10340	0,05877	<0,0001
tendência	0,01421	0,00545	<0,0001
sen12	0,03215	0,00466	<0,0001
cos12	0,12190	0,00718	<0,0001
sen6	-0,01089	0,00317	<0,0001
cos6	-0,01901	0,00403	<0,0001
MAG-AR(1)			
Variável	Estimativa	Erro Padrão	p-valor
MP2,5	0,10420	0,05885	<0,0001
tendência	0,01646	0,00587	<0,0001
sen12	0,03400	0,00480	<0,0001
cos12	0,12530	0,00812	<0,0001
sen6	-0,01230	0,00324	<0,0001
cos6	-0,01910	0,00404	<0,0001
ϕ_1	0,00091	0,00002	<0,0001

Fonte: Elaborado pelo autor.

Os modelos para MP2,5 μ m foram avaliados mediante o BIC e o teste de Box-Ljung para verificação da autocorrelação residual. Para MAG, o BIC foi de 170.097,8, com teste de Box-Ljung indicando ausência de autocorrelação significativa nos resíduos ($X^2 = 3,2753$, gl = 20, p = 1,000). A incorporação do termo AR(1) resultou em

melhoria substancial do ajuste, reduzindo o BIC para 167.935,3 e teve adequação na estrutura residual, com o teste de Box-Ljung não rejeitando a hipótese nula de ausência de autocorrelação ($X^2 = 8,1125$, $gl = 20$, $p = 0,9911$). A análise gráfica das funções ACF e PACF dos resíduos é mostrada na Figura 9.

Figura 9 – Análise Gráfica da ACF e PACF dos resíduos (MAG e MAG-AR(1)).



Fonte: Elaborado pelo autor.

Observou-se na comparação entre RR e IC 95% que ambos os modelos apresentaram estimativas similares para o efeito da variação interquartil das concentrações de $MP_{2,5\mu m}$ sobre o desfecho analisado (Tabela 13). Embora as estimativas pontuais indiquem aumento do risco associado à exposição ao material particulado fino, os amplos intervalos de confiança, que incluem a unidade, sugerem ausência de significância estatística em ambas as especificações. A incorporação do termo autorregressivo não alterou substancialmente a magnitude do efeito estimado, mantendo-se a consistência nas inferências entre os modelos (Tabela 14).

Tabela 14 - Estimativas de RR e IC 95% para a variação interquartil das concentrações de $MP_{2,5\mu m}$ entre os modelos MAG e MAG-AR(1).

$MP_{2,5\mu m}$	RR	IC 95%
MAG	4,40902	[0,84416 ; 23,02829]
MAG-AR(1)	4,46095	[0,85218 ; 23,35200]

Fonte: Elaborado pelo autor.

A conversão da série diária para a escala mensal mostrou-se metodologicamente adequada por três razões centrais. Primeiro, a agregação reduz substancialmente a autocorrelação de curto prazo, típica de séries diárias afetadas por padrões semanais, feriados, efeitos de calendário e flutuações intradiárias que não guardam relação direta com o processo epidemiológico de médio e longo prazo. Segundo, a escala mensal proporciona uma relação sinal–ruído mais favorável: a suavização das variações erráticas torna mais evidentes os ciclos sazonais e a tendência subjacente, aumentando a interpretabilidade dos resultados e sua coerência epidemiológica. Terceiro, a mudança de escala contribui para maior estabilidade estatística em modelos com termos de dependência temporal; na frequência mensal, tanto modelos MAG quanto MAG-AR(1) passam a apresentar resíduos compatíveis com ruído branco, ausência de autocorrelação significativa e adequada convergência do algoritmo de estimação.

A literatura recente reforça esses achados. Câmara *et al.* (2021), ao aplicarem um modelo MAG-ARMA a uma série temporal de morbidade respiratória em Belo Horizonte utilizando dados agregados mensalmente, também relataram ajuste satisfatório e resíduos sem autocorrelação, destacando que a incorporação do componente autorregressivo se torna mais estável e estatisticamente consistente quando a série é analisada nessa escala temporal.

