Anexo 2

UNISINOS

Programa de Pós-Graduação em

Computação Aplicada

Mestrado/Doutorado Acadêmico

Lucas Schroeder

Associação das queimadas com doeças respiratórias e complicações da COVID-19 no estado do Pará, Brasil.

São Leopoldo, 2023

UNIVERSIDADE DO VALE DO RIO DOS SINOS — UNISINOS
UNIDADE ACADÊMICA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA
NÍVEL MESTRADO

LUCAS SCHROEDER

ASSOCIAÇÃO DAS QUEIMADAS COM DOENÇAS RESPIRATÓRIAS E COMPLICAÇÕES DA COVID-19 NO ESTADO DO PARÁ, BRASIL

São Leopoldo
2023

Lucas Schroeder

ASSOCIAÇÃO DAS QUEIMADAS COM DOENÇAS RESPIRATÓRIAS E COMPLICAÇÕES DA COVID-19 NO ESTADO DO PARÁ, BRASIL

Proposta de dissertação apresentada para a obtenção do título de Mestre pelo Programa de Pós-Graduação em Computação Aplicada da Universidade do Vale do Rio dos Sinos — UNISINOS

Orientador:
Prof. Dr. Mauricio R. Veronez

Coorientador:
Prof. Dra. Clévia Rosset

São Leopoldo
2023

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)

Schroeder, Lucas

Associação das queimadas com doenças respiratórias e complicações da COVID-19 no Estado do Pará, Brasil / Lucas Schroeder — 2023.

69 f.: il.; 30 cm.

Dissertação (mestrado) — Universidade do Vale do Rio dos Sinos, Programa de Pós-Graduação em Computação Aplicada, São Leopoldo, 2023.

"Orientador: Prof. Dr. Mauricio R. Veronez, Unidade Acadêmica de Pesquisa e Pós-Graduação".

1. Incêndio. 2. Doenças Respiratórias. 3. K-means. 4. ARIMAX. 5. Análise de Séries Temporais. 6. SARS-Cov-2. 7. COVID-19. 8. ARIMAX. 9. Hospitalizações. 10. Taxa de Incidência. 11. Risco. I. Título.

CDU 004.732

Bibliotecária responsável: Fulana da Silva — CRB 12/3456

RESUMO

O Brasil tem enfrentado dois problemas simultâneos relacionados à saúde respiratória: os incêndios florestais e a alta taxa de mortalidade devido à pandemia de COVID-19. A floresta amazônica é um dos biomas brasileiros que mais sofre com incêndios causados pelo desmatamento ilegal. Esses incêndios podem trazer doenças respiratórias associadas à poluição do ar, sendo o Estado do Pará no Brasil o mais afetado. A pandemia de COVID-19 associada à poluição do ar pode aumentar potencialmente as hospitalizações e mortes relacionadas a doenças respiratórias. Aqui, objetivamos avaliar a associação das ocorrências de incêndio com as taxas de mortalidade por COVID-19 e internações por doenças respiratórias gerais no Estado do Pará, Brasil. Empregamos a técnica de aprendizado de máquina denominada *k-means* para agrupamento de focos de incêndios, acompanhada do método cotovelo utilizado para identificar a quantidade ideal de *clusters* para o algoritmo *k-means*. Agrupamos 10 grupos de cidades no Estado do Pará onde selecionamos os *clusters* com maior e menor ocorrência de incêndios de 2015 a 2019. Em seguida, foi proposto um modelo Auto-regressivo Integrado de Média Móvel Exógena (ARIMAX) para estudar a correlação serial de internações por doenças respiratórias e suas associações com ocorrências de incêndio. Em relação à análise do COVID-19, calculamos o risco de mortalidade e seu nível de confiança considerando a razão da taxa de incidência trimestral nos *clusters* com a alta e baixa exposição a incêndios. Adicionalmente, identificamos nos dois agrupamentos de cidades que o IDH (Índice de Desenvolvimento Humano) e PIB (Produto Interno Bruto) são semelhantes, porém possuem um comportamento diverso considerando as internações e incêndios florestais. A partir do modelo auto-regressivo e de média móvel (ARIMAX), foi possível mostrar que além da correlação da série temporal, as ocorrências de incêndios contribuem para o aumento das doenças respiratórias, com uma defasagem observada de seis meses após os incêndios para o agrupamento com alta exposição aos incêndios. Um destaque que merece atenção diz respeito à relação entre ocorrências de incêndios e mortes. Historicamente, o risco de mortalidade por doenças respiratórias é maior (cerca do dobro) em regiões e períodos com alta exposição ao fogo do que naqueles com baixa exposição ao fogo. O mesmo padrão se mantém no período da pandemia de COVID-19, onde o risco de mortalidade por COVID-19 foi 80% maior na região e período com alta exposição a incêndios. O processo de tomada de decisão é um problema crítico principalmente quando envolve políticas de controle ambiental e de saúde. As políticas ambientais são muitas vezes mais rentáveis como medidas de saúde do que o uso de serviços de saúde pública. Isso destaca a importância da análise de dados para apoiar a tomada de decisão e identificar a população que necessita de melhor infraestrutura devido a fatores ambientais históricos e ao conhecimento do risco à saúde associado. Os resultados sugerem que as ocorrências de incêndios contribuem para o aumento das internações por doenças respiratórias. A taxa de mortalidade relacionada ao COVID-19 foi maior no período com alta exposição a incêndios do que no período com baixa exposição a incêndios. As regiões com alta ocorrência de incêndios estão associadas a mais mortes por COVID-19, principalmente nos meses com maior número de incêndios.

Palavras-chave: Incêndio. Doenças Respiratórias. K-means. ARIMAX. Análise de Séries Temporais. SARS-Cov-2. COVID-19. ARIMAX. Hospitalizações. Taxa de Incidência. Risco.

ABSTRACT

Brazil has faced two simultaneous problems related to respiratory health: forest fires and the high mortality rate due to COVID-19 pandemics. The Amazon rain forest is one of the Brazilian biomes that suffers the most with fires caused by droughts and illegal deforestation. These fires can bring respiratory diseases associated with air pollution, and the State of Pará in Brazil is the most affected. COVID-19 pandemics associated with air pollution can potentially increase hospitalizations and deaths related to respiratory diseases. Here, we aimed to evaluate the association of fire occurrences with the COVID-19 mortality rates and general respiratory diseases hospitalizations in the State of Pará, Brazil. We employed machine learning technique for clustering k-means accompanied with the elbow method used to identify the ideal quantity of clusters for the k-means algorithm, clustering 10 groups of cities in the State of Pará where we selected the clusters with the highest and lowest fires occurrence from the 2015 to 2019. Next, an Auto-regressive Integrated Moving Average Exogenous (ARIMAX) model was proposed to study the serial correlation of respiratory diseases hospitalizations and their associations with fire occurrences. Regarding the COVID-19 analysis, we computed the mortality risk and its confidence level considering the quarterly incidence rate ratio in clusters with high and low exposure to fires. Using the k-means algorithm we identified two clusters with similar DHI (Development Human Index) and GDP (Gross Domestic Product) from a group of ten clusters that divided the State of Pará but with diverse behavior considering the hospitalizations and forest fires in the Amazon biome. From the auto-regressive and moving average model (ARIMAX), it was possible to show that besides the serial correlation, the fires occurrences contribute to the respiratory diseases increase, with an observed lag of six months after the fires for the case with high exposure to fires. A highlight that deserves attention concerns the relationship between fire occurrences and deaths. Historically, the risk of mortality by respiratory diseases is higher (about the double) in regions and periods with high exposure to fires than the ones with low exposure to fires. The same pattern remains in the period of the COVID-19 pandemic, where the risk of mortality for COVID-19 was 80% higher in the region and period with high exposure to fires. Regarding the SARS-COV-2 analysis, the risk of mortality related to COVID-19 is higher in the period with high exposure to fires than in the period with low exposure to fires. Another highlight concerns the relationship between fire occurrences and COVID-19 deaths. The results show that regions with high fire occurrences are associated with more cases of COVID deaths. The decision-make process is a critical problem mainly when it involves environmental and health control policies. Environmental policies are often more cost-effective as health measures than the use of public health services. This highlight the importance of data analyses to support the decision making and to identify population in need of better infrastructure due to historical environmental factors and the knowledge of associated health risk. The results suggest that the fires occurrences contribute to the increase of the respiratory diseases hospitalization. The mortality rate related to COVID-19 was higher for the period with high exposure to fires than the period with low exposure to fires. The regions with high fire occurrences is associated with more COVID-19 deaths, mainly in the months with high number of fires.

Keywords: Fire. Respiratory Diseases. K-means. ARIMAX. Time Series Analysis. SARS-Cov-2. COVID-19. Hospitalizations. Mortality. Incidence Rate Ratio. Risk.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE SIGLAS

INPE	Instituto Nacional de Pesquisas Espaciais
MODIS	*Moderate Resolution Imaging Spectroradiometer*
PM	Material Particulado
CID-10	Cadastro Internacional de Doenças (Edição 10)
KDD	*Knowledge Discovery of Databases*
HDR	Hospitalizações por Doenças Respiratórias
ARIMAX	*Autoregressive Integrated Moving Average with Explanatory Variable*
COVID-19	*(co)rona (vi)rus (d)isease*
OMS	Organização Mundial da Saúde
IC	Intervalo de Confiança
IBGE	Instituto Brasileiro de Geografia e Estatística
IDHM	Índice de Desenvolvimento Humano Municipal
PIB	Produto Interno Bruto
SUS	Sistema Único de Saúde
PNUD	Programa das Nações Unidas para o Desenvolvimento
NEO	NASA *Earth Observations*
WCSS	*Within Cluster Sum of Squares*
RTM	*Razão da Taxa de Mortalidade*
DP	*Desvio Padrão*
C1	*Cluster 1*
C4	*Cluster 4*

SUMÁRIO

1 INTRODUÇÃO

No Brasil, as mudanças nas práticas agrícolas e no uso da terra nos últimos anos afetaram diretamente o desenvolvimento econômico e social, especialmente na região que compreende o bioma Amazônia (CARRERO et al., 2020). Uma dessas práticas é a limpeza do solo com a utilização do fogo controlado(BRANDO et al., 2020). No entanto, seu uso indiscriminado aliado a políticas flexibilizadas elevou consideravelmente o número de incêndios na floresta amazônica, contribuindo para sua destruição acelerada (ESCOBAR, 2019; SMITH et al., 2014a). O número de incêndios entre janeiro e agosto (2020) foi 39% superior à média de incêndios dos últimos dez anos no mesmo período no Brasil (LIBONATI et al., 2021; RIBEIRO et al., 2018; CANO-CRESPO; TRAXL; THONICKE, 2021).

O Estado do Pará, no Brasil, que possui 24% da floresta amazônica acumulou 17% (174,903) do número total de incêndios (939,015) de 2015 a 2019, conforme registrado pelo Instituto Nacional de Pesquisas Espaciais (INPE, 2020). Essa destruição da floresta amazônica afeta diretamente o equilíbrio climático global. Além dos impactos globais, estudos anteriores demonstraram que a poluição do ar por incêndios na Amazônia é prejudicial à saúde humana local, independente de sua motivação ou causa (BUTT et al., 2021). Por exemplo, estudos em comunidades indígenas locais demonstraram que a fumaça de incêndios é uma das principais causas de hospitalizações respiratórias nessa população (MACHADO-SILVA et al., 2020; ALVES, 2020).

Em relação às doenças respiratórias, em dezembro de 2019 houve uma rápida disseminação mundial do beta-coronavírus 2019-nCOV, encontrado primeiro na cidade de Wuhan, na província de Hubei, na China, posteriormente identificado como SARS-COV-2 (síndrome respiratória aguda grave coronavírus 2) causando a doença denominada COVID-19 (PIERCE et al., 2020). A sintomatologia da COVID-19 envolve principalmente o sistema respiratório, variando de um quadro febril com sintomas respiratórios leves em alguns pacientes a pneumonia e sintomas mais graves em outros pacientes (XU et al., 2020). A rápida disseminação do COVID-19 em áreas brasileiras pode representar um desafio maior para as populações expostas à fumaça dos incêndios florestais. De fato, durante o surto respiratório agudo grave associado ao coronavírus 1 da síndrome respiratória aguda grave (SARS-Cov-1) em 2003, pacientes de áreas com altos níveis de poluição do ar exibiram um aumento de 200% no risco relativo de morte em comparação com as pessoas vivendo em áreas com baixo teor de poluição (A. Karan and K. Ali and S. Teelucksingh and S. Sakhamuri, 2020).

Com mais de 13 milhões de pessoas infectadas e 374 mil mortes até abril de 2021, o Brasil tem sofrido muito com essa doença. Só o Estado do Pará tem mais de 450 mil casos de COVID-19 e 11.000 óbitos, atingindo uma taxa de mortalidade de 138,3 óbitos por 100 mil habitantes (plataforma OPENDATASUS do Ministério da Saúde do Brasil) (NOTIFICAçõES DE SíNDROME GRIPAL , OPENDATASUS). Além disso, esse surto pode ser potencializado pela má qualidade do ar causada pela poluição do ar da cidade e incêndios florestais descontrolados (ZHOU et al., 2021). Porém, uma melhor análise dessa correlação ainda é um desafio devido

à grande quantidade de dados. Para superar esse desafio, abordagens de inteligência artificial como o *Knowledge Discovery of Databases* (KDD) (FAYYAD; HAUSSLER; STOLORZ, 1996; GRADY, 2016), poderia ser aplicado a este tipo de dados que requer 4Vs (volume, variedade, velocidade e veracidade) (CHANG; GRADY et al., 2015).

No entanto, o grande desafio é como compilar esses dados massivos de forma eficiente pois o volume, variedade e velocidade de dados digitais sobrecarregam facilmente as técnicas de análises convencionais. Neste caso, levanta-se a necessidade de abordagens específicas de inteligência artificial para manipular e garantir confiabilidade/veracidade em relação a tal quantidade de dados, como a Descoberta de Conhecimento em Bancos de Dados (KDD) (HAN J.; MILER, 2001). Um dos passos no KDD se refere aos métodos estatísticos de mineração de dados para extração rápida de informações úteis de dados espaciais/geográficos e temporais massivos e complexos (GUO D; MENNIS, 2009).

Assim, para lidar com 4 Vs (volume, variedade, velocidade e veracidade) associados a esses dados, propomos neste trabalho a utilização de um algoritmo relacionado a KDD denominado *k-means*, e uma análise de séries temporais (*Autoregressive Integrated Moving Average with Explanatory Variable* - ARIMAX) para identificar e analisar aglomerados de cidades mais e menos afetadas pelos incêndios florestais na Amazônia, e a relação entre esses cenários e as Hospitalizações por Doenças Respiratórias (HDR), e a taxa de mortalidade por SARS-COV-2 considerando o Estado do Pará no Brasil.

Diante disso, a motivação e a principal contribuição desta pesquisa é apresentar uma alternativa metodológica para o manuseio de uma base de dados de queimadas de grande proporção, objetivando identificar grupos de cidades distintas baseado no algoritmo K-means e posteriormente evidenciar qual o impacto destes aglomerados de cidades com a saúde respiratória da população residente.

Essa estratégia pode contribuir para a identificação de populações mais propensas a desenvolver complicações respiratórias durante as pandemias devido aos riscos ambientais, orientando a gestão da saúde pública, pois as melhorias ambientais costumam ser mais custo-efetivas como medidas de saúde do que o uso de serviços públicos de saúde (XU et al., 2020).

A hipótese é que a utilização de técnicas de agrupamentos de dados para identificação de cidades mais vulneráveis devido a fatores ambientais prévios, permitirá mapear regiões mais suscetíveis ao risco a mortalidade, principalmente no surgimento de pandemias.

Dada a hipótese, este trabalho tem como objetivo geral:

- Utilizar o algoritmo *k-means*, e uma análise de séries temporais (*Autoregressive Integrated Moving Average with Explanatory Variable* - ARIMAX) para identificar e analisar aglomerados de cidades mais e menos afetadas pelos incêndios florestais na Amazônia e a relação entre esses cenários e as Hospitalizações por Doenças Respiratórias e mortes por COVID-19.

Para o cumprimento do objetivo geral, foram estabelecidos os seguintes objetivos específicos:

- Identificar os grupos de cidades do estado do Pará com maior e menor incidência de queimadas com o algoritmo *K-means*;
- Verificar se há associação entre o número de queimadas com o aumento nos números de hospitalizações por doenças respiratórias através do modelo estatístico ARIMAX;
- Evidenciar se a taxa de mortalidade por SARS-Cov-2 teve maiores índices no grupo de cidades com maiores vulnerabilidades respiratórias decorrentes da poluição do ar (queimadas).

Esta proposta apresenta: uma revisão da literatura com conceitos sobre queimadas, doenças respiratórias, poluentes atmosféricos, pandemia associada ao SARS-COV-2 e descoberta de conhecimento em bases de dados conforme (Capítulo 2); trabalhos publicados alinhados aos temas abordados no trabalho (Capítulo 3); a descrição da metodologia aplicada no estudo (Capítulo 4); resultados encontrados para esta proposta (Capítulo 5); e discussão e considerações finais (Capítulo 6).

2 REFERENCIAL TEÓRICO

Este capítulo é destinado à apresentação de conceitos que serão utilizados no decorrer do trabalho. Nesta proposta será apresentada: uma revisão de literatura com conceitos sobre as queimadas e formas de aquisição dos dados, impactos das queimadas na saúde respiratória, Descoberta de Conhecimento em bases de dados e o surgimento da pandemia associada ao SARS-Cov-2.

2.1 Focos de Incêndios - INPE Queimadas

O fogo é um dos componentes essenciais para o desenvolvimento e manutenção de alguns recursos econômicos. No entanto, os humanos têm feito uso extensivo do fogo para limpar florestas, preparar e manter a terra para a agricultura (WERF et al., 2010; MORTON et al., 2008). Entre 2003 e 2012, cerca de 67 milhões de hectares de floresta foram queimados, principalmente na África e América do Sul (LIEROP et al., 2015).

Segundo (PIVELLO et al., 2021) em 2019 e 2020, as queimadas em diferentes biomas brasileiros receberam muita atenção na mídia e no debate público internacionalmente. Os incêndios na Amazônia simbolizam o imenso problema do desmatamento. No bioma Pantanal, as queimadas em 2020 marcaram o recorde de queima da maior área registrada nos últimos 20 anos: quase 30% da área do bioma estava em chamas. Os outros biomas tiveram anos com mais incêndios no passado, mas as duas últimas temporadas de incêndios também foram indicativas dos imensos desafios que o Brasil enfrenta em relação à conservação de seus ecossistemas naturais.

Os biomas e ecossistemas do Brasil diferem em sua resposta e vulnerabilidade ao fogo. Os regimes naturais de fogo foram modificados por atividades humanas, geralmente relacionadas a práticas de uso da terra ou devido a extremos climáticos ligados ao aquecimento global e às mudanças climáticas. Por Exemplo, em um estudo elaborado por (PACIFICO et al., 2015), observou-se que diferentes tipos de fogo têm diferentes causas. Embora a governança fraca possa levar a mais incêndios de desmatamento, as mudanças climáticas tornam as florestas mais quentes e secas, portanto, mais propensas a sustentar incêndios descontrolados (BRANDO et al., 2019).

Diferentes tipos de fogo também têm impactos diferentes. Por exemplo, incêndios descontrolados em áreas abertas podem matar o gado e destruir plantações e infraestrutura agrícola, enquanto mesmo incêndios florestais de baixa intensidade podem matar até 50% das árvores e reduzir o valor das florestas para a população local. Em contraste, as queimadas podem ser essenciais para a segurança alimentar e a subsistência de algumas das pessoas mais pobres da Amazônia. A falta de distinção entre os diferentes tipos de incêndio contribuiu para a incerteza em torno dos recentes incêndios na Amazônia e tem implicações importantes para as respostas políticas (BARLOW et al., 2012).

A prática de queimadas e desmatamento causam distúrbios no ecossistema, o que afeta tam-

bém a saúde humana. No Brasil, por exemplo, houve um aumento dos impactos do fogo no sistema respiratório humano devido ao aumento das emissões de aerossóis com degradação da qualidade do ar. Isso é mais evidente em crianças menores de cinco anos em municípios altamente expostos à seca (SMITH et al., 2014b). Como apontam os mesmos autores, o aerossol foi o principal fator de internações em municípios afetados pela seca durante 2005 na Amazônia brasileira. Nesse caso o Instituto Nacional de Pesquisas Espaciais (INPE) do Brasil vem monitorando incêndios e desmatamentos na Amazônia via satélite desde 1988. Esses dados têm sido importantes para medir a magnitude e distribuição espacial da degradação florestal e sua relação com doenças (FOLEY et al., 2007; NEPSTAD et al., 1999; PEREIRA et al., 2012).

O "Programa Queimadas"desenvolvido pelo INPE realiza e investe em pesquisa, desenvolvimento tecnológico e na inovação de produtos, processos e georreferenciamento para o monitoramento da propagação dos focos de incêndios ativos, com seu risco de gravidade e abrangência através da utilização de técnicas de Sensoriamento Remoto, Geoprocessamento e Modelagem Numérica. Os dados na plataforma são dispostos para a América, África e Europa, são atualizados diariamente, sendo o acesso às informações livres e gratuitos através de gráficos e tabelas. Os dados são obtidos com a utilização de sensoriamento remoto (imagens de satélite) para distinguir entre vegetação verde e superfícies queimadas (LIMONATI et al., 2011). De acordo com Limonati et al. (2015), o produto do satélite referência faz uso de observações de fogo ativo de vários sensores e depende de um Índice de Vegetação sensível à queima. Quanto a localização dos incêndios diários, a informação é indicada pelo sensor *Moderate Resolution Imaging Spectroradiometer* (MODIS), onde os focos são representados por 1 pixel, com resolução espacial de 1 km (PEREIRA et al., 2012).

A base de dados disponibilizada pelo INPE contém informações sobre as coordenadas geográficas dos focos de incêndio, bioma, estimativas de concentração de fumaça no ar, risco de incêndio, precipitação prevista, número de dias sem chuva, data, tempo, e cidade do foco observado. O monitoramento de queimadas e incêndios florestais em imagens de satélites se torna particularmente útil para toda abrangência territorial, condição esta que é capaz de refletir com precisão a situação geral do País. A detecção dos focos de queima de vegetação é padronizada em todas regiões, diariamente e com o passar dos anos, permitindo que pesquisadores elaborem análises temporais e espaciais da ocorrência do fogo possibilitando a comparação entre diferentes locais (INPE, 2020).

