

Deteccção de anomalias de Síndrome Respiratória Aguda Grave não especificada associada a fatores ambientais em Regiões Metropolitanas dos Biomas Amazônico e Cerrado no Brasil, 2019 a 2023.

Eucilene Alves Santana

Trabalho de Conclusão de Curso
MBA em Inteligência Artificial e Big Data

UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação

Detecção de anomalias de Síndrome
Respiratória Aguda Grave não especificada
associada a fatores ambientais em Regiões
Metropolitanas dos Biomas Amazônico e
Cerrado no Brasil, 2019 a 2023.

Eucilene Alves Santana

USP - São Carlos

2025

Eucilene Alves Santana

Detecção de anomalias de Síndrome Respiratória
Aguda Grave não especificada associada a fatores
ambientais em Regiões Metropolitanas dos Biomas
Amazônico e Cerrado no Brasil, 2019 a 2023.

Trabalho de conclusão de curso
apresentado ao Departamento de Ciências de
Computação do Instituto de Ciências Matemáticas
e de Computação, Universidade de São Paulo -
ICMC/USP, como parte dos requisitos para
obtenção do título de Especialista em Inteligência
Artificial e Big Data.

Área de concentração: Inteligência
Artificial.

Orientador: Prof. Dr. Francisco Aparecido Rodrigues

USP - São Carlos

2025

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

A474d Alves Santana, Eucilene
 Detecção de anomalias de Síndrome Respiratória Aguda
 Grave não especificada associada a fatores ambientais
 em Regiões Metropolitanas dos Biomas Amazônico e
 Cerrado no Brasil, 2019 a 2023. / Eucilene Alves
 Santana; orientador Francisco Aparecido Rodrigues. --
 São Carlos, 2025.
 45 p.

 Trabalho de conclusão de curso (MBA em Ciência de
 Dados) -- Instituto de Ciências Matemáticas e de
 Computação, Universidade de São Paulo, 2025.

 1. Epidemiologia. 2. Aprendizado de máquina. 3. Big
 Data. 4. Saúde. 5. Ambiental. I. Rodrigues, Francisco
 Aparecido, orient. II. Título.

Bibliotecários responsáveis pela estrutura de catalogação da publicação de acordo com a AACR2:
Gláucia Maria Saia Cristianini - CRB - 8/4938
Juliana de Souza Moraes - CRB - 8/6176

DEDICATÓRIA

*Aos grandes Homens da
minha vida, Rafael, Daniel e Isaac
pela compreensão, carinho e apoio
incansável.*

AGRADECIMENTOS

Agradeço à coordenação do curso pela oportunidade concedida por meio da bolsa de estudos, que possibilitou minha formação e ampliou minhas perspectivas profissionais e acadêmicas e ao meu orientador Dr. Francisco Aparecido Rodrigues pelas dicas de artigos interessantes para o desenvolvimento do trabalho.

À minha família, registro meu profundo reconhecimento pelo incentivo permanente, pela compreensão nas horas de estudo e pela valorização da educação desde minha formação inicial, ensinando-me que o conhecimento abre portas e transforma destinos.

Ao meu neto, Isaac Duarte Moraes, cuja ternura e entusiasmo acompanharam toda esta jornada. Suas palavras — *“Vovó fez a prova? Assistiu às aulas? Terminou seu MBA, vovó? Te ajudo se precisar”* — foram combustível diário para minha perseverança. Que este trabalho seja também um exemplo para você, como você foi inspiração para mim.

Manifesto também minha gratidão ao amigo Audêncio Rafael, por ter me apresentado o curso e incentivado minha participação no processo seletivo, gesto que contribuiu decisivamente para este percurso formativo.

Por fim, agradeço aos colegas do curso, cuja colaboração, troca de experiências e constante apoio mútuo fortaleceram a caminhada, reforçando que o aprendizado coletivo amplia conquistas e multiplica motivação.

A todos, meu sincero muito obrigada.

EPÍGRAFE

“Sou mais sábio do que este homem apenas neste ponto: não penso saber aquilo que não sei; ao passo que ele acredita saber o que não sabe.”

(Platão, Apologia de Sócrates 21d)

RESUMO

SANTANA, E.A. **Deteccção de anomalias de Síndrome Respiratória Aguda Grave não especificada associada a fatores ambientais em Regiões Metropolitanas dos Biomas Amazônico e Cerrado no Brasil, 2019–2023.** 2025 45f. Monografia (MBA em Ciências de Dados) – Centro de Ciências Matemáticas Aplicadas à Indústria, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2025.

O estudo analisou a incidência de Síndrome Respiratória Aguda Grave (SRAG) não especificada em Regiões Metropolitanas dos biomas Amazônico e Cerrado no Brasil, entre 2019 e 2023, com o objetivo de identificar padrões sazonais, espaciais e eventos anômalos associados a variáveis climáticas extremas, especialmente ondas de calor. Foi conduzido um estudo ecológico de séries temporais, integrando dados epidemiológicos do SIVEP-Gripe e indicadores ambientais. A abordagem metodológica incluiu imputação estrutural, decomposição de séries temporais por STL, análise de componentes principais, clusterização, aplicação de algoritmos de detecção de concept drift (Bai-Perron, CUSUM e Kolmogorov-Smirnov) e testes de associação, incluindo regressão logística binária. Os resultados evidenciaram padrão de transição epidemiológica, com picos residuais de SRAG não especificada durante 2020 e 2021 e estabilização após 2022. O modelo detectou eventos de drift predominantemente nas regiões amazônicas, com forte associação entre extremos térmicos e instabilidade epidemiológica (Belém: OR=2,76; Manaus: OR=11,64), enquanto no Cerrado os eventos refletiram majoritariamente a dinâmica local das internações por SRAG não especificada. A análise multivariada identificou heterogeneidade regional: A RM Vale do Rio Cuiabá apresentou comportamento isolado, Manaus e Belém dinâmicas opostas, e RIDE-DF e a RM de Goiânia padrões semelhantes. As variáveis originais demonstraram maior capacidade discriminatória para detecção de mudanças estruturais do que os componentes residuais. Conclui-se que extremos climáticos desempenham papel modulador na instabilidade respiratória, sobretudo na Amazônia, reforçando a necessidade de incorporar indicadores ambientais em sistemas de vigilância e desenvolver modelos preditivos sensíveis a variabilidades climáticas regionais.

Palavras-chave: epidemiologia; aprendizado de máquina; Big Data em saúde; vigilância sindrômica; clima e saúde; SRAG

ABSTRACT

SANTANA, E.A. **Detection of anomalies in unspecified Severe Acute Respiratory Syndrome in Metropolitan Regions of the Amazon and Cerrado biomes in Brazil, 2019–2023.** 45f. Monografia (MBA em Ciências de Dados) – Centro de Ciências Matemáticas Aplicadas à Indústria, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2025.

The study examined the incidence of unspecified Severe Acute Respiratory Syndrome (SARS) in metropolitan regions of the Amazon and Cerrado biomes in Brazil between 2019 and 2023, aiming to identify seasonal and spatial patterns as well as anomalous events associated with extreme climatic conditions, particularly heatwaves. An ecological time-series design was applied, integrating epidemiological data from the SIVEP-Gripe system with environmental indicators. The methodological framework encompassed structural imputation, seasonal-trend decomposition using Loess (STL), principal component analysis, clustering, concept-drift detection algorithms (Bai-Perron, CUSUM, and Kolmogorov-Smirnov), and association tests, including binary logistic regression. Findings indicated an epidemiological transition, with marked residual peaks of unspecified SARS during 2020 and 2021 followed by stabilization after 2022. Concept-drift events were predominantly detected in Amazonian regions, with strong associations between extreme thermal anomalies and epidemiological instability (Belém: OR=2.76; Manaus: OR=11.64), whereas in the Cerrado, drift primarily reflected local dynamics of unspecified SARS hospitalizations. Multivariate analysis revealed regional heterogeneity: the Cuiabá metropolitan region exhibited a distinct pattern, Manaus and Belém showed divergent dynamics, and the RIDE-DF and Goiânia regions clustered similarly. Original environmental and epidemiological variables demonstrated greater discriminatory ability for detecting structural changes compared with residual components. The study concludes that climatic extremes play a modulatory role in respiratory instability, particularly in the Amazon, underscoring the importance of incorporating environmental indicators into surveillance systems and developing predictive models capable of capturing region-specific climatic variability.

Keywords: epidemiology; machine learning; health Big Data; syndromic surveillance; climate and health; SRAS

LISTA DE ILUSTRAÇÕES

Figura 1. Resumo gráfico das etapas metodológicas para a detecção de anomalias de Síndrome Respiratória Aguda Grave (SRAG) e sua associação com alterações climáticas em estados dos Biomas Amazônico e de Cerrado no Brasil de 2019- 2023.	18
Figura 2. Dinâmica temporal de internações por síndrome respiratória grave com tendência, sazonalidade e resíduos nas regiões metropolitanas selecionadas entre 2019 e 2023.....	27
Figura 3.. Proporção da variância explicada pelos componentes principais.	28
Figura 4. Representação da variabilidade multivariada entre Regiões Metropolitanas brasileiras com base nos dois primeiros componentes principais, evidenciando padrões regionais distintos e estrutura de similaridade entre os territórios	30
Figura 5.Dinâmica temporal dos resíduos de SRAG e identificação de concept drift nas regiões metropolitanas estudadas entre 2019 e 2023.....	32
Figura 6. Comparação das séries padronizadas de SRAG residual e temperatura máxima (original e residual) com janelas de concept drift por RM entre 2019 e 2023.	33
Figura 7.Percentual de Variáveis Ambientais Significativamente Associadas ao Concept Drift por Região Metropolitana.....	34
Figura 8.Matriz de frequência de associações significativas por variável.	35
Figura 9 Hotmap: Associação entre variáveis e concept drifts por região Metropolitana.	36
Figura 10.Distribuição da força de associação (Cohen'd) por tipo de variável.	37
Figura 11.Razões de chances (OR) para ocorrência de concept drift segundo indicadores ambientais e de SRAG nas Regiões Metropolitanas selecionadas.	39

LISTA DE TABELAS

Tabela 1. Estimativas de razão de chances (OR) para a ocorrência de concept drift segundo variáveis climáticas e epidemiológicas nas Regiões Metropolitanas analisadas.....	40
---	----

SUMÁRIO

RESUMO	8
ABSTRACT	9
LISTA DE ILUSTRAÇÕES	10
LISTA DE TABELAS	11
1 INTRODUÇÃO	12
2 REVISÃO BIBLIOGRÁFICA	14
3 MÉTODOS.....	17
3.1.Coleta e preparação dos dados	18
3.1.1.Definição das janelas de análises.....	19
3.2.Tratamento de séries, imputação e decomposição (STL)	19
3.3. Análise de componentes principais e agrupamento	20
3.4. Concept drift	22
3.4.1.Métodos de Detecção.....	23
3.4.2.Testes de associações.....	24
4. RESULTADOS	26
4.1. Tratamento e Decomposição Temporal (STL).....	26
4.2. Análise de Componentes Principais (PCA).....	27
4.3. Detecção de Concept Drift	31
4.4 Testes de associações.....	34
4.5. Modelos de Regressão Logísticas.....	37
5 CONCLUSÃO.....	41
6. REFERÊNCIAS	43

1 INTRODUÇÃO

As mudanças climáticas têm se intensificado, causando impactos diretos à saúde da população humana. Entre os efeitos cita-se o aumento de casos de Síndrome Respiratória Aguda Grave (SRAG), cujas hospitalizações e mortalidade podem estar associadas à degradação da qualidade do ar (Ignotti et al., 2010), como no caso dos focos de queimadas que ocorrem anualmente no Brasil.