Após a conversão da base para série mensal ($n = 120$ observações), observou-se comportamento análogo: tanto o modelo MAG quanto o MAG-AR(1) apresentaram bom ajuste, resíduos aproximadamente aleatórios, independência temporal e testes de Ljung–Box não rejeitando a hipótese nula para múltiplos defasagens. Assim, a mudança de frequência foi decisiva para a obtenção de modelos estatisticamente bem comportados e, ao mesmo tempo, epidemiologicamente interpretáveis.

Por fim, cabe ressaltar que a modelagem de séries diárias de contagem apresenta desafios técnicos substanciais: forte autocorrelação de curto prazo, sobredispersão, efeitos semanais estruturais, flutuações irregulares e elevada sensibilidade a ruídos operacionais tornam o ajuste de modelos MAG, GLM ou semiparamétricos consideravelmente mais instável, frequentemente inviabilizando resíduos brancos e convergência adequada sem intervenções metodológicas adicionais.

6 CONCLUSÕES

Neste trabalho, foi proposta uma nova extensão de modelos, denominada MAG-ARMA, que possibilita o ajuste de modelos não paramétricos e semiparamétricos, incorporando covariáveis com relações lineares e não lineares com a variável resposta em dados de contagem com dependência temporal. A estrutura temporal é modelada por meio de componentes autorregressivos e de média móvel.

Os resultados reforçam a importância da escolha criteriosa dos métodos estatísticos na análise da associação entre poluentes atmosféricos e desfechos em saúde, especialmente em contextos urbanos com elevada carga de doenças cardiovasculares.

Além do avanço técnico, este trabalho fornece subsídios relevantes para a formulação de políticas públicas voltadas à mitigação dos efeitos da poluição do ar sobre a saúde cardiovascular. Ao orientar decisões sobre padrões aceitáveis de qualidade do ar, estratégias de monitoramento ambiental e alocação de recursos hospitalares em períodos críticos, estudos como este desempenham papel fundamental no planejamento preventivo de sistemas de saúde em grandes centros urbanos.

Futuras pesquisas devem explorar abordagens que integrem dados espaciais e temporais de alta resolução, análises estratificadas por grupos vulneráveis (como idosos ou pessoas com comorbidades), e métodos que considerem simultaneamente múltiplos poluentes e efeitos não lineares.

REFERÊNCIAS

ALIZADEH, G.; GHOLIPOUR, K.; AZAMI-AGHDASH, S.; DEHNAVIEH, R.; JAFARABADI, M. A.; AZMIN, M.; KHODAYARI-ZARNAQ, R. Social, Economic, Technological, and Environmental Factors Affecting Cardiovascular Diseases: A Systematic Review and Thematic Analysis. **International Journal of Preventive Medicine**, v. 13, 27 abr. 2022. DOI: 10.4103/ijpvm.IJPVM_105_20. PMID: 35706860; PMCID: PMC9188896.

ALVARES, Clayton Alcarde et al. Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, v. 22, n. 6, p. 711-728, 2013. Tradução. Disponível em: <https://doi.org/10.1127/0941-2948/2013/0507>. Acesso em: 30 set. 2024.

AL-OSH, M.; ALZAID, A. First order integer valued autoregressive (INAR (1)) process. **Journal of Time Series Analysis**, v. 8, p. 261–275, 1987.

BENJAMIN, M.; RIGBY, R.; STASINOPOULOS, D. Generalized autoregressive moving average models. **Journal of the American Statistical Association**, v. 98, p. 214–223, 2003.

BHASKARAN, K.; GASPARRINI, A.; HAJAT, S.; SMEETH, L.; ARMSTRONG, B. Time series regression studies in environmental epidemiology. **International Journal of Epidemiology**, v. 42, n. 4, p. 1187–1195, 2013. DOI: 10.1093/ije/dyt092.