2.2 Doenças Respiratórias e Poluentes Atmosféricos

O declínio da precipitação e as queimadas na região amazônica associadas às mudanças climáticas e ao desmatamento expõem as comunidades locais a uma qualidade do ar perigosa que pode levar a danos à saúde humana, como doenças do sistema respiratório (MACHADO-SILVA et al., 2020). Os incêndios na vegetação nos trópicos emitem partículas finas de material Particulado (PM 2,5) para a atmosfera, degradando a qualidade do ar regional e impactando a

saúde humana (BUTT et al., 2020).

A Amazônia, localizada na região norte do Brasil, exibe um forte ciclo sazonal de incêndios na vegetação e, consequentemente, concentrações de Material Particulado (PM) elevados (MARTIN et al., 2010). Os estados que compõem a região norte do Brasil, possuem dois períodos distintos sendo eles a estação chuvosa, quando há poucas observações de fogo, as concentrações de PM podem ser tão baixas quanto 1,5 *micron*, e o período de estação seca (agosto a outubro), quando há um grande número de incêndios e as concentrações médias de PM 2,5 da estação seca regional podem exceder 30 *microns* (ARTAXO et al., 2013).

Em relação às doenças respiratórias, de acordo com as organizações oficiais de proteção à saúde, as emissões de incêndios florestais podem ter implicações agudas ou de longo prazo na saúde das populações expostas (YOUSSOUF et al., 2014). A queima de biomassa produz poluentes gasosos e emissões de partículas finas, que produzem efeitos prejudiciais ao sistema respiratório (CARMO et al., 2010). Entre os principais componentes da fumaça dos incêndios florestais que podem afetar a qualidade do ar ambiente, é importante mencionar as partículas finas PM2.5 e PM10, cujos efeitos devem aumentar ainda mais quando suas concentrações ficam acima dos padrões de qualidade do ar (YOUSSOUF et al., 2014).

Segundo J.A. Foley G.P. Asner (2007), a emissão de gases de efeito estufa já afetam a medicina respiratória por meio de: aumento do número de mortes e morbidade aguda devido a ondas de calor, aumento da frequência de eventos cardiorrespiratórios devido às maiores concentrações de ozônio troposférico; mudanças na frequência de doenças respiratórias devido à poluição do ar e a distribuição espacial e temporal alterada de vetores de doenças infecciosas. Esses impactos não afetarão apenas aqueles com doenças respiratórias existentes, mas também podem influenciar a incidência e, portanto, a prevalência de doenças respiratórias.

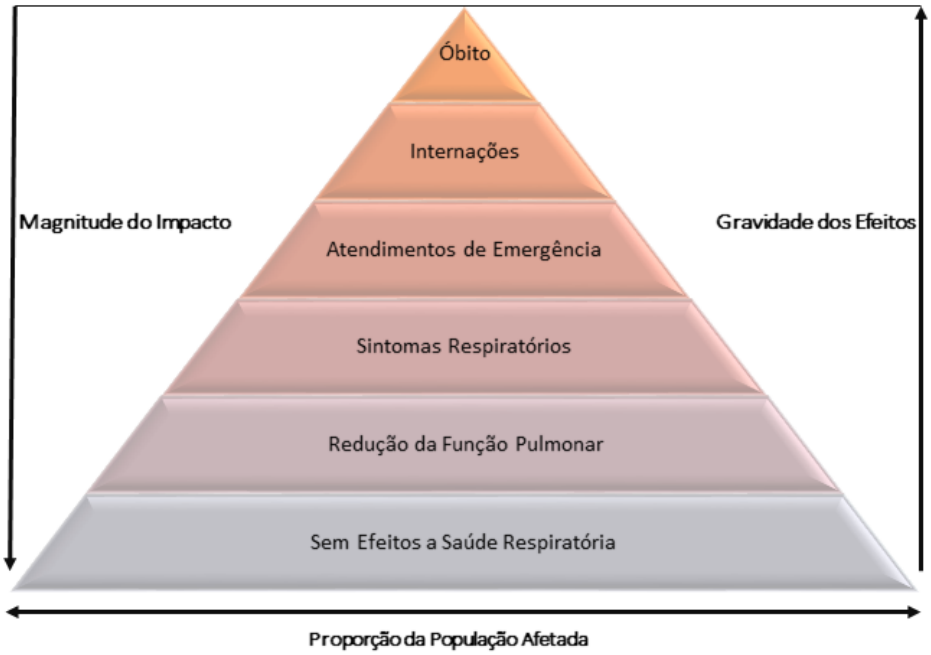
Em um estudo desenvolvido por Groot et al. (2019), a exposição ocupacional a incêndios florestais afeta a função pulmonar a curto prazo e pode aumentar o risco de hipertensão em longo prazo. A exposição queimadas também está associada a sintomas de estresse pós-traumático. O autor relata que mais pesquisas são necessárias para entender se a exposição ocupacional aos incêndios florestais resultam em impactos clinicamente significativos na função respiratória e para esclarecer melhor a relação entre a exposição ocupacional e pressão arterial, saúde mental e resultados de câncer.

No trabalho desenvolvido por Smith et al. (2014b), foi possível observar um aumento significativo (1,2%–267%) nas internações por doenças respiratórias em crianças menores de cinco anos em municípios altamente expostos à seca, sendo o aerossol o principal fator de hospitalizações em municípios afetados pela seca durante 2005, enquanto as condições de desenvolvimento humano atenuaram os impactos em 2010. Já Machado-Silva et al. (2020), demostrou um aumento de 27% nas hospitalizações por doenças do sistema respiratório em anos de seca. As chuvas e a umidade exercem um controle primário nas doenças do aparelho respiratório na região. Maiores índices de queimadas e fumaça concentram-se em agosto e setembro, e têm uma influência secundária do fogo nas hospitalizações por doenças do sistema respiratório. A

produção de fumaça está fortemente relacionada ao fogo e à temperatura, que estão associados ao aumento do número de internações durante a temporada de incêndios. Internações em crianças de até 5 anos e idosos acima de 60 anos são respectivamente 11 e 22 vezes maiores em relação as outras idades, confirmando a maior vulnerabilidade para crianças e idosos.

A Organização Mundial de Saúde (OMS, 1998) elaborou um documento relacionados a incêndios florestais, onde destaca a saúde como dependente de um ambiente saudável, evidenciando a necessidade em direcionar o controle de queimadas a um contexto global de mudanças. Abaixo a Figura 1 ilustra as principais situações relacionadas às queimadas e à parcela da população exposta. Em linhas gerais quanto mais sério o desfecho menor será a população atingida.

Figura 1 – Magnitude e a gravidade da saúde respiratória e efeitos da exposição aos poluentes atmosféricos.



Fonte: Adaptado de (OMS, 1998).

Conforme nota técnica da FIOCRUZ (2020), nos anos anteriores a 2020, foram realizados aproximadamente 8 mil internações mensais na região norte do Brasil. Deste montante, entre 8% e 13% representam internações associadas a doenças do sistema respiratório. já no ano de 2020, houve uma redução no número número total de internações na região, parcialmente explicado pelo atraso no envio de dados devido ao surgimento da Covid-19 (código B34.2 do capítulo I da CID10), doença também associada ao sistema respiratório.

Pesquisas relacionadas às queimadas devem considerar a influência de fatores exógenos abióticos que poderiam ter um ou mais efeitos diretos e indiretos com consequência aos ecossistemas, alterando o equilíbrio saúde/doença na região afetada. Ou seja, inúmeras variáveis que podem influenciar a ocorrência de agravos a saúde e separá-las para determinar o efeito isolado é bastante difícil, recomendando a existência de metodologias específicas para cada caso (ABATZOGLOU; WILLIAMS, 2016). Entretanto, mesmo com toda a literatura já disponível

sobre a relação saúde e poluentes atmosféricos em centros urbanos, poucos são os estudos que abordam os efeitos à saúde das populações expostas à fumaça das queimadas, principalmente na região Amazônica (OMS, 1998).

2.3 Pandemia Associada ao SARS-COV-2

A denominada COVID-19 é uma Síndrome Respiratória Aguda Grave (SARS), causada pelo SARS-CoV-2, um novo vírus que pertence à família Coronaviridae e foi relatado pela primeira vez em dezembro de 2019 na cidade de Wuhan, na China e logo depois, a doença se espalhou para todo o mundo. A partir de 23 de janeiro de 2020, a cidade de Wuhan foi bloqueada com todo o tráfego estritamente restrito e monitorado. Os cidadãos foram restringidos em casa para controlar a transmissão de humano para humano do SARS-CoV-2. As pessoas que estiveram em contato com pacientes com COVID-19 foram solicitadas a ficar em quarentena em casa ou foram levadas para instalações especiais de quarentena (WANG et al., 2020). Porém um dos maiores desafios do nosso tempo é a compreensão dos mecanismos de evolução e transmissão do SARS-CoV-2 (HE et al., 2020).

Em fevereiro de 2021, o SARS-CoV-2 infectou 112,20 milhões de pessoas e causou 2,49 milhões de mortes em todo o mundo. Embora a taxa de letalidade entre os pacientes SARS-CoV-2 seja menor (2,15%) do que seus parentes anteriores, SARS-CoV (9,5%) e MERS-CoV (34,4%), o SARS-CoV-2 tem sido observado como mais infeccioso e causou maior morbidade e mortalidade em todo o mundo (KUMAR et al., 2021). As pessoas infectadas com a COVID-19 apresentam sintomas principalmente associados ao sistema respiratório como: dificuldade para respirar, tosse dores de garganta, febre e alguns outros sintomas clínicos. Existe também portadores do vírus assintomáticos, os quais são de fundamental importância epidemiológica, pois se tornam potenciais transmissores (HE et al., 2020).

À medida que o SARS-CoV-2 continua sua rápida disseminação global, uma maior compreensão do nível de transmissão e a gravidade da infecção é crucial para orientar e planejar uma resposta à pandemia. Embora o teste de indivíduos com COVID-19 seja uma ferramenta essencial de saúde pública, a variabilidade nas capacidades de vigilância podem causar dificuldades na interpretação dos dados dos casos. Devido a relatórios mais completos, as mortes associadas ao COVID-19 são frequentemente vistas como um indicador mais confiável do tamanho da pandemia. Se relatado de forma confiável, o número de mortes associadas ao COVID-19 pode ser usado para inferir o número total de infecções por SARS-CoV-2 usando estimativas da taxa de letalidade (VERITY et al., 2020).

Segundo Guan et al. (2020), esta claro que a gravidade da infecção aumenta quando associada a idade populacional, porém, algumas questões importantes não são respondidas quanto a padrões diferentes de mortalidade entre os países. As heterogeneidades nas faixas etárias da população ou na prevalência de comorbidades podem contribuir para diferenças nos níveis de fatalidades observadas associadas a COVID-19. Além disso, ao analisar o número total de mor-

tes associadas ao COVID-19, pode ser difícil distinguir o nível de transmissão entre a população em geral de grandes surtos em populações vulneráveis, como as que vivem em lares de idosos e outros ambientes de cuidados de longo prazo. Sendo assim, concentrar-se nos dados de morte associados ao COVID-19 em indivíduos mais jovens, ou em toda a população, pode fornecer informações mais confiáveis sobre a natureza da transmissão e impactos da pandemia (CLARK et al., 2020).

O primeiro caso da COVID-19 no Brasil foi relatado em 26 de fevereiro de 2020 no estado de São Paulo. Já o primeiro óbito foi registrado na data de 17 de março e no més de abril o Brasil encontrava-se na décima colocação no ranking de pais com maior número de casos e óbitos (SANTOS et al., 2022). O Brasil é um pais com dimensões a nível continental e como consequência é um pais com muitas desigualdades econômicas, sociais e culturais internas, e o com o surgimento desta doença os impactos em diferentes regiões do pais podem ser controversos (ALVES et al., 2020).

Por exemplo, um estudo realizado no sul da amazônia na região noroeste do Brasil, elaborado por Andrade et al. (2021), foi possível identificar que o perfil predominante nas internações hospitalares e mortalidade por COVID-19, eram em homens com 60 anos, indígenas, pretos e pardos, com doenças e comorbidades pré existentes, sendo as mais prevalentes: Hipertensão arterial sistêmica, diabetes mellitus e obesidade. Já o estudo elaborado por Klokner et al. (2021), objetivou traçar o perfil epidemiológico, bem como analisar as variáveis preditivas de risco e proteção para COVID-19, na região sul do Brasil, nos estados do Rio Grande do Sul, Santa Catarina e Paraná. Os resultados deste estudo indicaram que a prevalência de óbitos provocados pela COVID-19 são maiores na faixa etária de 60 anos ou mais, e os grupos que não possuem comorbidades detém as maiores prevalências (RS: 95,7; SC: 96,4%) para a recuperação. Já para os grupos que não possuem comorbidades, a razão de chance de recuperação é de 1,149 (p=0,000 IC=1,143-1,155) no Rio Grande do Sul e, em Santa Catarina, é de 1,680 (p=0,000 IC=1,645-1,716). Ou seja, não apresentar comorbidade é uma variável preditora para o óbito, evidenciando assim que a população com menos comorbidades tende a ser considerado como um fator protetivo.

Na região nordeste, um estudo elaborado por Barros e Barros (2020), evidenciou que 73,96% dos óbitos confirmados por COVID-19 atingiam a faixa etária superior a 60 anos e apenas 24,88% tinham idade inferior a 60 anos. De acordo com sexo houve maior número de mortes no sexo masculino e em 66% dos óbitos totais os pacientes possuíam comorbidades. Nas regiões norte e centro-sul quase todos os sintomas foram mais comuns (febre, tosse, dor de garganta, falta de ar, desconforto respiratório, saturação arterial de oxigênio <95%, diarreia e vômito) em não sobreviventes no norte do que no centro-sul, sugerindo disparidades estruturais de saúde. Isso é ainda evidenciado por uma porcentagem substancialmente maior de não sobreviventes no norte do que no centro-sul e a proporção de pacientes internados que morreram revelou um padrão semelhante, com maior proporção no norte do que no centro-sul, sugerindo um efeito regional (BAQUI et al., 2020).

Como citado anteriormente, no estudo realizado por Baqui et al. (2020), indivíduos com COVID-19 na região norte apresentaram diversos sintomas associados diretamente ao sistema respiratório. Os estados que compõem esta região possuíram outros agravantes ao sistema respiratório, como exemplo, a prática de queimadas em sobreposição a pandemia associada ao SARS-CoV-2. Na maioria dos estudos citados, somente características populacionais foram consideradas, porém a precariedade de estudos que visam compreender outros fatores associados a pandemia, por exemplo, características ambientais, possibilitam respostas humanas aos impactos e desafios em relação as mudanças climáticas e a pandemia de COVID-19 (ZABANI-OTOU et al., 2020).

2.4 Descoberta de Conhecimentos em Banco de Dados e Mineração de Dados

Várias tecnologias fornecem conjuntos de dados que consistem em um grande número de pontos espaciais, comumente chamados de nuvens de pontos. Esses conjuntos de dados pontuais fornecem informações espaciais sobre o fenômeno a ser investigado, agregando valor através do conhecimento de formas e relações espaciais (MARCHI; PIROTTI; LINGUA, 2018). O volume crescente de dados geográficos digitais facilmente sobrecarregam técnicas de análise espacial convencionais. Estatística tradicional particularmente, têm altas cargas computacionais. Técnicas analíticas não podem descobrir facilmente novos padrões, tendências e relacionamentos que possam estar escondidos dentro de conjuntos de dados geográficos muito grandes e com diversas informações (ESTER; KRIEGEL; SANDER, 1997).

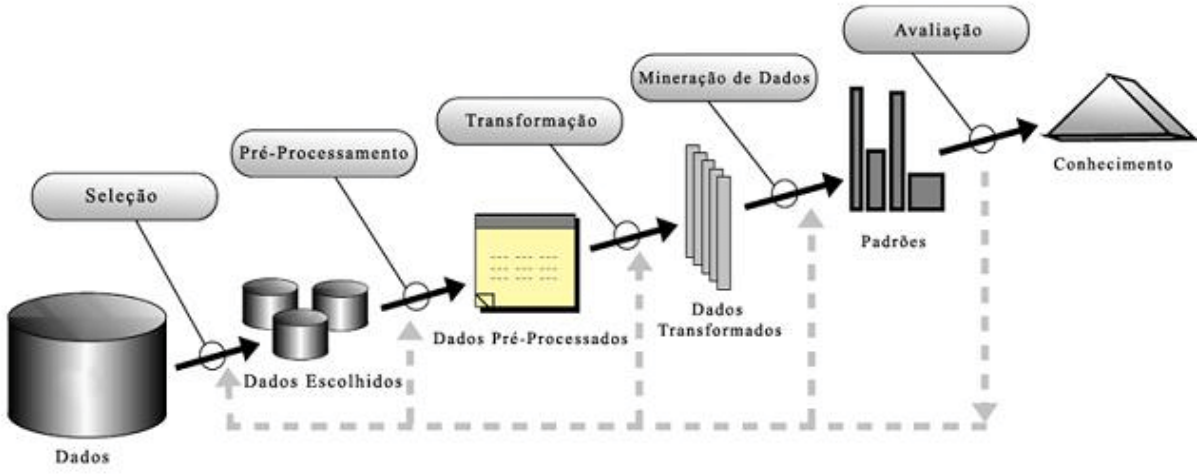
Os termos, Ciência de dados (*Data science*) e Inteligência artificial (*Artificial intelligence*) podem ser interpretadas como áreas de fundamental importância nos dias atuais. A Ciência de dados se refere a uma área de estudo muito ampla que se apropria de métodos científicos para retirar informações e percepções de conjuntos de dados, enquanto que a inteligência artificial refere-se ao que chamamos de inteligência das máquinas (UZINSKI; ABREU; OLIVEIRA, 2020).

A descoberta de conhecimento em bancos de dados, mais comumente conhecida como *Knowledge Discovery in Databases* (KDD) é uma resposta aos enormes volumes de dados que muitas vezes possuem muito mais informações escondidas do que as informações "superficiais" que são extraídas por análises tradicionais (PAZMIñO-MAJI; GARCíA-PEñALVO; CONDE-GONZáLEZ, 2017). Nesse caso, necessitam-se de cientistas que possam conhecer e manusear base de dados, utilizando-se de *Knowledge Dicovery from Databases* - (KDD) e técnicas de mineração de dados geográficos (MILLER; HAN, 2001).

Para (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996), o processo de KDD representa um procedimento composto de cinco principais tarefas que precisam ser realizadas de forma sequencial e uma após o término da outra, podendo ser repetido etapas anteriores. O objetivo destas etapas é descobrir informações relevantes para apoiar em tomadas de decisão e em decisões estratégicas. Na literatura as principais etapas do processo de KDD são: seleção de

dados, pré-processamento, transformação dos dados, mineração de dados e pós processamento ou avaliação conforme Figura 2.

Figura 2 – Fluxograma representando as etapas para a descoberta de conhecimento em bases de dados.



Fonte: Adaptado de (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996)

As subsecções seguintes irão explorar de forma sucinta os conceitos das etapas do processo de KDD representadas na Figura 2.

2.4.1 Seleção de Dados

O primeiro passo consiste na escolha e delimitação dos dados que serão utilizados na pesquisa. Geralmente, esses dados estão armazenados em bases de dados e disponibilizados através de diversas formas para a comunidade em geral (KUMAR; STEINBACH; TAN, 2009). No entanto, reunir essas informações em um banco de dados nem sempre é uma tarefa fácil, uma vez que este processo pode envolver a coleta de dados com ruídos, faltantes e valores ilegítimos surgindo então a necessidade de pré-processamento dos dados (ADRIAANS; ZANTINGE, 1996).

2.4.2 Pré-Processamento

Após a coleta de dados, a próxima etapa é a limpeza, sendo esta a mais importante etapa do pré-processamento. A quantidade de poluição que existe em bases de dados normalmente é muito grande. Por isso, uma boa ideia é passar algum tempo examinando os dados a fim de observar a possibilidades de problemas presentes na base de dados, embora isso possa na prática ser difícil e demorado, levando em consideração o trabalho com grandes conjuntos de dados (ADRIAANS; ZANTINGE, 1996).

A finalidade do pré-processamento é transformar os dados brutos e não padronizados em dados de formatos simples, como planilhas e tabelas. Esta etapa é considerada uma das etapas

mais trabalhosas e demoradas pois incluem técnicas de obtenção, fusão, limpeza, remoção de ruídos e remoção de dados duplicados (KUMAR; STEINBACH; TAN, 2009). Uma grande suposição feita pelas técnicas de mineração de dados é que o conjunto de dados está completo. A presença de valores omissos é, no entanto, muito comum nos processos de aquisição. Um valor nulo ou ausente significa um dado que não foi armazenado ou coletado devido a uma amostragem defeituosa, falha no processo ou limitações no processo de aquisição. Valores ausentes não podem ser evitados na análise de dados, e eles tendem a criar sérias dificuldades para os profissionais na interpretação dos resultados ou até mesmo não possibilitar a aplicação de técnicas de mineração (PAZMIñO-MAJI; GARCíA-PEñALVO; CONDE-GONZáLEZ, 2017).

2.4.3 Transformação

A etapa de transformação dos dados depende do objetivo da busca e do algoritmo a ser aplicado, pois é ele que possui as limitações a serem impostas a base de dados (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996). A melhor estruturação dos dados, como divisão de colunas e linhas, exclusão de dados que não serão minerados se torna importante para que haja um melhor resultado, garantindo assim uma melhor velocidade e qualidade na próxima etapa denominada Mineração de Dados (KUMAR; STEINBACH; TAN, 2009).