Estados como Mato Grosso, Amazonas e Pará localizados nos biomas do Cerrado e Amazônia lideram em número de áreas queimadas. Juntos, esses dois biomas concentram aproximadamente 86% da área queimada nos dos últimos 39 anos de monitoramento, em pelo menos em um ano, segundo dados do MapBioma de 1985 a 2023. Os estados, Mato Grosso e Pará lideram o ranking de queimadas, sendo considerados uma das áreas mais críticas do país. Esse cenário pode impactar diretamente a qualidade do ar da população que reside nesses territórios.

Dos 19.973 casos de SRAG registrados entre as Semanas Epidemiológicas: SE1 e SE 13 de 2023, no Brasil, 10.017 foram de SRAG não especificado, ou seja, sem identificação do agente etiológico (H1N1,H3N2, influenza A não subtipado, influenza B, vírus sincicial respiratório, SARS-CoV2, outros vírus respiratórios ou outros agentes etiológicos) conforme dados do informe publicado pelo Ministério da Saúde (2024), não diferente para as Unidades da Federação objetos de estudo.

Isso leva a questionar, se esse alto percentual (50,1%) de SRAG não especificado está relacionado, apenas à subnotificação e/ou ao atraso de notificação, ou se pode estar associado ao período de circulação de múltiplos vírus que não foram capturados em testes laboratoriais, ou ainda se pode estar associado às mudanças climáticas extremas afetando por exemplo, grupos populacionais específicos como idosos, crianças e indivíduos com comorbidades.

Estudos sobre a associação entre os eventos e a ocorrência de SRAG, ainda necessitam de análise aprofundada, especialmente no que diz respeito à detecção precoce de anomalias espaciotemporais a partir da análise dos dados monitorados pela vigilância das síndromes gripais, a partir dos dados registrados no Sistema de Informação de Vigilância Epidemiológica da Gripe (SIVEP-Gripe), sistema de gestão do Ministério da Saúde, em conjunto com as Secretarias Estaduais e Municipais de Saúde.

Essa situação justifica, por exemplo, a implementação de metodologias baseadas em séries temporais para identificar padrões incomuns na ocorrência de SRAG. Ressalta-se que a detecção de anomalias (outliers) em séries temporais por meio de técnicas de aprendizado de

máquina representa uma abordagem promissora para automatizar a identificação de períodos com aumentos ou reduções atípicas de casos, incluindo, por exemplo, registros classificados como SRAG não especificada. Segundo Chandola, Banerjee e Kumar (2009), métodos de detecção de anomalias são especialmente úteis para monitorar eventos raros ou inesperados em fluxos temporais de dados, como os observados em sistemas de vigilância epidemiológica.

Embora existam evidências crescentes da associação entre variáveis ambientais extremas e agravos respiratórios, a maioria dos estudos conduzidos até o momento apresenta limitações em integrar, de forma simultânea, dados epidemiológicos e ambientais com abordagem espaciotemporal. Essa lacuna metodológica compromete a capacidade de detecção precoce de padrões de anomalias, especialmente em contextos marcados por alta vulnerabilidade socioambiental, como os biomas Amazônico e do Cerrado.

Nesse sentido, o presente estudo propõe, de forma inovadora, o desenvolvimento de um modelo de detecção de anomalias espaciotemporais na incidência de SRAG, com foco na SRAG não especificada. Os objetivos específicos incluem: (i) analisar os padrões sazonais e territoriais da incidência de SRAG não especificada em Regiões Metropolitanas da Amazônia Legal e do Cerrado; (ii) correlacionar tais padrões com indicadores ambientais, como número de focos de calor e ondas de calor.

Parte-se da hipótese de que a incidência de SRAG não especificada apresenta padrão sazonal e espacial associado a eventos climáticos extremos, sendo mais pronunciada em áreas de maior fragilidade ambiental e social.

A análise foi focada em regiões metropolitanas localizada na Amazônia Legal e no Cerrado, por ser regiões historicamente afetadas por incêndios florestais e variações térmicas significativas. Ao integrar dados ambientais e epidemiológicos, espera-se contribuir para a formulação de estratégias de adaptação e mitigação, fortalecendo a vigilância em saúde ambiental e a resiliência dos sistemas de saúde frente às mudanças climáticas bem como subsidiar ações de vigilância e mitigação dos impactos à saúde, especialmente em populações em situação de vulnerabilidade.

A seguir, apresenta-se uma revisão da literatura científica com foco em estudos que exploram os efeitos de queimadas, ondas de calor e variabilidade climática sobre a incidência e internações por doenças respiratórias, com destaque para populações vulneráveis e regiões historicamente impactadas pela degradação ambiental.

2 REVISÃO BIBLIOGRÁFICA

Para embasar o desenvolvimento do modelo proposto, é necessário compreender as evidências existentes sobre a relação entre variáveis ambientais extremas e os agravos respiratórios, em especial a Síndrome Respiratória Aguda Grave (SRAG). Diversos estudos têm indicado que eventos climáticos extremos, como queimadas e ondas de calor, estão associados ao aumento da incidência de doenças respiratórias, sobretudo entre populações em situação de vulnerabilidade. Esses eventos, intensificados pelas mudanças climáticas, comprometem a qualidade do ar por meio do aumento da concentração de material particulado fino (PM_{2,5}), o que contribui para o agravamento de quadros respiratórios e cardiovasculares (Delfino et al., 2008; Ignotti et al., 2010).

De dez estudos que investigam a associação entre fatores ambientais e internações por doenças respiratórias, nove foram realizados no Brasil, com forte concentração nas regiões Norte e Centro-Oeste, áreas historicamente expostas a queimadas e variações térmicas extremas. Apenas um estudo foi conduzido fora do país, na Califórnia (EUA), também região frequentemente afetada por incêndios florestais (Delfino et al., 2008).

A análise dos padrões espaciais e temporais revela uma concentração de efeitos adversos em municípios próximos a áreas de queimada. Delfino et al. (2008) mostraram que as internações hospitalares aumentaram até 34% durante períodos de fumaça intensa e incêndios florestais. O padrão sazonal foi evidenciado em seis dos dez estudos, que indicaram a estação seca como o período crítico para hospitalizações.

A exposição ao material particulado fino (PM_{2,5}) foi diretamente associada ao aumento de internações em pelo menos quatro estudos. Ignotti et al. (2010) relataram aumento de 6% a 6,8% nas admissões por cada incremento de 10 µg/m³ de PM_{2,5} em crianças e idosos. Apesar da importância do PM_{2,5} como marcador de poluição atmosférica, apenas quatro estudos incluíram essa variável de forma quantitativa, o que aponta uma lacuna importante na literatura.

Em uma revisão sistemática realizada na tentativa de identificar doenças sensíveis ao clima no Brasil e no mundo, Sousa e colaboradores destaca a relação entre variáveis climáticas, como temperatura e precipitação, e a incidência ou agravamento de doenças respiratórias, essas foram uma das categorias mais estudadas, com um total de 24 publicações identificadas na revisão dos autores. Observou-se ainda, que as doenças respiratórias apresentaram altos coeficientes de clusterização, indicando que estão associadas a uma diversidade maior de variáveis climáticas (Souza et al., 2018).

A heterogeneidade dos fatores associados à SRAG evidencia a necessidade de mecanismos complementares de monitoramento. Nesse contexto, a vigilância epidemiológica baseada em casos hospitalizados fornece subsídios para o planejamento de respostas assistenciais, sendo complementada por abordagens analíticas que busquem detectar padrões de anomalias da distribuição dos casos ao longo do ano.

A Síndrome Respiratória Aguda Grave (SRAG) constitui um agravo de notificação compulsória no Brasil, conforme determina a Portaria GM/MS nº 6.734, DE 18 de março de 2025 (BRASIL, 2025). Um caso de SRAG é definida como o quadro clínico de um indivíduo hospitalizado com febre (ainda que referida), associada a tosse ou dor de garganta e que apresenta sinais de agravamento, como dispneia, saturação de oxigênio inferior a 95%, desconforto respiratório ou evolução para óbito por causa respiratória, independentemente de internação (Brasil, 2024).

A vigilância da SRAG hospitalizada é realizada por meio do Sistema de Informação da Vigilância Epidemiológica da Gripe (SIVEP-Gripe), que integra dados provenientes de todas as unidades de saúde, públicas e privadas, onde haja internação ou óbito por SRAG. O protocolo recomenda a coleta de amostras biológicas para diagnóstico etiológico, com posterior análise pelos Laboratórios Centrais de Saúde Pública (LACENs), priorizando a testagem para vírus respiratórios, como influenza A e B, SARS-CoV-2, vírus sincicial respiratório, rinovírus, metapneumovírus, entre outros (Brasil, 2024).

A classificação dos casos de SRAG depende da confirmação laboratorial. Quando não é possível identificar o agente etiológico, mesmo após a realização de testes com o painel viral respiratório, os casos são encerrados como SRAG não especificada, o que dificulta a compreensão da real carga de morbimortalidade associada a causas ambientais ou infecciosas emergentes. Essa limitação reforça a necessidade de abordagens complementares de análise, capazes de detectar padrões incomuns na ocorrência de SRAG.

Apesar de a vigilância da Síndrome Respiratória Aguda Grave (SRAG) dispor de uma definição de caso clínico padronizada e amplamente utilizada em âmbito nacional, a sua operacionalização ainda é fortemente orientada para a identificação do agente etiológico viral. Conforme conversa informal com a equipe técnica de vigilância, uma proporção significativa dos casos notificados, frequentemente superior a 50% é encerrada sem identificação laboratorial conclusiva, sendo classificados como SRAG não especificada.

Destaca-se ainda, que o sistema InfoGripe, desenvolvido pela Fundação Oswaldo Cruz (Fiocruz), realiza o monitoramento contínuo de casos de Síndrome Respiratória Aguda Grave (SRAG) no Brasil por meio de análises de séries temporais, incluindo projeções de curto prazo

com técnicas de nowcasting (GOMES, 2014). Embora represente um avanço importante para a vigilância epidemiológica em nível nacional, sua modelagem concentra-se predominantemente nos aspectos temporais da tendência de casos, sem incorporar explicitamente variáveis ambientais, como queimadas ou ondas de calor, nos seus modelos preditivos. Assim, permanece uma lacuna significativa quanto à compreensão da influência de eventos climáticos extremos sobre os padrões de ocorrência da SRAG, especialmente no que tange aos casos não especificados.

Essa limitação compromete a utilização plena dos dados de vigilância para análises epidemiológicas mais abrangentes, dificultando a compreensão dos determinantes envolvidos e a detecção de eventos atípicos não virais, como aqueles potencialmente associados a fatores ambientais. Tal cenário evidencia a necessidade de estratégias complementares de análise, capazes de explorar os padrões espaciotemporais desses casos não definidos, subsidiando ações mais eficazes tanto no campo da vigilância quanto da assistência à saúde.

Diante dessas lacunas, o presente estudo se propõe a desenvolver um modelo de detecção de anomalias espaciotemporais na incidência de SRAG, integrando dados ambientais e epidemiológicos. A proposta visa fortalecer a vigilância em saúde ambiental, aprimorar a capacidade de resposta diante de eventos extremos e contribuir para a formulação de estratégias de mitigação e adaptação.

Para alcançar os objetivos propostos e testar a hipótese, o presente estudo propõe a aplicação de técnicas de aprendizado de máquina (machine learning) voltadas à detecção de anomalias em séries temporais e dados georreferenciados. Essa abordagem visa identificar padrões incomuns na distribuição de casos, permitindo maior sensibilidade na vigilância de eventos respiratórios graves em contextos de exposição ambiental. A combinação entre dados epidemiológicos e indicadores ambientais poderá subsidiar, de forma inovadora, estratégias de alerta precoce e alocação de recursos em saúde. A descrição detalhada da abordagem metodológica, incluindo os critérios de seleção dos dados, as variáveis ambientais incorporadas e os modelos estatísticos e computacionais utilizados, será apresentada na seção de métodos.