BHATNAGAR, A. Environmental cardiology: studying mechanistic links between pollution and heart disease. **Circulation Research**, v. 99, p. 692-705, 2006. DOI: 10.1161/01.RES.0000243586.99701.cf.

BHATNAGAR, A. Environmental determinants of cardiovascular disease. **Circulation Research**, 2017 Jul 7; 121(2): 162-180. doi: 10.1161/CIRCRESAHA.117.306458. PMID: 28684622; PMCID: PMC5777598.

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

BROOK, R. D.; RAJAGOPALAN, S.; POPE, C. A.; BROOK, J. R.; BHATNAGAR, A.; DIEZ ROUX, A. V. et al. Particulate matter air pollution and cardiovascular disease: an update to the scientific statement from the American Heart Association. **Circulation**, v. 121, p. 2331-2378, 2010. DOI: 10.1161/CIR.0b013e3181d8e3e1.

CALEYACHETTY, R.; ECHOUFFO-TCHEUGUI, J. B.; SHIMBO, D.; ZHU, W.; MUENNIG, P. Cumulative social risk and risk of death from cardiovascular diseases and all causes. **International Journal of Cardiology**, v. 177, n. 3, p. 1106-1107, 2014. Disponível em: <https://doi.org/10.1016/j.ijcard.2014.08.088>.

CÂMARA, Ana Julia A.; FRANCO, Glaucia C.; REISEN, Valderio A.; BONDON, Pascal. Generalized additive model for count time series: an application to quantify the impact of air pollutants on human health. **Pesquisa Operacional**, v. 41, 2021. DOI: <https://doi.org/10.1590/0101-7438.2021.041.00241120>.

CANTERBURY, A.; ECHOUFFO-TCHEUGUI, J. B.; SHPILSKY, D. et al. Association between cumulative social risk, particulate matter environmental pollutant exposure, and cardiovascular disease risk. **BMC Cardiovascular Disorders**, v. 20, n. 76, 2020. DOI: <https://doi.org/10.1186/s12872-020-01329-z>.

DAVIS, R. A.; DUNSMUIR, W. T. M.; STREETT, S. B. Observation driven models for Poisson counts. **Biometrika**, Oxford, v. 90, p. 777–790, 2003.

DOMINICI, F. Time-series analysis of air pollution and mortality: a statistical review. **Research Report Health Effects Institute**, n. 123, p. 3–27, dez. 2004. Discussão p. 29–33. PMID: 15757000.

ERQOU, S.; ECHOUFFO-TCHEUGUI, J. B.; KIP, K. E.; AIYER, A.; REIS, S. E. Association of cumulative social risk with mortality and adverse cardiovascular disease outcomes. **BMC Cardiovascular Disorders**, v. 17, n. 1, p. 110, 2017. Disponível em: <https://doi.org/10.1186/s12872-017-0539-9>.

GAMERMAN, D.; SANTOS, T.; FRANCO, G. A non-Gaussian family of state-space models with exact marginal likelihood. **Journal of Time Series Analysis**, v. 34, p. 625–645, 2013.

GASPARRINI, A. Modelling lagged associations in environmental time series data: a simulation study. **Epidemiology**, v. 27, n. 6, p. 835–842, nov. 2016. DOI: 10.1097/EDE.0000000000000533.

GASPARRINI, A.; ARMSTRONG, B.; KENWARD, M. G. Distributed lag non-linear models. **Statistical Medicine**, v. 29, n. 21, p. 2224–2234, 2010.

GU, C. **Smoothing Spline ANOVA Models**. New York: Springer, 2002.

GUIA agenda 2030: integrando ODS, educação e sociedade. Disponível em: <https://bdce.unb.br/bibliodex/guia-agenda-2030-integrando-ods-educacao-e-sociedade/>. Acesso em: 20 set. 2024.

HARRELL, F. E. **Biostatistical modeling**. Nashville, TN, 2004.

HASTIE, T.; TIBSHIRANI, R. **Generalized additive models**. London: Chapman and Hall, 1990.

