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 GonzagaJr.[a,g] Marcelo Zagonel de Oliveira,[g,h] and Mauricio 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

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.

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

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).

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.

Funding No additional funding source was required for this study.

Keywords: Fire; Respiratory diseases; K-Means; ARIMAX; Time series analysis; SARS-Cov-2; COVID-19; Hospitalizations; Mortality; Incidence rate ratio; Risk

Research in context

Evidence before this study

The emergence of the SARS-Cov-2 related pandemic, and the high mortality rate associated with people with health vulnerabilities, require local studies that can identify groups most vulnerable to the impacts of the pandemic. The analysis of free and public data in studies relating environmental aspects, hospitalizations and current issues related to SARS-Cov-2 allows the preparation of an approach capable of understanding possible associations between historical exposure to air pollution and forest fires to deaths from COVID -19. The amount and variety of data requires the application of Knowledge Discovery (KDD) techniques like the k-means algorithm. This algorithm aimed to identify groups of cities in the State of Pará, with higher and lower incidences of fire. Burning data are provided by the National Institute for Space Research (INPE Queimadas) and the period used comprised the period (2015 to 2019). Populations residing in places with environmental problems tend to have greater impacts with the emergence of new diseases.

Added value of this study

The decision-making process is a critical problem, especially when it involves environmental and health control policies. The data regarding recurring fires in Brazil, especially in the states of the Legal Amazon, can be used to understand and enable the understanding of populations at risk from the emergence of pandemics. The results analyzing the data regarding the State of Pará, showed that, as fire occurrences increases the number of hospitalizations for respiratory diseases also increases. Additionally, the mortality risk related to COVID-19 is greater in the period of high fire exposure than in the period of low fire exposure. The regions with the highest occurrence of fires are associated with more deaths from COVID-19, especially in the months when fires are more intense.

Implications of all available evidence

With this study, we can verify that in no State of Pará there was an overlap of situations related to respiratory diseases. Populations residing in places with environmental problems tend to suffer more with the new respiratory diseases. Knowing the data regarding population health is fundamental the application of control and prevention policies. We were also able to correlate the pollution from forest fires in the Amazon to the increase of hospitalizations and deaths from COVID-19 using knowledge discovery techniques on the large amount of data available.

Introduction

In Brazil, changes in agricultural practices and land use in the last years have directly affected economic and social development, especially in the region comprising the Amazon biome.[1] One of these practices is the land clearing using controlled fire.[2] However, its indiscriminate use allied to relaxed policies has considerably raised the number of fires in the Amazon rain forest, contributing to its accelerated destruction.[3,4] The number of fires between January and August (2020) were 39% higher than the fire's average of the last ten years in the same period in Brazil.[5–7]

The State of Para in Brazil which owns 24% of Amazon rain-forest accumulated 17% (174,903) of the total number of fires (939,015) from 2015 to 2019, as registered by the National Institute for Space Research (INPE).[8] This destruction of the Amazon rain-forest directly affects the global climate balance.

In addition to the global impacts, previous studies have demonstrated that air pollution from Amazon fires is damaging to human local health.[9] For example, studies in local indigenous communities demonstrated that smoke from fires is one of the main causes of respiratory hospitalizations in this population.[10,11]

Regarding respiratory diseases, December 2019 saw a rapid spread of the beta-coronavirus 2019-nCOV, found first in the city of Wuhan in the Hubei province in China, later identified as SARS-COV-2 (Severe acute respiratory syndrome coronavirus 2) causing the COVID-19 disease.[12] COVID-19 symptomatology involves mainly the respiratory system, varying from a feverish condition with mild respiratory symptoms in some patients to, pneumonia and more severe symptoms in other patients.[13] The rapid spread of COVID-19 in Brazilian areas could represent a bigger challenge to the populations exposed to smoke from forest fires. Indeed, during the severe acute respiratory outbreak associated with the Severe acute respiratory syndrome coronavirus 1 (SARS-Cov-1) in 2003, patients from areas with high levels of air pollution exhibited a 200% increase in the relative risk of death compared with people living in areas with low pollution content.[14]

With more than 13 million people infected and 374 thousand deaths as in April 2021, Brazil has suffered greatly from this disease. The State of Pará alone has more than 450 thousand cases of COVID-19 and 11,000 deaths, reaching a mortality rate of 138.3 deaths per 100 thousand inhabitants (OPENDATASUS Brazilian Ministry of Health platform).[15]

Additionally, this outbreak can be potentialized by poor air quality caused by air city pollution and uncontrolled forest fires, however, a better analysis of this correlation is still a challenge due to the large amount of data. To overcome this challenge, artificial intelligence approaches like the Knowledge Discovery of Databases (KDD)[16,17] could be applied to this type of data which requires 4Vs (volume, variety, speed, and veracity).[18]

We propose in this work the use of a KDD related algorithm named k-means, and a time series analysis (Autoregressive Integrated Moving Average with Explanatory Variable - ARIMAX) to identify and analyze clusters of cities more and less affected by the Amazon forest fires, and the relationship between these scenarios and the hospitalizations for respiratory diseases (HRD), and hospitalizations and mortality rate by the SARS-COV-2 considering the State of Pará in Brazil.