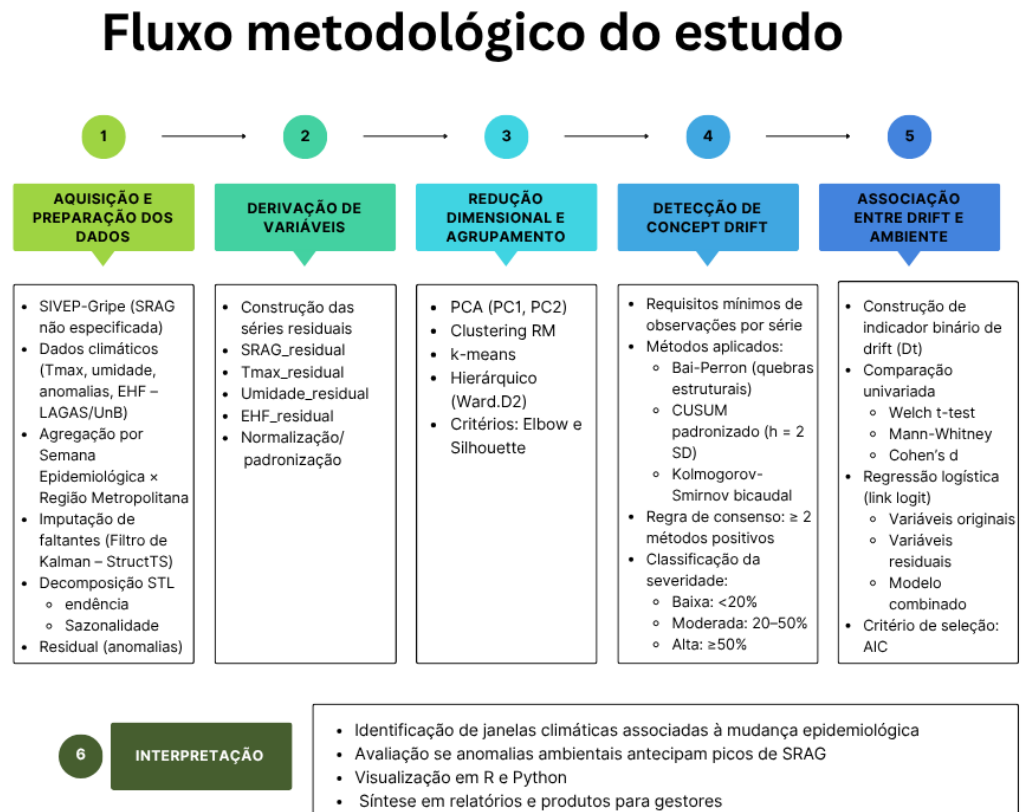
3 MÉTODOS

Trata-se de estudo ecológico de séries temporais com foco em interações por Síndrome Respiratória Aguda Grave (SRAG) sem especificação do agente etiológico e nas variáveis ambientais de cinco regiões Metropolitana Brasileira (RM), duas do Bioma Amazônico (EM Belém, RM Manaus) e três do Bioma Cerrado (RM Goiânia, RM Vale do Rio Cuiabá e a Região Integrada de Desenvolvimento do Distrito Federal e Entorno) no período de 2019 a 2023.

A metodologia utilizada é fundamentada nos princípios de aprendizado de máquina em ambientes não estacionários, por meio da detecção e compreensão da deriva conceitual (*concept drift*). Este fenômeno ocorre quando as distribuições estatísticas dos dados mudam ao longo do tempo, exigindo adaptação dinâmica dos modelos analíticos. Neste contexto, adotamos com base o trabalho de Lu et al. (2019), que evidencia o aprendizado sobre *drift* em três componentes: detecção, compreensão (quando, onde, e quão severa é a mudança) e adaptação. A abordagem é essencial para vigilância epidemiológica diante de eventos climáticos extremos, pois permite reconhecer padrões emergentes na incidência de SRAG não especificada associados a mudanças ambientais.

O Estudo foi desenvolvido em etapas conforme apresentado na Figura 1

Figura 1. Resumo gráfico das etapas metodológicas para a detecção de anomalias de Síndrome Respiratória Aguda Grave (SRAG) e sua associação com alterações climáticas em estados dos Biomas Amazônico e de Cerrado no Brasil de 2019- 2023.



Elaboração própria.

O detalhamento de cada etapa é apresentado a seguir:

3.1. Coleta e preparação dos dados

a. Dados de SRAG Sistema de Informação da Vigilância Epidemiológica da Gripe (SIVE-Gripe) com foco nas variáveis: data de início de sintomas dos casos hospitalizados, data de hospitalização, município de residência, município de hospitalização, SRAG não especificado (sem identificação do agente etiológico).

b. Dados ambientais: as variáveis de interesse selecionadas foram a Temperatura máxima, umidade média, anomalias de temperaturas e Fator de excesso de calor (EHF) estimada para as Regiões Metropolitanas. Utilizou-se os dados disponibilizados no repositório do Laboratório de Geografia, Ambiente e Saúde ([LAGAS](#)) da Universidade de Brasília.

3.1.1. Definição das janelas de análises

- c. Os dados foram divididos por períodos de análises fixos (semana epidemiológica por ano);
- d. Para cada período foi calculado medidas de distribuição (Distribuição de casos de SRAG não especificada por RM)
- e. Os dados ambientais foram agrupados por Semana epidemiológica ano para cada RM.

3.2. Tratamento de séries, imputação e decomposição (STL)

A presença de valores ausentes, comuns em análise de séries temporais é um problema frequente, para contornar essa limitação, optou-se pela utilização do filtro de Kalman estruturado para imputação de dados faltantes. Técnica implementada através da função *na.kalman()* pacotes *forecast* em R com o parâmetro *model= "StructTS"*, que ajusta um modelo de suavização estrutural para séries temporais, considerando sua estrutura subjacente de tendência e sazonalidade. Preservando também as propriedades estocásticas da série original (Hyndman; Athanasopoulos, 2018), especialmente para séries temporais com padrões complexos.

Para decompor cada série temporal das RM, aplicou-se o método Seasonal-Trend decomposition using Loess (STL), proposto por Cleveland et al., (1990). Para cada variável x_t sob análise, a decomposição STL foi implementada com os seguintes parâmetros: Frequência sazonal semanal: $f=52$ e a Opção robusta ativada. A partir da fórmula:

$$x_t = T_t + S_t + R_t$$

Onde:

T_t = representa o componente de tendência de longo prazo;

S_t = denota o componente sazonal periódico com frequência semanal

R_t = corresponde ao componente residual (remainder).

O STL é capaz de lidar com diferentes tipos de sazonalidade, incluindo aquelas que variam ao longo do tempo, sendo particularmente eficaz para séries com portamento não linear (Cleveland et al., (1990).

Após a decomposição STL, para cada série temporal foram criadas três variáveis derivadas:

- *_trend: representa o componente de tendência T_t
- *_seasonal: corresponde ao componente sazonal S_t ;
- *_residual: captura o componente residual R_t .

O componente residual R_t funciona como proxy para anomalias e flutuações não explicadas pelos componentes de tendência e sazonalidade. Este termo residual pode revelar os eventos atípicos de interesses, importante para detecção de padrões de anomalias nas séries temporais analisadas.

3.3. Análise de componentes principais e agrupamento

A técnica foi empregada para a redução de dimensionalidade e síntese de informação, permitindo a representação das variáveis analisadas em um espaço de dimensão reduzida. Constituiu-se dois conjuntos de dados distintos para finalidades analíticas específicas:

- Conjunto 1- análise temporal linha-a- linha: compreende as variáveis RM, temperatura_maxima_residual, umidade_media_residual, anomalias_temperatura e EHF, visando a síntese dimensional e posterior inclusão dos escores dos componentes principais (PC1, PC2) ao banco de dados principal.
- Conjunto2 – análise agregada por RM: Inclui médias e desvios-padrão, por Região Metropolitana, das variáveis temperatura_maxima_residual, umidade_media_residual, anomalias_temperatura_residual, EHF_residual e internacoes_SRAG_residual, destinando-se à comparação inter-RMs e procedimentos de clustering.

Todas as variáveis foram submetidas à padronização prévia à aplicação da PCA a partir da fórmula:

$$Z_{tj} = \frac{x_{tj} - \mu_j}{\sigma_j}$$

Onde:

Z_{tj} = valor padronizado da observação t da variável j ;

x_{tj} = representa cada observação das variáveis climáticas e de saúde antes da padronização;

μ_j = é a média da variável j ;

σ_j = desvio padrão da variável j .

A padronização garante que as variáveis com diferentes escalas de medida contribuam igualmente para análise, prevenindo que as variáveis com maior variância dominem artificialmente os componentes (Jolliffe, 2002).

A decomposição espectral da matriz de covariância (ou correlação, no caso de dados padronizados) é a base para a extração de autovalores λ_k e autovetores v_k do PCA. A variância é explicada por cada componente principal calculada por:

$$VarExp(PC_k) = \frac{\lambda_k}{\sum_{m=1}^p \lambda_m}$$

O processo de PCA gera dois produtos principais:

- Os escores dos componentes principais que identifica as coordenadas das observações no espaço reduzido, estes foram obtidos por meio da projeção dos dados padronizados Z nos autovetores correspondentes:

$$PC_k = Z_{vk}$$

- As cargas fatoriais (loadings), que são os coeficientes que mostram as correlações entre as variáveis originais e os componentes, permitindo a interpretação dos PCs dominantes, identificando quais variáveis contribuem mais significativamente para cada dimensão latente. Quanto maior o coeficiente, maior a correlação, podendo ser positiva ou negativa. Esse produto indica as direções das variáveis originais no espaço reduzido.

A partir dos componentes extraídos, realizou-se a análise de agrupamentos das Regiões Metropolitanas utilizando duas abordagens complementares: o *k-means* para particionamento e *agrupamento hierárquico* para validação da estruturação de cluster.

O algoritmo k-means foi implementado para particionamento das RM, baseando-se na minimização da soma dos quadrados intra-grupos por meio da fórmula:

$$\min \sum_{k=1}^K \sum_{i \in C_k} ||x_i - \underline{x}_{Ck}||^2$$

Onde:

C = representa o conjunto de clusters,

K = o número de clusters,

x_i = o vetor de características da RN i , e;

$\underline{x}_{Ck}$ = o Centróide do cluster C_k

A determinação do número ideal de cluster (K) baseou-se em dois critérios complementares:

- **Elbow Method:** Analisa a redução na Within-Cluster Sum of Squares (WCSS) em função do número de clusters, seleccionando o ponto onde ganhos marginais na compactação tornam-se desprezíveis.
- **Average Silhouette Method:** calcula para cada observação i o coeficiente de silhueta, a partir da fórmula:

$$s_i = \frac{b_i - a_i}{\max(a_i, b_i)}$$

Onde: a_i = representa a distância média da observação i às demais observações do mesmo cluster, e b_i = a menor distância média da observação i às observações de outro cluster. Valores próximos a 1 indicam cluster bem definidos (Kaufman; Rousseeuw, 2009).

De forma complementar, aplicou-se o método hierárquico aglomerativo utilizando a ligação Ward.D2, que minimiza a variância intra-cluster. O dendrograma resultante permitiu a inspeção visual da estrutura de agrupamento natural das RM, com cortes realizados no número K de cluster seleccionados pelos critérios anteriores,

3.4. Concept drift

As variáveis de interesse utilizadas foram SRAG residual e os residuais ambientais para temperatura máxima, umidade média, anomalias térmicas e EHF. A análise dos resíduos foi calculada por meio da fórmula:

$$r_t = X_t - \underline{X}_t$$

Onde: $\underline{X}_t$ representa o valor ajustado por tendência/sazonalidade.