HEINEN, A. Modelling Time Series Count Data: Autoregressive Conditional Poisson Model observations. **Munich Personal RePEc Archive**, 2003.

IBGE. Panorama São Paulo. 2022. <https://cidades.ibge.gov.br/brasil/sp/panorama>. Acesso em: 30 set. 2024. Atualizado em 22/12/2023.

INNESS, A. et al. The CAMS reanalysis of atmospheric composition. **Atmospheric Chemistry and Physics**, v. 19, n. 6, p. 3515-3556, 2019. DOI: 10.5194/acp-19-3515-2019.

JACOBS, D. R. Jr.; WOO, J. G.; SINAICO, A. R.; DANIELS, S. R.; IKONEN, J.; JUONALA, M. et al. Childhood cardiovascular risk factors and adult cardiovascular events. **New England Journal of Medicine**, v. 386, p. 1877-1888, 2022. DOI: 10.1056/NEJMoa2109191.

JINGESI, M.; LAN, S.; HU, J.; DAI, M.; HUANG, S.; CHEN, S. et al. Association between thermal stress and cardiovascular mortality in the subtropics. **International Journal of Biometeorology**, v. 67, n. 12, p. 2093-2106, dez. 2023. DOI: 10.1007/s00484-023-02565-6. Epub 2023 out. 25. PMID: 37878088.

KANNEL, W. B.; DAWBER, T. R.; KAGAN, A.; REVOTSKIE, N.; STOKES, J. 3rd. Factors of risk in the development of coronary heart disease six year follow-up experience. The Framingham Study. **Annals of Internal Medicine**, v. 55, p. 33-50, jul. 1961. DOI: 10.7326/0003-4819-55-1-33. PMID: 13751193.

KELSALL, J. E.; SAMET, J. M.; ZEGER, S. L.; XU, J. Air pollution and mortality in Philadelphia, 1974–1988. **American Journal of Epidemiology**, v. 146, p. 750–762, 1997.

MANSUR, A. P.; FAVARATO, D. Mortalidade por doenças cardiovasculares no Brasil e na região metropolitana de São Paulo: atualização 2011. **Arquivos Brasileiros de Cardiologia**, v. 99, p. 755–761, 2012.

MARTENS, W. Climate change, thermal stress and mortality changes. **Social Science and Medicine**, v. 46, p. 331-344, 1998.

MARTINS, L. C. et al. The effects of air pollution on cardiovascular diseases: lag structures. **Revista de Saúde Pública**, v. 40, n. 4, p. 677–683, ago. 2006.

MCKENZIE, E. Some simple models for discrete variate time series. **Water Resources Bulletin**, v. 21, p. 645–650, 1985.

METHLEY, A. M.; CAMPBELL, S.; CHEW-GRAHAM, C.; MCNALLY, R.; CHERAGHI-SOHI, S. PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. **BMC Health Services Research**, v. 14, n. 1, p. 1-10, 2014. DOI: 10.1186/s12913-014-0579-0.

MINISTÉRIO DOS TRANSPORTES. Frota de Veículos – 2024. Disponível em: <https://www.gov.br/transportes/pt-br/assuntos/transito/conteudo-Senatran/frota-de-veiculos-2024>. Acesso em: 30 set. 2024.

MÜNZEL, T.; HAHAD, O.; SØRENSEN, M.; LELIEVELD, J.; DUERR, G. D.; NIEUWENHUIJSEN, M.; DAIBER, A. Environmental risk factors and cardiovascular diseases: a comprehensive expert review. **Cardiovascular Research**, v. 118, n. 14, p. 2880-2902, nov. 2022. DOI: 10.1093/cvr/cvab316. PMID: 34609502; PMCID: PMC9648835.

NANDI, A.; GLYMOUR, M. M.; KAWACHI, I.; VANDERWEELE, T. J. Using marginal structural models to estimate the direct effect of adverse childhood social conditions

on onset of heart disease, diabetes, and stroke. **Epidemiology**, v. 23, n. 2, p. 223–232, 2012. DOI: <https://doi.org/10.1097/EDE.0b013e31824570bd>.