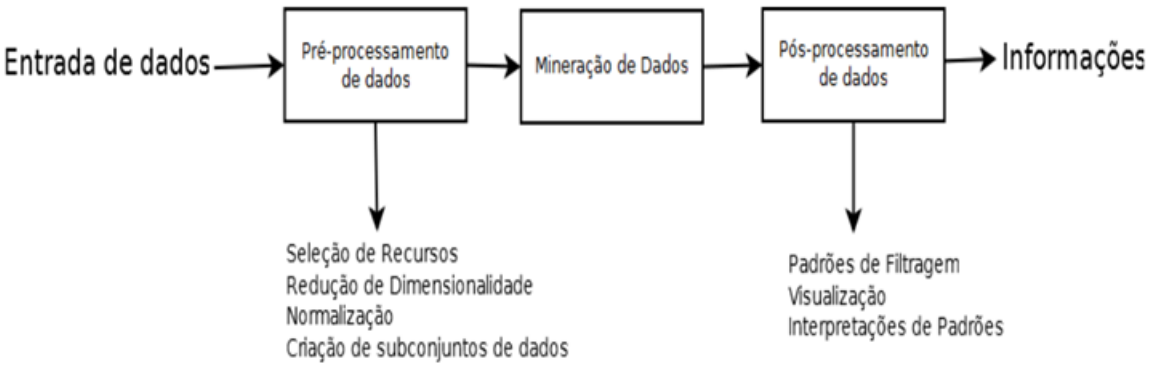
2.4.4 Mineração de Dados

Mineração de dados é o processo utilizado para descobrir padrões interessantes em grandes quantidades de dados. As fontes de dados podem incluir banco de dados, e outros repositórios que geram dados e são transmitidos para um sistema de forma dinâmica (HAN J.; KAMBER, 2011). Para Cortês S Da C; Porcaro (2002), a mineração de dados é um processo de interação entre homens e máquinas, onde o objetivo principal é a extração de padrões e relações entre variáveis de forma confiável em bases de dados de grande proporção.

Segundo Fayyad, Piatetsky-Shapiro e Smyth (1996), muitas pesquisadores tratam a mineração de dados como um sinônimo da descoberta de conhecimento. Alternativamente outros veem a mineração de dados como simplesmente uma etapa essencial no processo de descoberta de conhecimento, pois é nessa etapa em que dados brutos sem relação alguma, são minerados e transformados em dados com informações úteis. Já para Kumar, Steinbach e Tan (2009), o processo de mineração de dados consiste basicamente em duas etapas de transformação, sendo elas, pré-processamento e pós-processamento, conforme figura 3.

Nas abordagens de mineração de dados, uma das principais vantagens é a possibilidade de construir modelos de aprendizagem estatística interpretáveis, fornecendo compreensão qualitativa e quantitativa da relação entre as características de interesse. Esses modelos podem ser construídos para aprendizado supervisionado (a variável de resultado é prevista com base nas variáveis de entrada) ou não supervisionado (procura associações entre variáveis de entrada sem

Figura 3 – Etapas da Mineração de Dados.



Fonte: Adaptado de (KUMAR; STEINBACH; TAN, 2009)

uma medida de resultado) (T. Hastie and R. Tibshirani and J. Friedman, 2009).

Agrupamento ou *Clustering*, como é mais conhecido, é um método de agrupamento de linhas de dados que compartilham tendências e padrões semelhantes em grupos distintos. Os estudos de agrupamento não possuem uma variável independente como no processo de classificação, o que torna o agrupamento um método não supervisionado (R. Groth, 2000). O agrupamento de dados é baseado diretamente em cálculos de similaridade, que são fundamentais no cenário de agrupamento. Eles são os dados de entrada para o algoritmo *K-means* ou também conhecido como K-médias. Embora existam outras abordagens para identificar *clusters*, a detecção de agrupamentos pelo *K-means* está no cerne da mineração de dados espaciais. Assim, o *K-means* é uma forma de atingir nosso objetivo realizando agregação de dados em nós centrais (SOUZA; RAZENTE; BARIONE, 2014).

Em essência, a tarefa de *clustering* consiste em particionar os dados minerados em vários grupos de instâncias de dados, de tal forma que: (a) cada *cluster* tem instâncias muito semelhantes ou "próximas"umas das outras; e (b) as instâncias em cada *cluster* onde os mesmos são muito diferentes ou “distantes” dos outros *clusters* (SOUZA; RAZENTE; BARIONE, 2014). O conceito de similaridade ou distância é fundamental para lançar essas ideias no cenário de *clustering*. No caso de recursos contínuos, ou seja, variáveis, pode-se usar muitas funções de distância como medidas de similaridade. Cada uma dessas distâncias vem com sua própria geometria. A escolha da função distancia implica na geometria específica dos *clusters* formados (CIOS K. J.; KURGAN, 2007).

Muitas funções de distância podem ser usadas como medidas de similaridade. Cada uma dessas distâncias vem com sua própria geometria, o que implica no grupo formado. As medidas de distância comumente usadas são Distância de Manhattan (Equação 2.1), Distância Euclidiana (Equação 2.2) (CIOS K. J.; KURGAN, 2007).

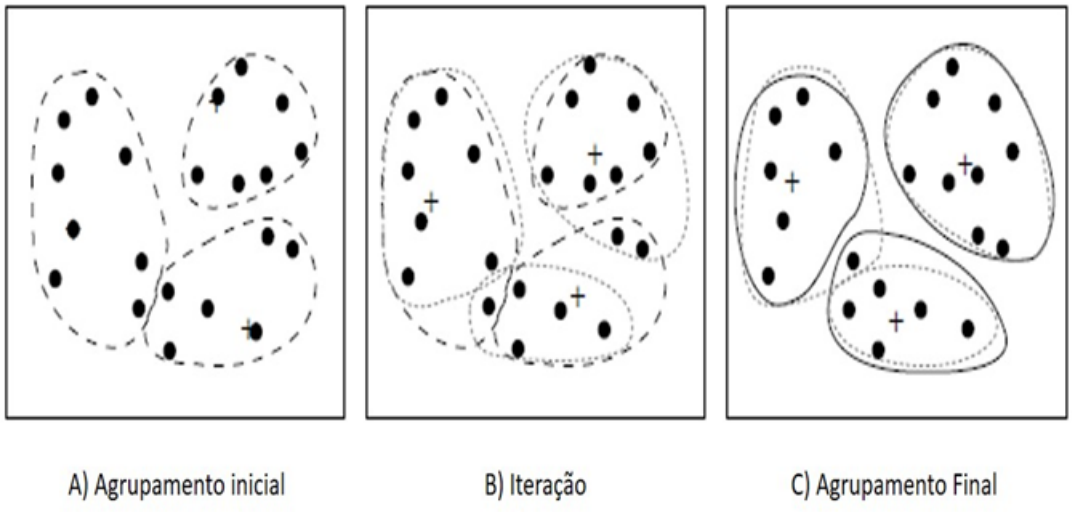
$$(x, y) = \sum_{j=1}^{n}(x_j - y_j) \tag{2.1}$$

$$(x, y) = \sqrt{\sum_{j=1}^{n}(x_j - y_j)^2} \tag{2.2}$$

onde x_j e y_j são as observações no plano cartesiano e n é o número total de observações.

A medida padrão utilizada na literatura cientifica para classificar e avaliar a qualidade de um agrupamento é através da distância média de todos os seus elementos. Quanto menores forem os valores da distância média, considera-se esse agrupamento como sendo o melhor e mais eficiente. Essa eficiência é medida através das médias dos cálculos de distância (SOUZA; RAZENTE; BARIONE, 2014). O método irá criar centroides c_i em cada agrupamento aleatoriamente e então as iterações serão feitas para ajustar o centroide, conforme ilustrado na Figura 4 (HAN J.; KAMBER, 2011).

Figura 4 – Etapas da formação dos *clusters*.



Fonte: Adaptado de (HAN J.; KAMBER, 2011)

A idéia básica por trás dos métodos de particionamento, como *K-means clustering*, é definir "k" *clusters* de tal forma que a variância total intra-*cluster* (ou soma total do quadrado dentro do cluster — WCSS) seja minimizada. Conceitualmente, o centroide de um *cluster* é seu ponto

central, calculado geralmente pela média, embora outras medidas possam ser usadas. A qualidade do *clustering* pode ser medida pelo quadrado da distância entre os pontos de amostra em cada *cluster* e o centroide, que é o WCSS (KUMAR; STEINBACH; TAN, 2009):

Resumindo, o método passo a passo é aplicado da seguinte forma:

- determine o número de k, ou seja, o número de centróides e *clusters* que serão criados;
- calcula a distância de cada observação amostral até o centroide;
- reposiciona a observação da amostra para o grupo cuja distância ao centróide é menor;
- recalcula a nova posição do centróide dentro de seu grupo;
- repete as iterações até que o centroide não mude de posição;

2.4.5 Pós-Processamento

De acordo com Kumar, Steinbach e Tan (2009), o pós-processamento apenas garante que dados úteis sejam incorporados na tomada de decisão e no processo de de descoberta de conhecimento. A visualização é um exemplo de pós-processamento, pois tenta retratar os dados obtidos na etapa de Mineração de Dados de uma maneira que o ser humano possa visualizar, explorar e entender melhor os resultados obtidos (HAN J.; MILER, 2001; KOUA; KRAAK, 2004).

O resultado do processo de Mineração de Dados, pode ser facilmente retratado através de representação visual. A saída deste processo computacional é representado usando gráficos como representação com a finalidade de facilitar percepção humana, pois oferece visualizações da estrutura geral do conjunto de dados, bem como a exploração de relacionamentos entre atributos (HAN J.; KAMBER, 2011). Várias representações gráficas fornecem formas de melhor representar a similaridade (padrões), relacionamentos, incluindo uma representação matricial de distância, 2D e projeções 3D, e componentes de visualização de planos. Estas diferentes formas de visualização integradas com outros gráficos acabam facilitando a análise exploratória e a descoberta de conhecimento (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

3 TRABALHOS RELACIONADOS

Este capítulo reúne trabalhos relevantes ao tema desta pesquisa, relacionados a aplicação de agrupamentos em dados de caráter geográficos (Seção 3.1) e trabalhos relacionados a queimadas, doenças respiratórias e COVID-19 (Seção 3.2) e considerações (Seção 3.2). Foram consideradas publicações relevantes nos últimos dez (10) anos com a finalidade de observar o estado da arte relativo a estes assuntos. A relevância esta associada a bons fatores de impacto e as bases de dados pesquisadas na *web* foram: *The Lancet Journal*, *Nature*, *ScienceDirect* e *MDPI Journals* e as buscas foram associadas as seguintes palavras chaves: *Fires, Respiratory Disease, Particulate Matter, clustering, K-means, Data Mining*, COVID-19 e SARS-COV-2.

3.1 Trabalhos relacionados a agrupamentos geográficos

A aplicação de técnicas de agrupamento foram abordadas por Schroeder et al. (2020), onde o algoritmo *K-means* foi empregado para identificar os grupos de cidades onde as ocorrências de incêndios são mais expressivas e um coeficiente de correlação de *Spearman* é calculado para inferir a dependência estatística entre as ocorrências de incêndio, doenças respiratórias, malária e leishmaniose. Com a identificação de pequenos grupos de cidades, foi possível identificar regiões com uma correlação temporal muito forte entre focos de incêndio e internações por doenças do sistema respiratório, malária e leishmaniose. Cada região apresentou comportamento suscetível a alguma doença, bem como algum grau de correlação com foco de incêndio. Foi detectado que a maior ocorrência de incêndios foi de julho a dezembro, período utilizado para análise de doenças respiratórias. O autor, (SCHROEDER et al., 2020) sugere investigações posteriores, onde outras abordagens de *cluster* de eventos pontuais podem ser aplicadas e comparadas. Porém, uma das limitações presentes neste estudo é que apenas a correlações foram analisadas em relação ao agrupamento com a maior incidência de queimadas, sendo assim, seria necessário uma comparação entre os agrupamentos de mais e menos quantidades de queimadas visando validar se de fato, grupos de cidades com menos queimadas apresentam um menor teor de hospitalizações.

Em um trabalho elaborado por Parente, Pereira e Tonini (2016), objetivou-se através de analises estatísticas espaço/tempo avaliar as mudanças no regime das queimadas em relação a diferentes tipos de clima e atividades de manejo do fogo em Portugal, país com maior incidência de incêndios do continente europeu. Portanto de forma resumida, os objetivos do estudo citado acima foram: (i) avaliar a influência das características do conjunto de dados dependência das estatísticas de varredura de permutação espaço-tempo (STPSS) para capturar fenômenos de agrupamento de focos de incêndio; (ii) identificar e caracterizar as condições atmosféricas associadas aos incêndios pertencentes a cada um dos *clusters* detectados. No estudo, as análises de *clusters* espaciais e temporais buscaram agrupar objetos mostrando uma superdensidade local no espaço e/ou no tempo. Como resultados, observou-se a capacidade do STPSS de identificar

corretamente os *clusters*, em relação ao seu número, localização e tamanho do espaço-tempo, apesar de eventuais divisões de espaço e/ou tempo dos conjuntos de dados. Por fim, a identificação dos agrupamentos possibilitou a confirmação do papel do clima nos dias em que os incêndios estavam ativos para as classes de pequenos, médios e grandes incêndios. No estudo citado, a identificação das localizações dos agrupamentos, permitiram a elaboração de análises de relação do clima em diferentes locais e seu impacto na incidência de incêndios.

Outra abordagem e aplicação do algoritmo K-meas foi abordada por Shafi e Waheed (2020), onde o autor visa a aplicação do K-Means para analisar as mudanças frequentes que ocorrem na poluição do ar da cidade de Southampton. Neste artigo, o K-Means e o método *elbow* foram empregados em um conjuntos de dados para mapear os níveis de PM 2.5 em um conjunto de dados de qualidade do ar. Os resultados experimentais permitiram identificar o número ideal de agrupamentos através do método *elbow* e a aplicação do algoritmo de agrupamento K-Means, mostra que as mudanças rápidas que ocorrem na qualidade do ar do nível mais baixo em pouco tempo atingem o nível tóxico mais alto no mesmo local devido aos focos de incêndio em apenas algumas horas. O autor sugere que no futuro, a qualidade do ar pode ser um assunto de maior preocupação, para que os sensores inteligentes e pesquisas especificas possam ser potencializados com ações apropriadas para emissão de alertas para quando a qualidade do ar atinge valores tóxicos.

Com o surgimento da pandemia global da COVID-19 e devido à sua novidade e rápida disseminação, os cientistas tiveram dificuldade em criar previsões precisas para esta doença e trabalhos com a aplicação de aprendizado de máquina começaram a surgir. Um trabalho elaborado por Nicholson et al. (2022), visou empregar métodos supervisionados e não supervisionados para identificar os fatores críticos demográficos, de mobilidade, clima, capacidade médica e de saúde relacionados ao município para estudar a propagação do COVID-19 antes da ampla disponibilidade de uma vacina. Foram utilizados recursos para agregar municípios em *clusters* significativos para apoiar esforços de análise de doenças mais refinados. Os autores propuseram o uso de 7 modelos de aprendizado de máquina e um novo método de previsão híbrido baseado em vizinhos mais próximos e k-means para prever as taxas de crescimento do COVID-19. O objetivo principal é descobrir as características mais importantes em nível de cidades relacionadas à propagação do COVID-19 e agregar cidades individuais em *clusters* com base nas características importantes em nível de cidade. Porém, a escolha do número de *clusters* foi elaborada de forma subjetiva. Para trabalhos futuros, os autores sugerem melhorar as características únicas de cada *cluster* para melhorar a previsão de séries temporais em nível regional e local para a previsão de doenças.

3.2 Trabalhos relacionados a queimadas, doenças respiratórias e COVID-19

Em um estudo elaborado por Machado-Silva et al. (2020), objetivou-se compreender os papéis desempenhados pelo declínio da precipitação e atividade do fogo, bem como as alterações

associadas aos parâmetros atmosféricos na incidência de internações por doenças respiratórias. O autor relata ser uma tarefa muito complexa, porém é provável que a importância relativa desses fatores dependa do tempo e variabilidade espacial utilizada na análise. No trabalho concluiu-se que houve um aumento de 27% em Hospitalizações por doenças respiratórias (exceto para Asma que diminuiu 75%) em anos de seca conforme descrito pela relação positiva (negativa para Asma) com a precipitação. As chuvas e a umidade exercem um controle primário nas doenças do aparelho respiratório na região. *Hotspots*, áreas queimadas e fumaça concentram-se em agosto e setembro, e têm uma influência secundária do fogo na hospitalizações. A produção de fumaça está fortemente relacionada ao fogo e à temperatura, que estão associados ao aumento do número de internações durante a temporada de incêndios. Com base nos dados observados no estudo, foi destacado o papel das tendências regionais de precipitação na condução de internações respiratórias, que são indicadores ecológicos cruciais para os seres humanos no contexto das mudanças climáticas. Porém algumas limitações foram observadas neste estudo, onde a área de estudo foi a cidade de Porto Velho. Os autores revelam que mais pesquisas devem se concentrar em uma cobertura espacial mais ampla sobre o região amazônica para permitir uma compreensão mais profunda de cada indicador.

No artigo elaborado por Smith et al. (2014b), considerando como área de estudo todo o território da amazônia legal, observou um aumento significativo (1,2%–267%) nas hospitalizações por doenças respiratórias em menores de cinco anos em municípios altamente expostos à seca. O aerossol foi o principal causa de hospitalizações em municípios afetados pela seca durante 2005, enquanto condições de desenvolvimento mitigaram os impactos em 2010. Os resultados demonstraram que os eventos de seca deterioraram a saúde respiratória das crianças, particularmente durante 2005, quando a seca foi mais concentrada geograficamente. Porem os autores relatam a necessidade de identificações de áreas críticas em uma escala regional para que haja um planejamento para maior demanda de serviços de saúde durante os períodos de seca. Já o trabalho elaborado por Rocha e Sant'Anna (2022), foram avaliados os efeitos da poluição do ar relacionada ao fogo na saúde da população na Amazônia brasileira. A estratégia de pesquisa foi baseada em um modelo de efeitos fixos município-a-mês, juntamente com uma abordagem de variáveis instrumentais que exploram a direção do vento e a poluição do ar nas áreas vizinhas, a fim de mudar exogenamente a exposição à poluição do ar na localidade. Concluiu-se que a exposição à poluição do ar, medida pelos níveis de concentração de PM2,5, está fortemente associada a um aumento nas internações hospitalares por problemas respiratórios. Os efeitos são maiores entre crianças e idosos e aumentam de forma não linear com os níveis de poluição.

Porém, estudos relacionados ao tema também ganham mais relevâncias em outras regiões do mundo. O artigo publicado por Meo et al. (2021), objetivou estudar os efeitos dos poluentes ambientais PM-2,5, monóxido de carbono e ozônio na incidência e mortalidade da infecção por SARS-COV-2 em dez cidades mais afetadas por incêndios florestais na Califórnia. Identificou-se no estudo que após o incêndio, a concentração de PM2,5 aumentou 220,71%; O3 em 19,56%; e a concentração de CO aumentou 151,05%. Após o incêndio, o número de casos e mortes por

COVID-19 ambos aumentaram respectivamente 56,9% e 148%. O incêndio florestal da Califórnia causou um aumento nas concentrações ambientais de poluentes tóxicos que foram temporariamente associados a um aumento na incidência e mortalidade da COVID-19. Contudo, os autores relatam a limitação do estudo sendo a escolha dos períodos para analise (períodos curtos), visto que este era o primeiro estudo relacionado ao tema elaborado na Califórnia. O primeiro período de tempo abrangeu a data do aparecimento do primeiro caso de (SARS-CoV-2) na Califórnia (que ocorreu em 19 de março de 2020) por meio de um recente surto de incêndio florestal (que ocorreu em 15 de agosto de 2020). O segundo período de tempo abrangeu o início do incêndio florestal (que foi 15 de agosto, 2020) a 22 de setembro de 2020 (que foi um período de quase 7 semanas, que foi suficiente para que o incêndio florestal afetasse a incidência de COVID-19). Outra limitação importante desse estudo foi que a modelagem estatística não levou em conta os muitos desafios inerentes a esses dados, como fatores de confusão, efeitos tardios, super-dispersão e heterogeneidade entre os municípios.

Outro estudo elaborado nos Estados Unidos, elaborado por Zhou et al. (2021), possibilitou aos autores adquirir e vincular dados diários disponíveis publicamente sobre PM 2.5, número de casos e mortes de COVID-19 e outros fatores de confusão para 92 cidades do oeste dos EUA que foram afetados pelos incêndios florestais de 2020. Foram estimados a associação entre a exposição de curto prazo ao PM 2,5 durante os incêndios florestais e a dinâmica epidemiológica dos casos e mortes por COVID-19. Foram ajustados vários fatores de confusão que variam no tempo (por exemplo: clima, sazonalidade, tendências de longo prazo, mobilidade e tamanho da população). Como resultados, foram identificados fortes evidências de que os incêndios florestais amplificaram o efeito da exposição de curto prazo ao PM 2,5 nos casos e mortes de COVID-19, embora com heterogeneidade substancial entre os municípios.

3.3 Considerações

Observando o estado da arte dos estudos realizados por Schroeder et al. (2020), Nicholson et al. (2022), Parente, Pereira e Tonini (2016), pode-se observar a aplicação de algoritmos de agrupamentos para a identificação de grupos de cidades, fator este que possibilita estudos de áreas de risco baseado em metodologias sólidas para identificação de áreas de risco, diferentemente da escolha aleatória. Todos os trabalhos, visaram agrupar cidades, ou seja, a geolocalização é um fator extremamente importante para realizar as medidas de similaridades/proximidades das observações.

Porém os estudos Nicholson et al. (2022), Parente, Pereira e Tonini (2016) possuem limitações quanto a metodologia de identificação do número ideal de *clusters*, pois a determinação do número ideal de *clusters* ocorreu de forma subjetiva. Já o estudo de Shafi e Waheed (2020), objetivou agrupar dados baseado na similaridade das observações de PM 2,5, onde o número ideal de *clusters* foi determinado através do método *elbow* e a área de estudo definida anteriormente. Todos os autores sugerem para trabalhos futuros, melhorar as características únicas

de cada *cluster* para melhorar a previsão de séries temporais em nível regional e local, para a previsão das relações das queimadas/poluentes com doenças.

Quanto aos estudos relacionados a doenças respiratórias, SARS-COV-2 e queimadas, foi possível compreender as metodologias aplicadas para elaboração destes estudos. O estudo de Machado-Silva et al. (2020), utilizou apenas uma cidade como local de estudo. O local de estudo foi escolhido estrategicamente pelos autores por ser uma das áreas urbanas mais densamente povoadas da floresta amazônica e a terceira cidade mais populosa da região Norte do Brasil com densidade populacional de cerca de 12,6 habitantes por quilômetro quadrado. No entanto, as queimadas geram impactos atmosféricos de maiores escalas regionais, considerar somente uma cidade pode não representar o real impacto das incidências de incêndios. Por outro lado, a escolha de uma única cidade altamente urbanizada também pode estar sendo impactada por poluições industriais.