Cada série temporal foi submetida a tratamento inicial que consistiu na remoção de valores ausentes. Estabeleceram-se critérios mínimos de qualidade: exigiu-se que cada série possuisse pelo menos 20 observações válidas para análise. Adicionalmente, para a execução do teste de Kolmogorov-Smirnov, requereu-se que cada metade da série (primeira e segunda parte) contivesse mínimo de 10 observações. Ressalta-se que nesta fase analítica não foram empregadas técnicas de imputação de dados.

3.4.1. Métodos de Detecção

Foram aplicados três métodos complementares nesta etapa. O teste de quebras estruturais baseado no algoritmo Bai-Perron que estima pontos de quebra T_j que minimizam a soma de quadrados dos resíduos por segmentos. A significância foi avaliada via estatística supF:

$$\sup F = \sup_{k \in K} F(k)$$

Onde: o $F(k)$ compara modelos com e sem quebra no ponto k . Segundo Zeileis et al., (2002) a existência de ao menos uma quebra ($m \geq 1$) indicam drift.

A implementação programática (CUSUM de resíduos Padronizados): monitora acumulação de desvios da média a partir das fórmulas:

$$\begin{aligned} S_t^{(+)} &= \max\{0, S_{t-1}^{(+)} + (y_t - \underline{y} - hs)\} \\ S_t^{(-)} &= \max\{0, S_{t-1}^{(-)} + (y_t - \underline{y} - hs)\} \end{aligned}$$

Considerou-se o limiar $h=2$ desvios-padrão. Os Índices com acúmulos persistentes (positivos ou negativos) são marcados como pontos de mudança (Brown; Durbin; Evans, 1975).

O Teste de Kolmogorov-Smirnov (KS) Bicaudal compara as distribuições da primeira e segunda metade da série pela fórmula:

$$D = \sup_x |F_1(x) - F_2(x)|$$

Onde F_1 e F_2 são CDFs empíricas das metades. Considerou-se Drift quando $p < 0,05$ conforme Kolmogorov (1993) e Smirnov (1939).

Utilizou-se como regra de consenso que para cada série residual, considerou-se drift detectado quando pelo menos dois dos três métodos sinalizaram mudança estrutural pela fórmula:

$$1\{drift\} = 1\{A + B + C \geq 2\}$$

Onde $A, B, C \in \{0,1\}$ representam, respectivamente, strucchange (quebras estruturais), CUSUM e KS.

A análise seguiu um laço iterativo por Região Metropolitana e variável residual, computando as Detecções individuais por método; o consenso entre métodos, o percentual de variáveis com drift por RM e a análise de componentes principais (PC1-PC5).

A classificação da severidade como Baixa <20%; moderada: entre 20 e 50% e Alta: $\geq 50\%$ das análises com drift foi definida ad hoc como critério operacional deste estudo, não existindo padronização consensual para tais percentuais na literatura de detecção de mudanças.

3.4.2. Testes de associações

A etapa de teste de associações teve por objetivo verifica-se os períodos de concept drift na série de interações por SRAG residuais estão associadas a variações nas variáveis ambientais (originais e residuais).

Para cada RM, definiu-se um indicador binário $D_t \in \{0,1\}$ para cada dat t , no qual $D_t = 1$ identifica janelas ao redor dos pontos de ruptura detectados por dois procedimentos complementares:

- (i) quebras estruturais em modelos de nível por segmentos, estimadas pelo método de Bai–Perron (Bai; Perron, 2003; Zeileis et al., 2002), e,
- (ii) estatística CUSUM bilateral com limiar $c=2$ desvios-padrão (Brown; Durbin; Evans, 1975). Operacionalmente, a janela foi expandida em $k=2$ observações em torno de cada ruptura, a fim de capturar o entorno imediato do evento de mudança.

Com o desfecho D_t definido, procedeu-se a testes univariados para cada variável ambiental X . Os valores de X foram particionados em dois grupos: período com drift ($D_t = 1$) e período sem drift $D_t = 0$. Para a comparação de médias utilizou o teste t de Welch (variâncias possivelmente desiguais), acompanhados do teste Mann-Whitney (robusto a desvios de normalidade) conforme descreveu Mann; Whitney, (1947).

Para quantificar a magnitude da diferença, calculou-se o tamanho de efeito de Cohen (d), definido como a diferença entre as médias dos grupos dividida pelo desvio-padrão conjunto (pooled), com interpretação padronizada : efeito pequeno ($|d| \approx 0,2$), médio ($|d| \approx 0,5$) e grande ($|d| \geq 0,8$) (Cohen, 1988). Os resultados foram considerados estatisticamente significativos quando $p < 0,05$, reportando-se conjuntamente o p =valor mínimo entre os testes paramétricos e não paramétrico e o respectivo $|d|$.

A avaliação da associação conjunta entre os preditores ambientais e a ocorrência de drift ajustou-se por RM, regressão logística binária com link logit (Dobson; Barnett, 2018), tendo D_t como variável resposta. Três especificações foram estimadas:

- (i) apenas variáveis originais;
- (ii) apenas variáveis residuais, e

(iii) modelo completo (originais mais residuais).

A comparação entre especificações utilizou o Akaike information Critério (AIC), selecionando-se o modelo com menor valor com o de melhor compromisso entre ajuste e parcimônia (Akaike, 1974). Coeficientes foram apresentados como odds ratios com intervalos de confiança de 95% (Hosmer; Lemeshow; Sturdivante, 2013).

Adicionalmente, para cada par temático (X , X^{res}) comparou-se a capacidade discriminatória por meio do $|d|$ dos testes univariados, computando-se $\Delta d = |d_{\text{res}}| - |d_{\text{orig}}|$. Classificou-se como “residual explica melhor” quando $\Delta d > 0,10$, a original explica melhor quando $\Delta d < -0,10$ e similar nos demais casos.

Essa comparação sintetiza, de forma padronizada, se os componentes residuais (anomalias de curto prazo) acrescentam informação relevante para a detecção de períodos de drift em relação às séries ambientais brutas.

Como controle de qualidade analítica, adotou-se:

- (a) tamanhos mínimos de amostra em cada grupo ($n \geq 5$) para execução dos testes;
- (b) inspeção de outliers e valores ausentes; e
- (c) padronização temporal por RM.

4. RESULTADOS

4.1. Tratamento e Decomposição Temporal (STL)

A série temporal semanal de internações por SRAG não especificada por Regiões Metropolitanas (Figura 2) permitiu identificar a estrutura temporal subjacente e distinguir padrões regulares de flutuações anômalas ao longo do período de 2019 a 2023 nas Regiões Metropolitanas analisadas nos biomas da Amazônia e Cerrado brasileiro.

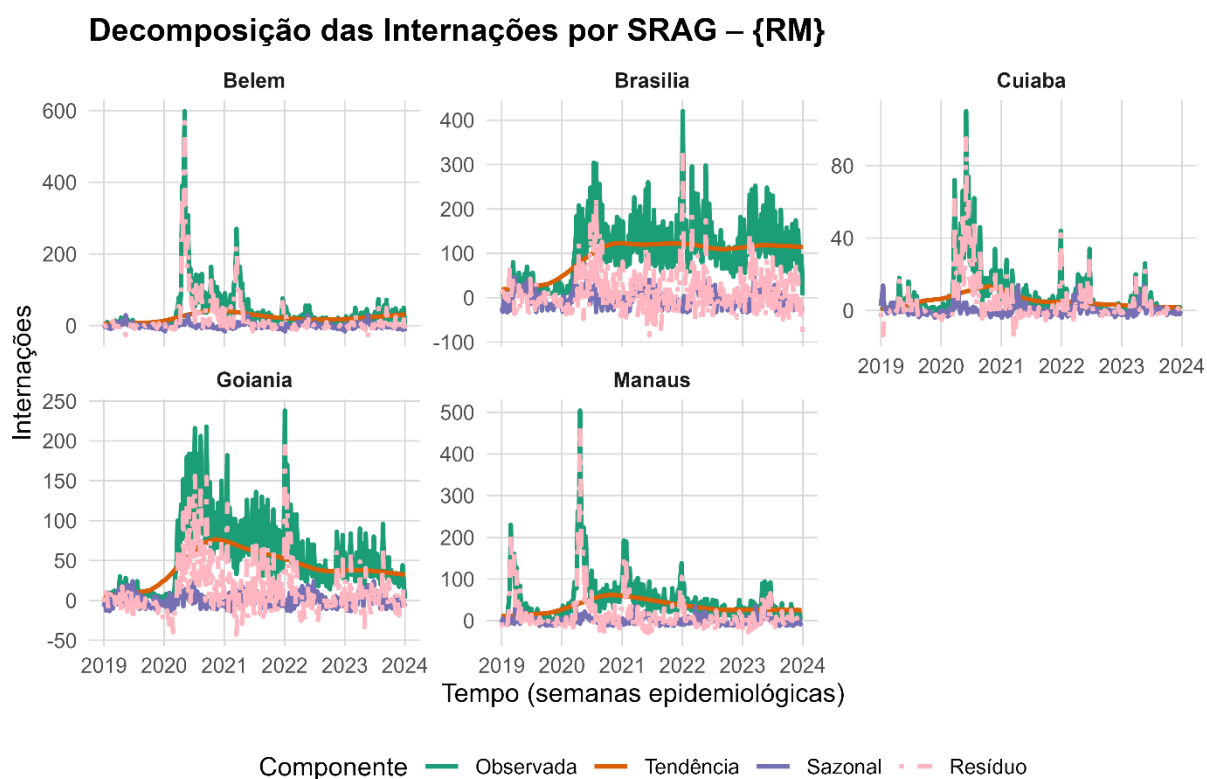
A linha de tendência evidencia elevações acentuadas no início do período, seguidas por declínio progressivo e estabilização em níveis inferiores após o pico epidemiológico. Esse comportamento indica transição de um regime epidêmico de alta intensidade para uma fase de circulação basal, compatível com um cenário de estabilização pós-evento agudo de saúde pública.

O componente sazonal demonstra oscilações regulares associadas à dinâmica respiratória anual, com maior amplitude durante o período inicial da série e redução gradual nos anos subsequentes. Tal padrão é consistente com a esperada sazonalidade respiratória observada em climas tropicais e subtropicais, sugerindo reemergência da estrutura sazonal após o período de maior perturbação epidemiológica (Tamerius et al., 2013; Schember et al., 2022; Fitter et al., 2021).

Os resíduos apresentam flutuações abruptas, sobretudo nos momentos de maior intensidade epidêmica, indicando desvios significativos em relação ao comportamento histórico esperado. Esses picos residuais foram mais pronunciados nas Regiões Metropolitanas de Belém e Manaus, enquanto Ride-DF, Goiânia e Cuiabá apresentaram variações atípicas de menor magnitude. A persistência de ruído residual, ainda que reduzida no período mais recente, reforça a necessidade de monitoramento contínuo para identificação precoce de eventos anômalos.

Em síntese, a decomposição STL demonstrou que, apesar do retorno progressivo da sazonalidade respiratória e da normalização da tendência de longo prazo, o período analisado foi marcado por episódios abruptos de alteração do padrão epidemiológico, compatíveis com surtos respiratórios intensos e subsequente estabilização estrutural. A observação conjunta dos componentes evidencia a utilidade da decomposição temporal como ferramenta para vigilância epidemiológica e suporte à detecção de desvios relevantes com potencial impacto em saúde pública.

Figura 2. Dinâmica temporal de internações por síndrome respiratória grave com tendência, sazonalidade e resíduos nas regiões metropolitanas selecionadas entre 2019 e 2023.