NELDER, J. A.; WEDDERBURN, R. W. M. Generalized linear models. **Journal of the Royal Statistical Society**. Series A (General), London, v. 135, p. 370–384, 1972.

OLIVER, S. L.; RIBEIRO, H. Water Supply, Climate Change and Health Risk Factors: Example Case of São Paulo—Brazil. In: LEAL FILHO, W.; AZEITEIRO, U.; ALVES, F. (Eds). **Climate Change and Health**. Cham: Springer, 2016. DOI: https://doi.org/10.1007/978-3-319-24660-4_25.

PATZ, J.; CAMPBELL-LENDRUM, D.; HOLLOWAY, T.; FOLEY, J. Impact of regional climate change on human health. **Nature**, v. 438, 2005.

PENG, Roger D.; DOMINICI, Francesca; LOUIS, Thomas A. Model choice in time series studies of air pollution and mortality. **Journal of the Royal Statistical Society Series A: Statistics in Society**, v. 169, n. 2, p. 179–203, mar. 2006. Disponível em: <https://doi.org/10.1111/j.1467-985X.2006.00410.x>.

POWELL-WILEY, T. M.; BAUMER, Y.; BAAH, F. O.; BAEZ, A. S.; FARMER, N.; MAHLOBO, C. T. et al. Social Determinants of Cardiovascular Disease. **Circulation Research**, v. 130, n. 5, p. 782-799, 4 mar. 2022. DOI: 10.1161/CIRCRESAHA.121.319811. PMID: 35239404; PMCID: PMC8893132.

REQUIA, W. J. et al. Health impacts of wildfire-related air pollution in Brazil: a nationwide study of more than 2 million hospital admissions between 2008 and 2018. **Nature Communications**, v. 12, n. 6555, p. 1-11, 2021. DOI: 10.1038/s41467-021-26822-7.

SØRENSEN, M.; PERSHAGEN, G. Transportation noise linked to cardiovascular disease independent from air pollution. **European Heart Journal**, v. 40, p. 604-606, 2019.

SABZMAKAN, L.; MOHAMMADI, E.; MOROWATISHARIFABAD, M. A.; AFAGHI, A.; NASERI, M. H.; MIRZAEI, M. Environmental determinants of cardiovascular diseases risk factors: a qualitative directed content analysis. **Iran Red Crescent Medical Journal**, v. 16, n. 5, p. e11573, maio 2014. DOI: 10.5812/ircmj.11573. Epub 2014 maio 5. PMID: 25031848; PMCID: PMC4082508.

SALDIVA, Paulo H. N. et al. Air pollution and mortality in elderly people: a time-series study in São Paulo, Brazil. **Archives of Environmental Health: An International Journal**, v. 50, n. 2, p. 159-163, 1995. DOI: 10.1080/00039896.1995.9940893.

SAMOLI, E.; ANALITIS, A.; TOULOUMI, G.; SCHWARTZ, J.; ANDERSON, H. R.; SUNYER, J.; BISANTI, L.; ZMIROU, D.; VONK, J. M.; PEKKANEN, J.; GOODMAN, P.; PALDY, A.; SCHINDLER, C.; KATSOUYANNI, K. Estimating the exposure-response relationships between particulate matter and mortality within the APHEA multicity project. **Environmental Health Perspectives**, v. 113, n. 1, p. 88–95, jan. 2005. DOI: 10.1289/ehp.7387.

SCHULZ, A. J.; KANNAN, S.; DVONCH, J. T.; ISRAEL, B. A.; ALLEN III, A.; JAMES, S. A.; HOUSE, J. S.; LEPKOWSKI, J. Social and physical environments and disparities in risk for cardiovascular disease: the Healthy Environments Partnership conceptual model. **Environmental Health Perspectives**, v. 113, n. 12, 2005.