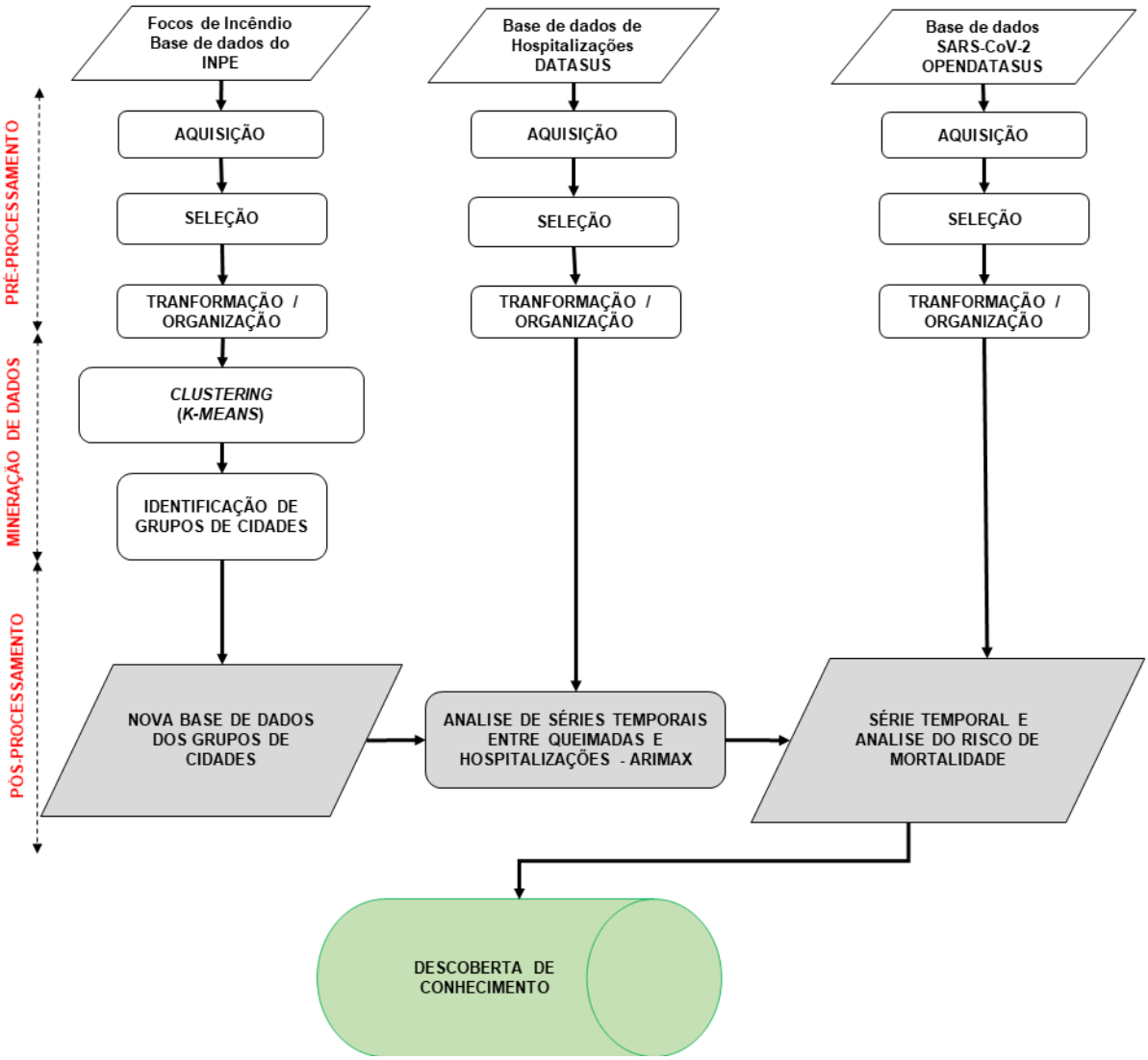
Já os estudos realizados por Smith et al. (2014b) e Rocha e Sant'Anna (2022), também evidenciam associações de aumento de hospitalizações com queimadas. Contudo ambos os estudos utilizaram uma grande área territorial, sendo ela todo o território da Amazônia legal. Porém os autores relatam a necessidade de identificações de áreas críticas em uma escala regional que possibilite um melhor e efetivo direcionamento de recursos.

Já os estudos elaborados nos Estados Unidos, especificamente em cidades do estado da Califórnia, elaborados por Meo et al. (2021) e Zhou et al. (2021), corroboram com a hipótese de associação ao aumento da mortalidade por COVID-19 em locais com maior exposição a poluentes atmosféricos potencializados por queimadas. Contudo, os estudos utilizaram municípios em nível individual sem verificar a possibilidade de queimadas oriundas do limite de um município estar prejudicando mais o município vizinho. Também o estudo utilizou pequenos períodos para validação de um impacto simultâneo de poluição oriunda das queimadas e mortalidade por COVID-19, ou seja, não utilizaram um período de 1 ano completo e nem compararam com anos sem a pandemia, podendo ter ocorrido a causalidade.

4 METODOLOGIA

Neste trabalho definiu-se a metodologia conforme fluxograma da Figura 5 e nas seções seguintes, foram descritas as etapas de pré-processamento, agrupamento, análise de séries temporais e pós-processamento até o objetivo final que é a descoberta de conhecimento.

Figura 5 – Fluxograma das etapas envolvidas desde o pré-processamento, agrupamento (K-means), pós-processamento até a descoberta do conhecimento (KDD).



Fonte: Elaborada pelo autor.

O fluxograma metodológico apresentado na Figura 5 esta estruturado de forma similar com o fluxograma publicado no artigo de Schroeder et al. (2020). Porém, existe a inclusão de mais variáveis e mais análises estatísticas. As próximas sessões foram destinadas para aprofundar e melhor explorar cada etapa presente no fluxograma apresentado.

4.1 Pré processamento

As etapas de pré-processamento apresentadas no fluxograma da Figura 5 incluem a preparação do banco de dados de incêndios florestais, internações, óbitos de COVID-19 e bancos de dados adicionais usados para entender os aspectos sociais da população estudada. Esses bancos de dados são brevemente apresentados a seguir.

4.1.1 Banco de Dados dos Focos de Incêndio

O banco de dados fornecido pelo INPE para dados de incêndios é denominado BDQUEIMADAS [1] onde é atualizado a cada três horas em toda a América Latina e contém informações sobre as coordenadas geográficas dos focos de incêndio, bioma, concentração de fumaça no ar, risco de incêndio, previsão de chuva, número de dias sem chuva, data, hora e cidade do foco observado (LIMONATI et al., 2015; INPE, 2020; **?**). Os dados de contagem de incêndios são obtidos por técnicas de sensoriamento remoto usando um sensor *Moderate Resolution Imaging Spectro-radiometer* (MODIS), integrado à missão do satélite TERRA/AQUA. Porém, neste estudo, os dados de interesse para a aplicação dos algoritmos de mineração de dados foram extraídos no formato de valores separados por vírgulas (.csv) e as informações de interesse para esse estudo são as coordenadas geográficas (Latitude e Longitude) de todos os focos de incêndios observados no Estado do Pará, durante o período de 2015 a 2020, conforme exemplo de tabela 1.

Tabela 1 – Exemplo da forma de disponibilização dos dados pelo Banco de Dados de Queimadas.

Data e Hora	**Estado**	**Município**	**Latitude**	**Longitude**
01/01/2017 16:39	Pará	Ulianópolis	-3,582	-47,044
04/01/2017 17:09	Pará	Santarém	-2,302	-54,482

Fonte: Elaborada pela autor.

4.1.2 Banco de Dados de Saúde: Hospitalizações por Doenças Respiratórias (HDR)

No Brasil, o Departamento de Informática do Sistema Único de Saúde (SUS) possui uma plataforma denominada TABNET[2] que disponibiliza dados relacionados a diversas doenças, criando assim, um projeto de interoperabilidade entre os sistemas públicos de saúde e a comunidade em geral. As causas de internação são classificadas de acordo com a Classificação Internacional de Doenças em sua 10ª Edição (CID-10), onde foram consideradas as internações relacionadas a doenças respiratórias (códigos J00-J99) (INFORMATION TECHNOLOGY DEPARTMENT OF THE PUBLIC HEALTH CARE SYSTEM-SUS , DATASUS). Na plataforma

[1] https://queimadas.dgi.inpe.br/queimadas/bdqueimadas
[2] https://datasus.saude.gov.br/informacoes-de-saude-tabnet/

do TABNET, as informações de internações são disponibilizadas de forma mensal e nos permite selecionar os registros de hospitalizações por local de Internação ou local de residência. Os dados de hospitalizações e óbitos por doenças respiratórias foram selecionados com base no município de residência durante o período de 2015 a 2020 para o Estado do Pará. A escolha de informação das hospitalizações por local de residencia possibilita entender o comportamento da população residente em um determinado local, em relação ao impacto dos efeitos das queimadas na saúde respiratória local.

4.1.3 Banco de Dados de Saúde: Mortes relacionadas à infecção por SARS-Cov-2

Devido ao surgimento da pandemia de SARS-Cov-2, o Ministério da Saúde implementou, por meio da Secretaria de Vigilância e Saúde, a vigilância da síndrome gripal leve, moderada e suspeitos de terem contraído SARS-Cov-2. Os dados de todos os casos de síndrome gripal são coletados pelo e-SUS NOTIFY e disponibilizados à comunidade por meio da plataforma OPENDATASUS [3], com dados de todos os pacientes e informações como data, estado, município, casos positivos, óbitos, local de residência e local de notificação (NOTIFICAçõES DE SíNDROME GRIPAL , OPENDATASUS). Para este estudo, utilizou-se somente dados de óbitos confirmados por local de residência no Estado do Pará, Brasil.

4.1.4 Bases de Dados Complementares

Para contribuir com as discussões, outras bases de dados foram incorporadas a este estudo. Os dados referentes ao Índice de Desenvolvimento Humano Municipal (IDHM) foram coletados do Programa das Nações Unidas para o Desenvolvimento (PNUD)[4], através do Plano Estratégico do PNUD, 2018-2021. Lembrando que o IDHM é baseado em três características básicas: vida longa e saudável, educação e um padrão de vida digno (HUMAN DEVELOPMENT INDEX , HDI). O Produto Interno Bruto (PIB) é uma medida da soma de todos os bens e serviços encontrados em um território em algum período. O PIB permite monitorar a atividade econômica e as diferenças econômicas entre as regiões permitindo analisar seu impacto no controle da pandemia (MCKEE; STUCKLER, 2020). O site que disponibiliza os dados do PIB de todos os municípios brasileiros é o Instituto Brasileiro de Geografia e Estatística (IBGE)[5].

Dados de aerossol (minúsculas partículas sólidas e líquidas suspensas na atmosfera) também foram observados. Exemplos de aerossóis incluem poeira levada pelo vento, cinzas vulcânicas, fumaça de incêndios e poluição industrial, que podem afetar o clima e a saúde das pessoas (MACHADO-SILVA et al., 2020). A *NASA Earth Observations* (NEO)[6] fornece dados para o sensor com um espectrorradiômetro de imagem de resolução moderada (MODIS) que está a

[3]https://opendatasus.saude.gov.br/
[4]https://www.br.undp.org/content/brazil/pt/home/idh0.html
[5]https://www.ibge.gov.br/estatisticas/downloads-estatisticas.html
[6]https://neo.sci.gsfc.nasa.gov/view.php?datasetId=MODAL2$_{MA}ER_OD$

bordo da NASA satélites Terra e Aqua e é usado para monitorar a espessura óptica do aerossol em um determinado dia ou ao longo de alguns dias.

4.2 Mineração de Dados

Como apontado por (GUO D; MENNIS, 2009), a mineração de dados espaciais não é uma tarefa simples. Na verdade, requer seleção meticulosa, pré-processamento e transformação/organização dos dados para garantir análises e resultados significativos. Além disso, também exige algoritmos computacionais eficientes para processar grandes conjuntos de dados e abordagens de visualização eficazes para apresentar e explorar padrões complexos. Primeiramente, foi realizado um aprendizado não supervisionado a partir de métodos de agrupamento para identificar regiões críticas (grupo de cidades) para posterior análise temporal/associação com doenças. As subseções a seguir descrevem esses métodos.

4.2.1 *Clustering (K-means)*

Neste trabalho utilizou-se um método de aprendizado de máquina não supervisionado de *cluster* chamado *K-means* (HAN J.; KAMBER, 2011; CORTêS S DA C; PORCARO, 2002; LIKAS; VLASSIS; VERBEEK, 2003; HARTIGAN; WONG, 1979). Este método utiliza os parâmetros de dados com dimensões e neste caso utilizamos a distância euclidiana, onde determinamos um número k de grupos. As Observações são agrupadas de acordo com a distância entre as coordenadas e seu centroide. Após cada iteração do algoritmo, a posição da observação é recalculada com base na posição do centroide de seus componentes. O algoritmo termina quando não ocorre nenhuma mudança significativa na posição (T. Hastie and R. Tibshirani and J. Friedman, 2009; SCHROEDER et al., 2020). O *K-means* nos permite adicionar pontos a nós centrais ou centroides, que são determinados como os grupos de interesse em nosso estudo.

Para determinar o número ideal de grupos para particionar um conjunto de dados, foi utilizado o método cotovelo (KODINARIYA; MAKWANA, 2013). O método do cotovelo consiste em definir o número de K-grupos de forma que a variância total dentro do grupo seja minimizada através da soma quadrada total dentro do grupo usando o *Within Cluster Sum of Squares* (WCSS). Sendo assim, a qualidade do *cluster* está ligada à distância média entre os as coordenadas geográficas dos focos de Incêndio e seu centroide para um determinado número de *clusters* formados (KUMAR; STEINBACH; TAN, 2009; HAN J.; KAMBER, 2011).

$$d(x, y) = \sqrt{\sum_{j=1}^{n}(x_j - y_j)^2} \tag{4.1}$$

onde x_j e y_j são as coordenadas geográficas (latitude e longitude) e n é o número de ocorrências de incêndios.

$$WCSS = \sum_{i=1}^{k} \sum_{j=1}^{n_j} d(x_j, c_i)^2. \tag{4.2}$$

onde k é o número de *clusters* e x_j é o ponto no espaço de um determinado objeto (coordenada de ocorrências de incêndio) e c_i é o centroide do cluster e n é o número de ocorrências de incêndio.

A utilização do *k-means* ocorre por ser um algoritmo de agrupamento de dados baseado em partição, ou seja, cada partição deve conter pelo menos um objeto, e cada objeto deve pertencer somente a um grupo. Essa partição ocorre baseado nas distâncias entre as observações, sendo assim, tratando se de coordenadas geográficas torna-se um algoritmo mais eficiente (ESLING; AGON, 2012). Existem outras abordagens técnicas de agrupamentos, como por exemplo, agrupamentos baseados em hierarquia onde o algoritmo busca correlações entre o conjunto de dados, onde inicialmente cada cada objeto pertence ao seu próprio grupo, e a cada iteração, os dois grupos mais parecidos são unidos formando assim um novo grupo (AGHABOZORGI; Seyed Shirkhorshidi; Ying Wah, 2015).

As vantagens da utilização *K-means* para este estudo:

- O armazenamento em *cluster* hierárquico não pode manipular dados de grandes proporções, já o *K-means* pode. Isso ocorre porque a complexidade do tempo de cálculo do *K-means* é linear (distância em linha reta entre as coordenadas geográficas) e a clusterização hierárquica é quadrática;
- K-means funciona bem quando a forma dos *clusters* é hiper esférica como círculo em 2D, esfera em 3D (dados de coordenadas geográficas latidude e longitude são 2D);
- Clusterização K-means requer conhecimento prévio de K ou seja, número se clusters desejados para dividir os dados. Já no *clustering hierárquico*, é necessário interpretar o dendrograma;
- Densidades variáveis dos pontos de dados não afetam o algoritmo de agrupamento *K-means*.);

4.3 Pós Processamento

Para mostrar o impacto dos incêndios na saúde respiratória, foram considerados dois grupos: grupos de municípios com maior número de incêndios e o grupo de municípios com menor número de incêndios no Estado do Pará. Após a identificação desses grupos por meio do algoritmo *K-means*, realizamos uma comparação entre a taxa de hospitalizações por doenças respiratórias gerais (HDR) entre os anos de 2015 a 2019. A taxa de internação é dada por

$$\text{Taxa de Hospitalização} = \frac{\text{Total de Casos de Hospitalizações Mensais} \quad \times \quad 100,000}{\text{População Total}}. \tag{4.3}$$

Assim, serão utilizadas análises estatísticas e ferramentas gráficas integrando os resultados da mineração de dados (agrupamento) com dados de saúde para investigar as relações entre as queimadas e as doenças respiratórias.

4.3.1 Análise de Séries Temporais - ARIMAX

Para levar em conta a correlação serial das séries temporais e verificar o quanto as ocorrências de incêndio influenciam no HDR, foi construído um modelo ARIMAX. Este modelo é uma extensão do modelo ARIMA (*Auto-regressive Integrated Moving Average*) permitindo a inclusão de variáveis exógenas. O modelo ARIMA é utilizado quando se pretende modelar dados do passado visando compreender ou prever acontecimentos futuros de uma única variável. Neste estudo, as séries temporais de HDR possuem influência de outros fatores, sendo assim, existe uma co-variável chamada queimadas que possui influencia no comportamento da série temporal, sendo necessário a inclusão desta variável no modelo, ou seja, variável Exógena (X). São aplicados os procedimentos usuais de identificação, estimativa e validação. Detalhes sobre o ARIMAX podem ser obtidos em (BOX; JENKINS; REISEL, 2008).

Ao analisar séries temporais para identificar tendências utilizando diferenciação sucessiva (observação no momento t menos a observação no momento $t-1$) é possível induzir uma média constante e, consequentemente, a estacionaridade. Sejam F_t e H_t as séries temporais de ocorrências de incêndio e HDR após diferenças de ∇^d para induzir a estacionariedade, se necessário. Assim, o modelo $ARIMAX(p,d,q)$ pode ser escrito como:

$$H_t = \gamma_1 F_{t-\ell} + \alpha_1 H_{t-1} + \alpha_2 H_{t-2} + \cdots + \alpha_p H_{t-p} + \tag{4.4}$$

$$+ \ \epsilon_t + \beta_1 \epsilon_{t-1} + \beta_2 \epsilon_{t-2} + \cdots + \beta_q \epsilon_{t-q} +$$

onde ϵ_t é o ruído branco e p e q são as ordens auto-regressiva e média móvel, respectivamente. Essas ordens são identificadas a partir de funções de autocorrelação. De fato, a correlação cruzada pode ser avaliada para identificar a defasagem ℓ da variável incêndio para explicar o HDR. Para avaliar a contribuição da variável independente F, são considerados o Critério de Informação de Akaike (AIC) e o teste de taxa de verossimilhança. Os pressupostos residuais usuais de correlação serial, normalidade e variância constante foram verificados após a estimação do modelo.

4.3.2 Analises SARS-Cov-2

Com a análise e confirmação do HDR estar sendo potencializado pelos incêndios, será realizada a análise da taxa de mortalidade por SARS-Cov-2 por local de residência em ambos os *clusters* no estado do Pará. Os dados do SARS-Cov-2 foram extraídos do site OPENDATASUS para o ano de 2020. Consideramos a taxa de mortalidade para cada *cluster* por 100 mil habitantes e a equação é dada por

$$\text{Taxa de Mortalidade} = \frac{\text{Total de Óbitos por SARS-Cov-2} \quad \times \quad 100.000}{\text{População Total}}. \tag{4.5}$$

ao invés da utilização da taxa de letalidade (porcentagem) ou taxa de letalidade (proporção) que exigem o número de casos positivos de SARS-Cov-2 que são difíceis de estimar devido à falta de um teste amplo na população. Diversos estudos utilizaram a taxa de letalidade em estudos clínicos, porém Os

dados da taxa de letalidade podem ser tendenciosos, pois no Brasil os casos notificados de COVID-19 são altamente dependentes da qualidade e quantidade de testes aplicados em cada região, políticas de testagem, sub-notificação de casos e até casos duplicados devido a reinfecções.

Para calcular a razão da taxa de mortalidade (RTM) comparamos o risco de mortalidade entre os dois *clusters*, um incluindo as cidades com maior número de incêndios *(High Fires)* (HF) e outro com menor número de incêndios *(Low Fires)* (LF) no Estado do Pará. Agregamos os dados em períodos trimestrais e usamos a seguinte expressão:

$$\widehat{RTM} = \frac{\text{Taxa de Mortalidade no } \textit{cluster} \text{ HF}}{\text{Taxa de Mortalidade no } \textit{cluster} \text{ LF}} = \frac{\text{Óbitos}_{HF}/\text{População}_{HF}}{\text{Óbitos}_{LF}/\text{População}_{LF}}. \tag{4.6}$$

Para avaliar se o $\widehat{RTM}$ estimado indica risco ou não, o intervalo de confiança (IC) precisa ser calculado. Assim como a razão de risco, a RTM não é normalmente distribuída, mas seu logaritmo natural é. Com o desvio padrão (DP) da razão de taxa de log, $\widehat{DP}[ln(\widehat{TI})] = \left(\frac{1}{Obitos}_{HF} + \frac{1}{Obitos}_{LF}\right)^{\frac{1}{2}}$, é possível calcular o IC de 95% de RTM como:

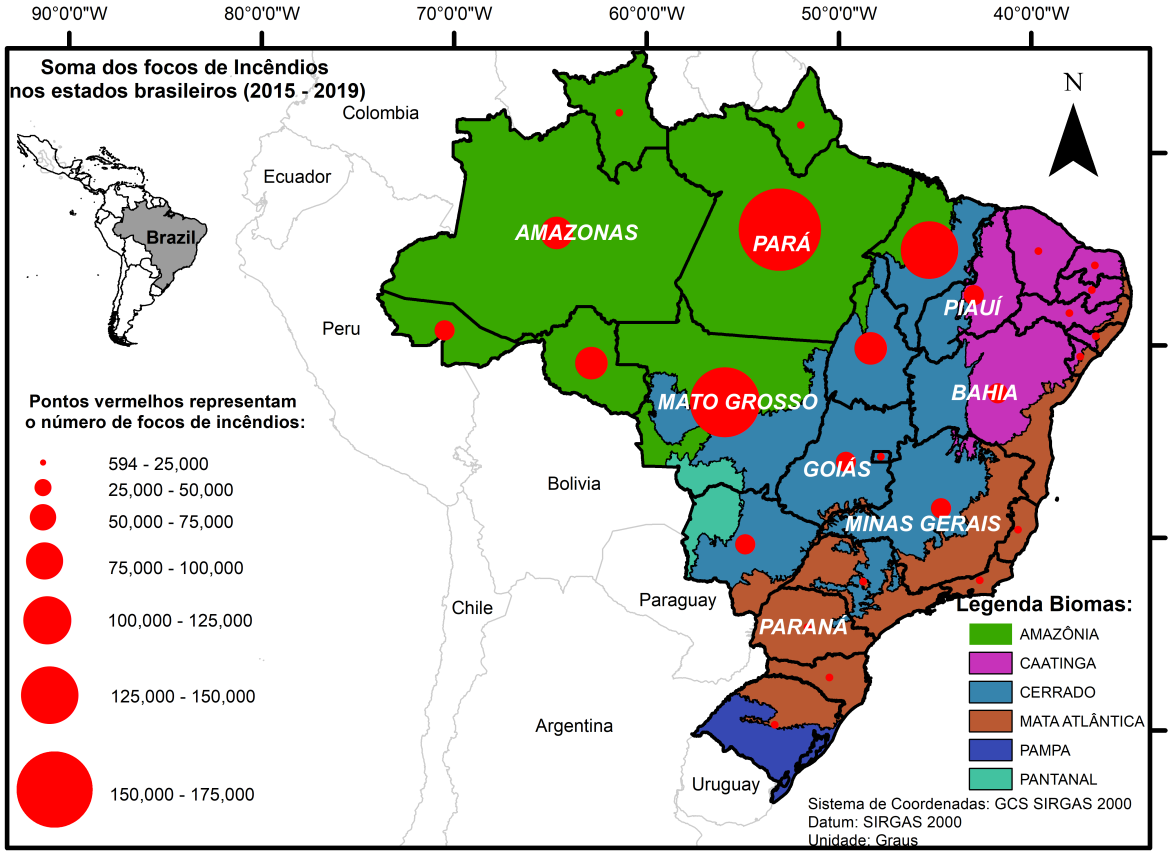
$$e^{ln(\widehat{RTM}) \pm 1.96 \widehat{DP}[ln(\widehat{RTM})]}. \tag{4.7}$$

5 RESULTADOS

5.1 Distribuição e compreensão das ocorrências de Incêndios no Brasil

A figura 6 ilustra a distribuição espacial dos focos de incêndios registrados no banco de dados do INPE em todos os estados brasileiros de forma cumulativa entre o período de 2015 a 2019.