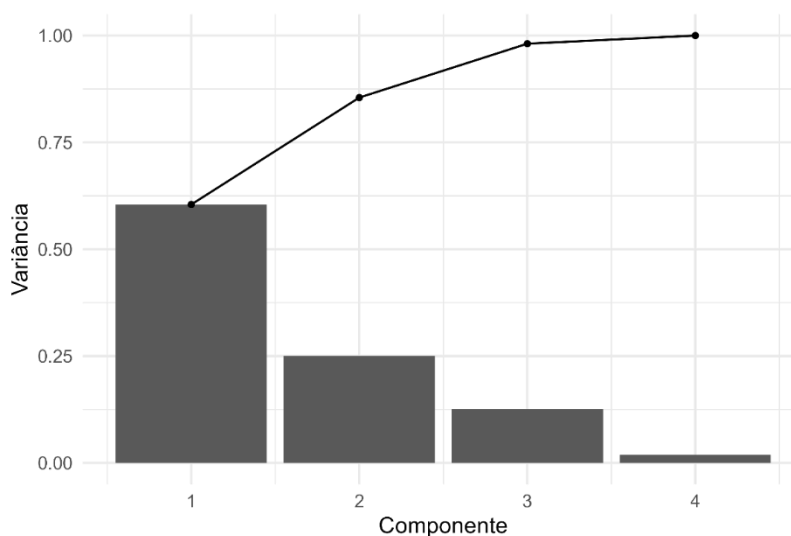


Fonte: Sivep-Gripe. Elaboração própria.

4.2. Análise de Componentes Principais (PCA)

O gráfico de variância (Figura 3) indica que o primeiro componente principal (PC1) responde pela maior fração da variância total (>55%), seguido pelo segundo componente (PC2), que explica aproximadamente 25%. Assim, juntos, PC1 e PC2 retêm cerca de 80% da variação total dos dados, justificando sua interpretação conjunta para caracterização do sistema.

Figura 3.. Proporção da variância explicada pelos componentes principais.



Elaboração própria.

No espaço fatorial (Figura 4A), observa-se que a RM Cuiabá (Vale do Rio Cuiabá) apresenta posição isolada, com valores negativos em ambos os eixos, sugerindo estrutura epidemiológica e ambiental distinta das demais regiões, potencialmente associada ao clima tropical sazonal seco e às dinâmicas respiratórias particulares do Centro-Oeste continental.

A RM de Manaus destaca-se com valores positivos elevados no segundo componente, indicando comportamento diferenciado, possivelmente influenciado pela combinação entre clima tropical úmido, densidade populacional concentrada e maior carga de eventos respiratórios no período analisado. A RM de Belém, embora também localizada na Amazônia, posiciona-se de forma oposta a RM de Manaus no segundo componente, sugerindo padrões distintos dentro da mesma macrorregião, indicando variabilidade intra-regional relevante.

As RM Goiânia e de Brasília (RIDE-DF) situam-se mais próximas ao centro, com valores moderados em ambos os eixos, refletindo maior homogeneidade e dinâmica intermediária no conjunto de indicadores, o que sugere comportamento respiratório mais estável e menos polarizado.

O dendrograma (Figura 4B) corrobora esses achados, ao demonstrar dois agrupamentos principais:

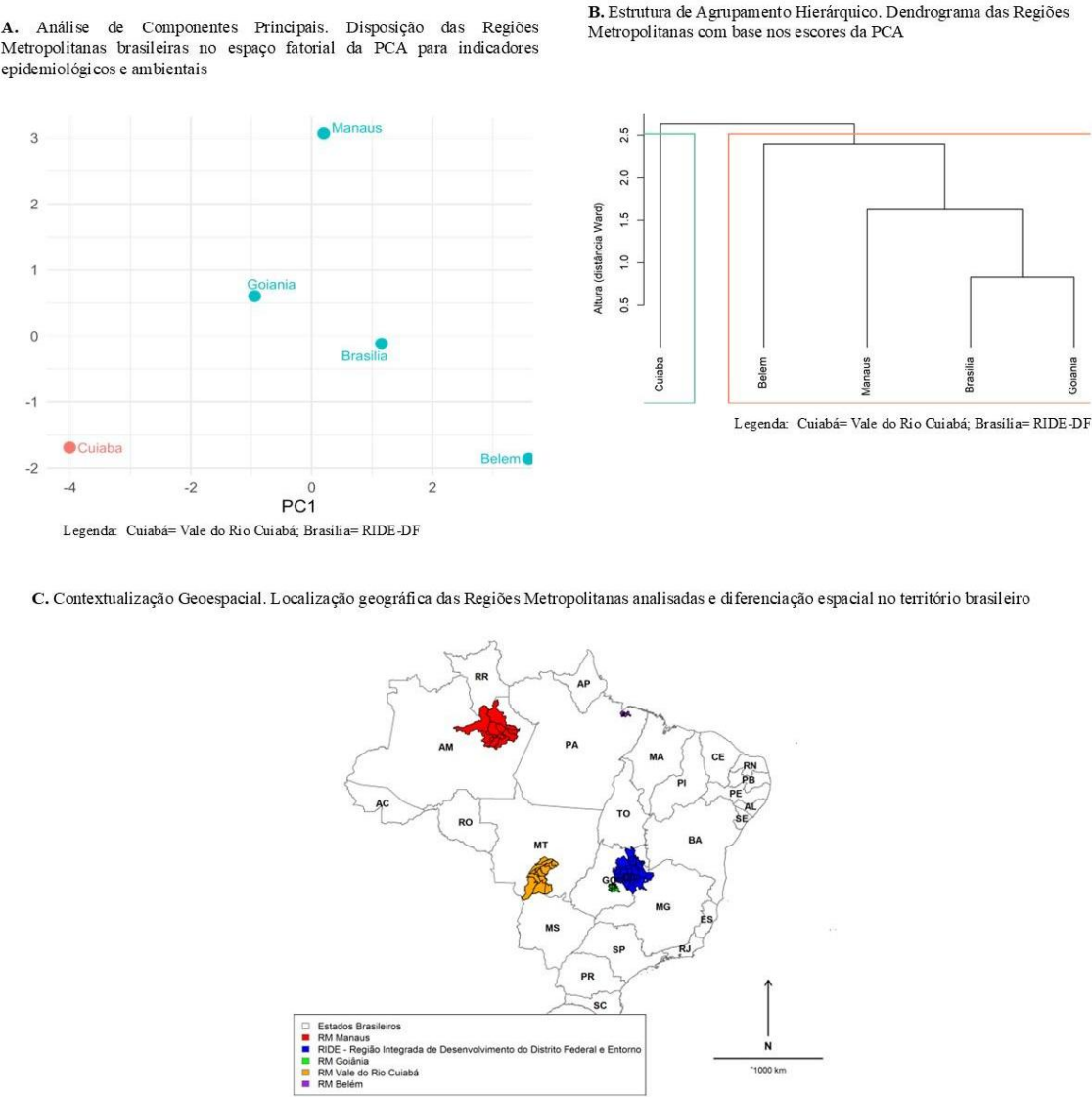
(i) Brasília (RIDE-DF) e a RM de Goiânia, compondo um cluster de comportamento similar;

(ii) RM de Manaus e a RM de Belém, porém com maior distância entre si do que entre Brasília (RIDE-DF) e a RM de Goiânia, reforçando a especificidade amazônica; e RM de Cuiabá (Vale do Rio Cuiabá) permanecendo como unidade isolada, confirmando sua singularidade estrutural.

O mapa espacial (Figura 4C) complementa a interpretação ao contextualizar a distribuição geográfica e climática das áreas, destacando diferenças macroambientais que ajudam a explicar a separação multivariada observada.

Em conjunto, os resultados indicam heterogeneidade regional relevante, sugerindo que análises e políticas de vigilância em saúde ambiental e respiratória devem considerar as especificidades climáticas, demográficas e ecossistêmicas de cada região, especialmente para modelagem de risco, detecção de anomalias e planejamento de resposta epidemiológica.

Figura 4. Representação da variabilidade multivariada entre Regiões Metropolitanas brasileiras com base nos dois primeiros componentes principais, evidenciando padrões regionais distintos e estrutura de similaridade entre os territórios



Elaboração própria.

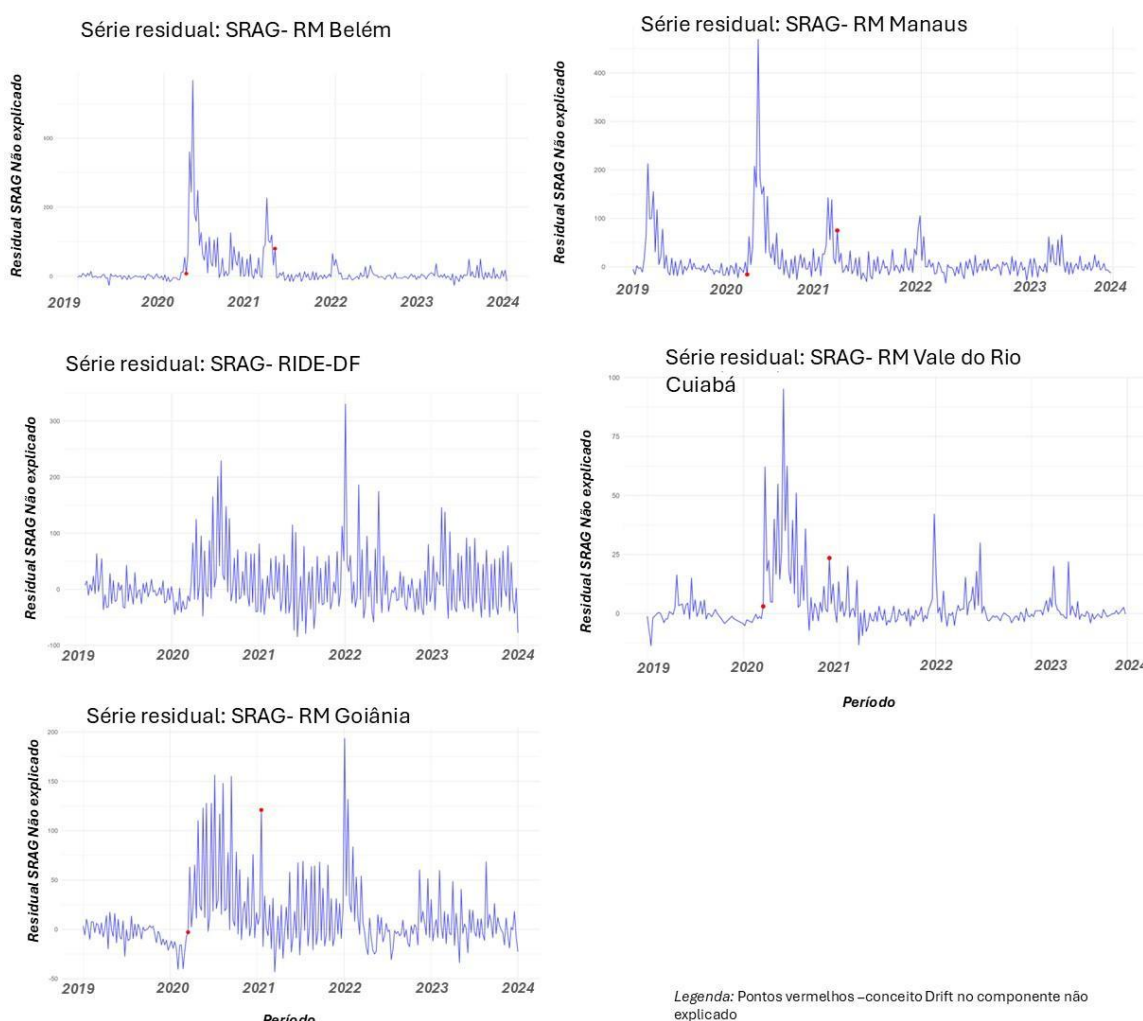
4.3. Detecção de Concept Drift

As séries temporais residuais de SRAG nas cinco regiões metropolitanas avaliadas evidenciam episódios discretos de concept drift, representados por pontos vermelhos que denotam mudanças não explicadas pelo modelo ajustado (Figura 5). Esses eventos caracterizam períodos nos quais a dinâmica da série se altera significativamente em relação ao padrão histórico previsto pelos componentes explicativos, refletindo perturbações estruturais no processo epidemiológico.