SCHWARTZ, J. Nonparametric smoothing in the analysis of air pollution and respiratory illness. **Canadian Journal of Statistics**, v. 22, p. 471–488, 1994.

SHEPHARD, N. **Generalized linear autoregressions**. Technical Report. Oxford: Nuffield College, 1995.

TATEISHI, K.; HMOUD, H.; DE GREGORIO, L.; DE GREGORIO, I.; ASSELIN, C. Y.; DE GREGORIO, J. Impact of Cardiac and Cerebrovascular Complications During Hospitalization on Long-Term Prognosis in Patients With COVID-19. **American Journal of Cardiology**, v. 209, p. 114-119, 15 dez. 2023. DOI: 10.1016/j.amjcard.2023.09.083. Epub 2023 out. 12. PMID: 37839464.

TOMOV, L.; CHERVENKOV, L.; MITEVA, D. G.; BATSELOVA, H.; VELIKOVA, T. Applications of time series analysis in epidemiology: literature review and our experience during COVID-19 pandemic. **World Journal of Clinical Cases**, v. 11, n. 29, p. 6974–6983, 16 out. 2023. DOI: 10.12998/wjcc.v11.i29.6974.

VIRANI, S. S.; ALONSO, A.; APARICIO, H. J.; BENJAMIN, E. J.; BITTENCOURT, M. S.; CALLAWAY, C. W.; CARSON, A. P.; CHAMBERLAIN, A. M.; CHENG, S.; DELLING, F. N. et al. Heart disease and stroke statistics—2021 update. **Circulation**, v. 143, p. e254–e743, 2021. DOI: 10.1161/CIR.0000000000000950.

WOOD, S. **Generalized Additive Models: An Introduction With R**. Boca Raton, FL: Chapman and Hall/CRC Press, 2006.

APÊNDICE A – Imagem de satélite das estações de monitoramento do ar no município de São Paulo, Brasil.



Fonte: CETESB

APÊNDICE B – Algoritmo de Ajuste MAG-ARMA (Um Único Poluente)

Algoritmo de Ajuste MAG-ARMA (Um Único Poluente)

Passo 1: Preparação da Base de Dados

- Carregar dados de internações, poluente atmosférico e temperatura média.
- Remover observações com valores ausentes.
- Definir vetor temporal $t = 1, \dots, n$.

Passo 2: Análise Gráfica da Decomposição Temporal e Construção dos Termos Temporais

- Criar termo de tendência linear.
- Criar harmônicos anuais e ou semestrais.

Passo 3: Seleção dos Graus de Liberdade para Spline da Variável de Confusão

- Cálculo do número de gl pelo método híbrido (VCG e AIC)
- Construir spline natural da temperatura com gl .
- Ajustar modelo MAG sem componentes ARMA.
- Calcular o AIC correspondente.
- Selecionar o valor de gl com menor AIC.

Passo 4: Ajuste do Modelo MAG

- Construir matriz de covariáveis com: poluente, spline da confundidora, tendência e harmônicos.
- Ajustar modelo MAG assumindo distribuição de Poisson.
- Calcular o BIC do modelo ajustado.
- Obter resíduos de Pearson.

Passo 5: Avaliação dos Resíduos

- Analisar gráficos de ACF e PACF dos resíduos.
- Se não houver autocorrelação (resíduos com ruído branco):
- Considerar o modelo MAG como adequado.
- Caso contrário (autocorrelação presente):
- Prosseguir para o Passo 6.

Passo 6: Ajuste do Modelo MAG-ARMA

- Ajustar modelo MAG-ARMA incorporando termos AR e/ou MA.
- Definir ordens AR e MA com base nos gráficos de ACF e PACF do Passo 5.

Passo 7: Verificação do Modelo MAG-ARMA

- Reavaliar ACF e PACF dos novos resíduos.
- Calcular o BIC do modelo ajustado.

Passo 8: Comparação de Modelos

- Comparar os valores de BIC entre MAG e MAG-ARMA.
- Selecionar o modelo final com menor BIC e com ruído branco.