Figura 6 – Distribuição espacial da soma do número de incêndios (tamanhos dos pontos vermelhos) nos últimos 5 anos (2015-2019) antes da pandemia nos estados do Brasil.



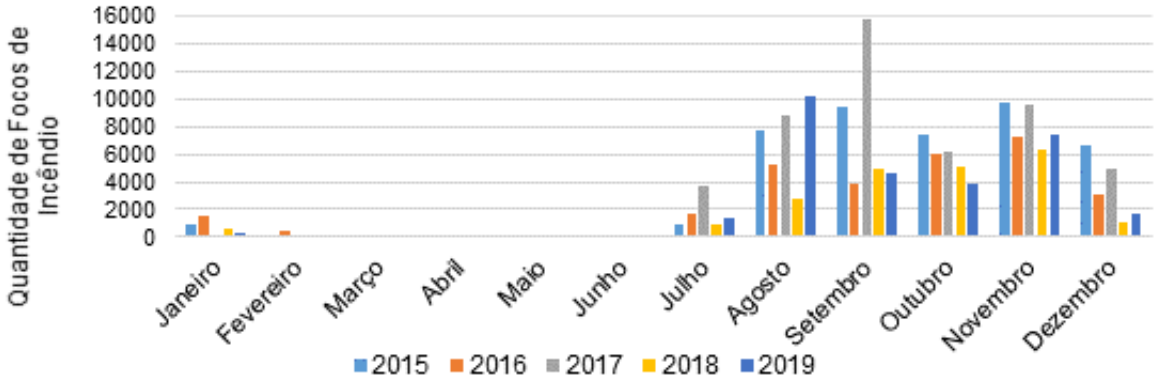
Fonte: Elaborada pelo autor.

O Estado do Pará, localizado na Amazônia brasileira, apresenta a maior incidência de incêndios nos últimos 5 anos, com mais de 170,000 ocorrências de incêndios. O Pará possui 20% da floresta amazônica em seu território e 24% dos incêndios registrados no Brasil ocorrem neste estado. Este estado possui a predominância do clima Equatorial Quente e úmido, porém possui a segunda maior extensão territorial do Brasil, sendo assim, existem algumas variabilidades nos tipos climáticos do estado.

Assim, o estado foi selecionado para realizar uma análise mais detalhada, pois há a necessidade de compreender o comportamento das queimadas dentro do Estado do Pará por meio de uma distribuição espacial e também durante os meses do ano. Pela sua localização geográfica, o Pará encontra se em uma região do planeta com maior incidência solar, atingindo elevadas temperaturas. Porém, neste estado, também ocorre a baixa amplitude térmica anual, ou seja, existem grandes diferenças entre as temperaturas mais baixas e mais altas, não havendo estações do ano bem definidas. Para melhor compreender o comportamento mensal das queimadas no estado do Pará, elaborou-se um histograma mensal

dos quantitativos de focos de incêndios entre o período de 2015 a 2019 conforme Figura 7.

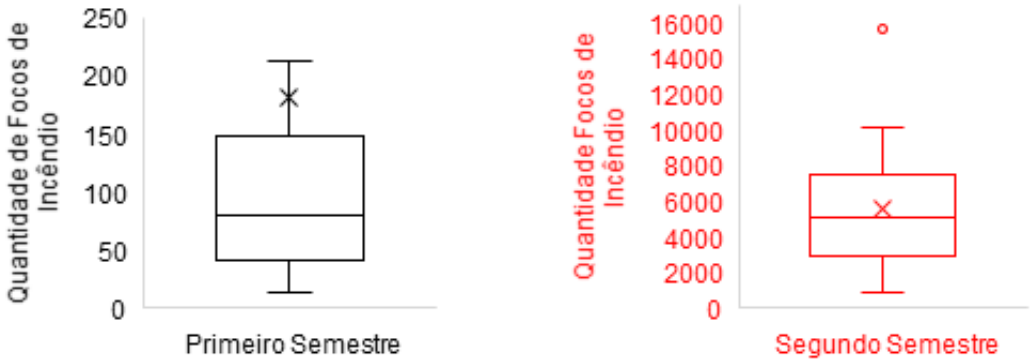
Figura 7 – Distribuição mensal do número de incêndios nos últimos 5 anos (2015-2019), no estado do Pará.



Fonte: Elaborada pelo autor.

Observando a Figura 7 identifica-se que a quantidade de queimadas tem um aumento significativo a partir do mês de Julho até o mês de dezembro (segundo semestre do ano). Durante o período analisado, o segundo semestre possui mais de 96% das queimadas que ocorrem em todo o território paraense. No estado do Pará, por não haver uma definição clara quanto as estações do ano, o segundo semestre corresponde ao período de seca, ou seja, pouca precipitação e temperaturas elevadas. Já no primeiro semestre (janeiro a junho), ocorrem poucas queimadas devido ao aumento da precipitação e consequentemente redução das temperaturas. A Figura 8, foi elaborada para obter uma perspectiva sobre o caráter dos dados a nível semestral.

Figura 8 – Gráfico *Boxplot* representando os dados de queimadas no primeiro e segundo semestre durante o período de 2015 a 2019.



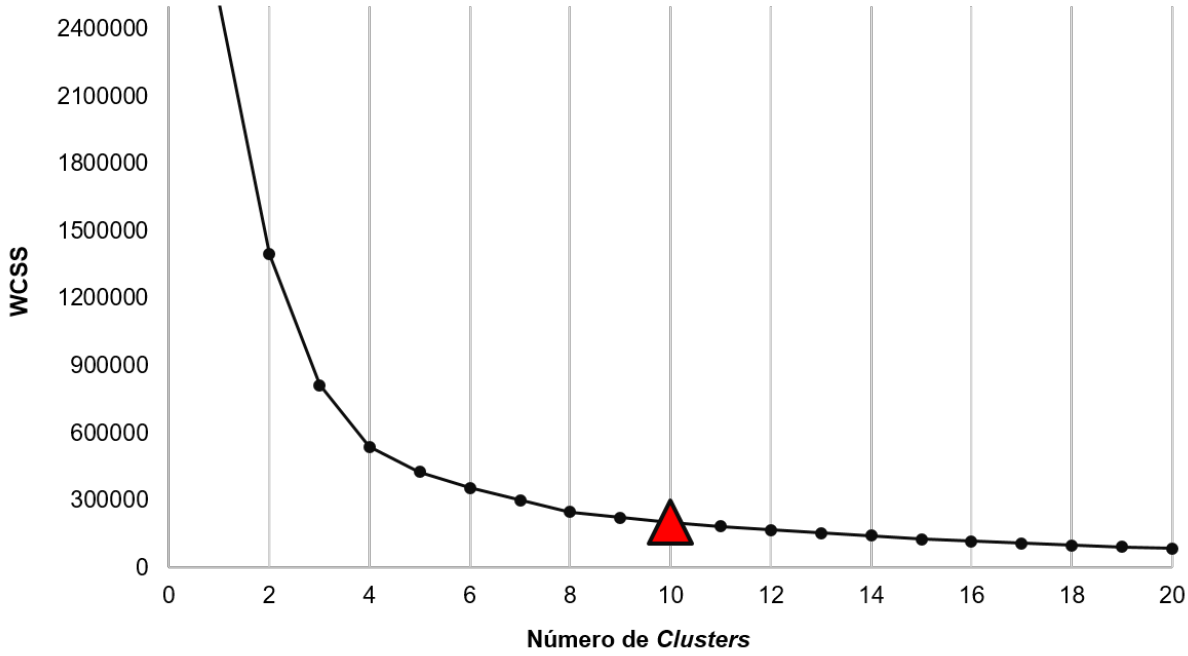
Fonte: Elaborada pelo autor.

A Figura 8, apresenta que a média de queimadas dos meses no primeiro semestre é de aproximadamente 180 focos de incêndios mensais. Já o *Boxplot*, do segundo semestre (gráfico vermelho), apresenta um valor médio de 5,600 focos de incêndios mensais e mediana de 5,106 focos de incêndio. O gráfico do segundo semestre também apresenta um *outlier*. O *outleir* representa a ocorrência de incêndio no mês de setembro de 2017, ano recorde de queimadas no estado do Pará.

5.2 Encontrando *Clusters* Baseado na Similaridade das Ocorrências de Incêndios

Identificado o Estado do Pará como estado líder em quantitativos de queimadas, buscou-se identificar se os incêndios florestais ocorrem com maior frequência em algum grupo de cidades do Estado do Pará e qual a sua localização geográfica, sendo assim, o algoritmo *k-means* foi implementado juntamente com o método cotovelo. A Figura 9 ilustra os resultados do método do cotovelo indicando um número ideal de aglomerados.

Figura 9 – Gráfico de cotovelo com a Soma de Quadrados Dentro do *Cluster* (WCSS) (eixo vertical) para o Estado do Pará e o número de *clusters* (eixo horizontal) criados de 1 a 20. O triângulo vermelho indica o número ideal de *clusters*.

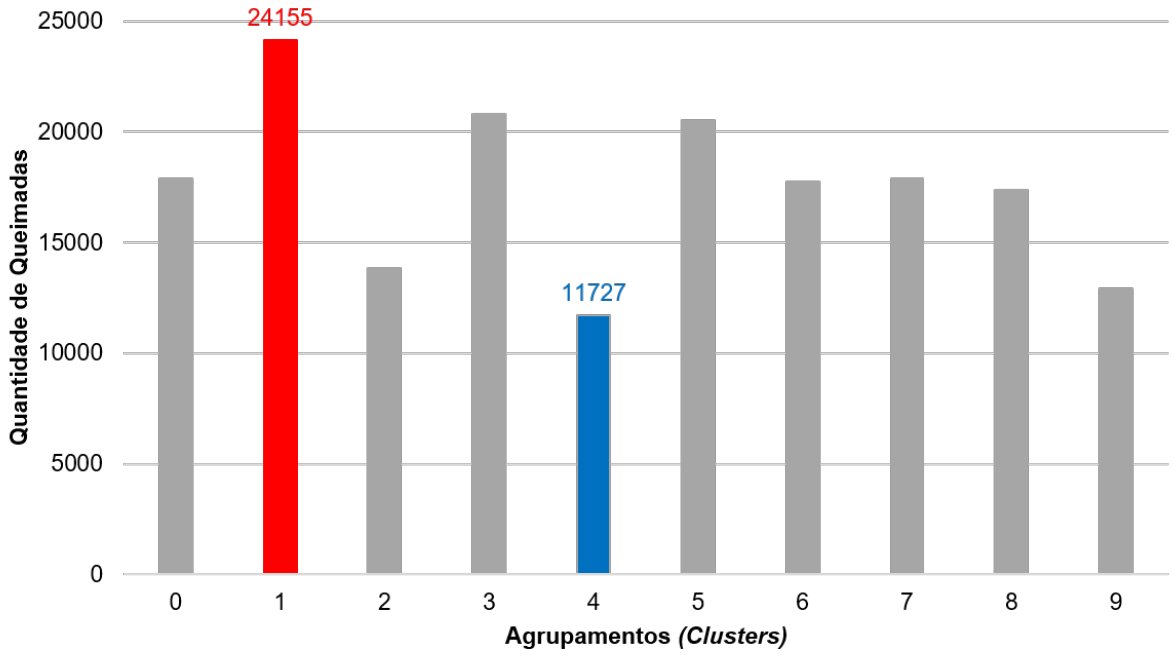


Fonte: Elaborado pelo autor.

A Figura 9 representa o comportamento das coordenadas geográficas das ocorrências de incêndios de acordo com o número de *clusters*, onde determinou-se o número ideal de *clusters* para o Estado do Pará igual a dez (Triangulo Vermelho). Neste caso, foi estipulado para o algoritmo um número máximo de 20 *clusters*, pois aumentar a quantidade de *clusters* no algoritmo *K-means*, permite-nos achar um equilíbrio, onde as observações que formam cada agrupamento sejam o mais homogêneas possíveis e que os agrupamentos formados sejam o mais diferentes um dos outros. Analisando a parte matemática, não escolhemos o número ideal de agrupamentos igual a 4 por exemplo, pois estamos buscando uma quantidade de agrupamentos em que a soma dos quadrados intra-*clusters (WCSS)* seja a menor possível, sendo zero considerado o resultado ótimo. Também não escolhemos o número ideal de agrupamentos igual a 20 pois a partir de 10 *clusters* indicado pelo "cotovelo", observa-se que não existe ganho em relação ao aumento de *clusters*, ou seja, possuem baixo ganho para aumentar a diferenciação dos demais agrupamentos. Através da Figura 10, ilustramos os quantitativos de incêndios para cada um dos 10 aglomerados de focos de incêndio e a Figura 11, ilustra as posições geográficas dos *clusters* identificados

para o Estado do Pará.

Figura 10 – Gráfico com a distribuição da quantidade de incêndios por *cluster* no período de 2015 a 2019. *Cluster* 1 (vermelho) maior frequência de incêndios e *Cluster* 4 (azul) menor frequência de incêndios.

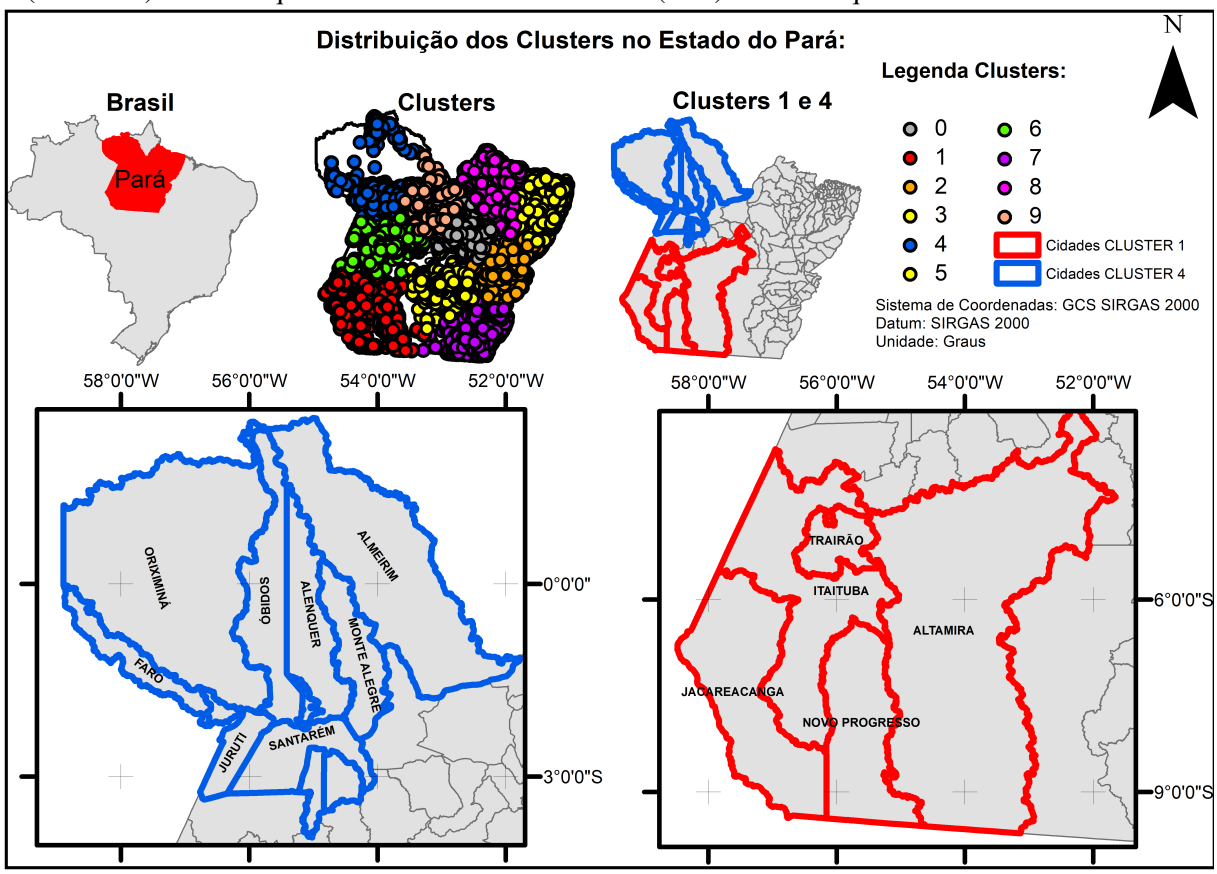


Fonte: Elaborado pelo autor.

Com a referida aplicação da técnica de agrupamento *K-means*, identificamos os grupos de coordenadas onde se concentram as maiores e menores quantidades de ocorrências de incêndio no Estado do Pará. Cada grupo representa ocorrências de incêndio próximas umas das outras, ou seja, existe uma similaridade entre as distancias médias dos focos de queimadas e o centroide de cada *cluster*, uma vez que a distância euclidiana é a medida de similaridade entre as coordenadas. Se tivéssemos considerado um método simples como apenas a contagem de incêndios por municípios, seriam identificados apenas os municípios com mais ocorrências de incêndio, sendo assim, não saberíamos qual a distribuição e proximidade geográfica entre os *clusters* e entre os focos de incêndio intra-*cluster*. Por exemplo, se as ocorrências de incêndio estão próximas à periferia de um município, as consequências do incêndio podem ser ainda mais intensas para o município vizinho, dependendo da proximidade de alguma cidade ou do tamanho do município. Para minimizar este problema, foram considerados todos os municípios que tiveram alguma ocorrência de incêndio no aglomerado identificado.

É possível destacar na Figura 10 que os *clusters* 1 e 4 tiveram o maior número de incêndios e o menor número de incêndios respectivamente. O agrupamento 1 (identificado em vermelho) é formado por cinco municípios com população estimada em 268.831 habitantes no ano de 2020, sendo eles: Altamira, Itaituba, Jacareacanga, Novo Progresso e Trairão; O *cluster* 4 é formado por um conjunto de doze cidades: Alenquer, Almeirim, Belterra, Curuá, Faro, Juruti, Mojuí dos Campos, Monte Alegre, Óbidos, Oriximiná, Santarém e Terra Santa, com população estimada em 710.867 pessoas. Para facilitar a compreensão da distribuição geográfica dos agrupamentos identificados pelo algoritmo, plotamos a Figura 11.

Figura 11 – Mapa com a distribuição geográfica e identificação das cidades dos *clusters* 1 e 4. O *Cluster* 1 (vermelho) maior frequência de incêndios e *Cluster* 4 (azul) menor frequência de incêndios.



Fonte: Elaborado pelo autor.