Observou-se maior intensidade e frequência de drifts entre 2020 e 2021, principalmente nas RM de Belém, RM de Manaus, RIDE-DF e RM de Goiânia. A RM Vale do Rio Cuiabá apresentou menor amplitude residual, apesar de episódios pontuais de drift no mesmo período, sugerindo resposta epidemiológica heterogênea entre os territórios amazônicos e do Centro-Oeste (Figura 5).

Após 2022, nota-se redução substancial na magnitude e frequência dos drifts, com estabilização próxima ao nível basal (Figura 5). Esse comportamento indica reestruturação do padrão epidemiológico e retorno a uma dinâmica mais previsível do componente residual, coerente com fase endêmica e adaptação comportamental, imunológica e assistencial ao SARS-CoV-2.

Figura 5. Dinâmica temporal dos resíduos de SRAG e identificação de concept drift nas regiões metropolitanas estudadas entre 2019 e 2023.



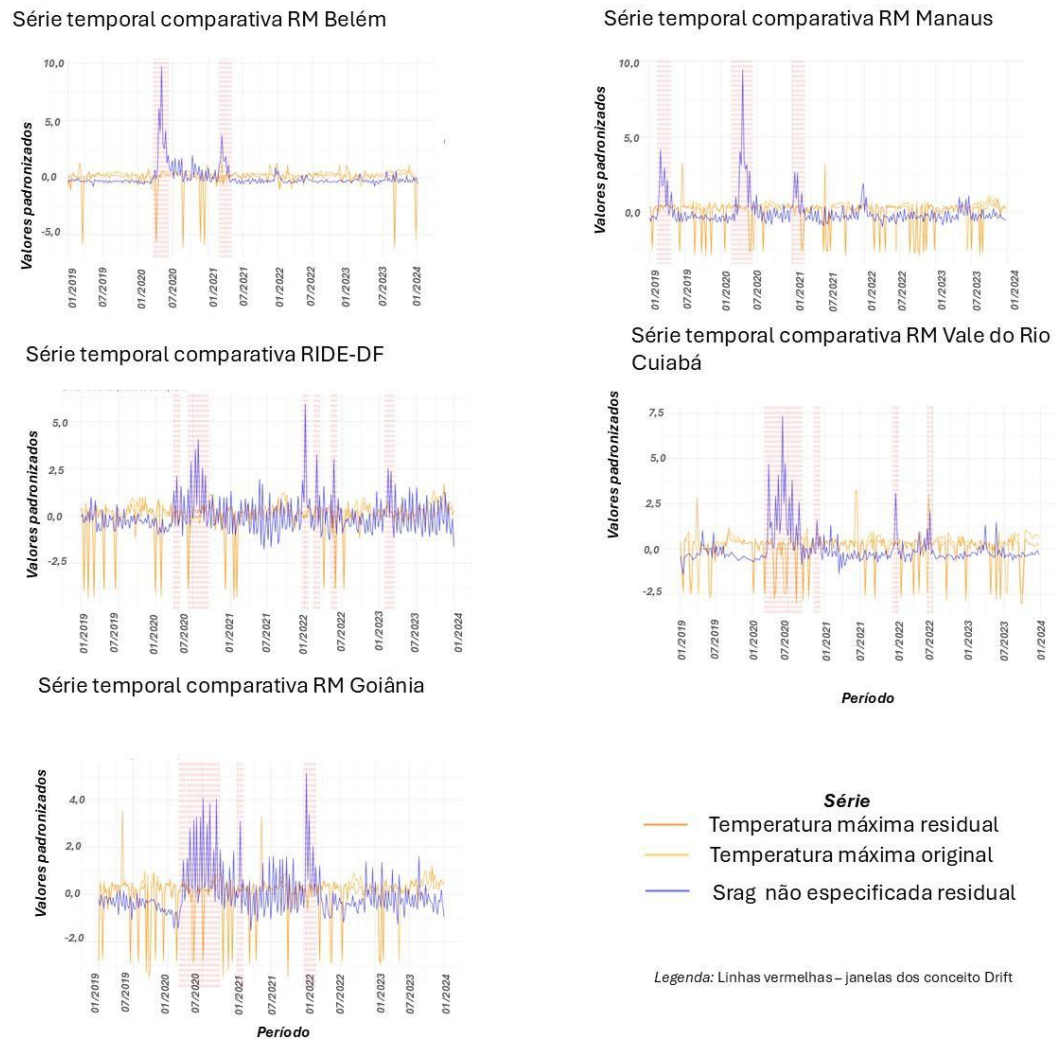
Elaboração própria.

Nas séries comparativas padronizadas, linhas verticais vermelhas identificam janelas temporais associadas aos episódios de drift (Figura 6). Em diversos casos, esses segmentos coincidem ou ocorrem próximos a períodos de flutuação acentuada na temperatura máxima residual, especialmente para a RM de Manaus, RIDE-DF e RM de Goiânia (Figura 6). Tal sinal sugere possível papel modulador de anomalias térmicas extremas no comportamento inesperado dos casos residuais de SRAG não especificado.

Contudo, observa-se que nem todos os eventos de drift acompanham variações térmicas relevantes, e vice-versa, reforçando que o fenômeno é multifatorial. Fatores epidemiológicos sociais e ambientais atuam conjuntamente, e a temperatura representa apenas um dos determinantes potencialmente envolvidos.

A coincidência temporal seletiva entre drifts e oscilações térmicas reforça a hipótese exploratória de que extremos climáticos podem amplificar a instabilidade epidemiológica em determinados cenários urbanos.

Figura 6. Comparação das séries padronizadas de SRAG residual e temperatura máxima (original e residual) com janelas de concept drift por RM entre 2019 e 2023.

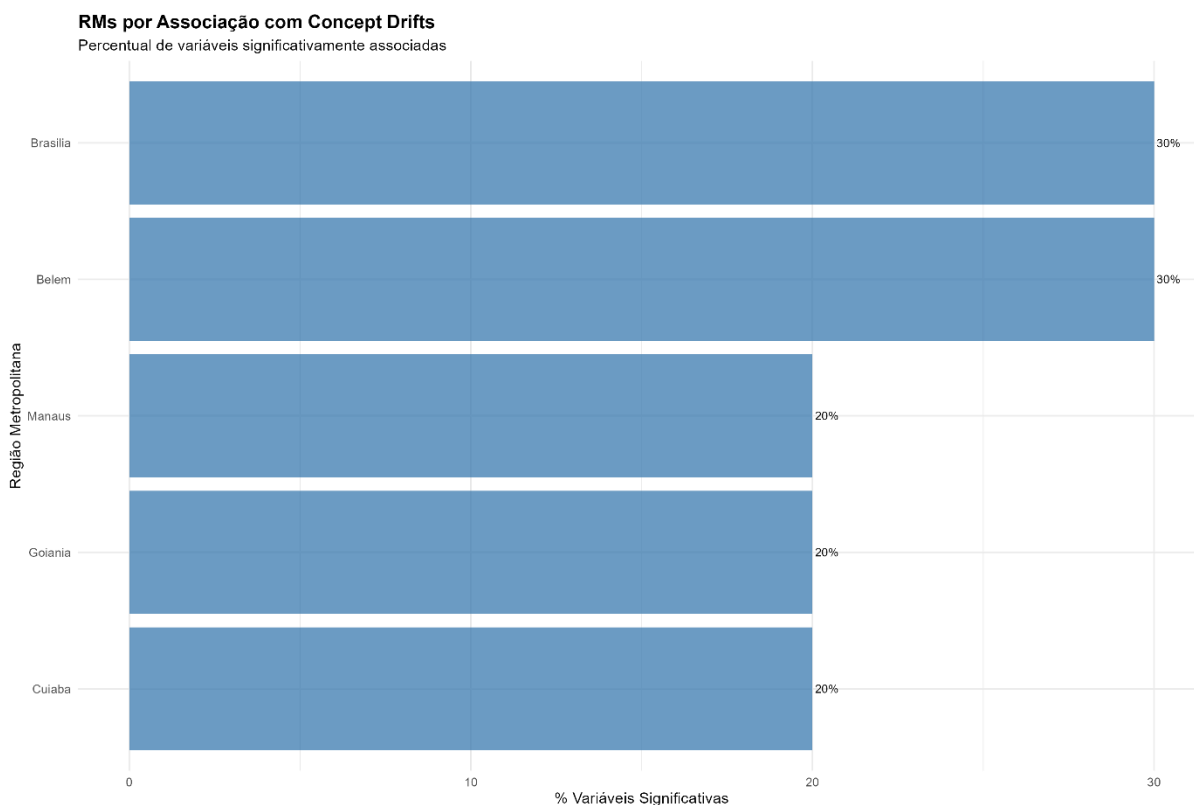


Elaboração própria.

4.4 Testes de associações

O gráfico a seguir apresenta a distribuição do número de observações analisadas por Região Metropolitana, assegurando comparabilidade entre as áreas estudadas. Observa-se que as cinco Regiões Metropolitanas apresentaram tamanhos amostrais elevados, com maior disponibilidade de observações em Brasília (RIDE-DF) e Belém, seguidas por Manaus, Goiânia e Cuiabá (Vale do Rio Cuiabá) (Figura 7). A distribuição equilibrada de semanas analisadas reforça a robustez da avaliação das séries temporais ao longo do período estudado.

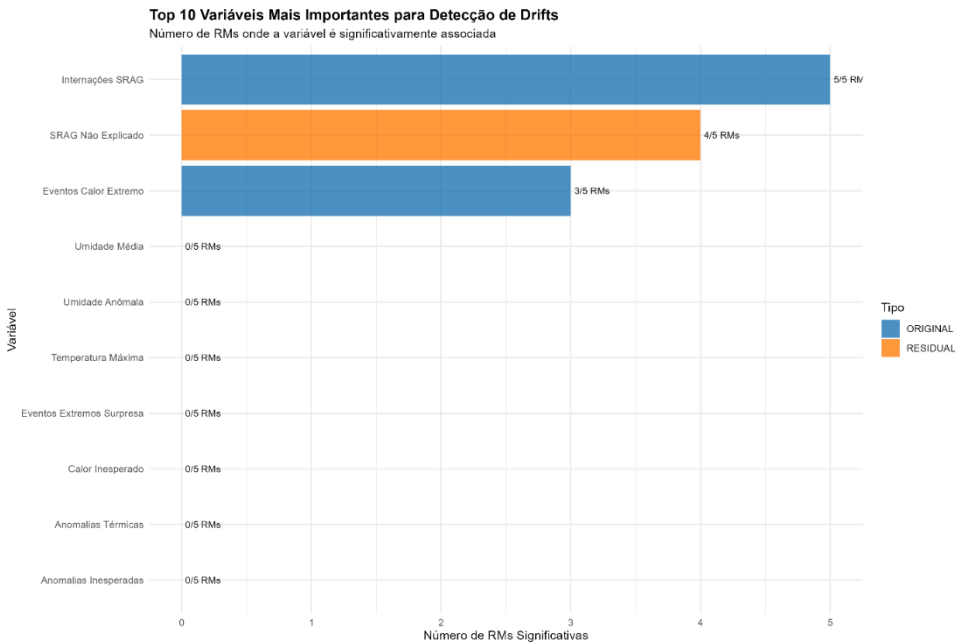
Figura 7. Percentual de Variáveis Ambientais Significativamente Associadas ao Concept Drift por Região Metropolitana.



O gráfico de frequência de associações significativas evidência que as interações por SRAG não especificado foram o indicador mais sensível para caracterizar períodos de concept drift em todas as regiões, seguidas pelos resíduos da série de SRAG não especificado, indicando instabilidade epidemiológica não explicada por tendência e sazonalidade. Entre as variáveis ambientais, o EHF destacou-se como o marcador climático mais associado ao drift, sobretudo

na Amazônia (Figura 8), reforçando o papel das ondas de calor como gatilho de instabilidade respiratória em contextos tropicais.