Passo 9: Estimativa do Efeito do Poluente

- Calcular RR e intervalo de confiança de 95% para a variação interquartil do poluente.

Passo 10: Repetição para o Segundo Poluente

- Repetir todo o procedimento (Passos 1 a 9) para o outro poluente atmosférico.
- Comparar os efeitos estimados (RR) entre os dois poluentes.

APÊNDICE C – Código em R Modelo MAG-ARMA

```

# =====
# Bibliotecas e leitura dos dados
# =====
library(glarma)
library(dplyr)
library(splines)
library(stats)
data(df_mensal)
# =====
# 1. Matriz Desenho (poluente + spline(Temp) + harmônicos + tendência
# =====
XI_mensal <- cbind(
  MP2_5 = df_mensal$MP2_5,
  sTemp,
  trend = df_mensal$trend_mensal,
  ano_sin1 = df_mensal$ano_sin1,
  ano_cos1 = df_mensal$ano_cos1,
  semi_sin1 = df_mensal$semi_sin1,
  semi_cos1 = df_mensal$semi_cos1)
# =====
# 2. Construção do spline para temperatura (relação não linear)
# =====
sTemp <- as.data.frame(ns(df_mensal$Temp, df = 4))
colnames(sTemp) <- paste0("sTemp_", 1:ncol(sTemp))
# =====
# 3. Ajuste do modelo MAG (sem estrutura ARMA)
# =====
GAM_2_5 <- glarma(
  y = df_mensal$internacao_mensal,
  X = as.matrix(XI_mensal),
  phiLags = NULL, # sem AR
  thetaLags = NULL, # sem MA
  residuals = "Pearson",
  type = "Poi",
  method = "NR",
  maxit = 100)
summary(GAM_2_5)
# =====
# 3. Ajuste do modelo MAG-ARMA --- AR(1)
# =====
GAM_AR1_2_5 <- glarma(
  y = df_mensal$internacao_mensal,
  X = as.matrix(XI_mensal),
  phiLags = c(1), # componente AR(1)
  thetaLags = NULL, # sem MA
  residuals = "Pearson",
  type = "Poi",
  method = "NR",
  maxit = 100)
# =====
# 3. Cálculo do BIC — GOF para modelos GLARMA
# =====
calc_BIC_glarma <- function(model, n_obs) {
  ll <- as.numeric(logLik(model, deriv = 0)) # log-verossimilhança
  k <- length(model$coefficients) # coeficientes beta
  if (!is.null(model$phiLags)) k <- k + length(model$phiLags)
  if (!is.null(model$thetaLags)) k <- k + length(model$thetaLags)
  bic <- -2 * ll + k * log(n_obs)
  return(bic)}

```

```

n_obs <- nrow(df_mensal)
bic_GAM <- calc_BIC_glarma(GAM_2_5, n_obs)
cat("BIC (GAM - sem ARMA): ", round(bic_GAM, 3), "\n")
# =====
# 4. Diagnóstico: ACF e PACF dos resíduos e teste Ljung-Box
# =====
acf(GAM_2_5$residuals, lag.max = 50, main = "ACF dos resíduos - GAM")
pacf(GAM_2_5$residuals, lag.max = 50, main = "PACF dos resíduos - GAM")
# Teste de Ljung-Box
Box.test(GAM_2_5$residuals, lag = 20, type = "Ljung-Box")
# =====
# 5. Estimativa de risco relativo (RR) associado ao MP2,5
# =====
beta_MP2_5 <- coef(GAM_2_5, type = "beta")["MP2_5"]
se_MP2_5 <- 0.05877 # erro padrão informado no summary
xi <- IQR(df_mensal$MP2_5, na.rm = TRUE)
z <- qnorm(0.975)
(RR <- exp(beta_MP2_5 * xi))
(lowerIC <- exp(beta_MP2_5 * xi - z * se_MP2_5 * xi))
(upperIC <- exp(beta_MP2_5 * xi + z * se_MP2_5 * xi))

```