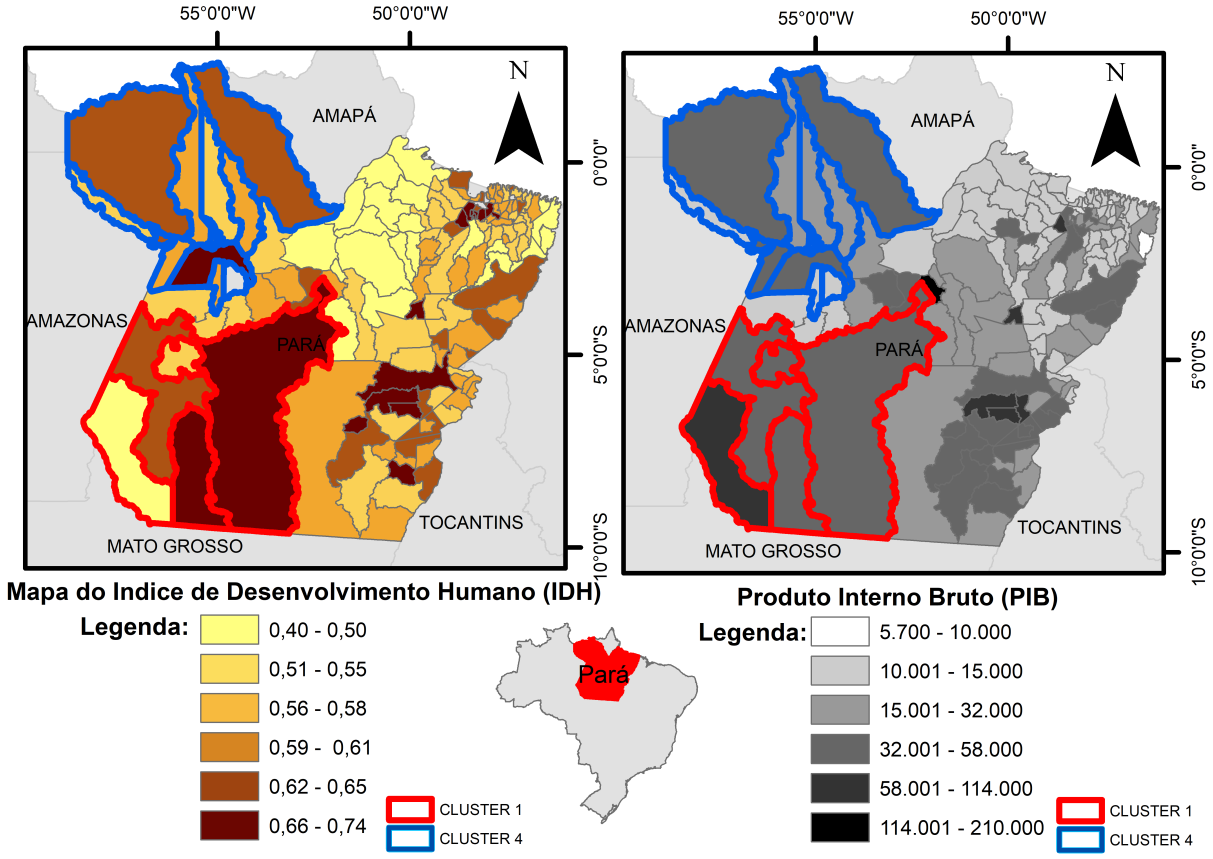
Observando as localizações geográficas dos *clusters* identificados no estado do Pará, observa-se que o agrupamento 1, esta localizado ao SUL e o agrupamento 4 esta localizado ao NORTE do estado. Em média, um município do Pará possui área de 8 651,881 km², porém as cidades que compõem o *cluster* 1 e 4 são municípios de grandes proporções territoriais. O agrupamento de cidades 1, por exemplo, totaliza uma área de 325 033,462 km², representando 26,06% do estado do Pará. Já o *cluster* 4, possui uma área total de 466 567,004 km², representando 35,61% da área total do estado. Portanto, os agrupamentos gerados pelo algoritmo *K-means* nos permitirá realizar as comparações entre os dois *clusters*, visto que, ambos totalizam mais de 50% do estado do Pará, possuem extensas áreas territoriais e estão localizados próximos geograficamente.

5.3 Observando o Índice de Desenvolvimento Humano (IDH) e Produto Interno Bruto (PIB) nos *clusters*

Após a identificação dos grupos de cidades com maior e menor índice de incêndios no Estado do Pará (cluster 1 e 4, respectivamente), analisou-se o Índice de Desenvolvimento Humano (IDH) e o Produto Interno Bruto (PIB) de ambos os *clusters*. Essas variáveis são extremamente importantes para analisar e comparar dois grupos de cidades em relação ao comportamento a doenças, sendo necessário compreender se possuem características sociais e econômicas similares, pois estas variáveis podem estar diretamente

associadas a um maior impacto das doenças. Na Figura 12 a distribuição do IDH e do PIB no Estado do Pará.

Figura 12 – IDH e PIB dos municípios do Estado do Pará.



Fonte: Elaborado pelo autor.

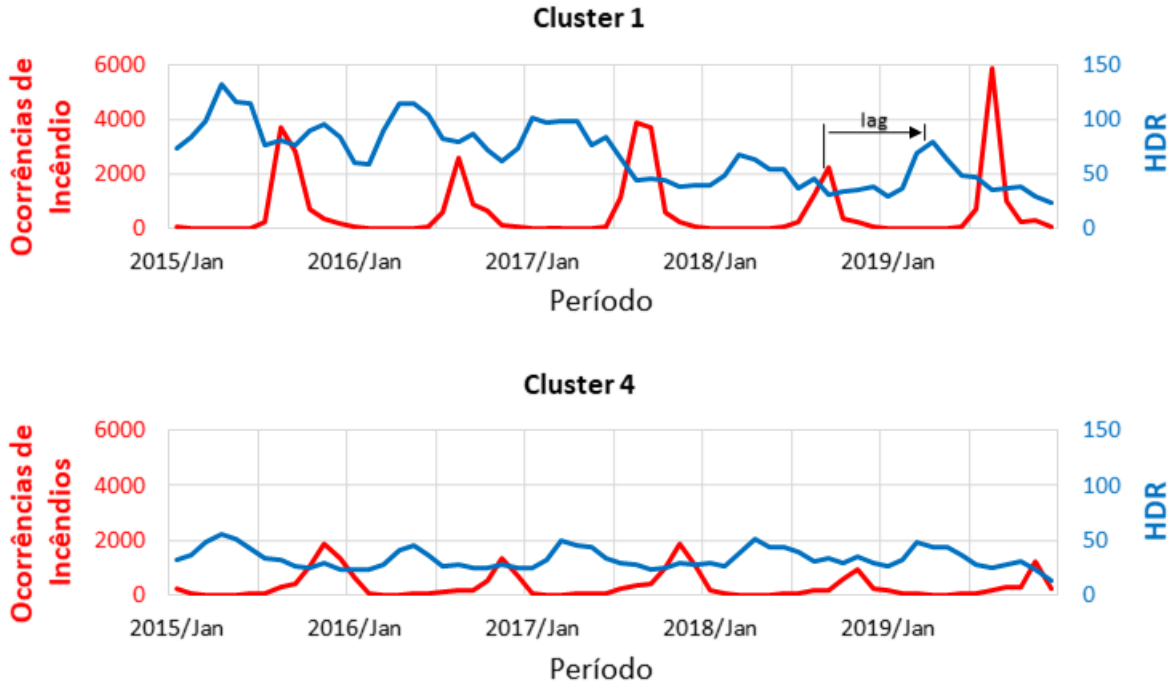
Os dados da Figura 12, indicam através do mapa de calor que o *cluster* 1 possui um maior IDH e PIB do Pará, embora a densidade demográfica seja diferente entre eles. O grupo de cidades que compõem o agrupamento 1 possui um valor médio de IDH igual a 0,609, PIB médio de 24.634 e o agrupamento 4 apresenta uma valor médio de IDH igual a 0,604 e PIB médio igual a 14.924. Ambos os *clusters* possuem grandes áreas territoriais com IDH e PIB elevados em relação aos demais *clusters* do Estado do Pará, contribuindo assim, para uma comparação justa entre ambos os grupos de cidades identificados.

5.4 Análise de séries temporais para investigar a influência das ocorrências de incêndio nas HDR

Identificados os agrupamentos a serem analisados e comparados, foram coletados os dados de HDR para compor as as bases de dados. Antes de construir o modelo de série temporal para analisar a correlação entre as queimadas e HDR, a função de correlação cruzada foi investigada para identificar o tempo de defasagem entre os picos de ocorrência de incêndios e os picos de HDR, pois ambos os picos não ocorrem de forma simultânea. Observando os dados de queimadas em ambos os *clusters* é possível identificar que os picos das ocorrências de incêndios em ambos os agrupamentos não ocorrem exatamente no mesmo período. No *cluster* 1 ocorrem entre os meses de julho a setembro. Já no *cluster* 4 o pico de queimadas ocorrem no mês de outubro a dezembro. Sendo assim, identificamos uma defasagem de

aproximadamente 6 meses que pode ser visualizada na Figura 13, juntamente com os dados de HDR e ocorrências de incêndio considerando os *clusters* 1 e 4 respectivamente.

Figura 13 – HDR e séries temporais de ocorrências de incêndio para os *clusters* 1 e 4 respectivamente.



Fonte: Elaborada pelo autor.

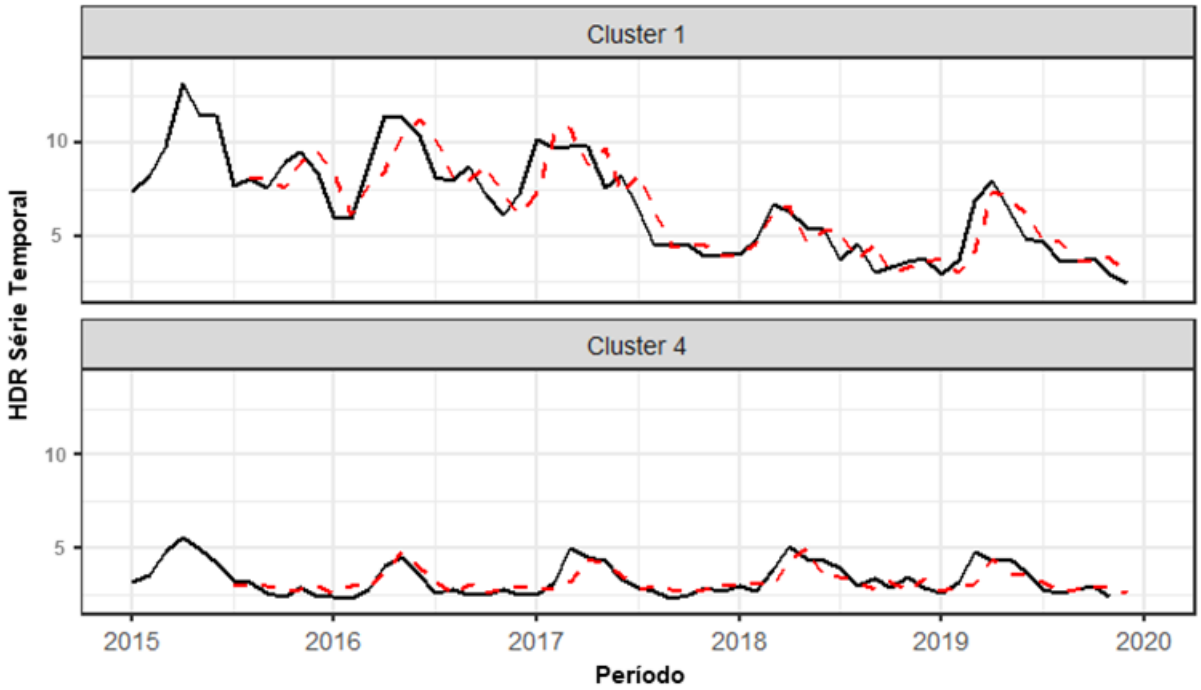
Observando a Figura 13, em relação a série temporal das queimadas em ambos os *clusters* (linhas vermelhas), observa-se que diferentemente do gráfico inicial, representado na Figura 7 de distribuições mensais de queimadas em todo o estado do Pará, as queimadas nos agrupamentos 1 e 4 não ocorrem durante todo o semestre. Porém como já identificado, as queimadas não ocorrem de forma simultâneas.

Analisando as series temporais dos dois agrupamento em relação as HDR (linhas azuis), podemos observar que os picos das hospitalizações também possuem um diferenciação em relação ao mês em que os picos ocorrem. Como já é conhecido cientificamente, cada estado/região do Brasil apresenta variação nos períodos em que as doenças respiratórias se manifestam com maior incidência, estando diretamente associadas com as variações climáticas de cada região. Porém, analisando as séries temporais, podemos observar que no *cluster* 4, região com menos incidência de queimadas, possui um padrão de comportamento parecido em relação ao espaço/tempo analisados, ou seja, as queimadas e hospitalizações ocorrem todos os anos, praticamente nos mesmos períodos e com intensidades similares. Já o agrupamento 1, também podemos observar que os picos de queimadas e HDR ocorrem nos mesmos períodos, porém com uma maior variação de intensidades aos anos analisados. Também podemos observar que na série temporal 1, ocorre um pequeno aumento de HDR logo após o pico de queimadas, aspecto esse que não ocorre no *cluster* 4.

Na Figura 14 são apresentados os modelos HDR (linhas pretas), bem como os modelos ARIMAX para os *clusters* 1 e 4 respectivamente (linhas vermelhas tracejadas). Os parâmetros estimados para avaliação dos modelos estão identificados na na Tabela 2. A variável relacionada aos aerossóis também poderia ser incluída no modelo, mas por ser altamente correlacionada com a ocorrência de incêndios,

não foi adicionada para evitar a multicolinearidade.

Figura 14 – Séries temporais de taxa de internação por doenças respiratórias gerais (HDR) (linha preta) e modelos ARIMAX (linhas tracejadas vermelhas) nos *clusters* 1 e 4. Os dados de HDR no eixo y representam as taxas de hospitalizações por 10 mil habitantes.



Fonte: Elaborado pelo autor.

Tabela 2 – Parâmetros Estimados, Desvio Padrão (DP), e Valor-p para o modelo ARIMAX para as séries temporais de HRD do C1 e C4 no estado do Pará. Quanto ao cluster 4, a série temporal de HDR não foi diferenciada $(d = 0)$ e possui coeficientes de média móvel (β_1), um termo constante (média de HDR) precisa ser incluído neste modelo.

		Estimado	Desvio Padrão (DP)	Valor-p
Cluster 1	Ocorrências de Incêndios	0.0005	0.0002	0.0189
Cluster 4	μ	2.8545	0.1338	<0.0001
	β_1	0.5811	0.1119	<0.0001
	Ocorrências de Incêndios	0.0009	0.0003	<0.0001

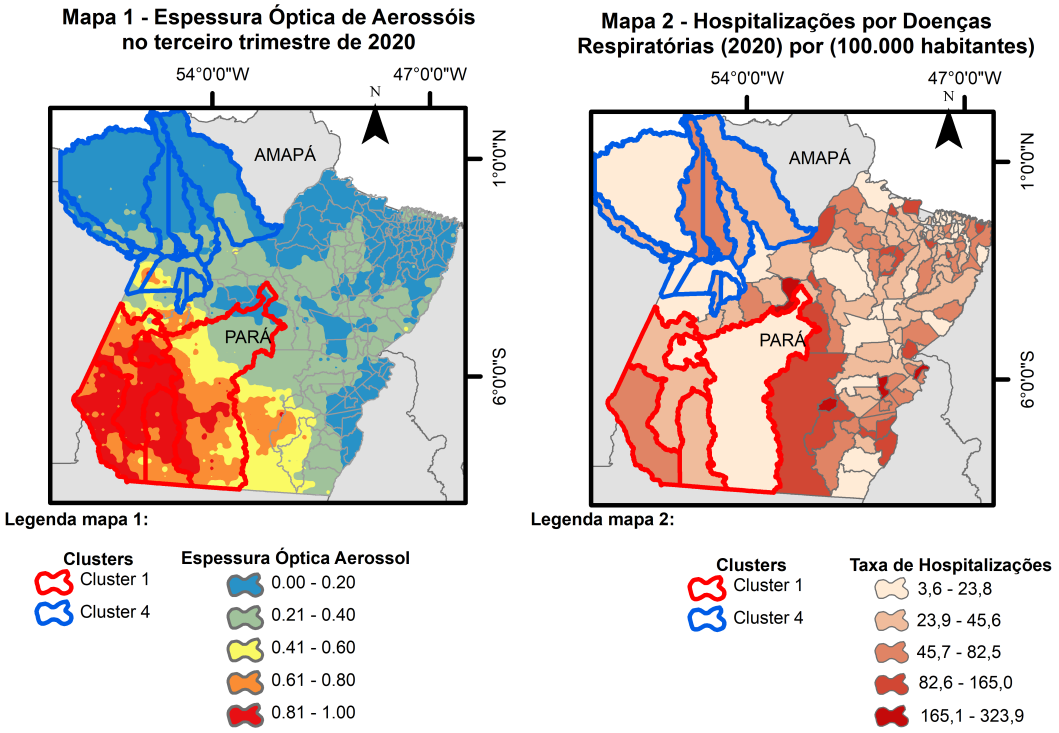
Verifica-se que na Figura 14 que os modelos propostos pelo ARIMAX (linhas tracejadas vermelhas) explicam bem os dados de HDR. A partir da Tabela 2 confirmamos que as ocorrências de incêndios com a defasagem em cerca de 6 meses influenciam no aumento do HDR em ambos os *clusters*. Para identificação da defasagem ideal, computamos através da correlação cruzada as defasagens de -4 a -8 meses, onde -6 meses representou a melhor correlação. Esse aumento pode ser interpretado através dos valores de estimativas positivas na Tabela 2, ou seja, a aumento de queimadas induz ao aumento de HDR. Observando a Figura 14, apenas a série temporal HDR em C1 apresentou uma tendência (de

diminuição), ou seja, é uma série temporal não estacionária exigindo diferenciação. Uma diferença ($d = 1$) foi suficiente para induzir a estacionariedade. Por isso, a média do HDR não foi estimada (muito próxima de zero após a diferenciação) em C1. Para C4, a média de HDR estimada durante o período investigado foi de 2,85 (Tabela 2) e o parâmetro de média móvel representa a correlação temporal da série temporal de HDR.

5.5 Associação de incêndios com mortalidade por SARS-Cov-2 e Doenças Respiratórias

Para verificar o impacto do SARS-Cov-2 nas famílias residentes nos *clusters* 1 e 4, foram elaborados mapas relacionados ao comportamento dessas duas regiões durante o ano de 2020. Na Figura 15, o mapa 1 representa a distribuição de poluentes atmosféricos no Estado do Pará durante o ano de 2020 (os *clusters* 1 e 4 estão em destaque).

Figura 15 – O mapa 1 representa a espessura óptica (menos de 0,1 indica um céu cristalino com alta visibilidade; 1 indica a presença de aerossóis densos). O Mapa 2 representa o HDR em 2020.



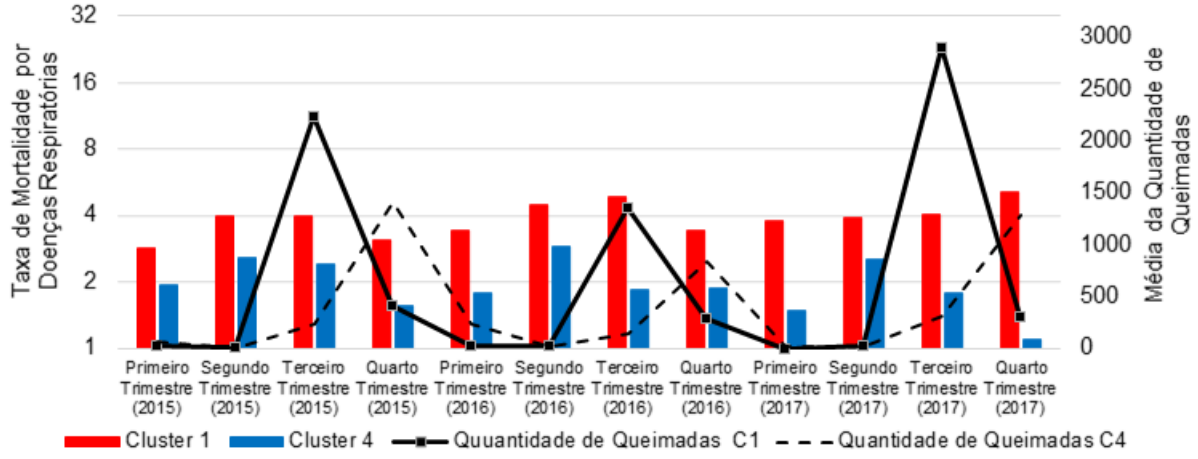
Fonte: Elaborado pelo autor.

Constatou-se no mapa 1, que o *cluster* 1 apresenta maiores valores de poluentes atmosféricos considerando o terceiro trimestre do ano de 2020, o que está de acordo com o padrão de queimadas anuais nesta região. Essa ilustração colabora com a validação dos agrupamentos identificados no algoritmo K-means, onde observa se o *cluster 4*, com o ar extremamente límpido e o ar do *cluster* 1 extremamente poluído. No mapa 2 da Figura 15 os HDR no ano de 2020 em todo o Estado do Pará estão representados mostrando uma redução no número de HDR para o ano de 2020 com um valor médio de 37 hospitalizações por 100,000 mil habitantes no *cluster 1*, e valor médio de 31 hospitalizações no *cluster* 4. Quando comparado ao HDR médio de 2015-2019 (67 hospitalizações mensais por 100,000 mil habitantes no

cluster 1 e 32 no *cluster* 4), observa-se numericamente essa redução. Porém, os valores de hospitalizações para o ano de 2020 podem ter sido afetados pelo surgimento da pandemia, onde pessoas com sintomas deixaram de procurar assistência hospitalar devido as restrições quanto a pandemia.

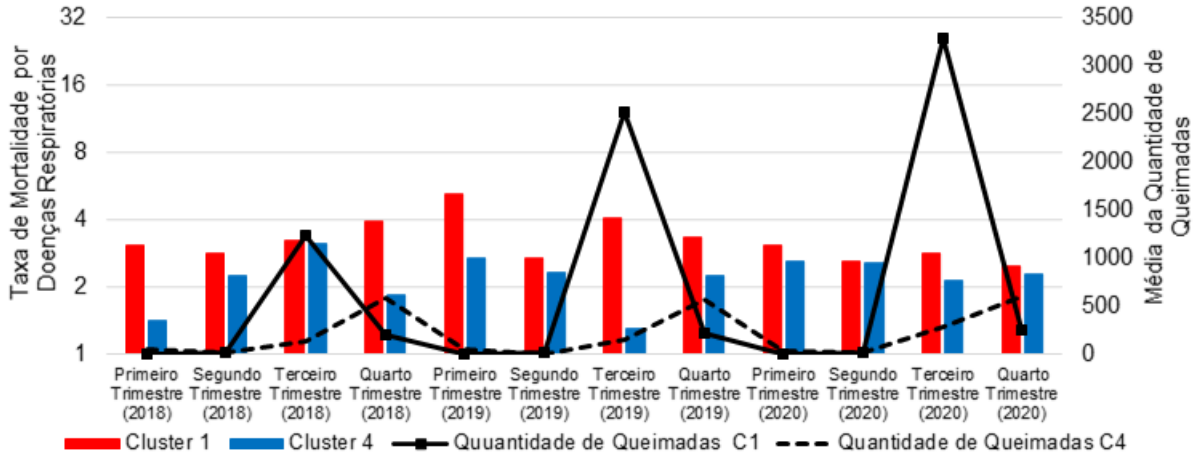
Para ver o padrão histórico de mortalidade por doenças respiratórias e COVID-19, elaborou-se as Figuras 16 e 17 que mostram as taxas de mortalidade por doenças respiratórias para os trimestres dos anos a partir de 2015 a 2020, e as Figuras 18 e 19 referentes as taxas de mortalidade por COVID-19 por trimestre no ano de 2020. As ocorrências de incêndio também são mostradas nestas figuras para os outros trimestres.