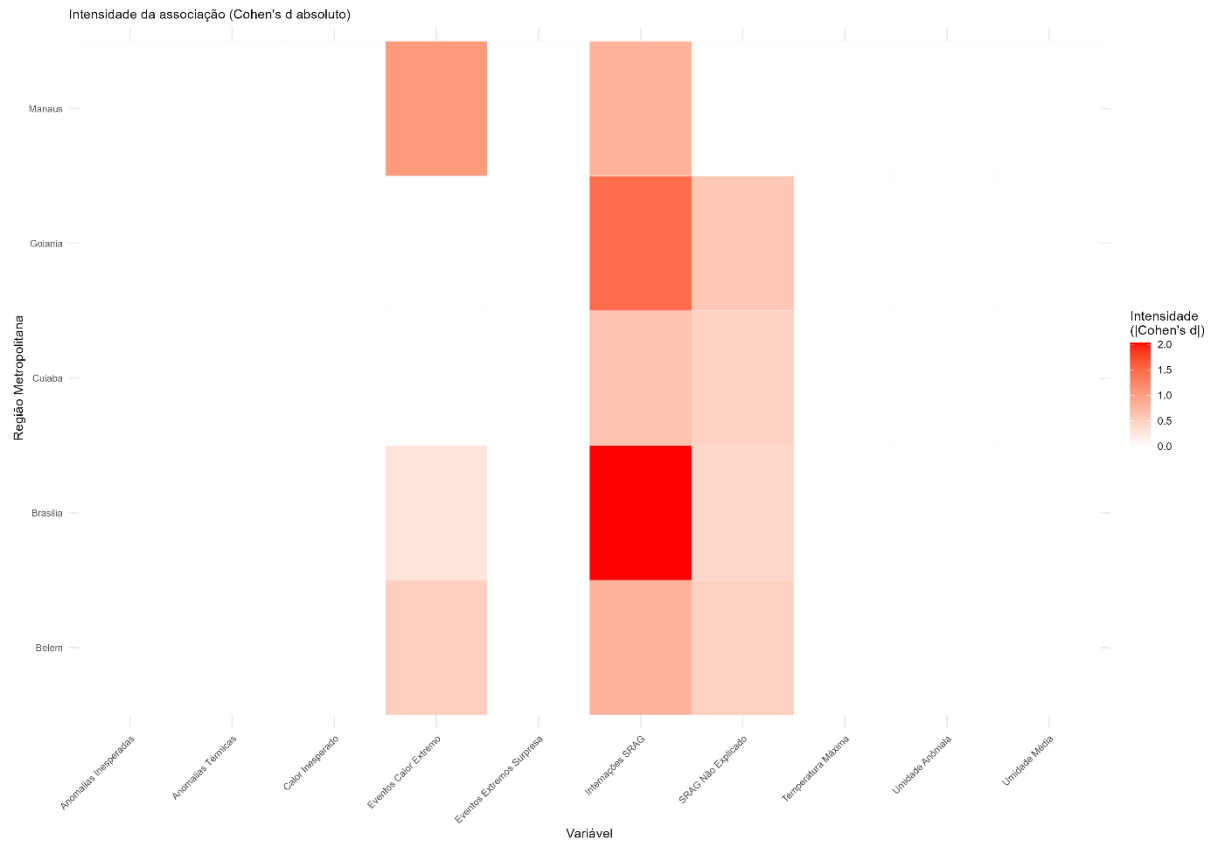
Figura 8. Matriz de frequência de associações significativas por variável.



Elaboração própria.

O heatmap (Figura 9) evidência maior frequência de associações significativas entre eventos de calor extremo (EHF) e períodos de concept drift nas regiões da Amazônia (Belém e Manaus). Nas regiões do Cerrado, a associação estatística foi menos consistente, com predominância de associação apenas com o aumento de interações por SRAG não específica. Esse padrão sugere maior sensibilidade climática respiratória no bioma Amazônico.

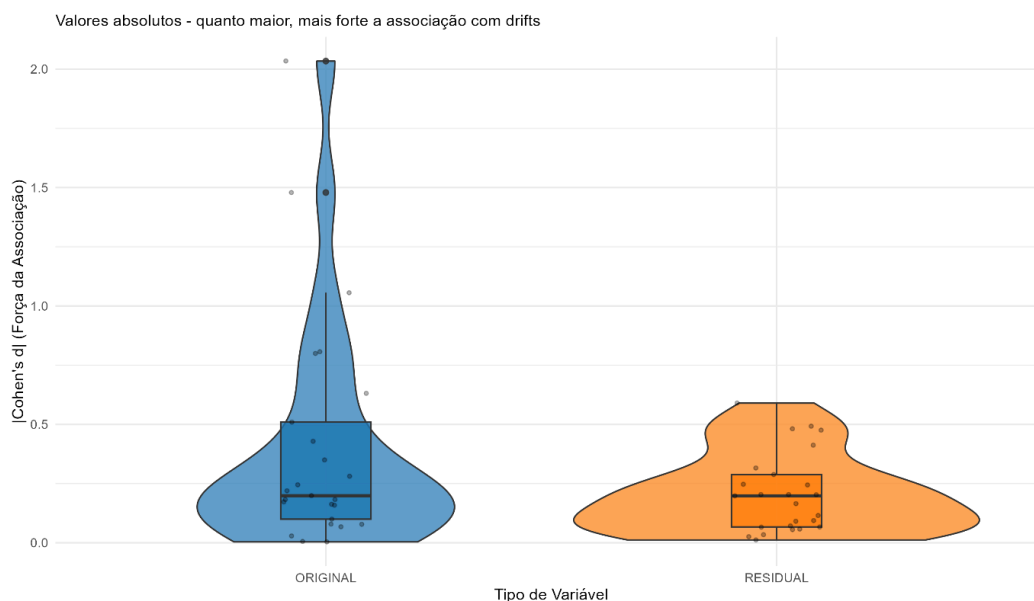
Figura 9 Hotmap: Associação entre variáveis e concept drifts por região Metropolitana.



Elaboração própria.

As distribuições de *Cohen's d* evidenciam maior magnitude de efeito para variáveis originais, com valores mais altos e ampla dispersão, incluindo episódios de forte impacto ($d > 1$). As variáveis residuais apresentaram distribuição mais concentrada e magnitudes mais baixas, sugerindo menor capacidade discriminatória sistemática, embora forneçam sinais adicionais em eventos específicos (Figura 10). Esses achados indicam que os indicadores brutos capturam mais intensamente as alterações estruturais associadas ao concept drift, enquanto o componente residual atua como marcador complementar de instabilidade de curta duração.

Figura 10. Distribuição da força de associação (Cohen's d) por tipo de variável.



Elaboração própria.

4.5. Modelos de Regressão Logísticas

Os forest plots (Figura 11) sintetizam os efeitos estimados pelos modelos de regressão logística binária aplicados para avaliar a associação entre variáveis ambientais e indicadores de saúde com a ocorrência de concept drift nas séries temporais de internações por SRAG não especificado nas regiões metropolitanas analisadas. Cada ponto representa a razão de chances (odds ratio, OR), enquanto as barras horizontais correspondem aos intervalos de confiança de 95% (IC95%). Valores de OR superiores a 1 indicam aumento na probabilidade de detecção de drift associado à unidade de incremento da variável, enquanto OR inferiores a 1 sugerem efeito protetor ou redução na probabilidade de drift. A significância estatística foi considerada quando o IC95% não incluiu o valor nulo (OR=1), denotando associação robusta.

De forma geral, observou-se que valores mais elevados de internações por SRAG não especificados estiveram consistentemente associados a maiores chances de ocorrência de drifts nas RMs de Belém, Brasília, Goiânia e Manaus, com OR variando entre 1,08 e 1,23 e IC95% sem sobreposição com a unidade.

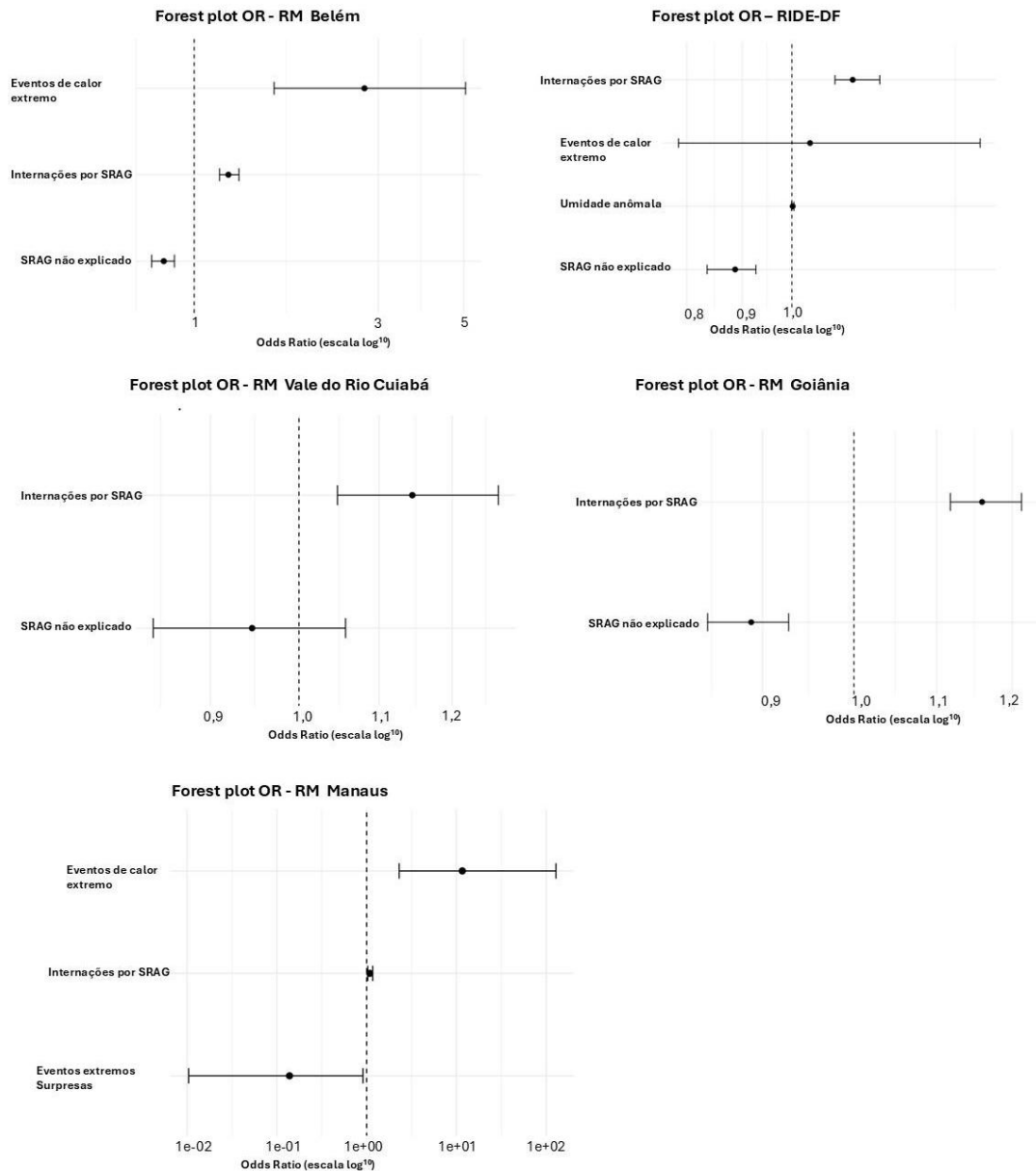
No contexto climático, eventos extremos de calor apresentaram associação significativa para a RM de Belém (OR=2,76; IC95% 1,61–5,05) e, especialmente, para a RM de Manaus (OR=11,64; IC95% 2,30–128,46), indicando que períodos com calor extremo aumentaram substancialmente a chances de concept drift nessas localidades.