Figura 16 – Série temporal da taxa de mortalidade por doenças respiratórias período de 2015 a 2017. As barras azul e vermelha referem-se ao *Cluster* 1 e 4, respectivamente. Linhas pretas contínuas e tracejadas referem-se a ocorrências de Incêndio nos *clusters* 1 e 4, respectivamente.



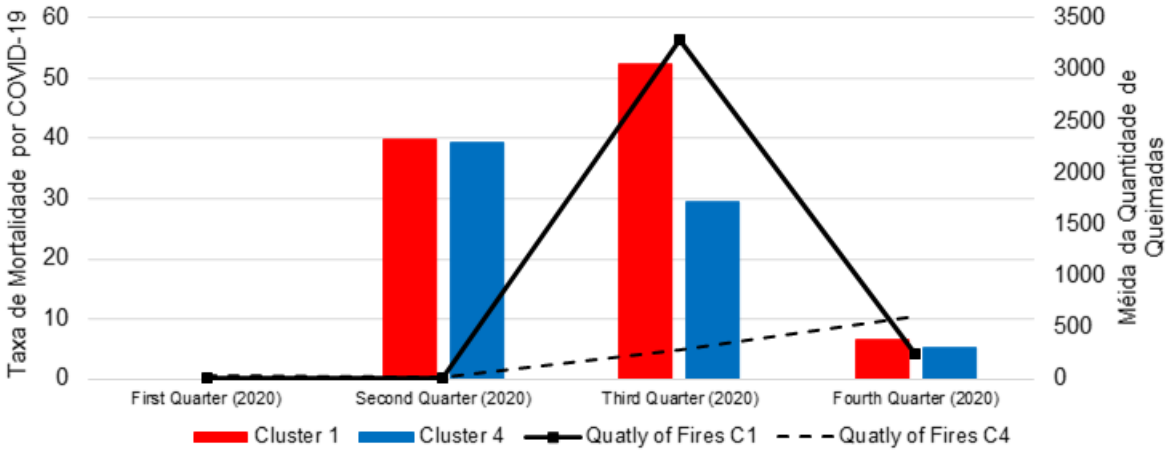
Fonte: Elaborado pelo autor.

Figura 17 – Série temporal da taxa de mortalidade por doenças respiratórias período de 2018 a 2020. As barras azul e vermelha referem-se ao *Cluster* 1 e 4, respectivamente. Linhas pretas contínuas e tracejadas referem-se a ocorrências de Incêndio nos *clusters* 1 e 4, respectivamente.



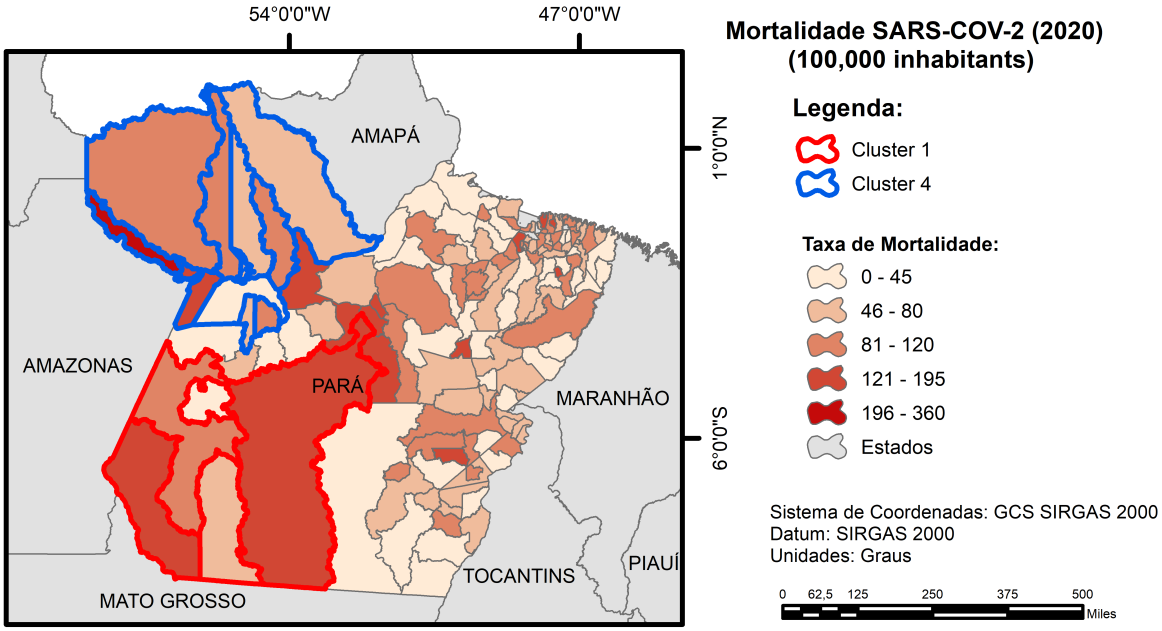
Fonte: Elaborado pelo autor.

Figura 18 – Série temporal da taxa de mortalidade por COVID-19 no ano de 2020.



Fonte: Elaborado pelo autor.

Figura 19 – Taxa de mortalidade por SARS-Cov-2 no ano de 2020 no Estado do Pará.



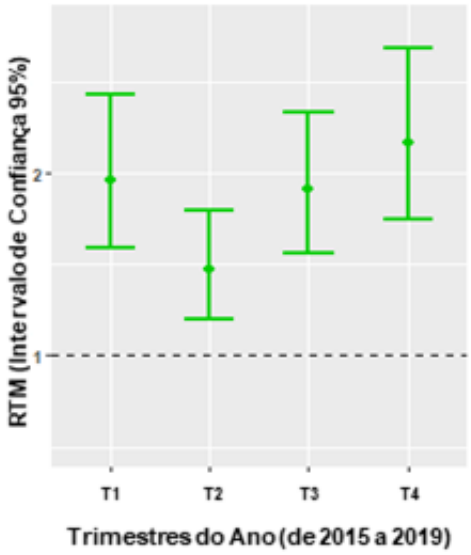
Fonte: Elaborado pelo autor.

Na Figura 19, pode-se observar através do mapa de calor, que as taxas de mortalidade mais altas (cidades com preenchimento em vermelho forte) são observadas no *cluster 1*, ou seja, região também identificada com altos índices de incêndios florestais.

É possível identificar que a taxa de mortalidade por doenças respiratórias é maior no *cluster* 1 onde há mais ocorrências de incêndios independentemente do período do ano (período de 2015 a 2020). Em relação à mortalidade por COVID-19, a taxa também é maior no *cluster* 1 do que no *cluster* 4, principalmente no terceiro trimestre quando as ocorrências de incêndio são maiores do que nos demais trimestres. Assim, permite levantar uma possível vulnerabilidade a complicações respiratórias em populações residentes no *cluster* 1.

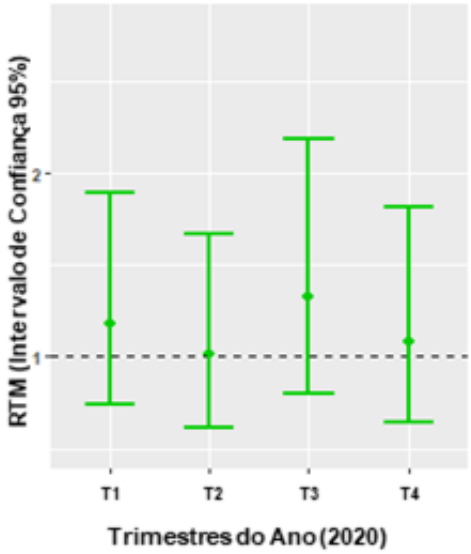
Considerando os dados apresentados nas Figuras 16, 17 e 18, foram computadas as RTMs (Equação 4.6) e seus IC (Equação 4.7). Para as taxas de mortalidade por doenças respiratórias no período de 2015 a 2019 (antes da pandemia), apresentamos as RTMs na Figura 20. Em 2020, as RTMs podem ser vistas separadamente para a mortalidade por doenças respiratórias e mortalidade por COVID-19 nas Figuras 21 e 22, respectivamente. Nas figuras o ícone verde representa o intervalo de confiança (IC de 95%) e o ponto verde representa a RTM. Já a linha preta tracejada (valor 1 do eixo y), corresponde a não variação (maior mortalidade) nas taxas de mortalidade entre os *clusters*.

Figura 20 – RTM para a mortalidade por doenças respiratórias nos trimestres compreendidos entre os anos de 2015 e 2019.



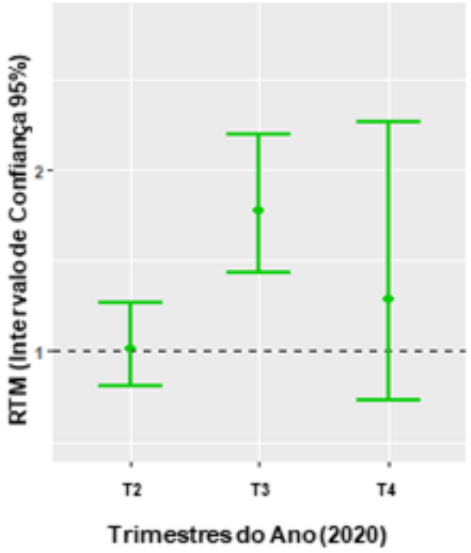
Fonte: Elaborado pelo autor.

Figura 21 – RTM para a mortalidade por doenças respiratórias nos trimestres do ano de 2020, sem considerar a mortalidade por COVID-19.



Fonte: Elaborado pelo autor.

Figura 22 – RTM por COVID-19 nos trimestres de 2020.



Fonte: Elaborado pelo autor.

A Figura 20 demonstra que a taxa de mortalidade foi aproximadamente duas vezes maior no *cluster* 1 em todos os trimestres, exceto no trimestre 2, durante o período de 2015-2019, ou seja, a população residente do C1 possui um risco a mortalidade superior em odos os trimestres do ano em relação ao agrupamento C4, sendo isso possivelmente associado a poluição e fragilidade respiratória oriunda das frequentes queimadas florestais. A Figura 21 mostra que as taxas de mortalidade por doenças respiratórias não foram estatisticamente diferentes entre os *clusters* ao longo do ano de 2020, podendo ser observado pelo valor 1 estar contido dentro do intervalo de confiança.

Já para a Figura 22, o segundo e o quarto trimestre não são estatisticamente significativos em relação a uma maior mortalidade por COVID-19 entre os *clusters*. Porém, observa-se que a mortalidade por COVID- 19 foi 1,8 vezes maior no *cluster* 1 no trimestre 3, período em que a ocorre a maior incidência de queimadas, indicando que o risco de mortalidade no terceiro trimestre foi cerca de 80% maior no *cluster* 1 em comparação com o *cluster* 4.

6 DISCUSSÃO E CONSIDERAÇÕES FINAIS

O estudo da sobreposição entre incêndios florestais e a pandemia relacionada ao SARS-Cov-2 mostrou a necessidade urgente de determinar e compreender a relação entre incêndios e doenças respiratórias regionais e, finalmente, a relação com complicações respiratórias da COVID-19. Isso é extremamente fundamental para orientar as ações de saúde pública em diferentes localidades considerando a vulnerabilidade da população específica. O Estado do Pará é o segundo maior estado do Brasil e o maior em número de queimadas (INSTITUTO BRASILEIRO DE GEOGRAFIA ESTATÍSTICA - IBGE - 2020).

A partir de 2015, foi identificada a intensificação do fenômeno El Niño, causando o aquecimento das águas na superfície do Pacífico, levando à supressão das chuvas no leste da Amazônia e aumento do risco de queimadas, principalmente nos Estados do Maranhão, Mato Grosso e Pará, causando secas prolongadas potencializando a abrangência territorial dos incêndios florestais (JIMÉNEZ-MUÑOZ et al., 2016). Além disso, no ano de 2020 foram registrados as maiores ocorrências de incêndios florestais no território amazônico, em relação aos últimos dez anos, afetando diretamente a saúde respiratória da população e agravando a vulnerabilidade da população para o confronto ao surgimento de novas doenças respiratórias (JUNIOR et al., 2020).

A ocorrência de grandes incêndios no Brasil e no mundo tem chamado principalmente a atenção para o problema ambiental, mas as medidas de prevenção e controle de seus impactos ainda são insuficientes (KAREN; W., 2012). A queima de biomassa produz poluentes gasosos e partículas finas, que produzem efeitos prejudiciais ao sistema respiratório (CARMO et al., 2010). Entre os principais componentes da fumaça oriunda dos incêndios florestais que podem afetar a qualidade do ar, é importante mencionar as partículas finas PM2,5 e PM10 (partículas com diâmetro inferior a 2,5 e 10 micrômetros respectivamente), cujos efeitos devem aumentar ainda mais quando suas concentrações ficam acima dos padrões de qualidade do ar estabelecidos (YOUSSOUF et al., 2014).

A identificação das populações em risco de complicações respiratórias é de fundamental importância para o controle da pandemia de SARS-Cov-2 e outras futuras pandemias (SOLIMINI et al., 2021). Estudos epidemiológicos relacionados a poluentes do ar e mortes por infecção por SARS-Cov-2 estão recebendo muita atenção científica. Novas metodologias de interpretação de variáveis que podem estar associadas para com o sistema respiratório e que podem exacerbar os sintomas da COVID-19 e aumentar o risco de cobertura, permitem que órgãos públicos e ambientais tomem medidas preventivas (AVEYARD et al., 2021). O coronavírus (SARS-Cov-2) está gerando maior taxa de mortalidade e óbitos principalmente em pessoas com comorbidades. Por exemplo, segundo Abelsohn e Stieb (2011), a poluição do ar causa morbidade e mortalidade significativas, podendo afetar o sistema respiratório (exacerbação da asma e doença pulmonar obstrutiva crônica) e o sistema cardiovascular (arritmia de compensação, insuficiência cardíaca e derrames).

A poluição gerada por incêndios florestais gera partículas conhecidas como aerossóis e a exposição a longo prazo a essas partículas aumenta a gravidade dos resultados associados a hospitalizações e mortalidade associada ao COVID-19, conforme também identificado por Wu et al. (2020). A pandemia global de SARS-Cov-2 tem sido associada a infecções e mortes entre pessoas em ambientes poluídos. Os aerossóis inalados por pessoas com doenças respiratórias têm mostrado significância com o aumento das internações segundo Fennelly (2020), e os incêndios florestais contribuem muito para esse fator. Além disso, vários estudos relataram que o vírus SARS-Cov-2 tem um impacto maior em pessoas que já apre-

sentam alguma comorbidade, por exemplo. asma, que é mais prevalente em regiões poluídas (HU et al., 2021).

A análise de dados gratuitos e públicos relacionados a aspectos ambientais, internações e temas atuais relacionados ao SARS-Cov-2 permitem a elaboração de uma abordagem para estudar potenciais associações entre exposição histórica à poluição atmosférica e o aumento da vulnerabilidade às mortes por COVID-19. Uma das limitações no desenvolvimento dos estudos é a falta de detalhes mais específicos das internações por doenças respiratórias e dados sobre vulnerabilidades pré-existentes em óbitos por SARS-Cov-2. A análise regionalizada proposta neste estudo não garante a replicabilidade dos resultados para grandes populações representativas, ou mesmo para outras cidades do Estado do Pará. Nosso estudo utiliza abordagens de espaço e tempo para analisar eventos regularmente aplicados em diversas áreas de pesquisa. Usando nosso estudo como exemplo, resumimos as limitações não mensuradas e os fatores de confusão para pesquisas futuras.

Entre as principais limitações ao trabalhar com dados de saúde pública relativos a óbitos por SARS-Cov-2, praticamente em tempo real, estão as restrições de dados e fatores de risco individuais como idade, raça, comorbidades e dados sobre tabagismo, por exemplo. Considerando a escala espacial, não foi possível realizar neste trabalho análises em nível individual, o que pode afetar a associação de dados entre grupos de cidades. Além disso, os dados sobre o diagnóstico de COVID-19 ou a taxa de letalidade carecem de precisão devido às políticas de testagem irregulares ou informações mais precisas disponíveis em alguns casos, o que pode tornar os resultados sistematicamente tendenciosos. Apesar disso, este trabalho traz resultados muito importantes contribuindo para uma melhor compreensão do comportamento das doenças respiratórias e por SARS-Cov-2 em áreas com alta incidência de queimadas, destacando a importância de cuidados específicos nas áreas mais vulneráveis.

Com relação à aplicação de técnicas de aprendizado de máquina para descoberta de conhecimento em uma grande quantidade de dados, o método proposto combinando o algoritmo *k-means*, o método cotovelo e a análise de séries temporais delineou com sucesso e corroborou com a hipótese anterior de que a população exposta à poluição dos incêndios florestais geram maiores taxas de hospitalizações e e suscetibilidade a óbitos durante a pandemia de SARS-COV-2. Trabalhos anteriores tentaram fazer essa correlação ou sugerir técnicas de aprendizado de máquina para extrair tais informações como no trabalho de Al Ferdous et al. (2020) que apenas sugere várias técnicas de clusterização para obter *insights* relacionados ao COVID-19, o trabalho de Parente, Pereira e Tonini (2016) que realizou uma análise do espaço-tempo sobre dados de incêndios florestais, e o trabalho de Shafi e Waheed (2020) que empregou o algoritmo *k-means* para agrupar e analisar dados de poluição do ar. No entanto, nenhum desses trabalhos trouxe *insights* sobre doenças respiratórias e o SARS-COV-2, o que destaca a importância de nossos achados em relação a políticas públicas relacionadas à gestão da saúde e meio ambiente.

Estudos elaborados por Dash, Sethi e Dash (2021) e ROMERO (2020) apresentam que populações fragilizadas com baixa concentração de renda, de infraestrutura e serviços (variáveis diretamente associadas ao PIB e IDHM baixos) proporcionam condições para uma maior ocorrência de doenças infecciosas e consequentemente impactos mais elevados na saúde populacional. Observando o IDH e PIB dos agrupamentos de cidades identificados neste estudo, entende-se que as regiões analisadas são similares e com valores socio-econômicos elevados, apenas diferindo no número de queimadas e poluição atmosférica, aspecto esse, que pode estar diferenciando os grupos em relação aos impactos das doenças respiratórias e a mortalidade no surgimento da pandemia.

O estudo de séries temporais com o modelo ARIMAX permitiu compreender a validar a fragilidade populacional dos residentes no C1 e compreender comportamento sazonal, entre o pico das queimadas e o pico das hostilizações por doenças respiratórias. Observa-se que o período critico para as queimadas esta no terceiro trimestre (Julho,agosto e setembro de 2015 a 2019) e o mesmo ocorreu no ano de 2020, anos do surgimento da pandemia associada ao SARS-COV-2. Compreender os períodos em que os picos e incrementos na incidência das doenças ocorrem, permite uma melhor preparação da área da saúde visando o enfrentamento das doenças.

Durante o surto respiratório agudo grave associado ao coronavírus 1 da síndrome respiratória aguda grave (SARS-Cov-1) em 2003, pacientes de áreas com altos níveis de poluição do ar exibiram um aumento de 200% no risco relativo de morte em comparação com as pessoas vivendo em áreas com baixo teor de poluição (A. Karan and K. Ali and S. Teelucksingh and S. Sakhamuri, 2020). Em um estudo realizado na região do oeste dos Estados Unidos verificou-se a associação entre a exposição de curto prazo ao PM 2,5 durante os incêndios florestais e a dinâmica epidemiológica dos casos e óbitos por COVID-19 e fortes evidências de que os incêndios florestais amplificaram o efeito da exposição de curto prazo ao PM 2,5 nos casos e mortes por COVID-19 (ZHOU et al., 2021).

O estudo proposto, permitiu compreender o comportamento da população residente em relação a doenças respiratórias, antes da pandemia (2015-2019) e seus impactos no primeiro ano de pandemia (2020). A metodologia proposta, permitiu encontrar grupos de cidades baseado nas ocorrências de incêndios, desviando de unidades territoriais pré estabelecidas pelo IBGE. Observou-se de fato um aumento de 80% no risco a mortalidade por COVID-19 no grupo de cidades mais afetados pelos incêndios florestais no terceiro trimestre de acordo com a hipótese pré-estabelecida.

Com este estudo, podemos perceber que no Estado do Pará houve uma sobreposição de situações relacionadas às doenças respiratórias. Populações que residem em locais com problemas ambientais tendem a ter maiores impactos no surgimento de novas doenças. Conhecer as populações é de fundamental importância para a aplicação de políticas de controle e prevenção. Essa metodologia proposta, visa facilitar a identificação de áreas de risco para melhores tomadas de decisões. Aplicar somente à COVID-19 toda a responsabilidade pelo alto índice de mortalidade pode não ser a forma mais correta de interpretar os impactos da pandemia, pois existem fatores potencializadores, neste caso, o alto índice de internações por doenças respiratórias associadas à prática de queimadas.

REFERÊNCIAS

A. Karan and K. Ali and S. Teelucksingh and S. Sakhamuri. The impact of air pollution on the incidence and mortality of covid-19. **Glob. Health Res. Policy,**, [S.l.], 2020.

ABATZOGLOU, J. T.; WILLIAMS, A. P. Impact of anthropogenic climate change on wildfire across western us forests. **Proceedings of the National Academy of Sciences**, [S.l.], v. 113, n. 42, p. 11770–11775, 2016.

ABELSOHN, A.; STIEB, D. Health effects of outdoor air pollution: approach to counseling patients using the air quality health index. **Can Fam Physician**, [S.l.], v. 57, n. 8, p. 881–887, 2011.

ADRIAANS, P.; ZANTINGE, D. **Data mining**. 1nd. ed. Harlow-England: Addison-Wesley Professional, 1996.

AGHABOZORGI, S.; Seyed Shirkhorshidi, A.; Ying Wah, T. Time-series clustering – a decade review. **Information Systems**, [S.l.], v. 53, p. 16–38, 2015.

AL FERDOUS, F. et al. A conceptual review on different data clustering algorithms and a proposed insight into their applicability in the context of covid-19. **Journal of Advances in Technology and Engineering Research**, [S.l.], v. 6, n. 2, p. 58–68, 2020.

ALVES, H. J. d. P. et al. The covid-19 pandemic in brazil: an application of the k-means clustering method. **Research, Society and Development**, [S.l.], v. 9, n. 10, p. e5829109059, Oct. 2020.