Em contraposição, o componente residual de SRAG (SRAG não explicado), ao isolar variação não capturada pelas covariáveis ambientais, apresentou valores de OR abaixo de 1 em

todas as regiões, evidenciando possível tendência estabilizadora da série quando efeitos ambientais dominantes são removidos do sinal. Essa observação fortalece a hipótese de que parte relevante do drift decorre de flutuações ambientais capturadas no modelo, enquanto o componente residual representa ruído menos sistemático.

Os resultados destacam o papel da Amazônia, onde extremos térmicos apresentaram maior magnitude de efeito, sugerindo uma interação potencialmente mais sensível entre ambiente, vulnerabilidade populacional e demanda assistencial.

Figura 11. Razões de chances (OR) para ocorrência de concept drift segundo indicadores ambientais e de SRAG nas Regiões Metropolitanas selecionadas.



Elaboração própria.

A Tabela 1 apresenta as estimativas de razão de chances (OR) e intervalos de confiança de 95% para a associação entre variáveis climáticas e epidemiológicas e a ocorrência de concept drift. Os mesmos resultados são representados graficamente nos forest plots (Figura 11), permitindo visualização comparativa da magnitude e direção dos efeitos entre regiões metropolitanas.

Tabela 1. Estimativas de razão de chances (OR) para a ocorrência de concept drift segundo variáveis climáticas e epidemiológicas nas Regiões Metropolitanas analisadas.

M	R	Rotulo	OR (IC95%)	p-valor	Situação.
Belém		Eventos de Calor Extremo	2,76 (1,61–5,05)	0,0005	Sim
Belém		Internações por SRAG	1,23 (1,16–1,31)	0,0000	Sim
Belém		SRAG Não Explicado	0,83 (0,78–0,89)	0,0000	Sim
RIDE-DF		Eventos de Calor Extremo	1,04 (0,79–1,49)	0,8155	Não
RIDE-DF		Internações por SRAG	1,14 (1,10–1,20)	0,0000	Sim
RIDE-DF		Umidade Anômala	1,00 (1,00–1,00)	0,1996	Não
RIDE-DF		SRAG Não Explicado	0,89 (0,84–0,93)	0,0000	Sim
Vale do Rio Cuiabá		Internações por SRAG	1,14 (1,05–1,27)	0,0055	Sim
Vale do Rio Cuiabá		SRAG Não Explicado	0,95 (0,84–1,06)	0,3364	Não
Goiânia		Internações por SRAG	1,16 (1,12–1,21)	0,0000	Sim
Goiânia		SRAG Não Explicado	0,89 (0,84–0,93)	0,0000	Sim
Manaus		Eventos de Calor Extremo	11,64 (2,30–128,46)	0,0164	Sim
Manaus		Internações por SRAG	1,08 (1,03–1,17)	0,0116	Sim
Manaus		Eventos Extremos Surpresa	0,14 (0,01–0,91)	0,0825	Não

Elaboração própria

5 CONCLUSÃO

Com base na análise STL realizada, conclui-se que a série temporal de internações por SRAG nas Regiões Metropolitanas dos biomas Amazônia e Cerrado evidenciou uma transição epidemiológica caracterizada pela estabilização progressiva da tendência de longo prazo após um período de elevada intensidade, com reestabelecimento gradual do padrão sazonal respiratório característico de climas tropicais e subtropicais. A decomposição temporal identificou ainda a ocorrência de eventos anômalos significativos, particularmente nas regiões de Belém e Manaus, com resíduos de maior magnitude durante os picos epidêmicos, reforçando a utilidade da ferramenta para vigilância epidemiológica e detecção precoce de desvios com potencial impacto em saúde pública.

A análise fatorial e de agrupamentos revelou padrões epidemiológicos distintos para as SRAG, com Cuiabá como caso singular, Manaus e Belém com dinâmicas opostas na Amazônia, e Goiânia/Brasília (RIDE-DF) formando um perfil homogêneo. Evidenciando que a vigilância em saúde e ambiente deve incorporar essas especificidades regionais para o planejamento de ações eficazes.

De forma integrada, os resultados sugerem que a ocorrência de concept drift na série de internações por SRAG está associada a extremos térmicos, sobretudo nas regiões amazônicas, enquanto no Cerrado o fenômeno parece refletir predominantemente a dinâmica epidemiológica local. A combinação de testes estatísticos e visualizações reforça a validade da abordagem proposta para vigilância sindrômica sob condições de mudança climática,

A análise de concept drift revelou que as perturbações epidemiológicas em SRAG, mais intensas em 2020-2021, estabilizaram após 2022, indicando transição para uma fase endêmica. A coincidência seletiva entre esses drifts e anomalias térmicas em Manaus, RIDE-DF e Goiânia sugere que extremos climáticos podem modular a instabilidade respiratória. Contudo, a associação incompleta reforça a natureza multifatorial do fenômeno, envolvendo interações complexas entre fatores ambientais, sociais e imunológicos.

Os resultados indicam que o concept drift em SRAG possui dinâmica distinta por bioma: na Amazônia está fortemente associado a extremos térmicos (EHF), enquanto no Cerrado reflete principalmente a dinâmica epidemiológica local. Indicadores brutos mostraram maior sensibilidade que variáveis residuais na detecção dessas alterações, sustentando a utilidade da abordagem para vigilância sindrômica em contextos climáticos diversos.

Com base na análise dos forest plots, conclui-se que a ocorrência de concept drift nas séries de SRAG não especificada apresenta padrões regionais distintos. O aumento nas

internações por SRAG elevou consistentemente as chances de drift na maioria das regiões, enquanto os extremos de calor demonstraram efeito particularmente acentuado nas regiões amazônicas, com odds ratios de 2,76 em Belém e 11,64 em Manaus. Em contrapartida, o componente residual de SRAG apresentou efeito estabilizador em todas as regiões estudadas. Esses resultados confirmam que a instabilidade epidemiológica é modulada por condições específicas de cada região, com destaque para a maior vulnerabilidade das áreas amazônicas aos extremos térmicos.

Diante do exposto recomenda-se que sistemas de vigilância em saúde incorporem alertas baseados em extremos climáticos, particularmente para as regiões metropolitanas da Amazônia, e desenvolvam modelos preditivos que contemplem as heterogeneidades regionais identificadas.

Ressalta-se, contudo, que a análise esteve limitada à disponibilidade de dados secundários, sendo pertinente a inclusão de variáveis socioeconômicas e de capacidade assistencial em investigações futuras para aprimorar a compreensão dos mecanismos subjacentes aos drifts observados

6. REFERÊNCIAS

AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, v. 19, n. 6, p. 716-723, 1974.

BAI, J.; PERRON, P. Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, v. 18, n. 1, p. 1-22, 2003.

BRASIL. Lei nº 7.000, de 20 de dezembro de 1990. Dispõe sobre a proibição da pesca. *Diário Oficial da União, Brasília, DF*, 21 jan. 1991. Seção 1, p. 51.

BRASIL. Ministério da Saúde. Gabinete do Ministro. Portaria GM/MS nº 6.734, de 18 de março de 2025. Altera o Anexo 1 do Anexo V da Portaria de Consolidação GM/MS nº 4, de 28 de setembro de 2017. *Diário Oficial da União, Brasília, DF*, n. 61, p. 82, 31 mar. 2025. Disponível em: https://bvsms.saude.gov.br/bvs/saudelegis/gm/2025/prt6734_31_03_2025.html. Acesso em: 21 abr. 2025.

BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde e Ambiente. Guia de vigilância integrada da covid-19, influenza e outros vírus respiratórios de importância em saúde pública [recurso eletrônico]. Brasília: Ministério da Saúde, 2024. 146 p. Disponível em: <https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/svsa/covid-19/guia-vigilancia-integrada-da-covid-19>. Acesso em: 21 abr. 2025.

BROWN, R. L.; DURBIN, J.; EVANS, J. M. Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society, Series B (Methodological)*.

CHANDOLA, V.; BANERJEE, A.; KUMAR, V. Anomaly detection: A survey. *ACM Computing Surveys*, v. 41, n. 3, art. 15, p. 1–58, 2009. DOI: 10.1145/1541880.1541882.

CLEVELAND, R. B.; CLEVELAND, W. S.; MCRAE, J. E.; TERPENNING, I. STL: A Seasonal-Trend Decomposition Procedure Based on Loess. *Journal of Official Statistics*, v. 6, n. 1, p. 3-73, 1990.

DELFINO, R. J. et al. The relationship of respiratory and cardiovascular hospital admissions to the southern California wildfires of 2003. *Occupational and Environmental Medicine*, v. 65, n. 3, p. 189-197, 2008. DOI: 10.1136/oem.2008.041376.

DOBSON, A. J.; BARNETT, A. *An Introduction to Generalized Linear Models*. 4. ed. Boca Raton: CRC Press, 2018.

FITTER, D. L. et al. The impact of a global respiratory pandemic on circulation patterns of other respiratory viruses. *Journal of Infectious Diseases*, v. 224, p. S343–S352, 2021.

GOMES, M. F. C. InfoGripe: Monitoramento de casos de Síndrome Respiratória Aguda Grave (SRAG). Rio de Janeiro: Fiocruz, [2014]. Disponível em: <https://gitlab.fiocruz.br/marcelo.gomes/infogripe>. Acesso em: 21 abr. 2025.

HOSMER, D. W.; LEMESHOW, S.; STURDIVANT, R. X. *Applied Logistic Regression*. 3. ed. Hoboken: Wiley, 2013.

HYNDMAN, R. J.; ATHANASOPOULOS, G. *Forecasting: Principles and Practice*. 2. ed. Melbourne: OTexts, 2018.

IGNOTTI, E. et al. Poluição do ar e internações por doenças respiratórias em crianças e idosos em Mato Grosso, Brasil. *Revista de Saúde Pública*, v. 44, n. 1, p. 121-130, 2010.

JOLLIFFE, I. T. *Principal Component Analysis*. 2. ed. New York: Springer-Verlag, 2002.

KAUFMAN, L.; ROUSSEAU, P. J. *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken, NJ: John Wiley & Sons, 2009.

KOLMOGOROV, A. N. Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Istituto Italiano degli Attuari*, v. 4, p. 83–91, 1933.

MANN, H. B.; WHITNEY, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, v. 18, n. 1, p. 50-60, 1947.

MAPBIOMAS. Mapeamento das áreas queimadas no Brasil entre 1985 e 2023 - Coleção 3. Disponível em: <https://plataforma.brasil.mapbiomas.org/hotsite-fogo>. Acesso em: 21 abr. 2025.

MINISTÉRIO DA SAÚDE. Secretaria de Vigilância em Saúde. Boletim Epidemiológico nº 162 - COE Coronavírus. Vigilância da covid-19 no Brasil – abril 2024. Disponível em: <https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/boletins/epidemiologicos/covid-19/2024/boletim-epidemiologico-no-162-coe.pdf/view>. Acesso em: 10 jan. 2024.

SCHEMBER, E. et al. Reemergence of Influenza Virus Circulation After Relaxation of COVID-19 Mitigation Measures. *Morbidity and Mortality Weekly Report*, v. 71, n. 29, p. 917–922, 2022.

SMIRNOV, N. V. On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Bulletin Moscow University*, v. 2, p. 3–16, 1939.

SOUZA, T. C. M.; AMANCIO, F.; HACON, S. S.; BARCELOS, C. Doenças sensíveis ao clima no Brasil e no mundo: revisão sistemática. *Revista Panamericana de Salud Pública*, v. 42, e85, 2018.

TAMERIOUS, J. D. et al. Environmental Predictors of Seasonal Influenza Epidemics across Temperate and Tropical Climates. *PLoS Pathogens*, v. 9, n. 3, e1003194, 2013.

ZEILEIS, A. et al. strucchange: An R package for testing for structural change in linear regression models. *Journal of Statistical Software*, v. 7, n. 2, p. 1–38, 2002.