ALVES, L. Amazon fires coincide with increased respiratory illnesses in indigenous populations. **The Lancet Respir. Med.**, [S.l.], v. 8, p. e84, 2020.

ANDRADE, G. D. et al. Mortality profile associated with pandemic infection by sars-cov-2 in a public hospital in the southern region o western amazonia. **Research, Society and Development**, [S.l.], v. 10, n. 13, p. e288101321359, Oct. 2021.

ARTAXO, P. et al. Atmospheric aerosols in amazonia and land use change: from natural biogenic to biomass burning conditions. **Faraday Discuss.**, [S.l.], v. 165, p. 203–235, 2013.

AVEYARD, P. et al. Association between pre-existing respiratory disease and its treatment, and severe covid-19: a population cohort study. **The Lancet**, [S.l.], 2021.

BAQUI, P. et al. Ethnic and regional variations in hospital mortality from covid-19 in brazil: a cross-sectional observational study. **The Lancet Global Health**, [S.l.], v. 8, n. 8, p. e1018–e1026, 2020.

BARLOW, J. et al. The critical importance of considering fire in redd+ programs. **Biological Conservation**, [S.l.], v. 154, p. 1–8, 2012. REDD+ and conservation.

BARROS, G. M. d.; BARROS, G. M. d. Covid-19 in northeast brazil: preliminary characteristics of deaths. **Research, Society and Development**, [S.l.], v. 9, n. 11, p. e89291110166, Dec. 2020.

BOX, G.; JENKINS, G.; REISEL, G. **Time series analysis-wiley series in probability and statistics**. [S.l.]: New Jersey: John Wiley & Sons, 2008.

BRANDO, P. M. et al. Droughts, wildfires, and forest carbon cycling: a pantropical synthesis. **Annual Review of Earth and Planetary Sciences**, [S.l.], v. 47, n. 1, p. 555–581, 2019.

BRANDO, P. M. et al. The gathering firestorm in southern amazonia. **Science Advances**, [S.l.], v. 6, n. 2, 2020.

BUTT, E. W. et al. Large air quality and human health impacts due to amazon forest and vegetation fires. **Environmental Research Communications**, [S.l.], v. 2, n. 9, p. 095001, sep 2020.

BUTT, E. W. et al. Large air quality and public health impacts due to amazonian deforestation fires in 2019. **GeoHealth**, [S.l.], v. 5, n. 7, p. e2021GH000429, 2021. e2021GH000429 2021GH000429.

CANO-CRESPO, A.; TRAXL, D.; THONICKE, K. Spatio-temporal patterns of extreme fires in amazonian forests. **The European Physical Journal Special Topics**, [S.l.], p. 1–12, 2021.

CARMO, C. do et al. Associação entre material particulado de queimadas e doenças respiratórias na região sul da amazônia brasileira. **Rev Panam Salud Publica**, [S.l.], v. 27, p. 10–16, 2010.

CARRERO, C. et al. Deforestation trajectories on a development frontier in the brazilian amazon: 35 years of settlement colonization, policy and economic shifts, and land accumulation. **International Journal of Environmental Management**, [S.l.], v. 66, p. 966–984, 2020.

CHANG, W. L.; GRADY, N. et al. Nist big data interoperability framework: volume 1, big data definitions. , [S.l.], 2015.

CIOS K. J.; KURGAN, L. A. P. W. S. R. W. **Data mining**: a knowledge dicovery approach. USA: Springer, 2007.

CLARK, A. et al. Global, regional, and national estimates of the population at increased risk of severe covid-19 due to underlying health conditions in 2020: a modelling study. **The Lancet Global Health**, [S.l.], v. 8, n. 8, p. e1003–e1017, 2020.

CORTêS S DA C; PORCARO, R. M. L. S. Mineração de dados funcionalidades, técnicas e abordagenss. **Pontifícia Universidade Católica do Rio de Janeiro**, [S.l.], 2002.

DASH, D. P.; SETHI, N.; DASH, A. K. Infectious disease, human capital, and the brics economy in the time of covid-19. **MethodsX**, [S.l.], v. 8, p. 101202, 2021.

ESCOBAR, H. Amazon fires clearly linked to deforestation, scientists say. **Science**, [S.l.], v. 365, n. 6456, p. 853–853, 2019.

ESLING, P.; AGON, C. Time-series data mining. **ACM Comput. Surv.**, New York, NY, USA, v. 45, n. 1, dec 2012.

ESTER, M.; KRIEGEL, H.-P.; SANDER, J. Spatial data mining: a database approach. In: ADVANCES IN SPATIAL DATABASES, 1997, Berlin, Heidelberg. **Anais...** Springer Berlin Heidelberg, 1997. p. 47–66.

FAYYAD, U. M.; HAUSSLER, D.; STOLORZ, P. E. Kdd for science data analysis: issues and examples. In: KDD, 1996. **Anais...** [S.l.: s.n.], 1996. p. 50–56.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. **AI Magazine**, [S.l.], v. 17, n. 3, p. 37, Mar. 1996.

FENNELLY, K. Particle sizes of infectious aerosols: implications for infection control. **The Lancet**, [S.l.], v. 8, n. 9, p. 914–924, 2020.

FIOCRUZ. **Covid-19 e queimadas na amazônia legal e no pantanal**: aspectos cumulativos e vulnerabilidades. Disponível em: <https://climaesaude.icict.fiocruz.br/sites/climaesaude.icict.fiocruz.br>. Acesso em: DEZ. 2020.

FOLEY, J. A. et al. Amazonia revealed: forest degradation and loss of ecosystem goods and services in the amazon basin. **Frontiers in Ecology and the Environment**, [S.l.], v. 5, n. 1, p. 25–32, 2007.

GRADY, N. W. Kdd meets big data. In: IEEE INTERNATIONAL CONFERENCE ON BIG DATA (BIG DATA), 2016., 2016. **Anais...** [S.l.: s.n.], 2016. p. 1603–1608.

GROOT, E. et al. A systematic review of the health impacts of occupational exposure to wildland fires. **International Journal of Occupational Medicine and Environmental Health**, [S.l.], v. 32, n. 2, p. 121–140, 2019.

GUAN, W.-j. et al. Clinical characteristics of coronavirus disease 2019 in china. **New England Journal of Medicine**, [S.l.], v. 382, n. 18, p. 1708–1720, 2020.

GUO D; MENNIS, J. Spatial data mining and geographic knowledge discovery—an introduction. **Computers, Environment and Urban Systems**, [S.l.], v. 33, p. 403–408, 2009.

HAN J.; KAMBER, M. P. J. **Data mining**: concepts and techniques. 3rd. ed. Waltham,USA: Elsevier, 2011.

HAN J.; MILER, H. J. **Geographic data mining and knowledge discovery**. 1nd. ed. London: Taylor e Francis, 2001.

HARTIGAN, J. A.; WONG, M. A. Algorithm as 136: a k-means clustering algorithm. **Journal of the royal statistical society. series c (applied statistics)**, [S.l.], v. 28, n. 1, p. 100–108, 1979.

HE, J. et al. Molecular mechanism of evolution and human infection with sars-cov-2. **Viruses**, [S.l.], v. 12, n. 4, 2020.

HU, B. et al. Characteristics of sars-cov-2 and covid-19. **Nature**, [S.l.], v. 19, p. 141–154, 2021.

Human Development Index (HDI) Ranking. , [S.l.]. Accessed on: 10.10.2020.

Information Technology Department of the Public Health Care System-SUS (DATASUS). , [S.l.].

INPE. **Inpe queimadas**. 2020.

J.A. FOLEY G.P. ASNER, M. C. M. C. R. D. H. G. E. H. S. O. J. P. N. R. P. S. Amazonia revealed: forest degradation and loss of ecosystem goods and services in the amazon basin. **Frontiers in Ecology and the Environment**, [S.l.], v. 5, n. 1, p. 25–32, 2007.

JIMÉNEZ-MUÑOZ, J. C. et al. Record-breaking warming and extreme drought in the amazon rainforest during the course of el niño 2015–2016. **Scientific reports**, [S.l.], v. 6, n. 1, p. 1–7, 2016.

JUNIOR, C. H. L. S. et al. The brazilian amazon deforestation rate in 2020 is the greatest of the decade. **Nature Ecology & Evolution**, [S.l.], v. 5, n. 2, p. 144–145, Dec. 2020.

KAREN, B.; W., M. S. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. **The Lancet**, [S.l.], v. 380, n. 4836, p. 37–43, 2012.

KLOKNER, S. G. M. et al. Epidemiological profile and risk factors predictors of covid-19 in southern brazil. **Research, Society and Development**, [S.l.], v. 10, n. 3, p. e17710313197, Mar. 2021.

KODINARIYA, T. M.; MAKWANA, P. R. Review on determining number of cluster in k-means clustering. **International Journal**, [S.l.], v. 1, n. 6, p. 90–95, 2013.

KOUA, E.; KRAAK, M. Geovisualization to support the exploration of large health and demographic survey data. **International Journal of Health Geographics**, [S.l.], 2004.

KUMAR, A. et al. Wuhan to world: the covid-19 pandemic. **Frontiers in Cellular and Infection Microbiology**, [S.l.], v. 11, 2021.

KUMAR, V.; STEINBACH, M.; TAN, P.-N. **Introduction to data mining**. 1st ed. ed. [S.l.]: Addison Wesley, 2009.

LIBONATI, R. et al. Twenty-first century droughts have not increasingly exacerbated fire season severity in the brazilian amazon. **Scientific reports**, [S.l.], v. 11, n. 1, p. 1–13, 2021.

LIEROP, P. V. et al. Global forest area disturbance from fire, insect pests, diseases and severe weather events. **Forest Ecology and Management**, [S.l.], v. 352, p. 78–88, 2015.

LIKAS, A.; VLASSIS, N.; VERBEEK, J. J. The global k-means clustering algorithm. **Pattern recognition**, [S.l.], v. 36, n. 2, p. 451–461, 2003.

LIMONATI, R. et al. On a new coordinate system for improved discrimination of vegetation and burned áreas using mir/nir information. **Remote Sensing of Environment**, [S.l.], v. 115, p. 1464–1477, 2011.

LIMONATI, R. et al. An algorithm for burned area detection in the brazilian cerrado using 4 μm modis imagery. **Remote Sens.**, [S.l.], v. 7, p. 15782–15803, 2015.

MACHADO-SILVA, F. et al. Drought and fires influence the respiratory diseases hospitalizations in the amazon. **Ecological Indicators**, [S.l.], v. 109, p. 105817, 2020.

MARCHI, N.; PIROTTI, F.; LINGUA, E. Airborne and terrestrial laser scanning data for the assessment of standing and lying deadwood: current situation and new perspectives. **Remote Sensing**, [S.l.], v. 10, n. 9, 2018.

MARTIN, S. T. et al. Sources and properties of amazonian aerosol particles. **Reviews of Geophysics**, [S.l.], v. 48, n. 2, 2010.

MCKEE, M.; STUCKLER, D. If the world fails to protect the economy, covid-19 will damage health not just now but also in the future. **Nature Medicine**, [S.l.], v. 26, n. 5, p. 640–642, 2020.

MEO, S. A. et al. Effect of environmental pollutants pm-2.5, carbon monoxide, and ozone on the incidence and mortality of sars-cov-2 infection in ten wildfire affected counties in california. **Science of The Total Environment**, [S.l.], v. 757, p. 143948, 2021.

MILLER, H. J.; HAN, J. **Geographic data mining and knowledge discovery**. USA: Taylor Francis, Inc., 2001.

MORTON, D. C. et al. Agricultural intensification increases deforestation fire activity in amazonia. **Global Change Biology**, [S.l.], v. 14, n. 10, p. 2262–2275, 2008.

NEPSTAD, D. C. et al. Large-scale impoverishment of amazonian forests by logging and fire. **Nature**, [S.l.], v. 398, n. 6727, p. 505–508, 1999.

NICHOLSON, C. et al. A machine learning and clustering-based approach for county-level covid-19 analysis. **PLOS ONE**, [S.l.], v. 17, n. 4, p. 1–24, 04 2022.

Notificações de Síndrome Gripal (OPENDATASUS). , [S.l.].

OMS. **Indicadores para o estabelecimento de políticas e a tomada de decisão em saúde ambiental**. Disponível em: <https://www.who.int/docstore/peh/Vegetation$_fires/Executive_Summary.pdf >$ $.Acessoem : DEZ$. 2021.

PACIFICO, F. et al. Biomass burning related ozone damage on vegetation over the amazon forest: a model sensitivity study. **Atmospheric Chemistry and Physics**, [S.l.], v. 15, n. 5, p. 2791–2804, 2015.

PARENTE, J.; PEREIRA, M. G.; TONINI, M. Space-time clustering analysis of wildfires: the influence of dataset characteristics, fire prevention policy decisions, weather and climate. **Science of The Total Environment**, [S.l.], v. 559, p. 151–165, 2016.

PAZMIñO-MAJI, R.; GARCíA-PEñALVO, F.; CONDE-GONZáLEZ, M. Statistical implicative analysis approximation to kdd and data mining: a systematic and mapping review in knowledge discovery database framework. In: 2016 , 2017. **Anais...** [S.l.: s.n.], 2017.

PEREIRA, A. et al. Validação de focos de calor utilizados no monitoramento orbital de queimadas por meio de imagens tm. **Cerne**, [S.l.], v. 18, p. 335–343, 2012.

PIERCE, C. A. et al. Immune responses to sars-cov-2 infection in hospitalized pediatric and adult patients. **Science Translational Medicine**, [S.l.], v. 12, n. 564, 2020.

PIVELLO, V. R. et al. Understanding brazil's catastrophic fires: causes, consequences and policy needed to prevent future tragedies. **Perspectives in Ecology and Conservation**, [S.l.], v. 19, n. 3, p. 233–255, 2021.

R. Groth. Data mining: building competitive advantage. **Prentice Hall PTR - 2000 - New Jersey, USA**, [S.l.], 2000.

RIBEIRO, I. et al. Biomass burning and carbon monoxide patterns in brazil during the extreme drought years of 2005, 2010, and 2015. **Environmental Pollution**, [S.l.], v. 243, p. 1008–1014, 2018.

ROCHA, R.; SANT'ANNA, A. A. Winds of fire and smoke: air pollution and health in the brazilian amazon. **World Development**, [S.l.], v. 151, p. 105722, 2022.

ROMERO, J. A. R. RELAÇÃO ENTRE AS CONDIÇÕES SOCIOECONÔMICAS E A INCIDÊNCIA DA PANDEMIA DA COVID-19 NOS MUNICÍPIOS DO CEARÁ. **Boletim de Conjuntura (BOCA)**, [S.l.], v. 3, n. 7, July 2020.

SANTOS, D. F. dos et al. Epidemiological analysis of 9,897 deaths from covid-19 in the period from march to may 2020, in brazil. **Journal of Evidence-Based Healthcare**, [S.l.], v. 3, p. e3447, Jan. 2022.

SCHROEDER, L. et al. Respiratory diseases, malaria and leishmaniasis: temporal and spatial association with fire occurrences from knowledge discovery and data mining. **International Journal of Environmental Research and Public health**, [S.l.], v. 17, n. 10, p. 3718–3742, 2020.

SHAFI, J.; WAHEED, A. K-means clustering analysing abrupt changes in air quality. In: INTERNATIONAL CONFERENCE ON ELECTRONICS, COMMUNICATION AND AEROSPACE TECHNOLOGY (ICECA), 2020., 2020. **Anais...** [S.l.: s.n.], 2020. p. 26–30.

SMITH, L. T. et al. Drought impacts on children's respiratory health in the brazilian amazon. **Scientific reports**, [S.l.], v. 4, n. 1, p. 1–8, 2014.

SMITH, L. T. et al. Drought impacts on children's respiratory health in the brazilian amazon. **Scientific Reports**, [S.l.], v. 4, n. 1, p. 3726, 2014.

SOLIMINI, A. et al. A global association between covid-19 cases and airborne particulate matter at regional level. **Nature**, [S.l.], v. 11, 2021.

SOUZA, J. D.; RAZENTE, H.; BARIONE, M. Optimizing metric access methods for querying and mining complex data types. **Journal of the Brazilian Computer Society.**, [S.l.], v. 20, p. 47–57, 2014.

T. Hastie and R. Tibshirani and J. Friedman. The elements of statistical learning: data mining, inference, and prediction. **Yale University Press - 2009**, [S.l.], 2009.

UZINSKI, J. C.; ABREU, C. C. E. de; OLIVEIRA, B. R. de. **Aplicações de inteligência artificial e ciência de dados**. Nova Xavantina, MT: Pantanal: Pantanal Editora, 2020.

VERITY, R. et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. **The Lancet Infectious Diseases**, [S.l.], v. 20, n. 6, p. 669–677, 2020.

WANG, Z. et al. Household transmission of sars-cov-2. **Journal of Infection**, [S.l.], v. 81, n. 1, p. 179–182, 2020.

WERF, G. R. Van der et al. Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997â€“2009). **Atmospheric Chemistry and Physics**, [S.l.], v. 10, n. 23, p. 11707–11735, 2010.

WU, X. et al. Air pollution and covid-19 mortality in the united states: strengths and limitations of an ecological regression analysis. **Science Advances**, [S.l.], v. 6, n. 45, 2020.

XU, Z. et al. Pathological findings of covid-19 associated with acute respiratory distress syndrome. **The Lancet**, [S.l.], v. 8, p. 420–422, 2020.

YOUSSOUF, H. et al. Non-accidental health impacts of wildfire smoke. **International Journal of Environmental Research and Public Healtht**, [S.l.], v. 11, p. 11772–11804, 2014.

ZABANIOTOU, A. et al. From multidisciplinarity to transdisciplinarity and from local to global foci: integrative approaches to systemic resilience based upon the value of life in the context of environmental and gender vulnerabilities with a special focus upon the brazilian amazon biome. **Sustainability**, [S.l.], v. 12, n. 20, 2020.

ZHOU, X. et al. Excess of covid-19 cases and deaths due to fine particulate matter exposure during the 2020 wildfires in the united states. **Science Advances**, [S.l.], v. 7, n. 33, p. eabi8789, 2021.

ANEXO A – ARTIGOS PUBLICADOS

Figura 23 – Artigo publicado referente a este estudo.

Fire association with respiratory disease and COVID-19 complications in the State of Pará, Brazil

Lucas Schroeder,[a,g] Eniuce Menezes de Souza,[b] Clévia Rosset,[d] Ademir Marques Junior,[a,g] Juliano André Boquett,[e,f] Vinicius Francisco Rofatto,[c] Diego Brum,[a,g] Luiz Gonzaga Jr.,[a,g] Marcelo Zagonel de Oliveira,[g,h] and Maurício Roberto Veronez,[a,g,h,*]

[a]Post Graduate Program in Applied Computing, Unisinos University. São Leopoldo, RS, Brazil
[b]Department of Statistics, State University of Maringá. Maringá, PR, Brazil
[c]Institute of Geography, Federal University of Uberlândia. Monte Carmelo, MG, Brazil
[d]Medical Genomics Laboratory, Hospital de Clínicas de Porto Alegre. Porto Alegre, RS, Brazil
[e]Post-graduate Program in Genetics and Molecular Biology, Institute of Biosciences, Federal University of Rio Grande do Sul. Porto Alegre, RS, Brazil
[f]Post-graduate Program in Child and Adolescent Health, Faculty of Medicine, Universidade Federal do Rio Grande do Sul. Porto Alegre, RS, Brazil
[g]Vizlab | X-Reality and Geoinformatics Lab, Unisinos University. São Leopoldo, RS, Brazil
[h]Graduate Program in Biology, Unisinos University. São Leopoldo, RS, Brazil

The Lancet Regional Health - Americas
2022;6: 100102
Published online 3 November 2021
https://doi.org/10.1016/j.lana.2021.100102

Summary

Background Brazil has faced two simultaneous problems related to respiratory health: forest fires and the high mortality rate due to COVID-19 pandemics. The Amazon rain forest is one of the Brazilian biomes that suffers the most with fires caused by droughts and illegal deforestation. These fires can bring respiratory diseases associated with air pollution, and the State of Pará in Brazil is the most affected. COVID-19 pandemics associated with air pollution can potentially increase hospitalizations and deaths related to respiratory diseases. Here, we aimed to evaluate the association of fire occurrences with the COVID-19 mortality rates and general respiratory diseases hospitalizations in the State of Pará, Brazil.

Methods We employed machine learning technique for clustering k-means accompanied with the elbow method used to identify the ideal quantity of clusters for the k-means algorithm, clustering 10 groups of cities in the State of Pará where we selected the clusters with the highest and lowest fires occurrence from the 2015 to 2019. Next, an Auto-regressive Integrated Moving Average Exogenous (ARIMAX) model was proposed to study the serial correlation of respiratory diseases hospitalizations and their associations with fire occurrences. Regarding the COVID-19 analysis, we computed the mortality risk and its confidence level considering the quarterly incidence rate ratio in clusters with high and low exposure to fires.

Findings Using the k-means algorithm we identified two clusters with similar DHI (Development Human Index) and GDP (Gross Domestic Product) from a group of ten clusters that divided the State of Pará but with diverse behavior considering the hospitalizations and forest fires in the Amazon biome. From the auto-regressive and moving average model (ARIMAX), it was possible to show that besides the serial correlation, the fires occurrences contribute to the respiratory diseases increase, with an observed lag of six months after the fires for the case with high exposure to fires. A highlight that deserves attention concerns the relationship between fire occurrences and deaths. Historically, the risk of mortality by respiratory diseases is higher (about the double) in regions and periods with high exposure to fires than the ones with low exposure to fires. The same pattern remains in the period of the COVID-19 pandemic, where the risk of mortality for COVID-19 was 80% higher in the region and period with high exposure to fires. Regarding the SARS-COV-2 analysis, the risk of mortality related to COVID-19 is higher in the period with high exposure to fires than in the period with low exposure to fires. Another highlight concerns the relationship between fire occurrences and COVID-19 deaths. The results show that regions with high fire occurrences are associated with more cases of COVID deaths.

Interpretation The decision-make process is a critical problem mainly when it involves environmental and health control policies. Environmental policies are often more cost-effective as health measures than the use of public health services. This highlight the importance of data analyses to support the decision making and to identify population in need of better infrastructure due to historical environmental factors and the knowledge of associated health risk. The results suggest that The fires occurrences contribute to the increase of the respiratory diseases

*Corresponding author.
E-mail address: veronez@unisinos.br (M.R. Veronez).

Fonte: Elaborado pelo autor.