This strategy may contribute to the identification of populations that are more prone to develop respiratory complications during pandemics given the environmental risks, guiding public health management, as environmental improvements are often more cost-effective as health measures than the use of public health services.[13]

Methods

This work methodology, as presented in Fig. 1 and following sections, focuses on the knowledge discovery steps for pre-processing, clustering, time series analysis, and post-processing.

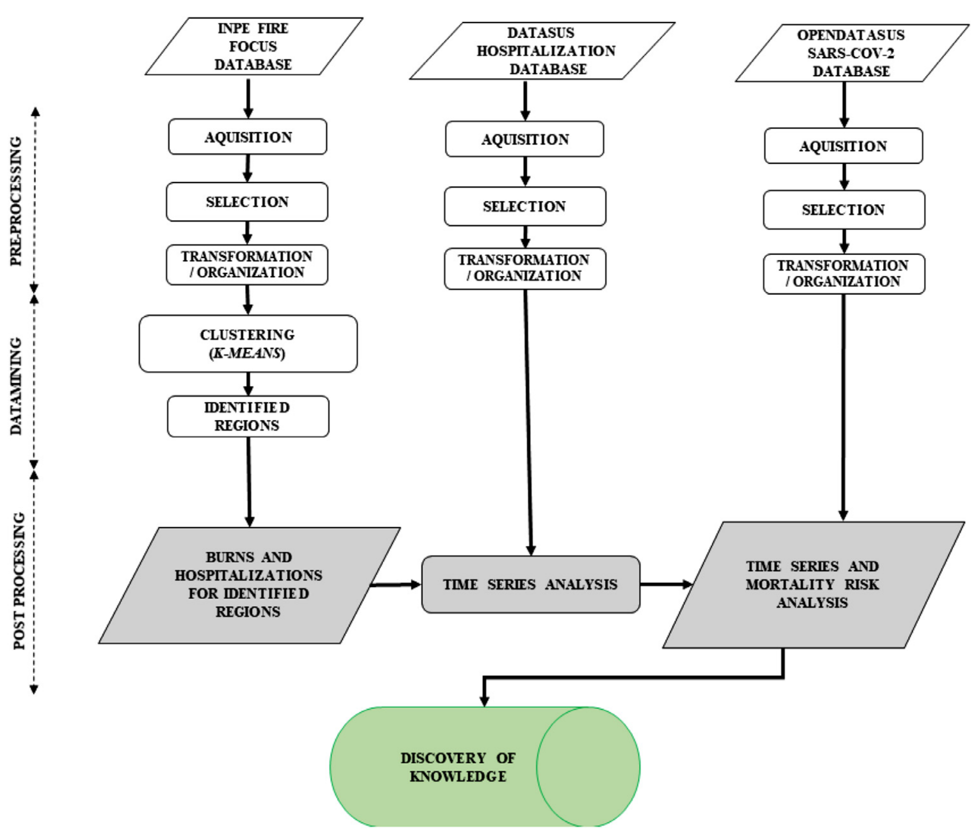


Figure 1. Flowchart of the involved steps from pre-processing, clustering (K-means), post-processing until the knowledge discovery (KDD).

Pre-processing

The pre-processing steps as presented in the flowchart of Fig. 1 include the data preparation of databases of forest fires, hospitalizations, COVID-19 cases and deaths, and additional databases used to understand social aspects of the population studied. These databases are briefly presented bellow.

Forest fires database. The database provided by INPE for fires data named BDQUEIMADAS[1] is updated every three hours across Latin America and contains information on the geographic coordinates of the outbreaks of fire, biome, smoke concentration in the air, risk of fire, expected rain, number of days without rain, date, time and city of the observed focus.[8,19,20] The fire count data is obtained by remote sensing techniques using a Moderate Resolution Imaging Spectro-radiometer (MODIS) sensor, integrated into the satellite mission TERRA/AQUA. In this work we selected the data from 2015 to 2020 in the State of Pará, Brazil.

Health database: Hospitalizations for Respiratory Disease (HRD). In Brazil, the Information Technology Department of the Unified Health System (SUS) has a platform named TABNET[2] that makes data available, creating an interoperability project between public health systems and the general community. The causes of hospitalization are classified according to the International Classification of Diseases in its 10th iteration (ICD-10), where we considered hospitalizations related to respiratory diseases (codes J00-J99). The cases were selected based on the municipality of residence and time period of hospitalization.[21]

Health database: Cases and deaths related to SARS-Cov-2 infection. Due to the emergence of the SARS-Cov-2 pandemics, the Ministry of Health implemented through the Secretary for Surveillance and Health, the surveillance of mild, moderate flu syndrome and suspects of having contracted SARS-Cov-2. Data for all cases of influenza syndrome are collected by e-SUS NOTIFY and made available to the community through the platform OPENDATASUS,[3] with data from all patients and information such as date, state, municipality, positive cases, deaths, place of residence and place of notification.[15]

Complementary databases. To contribute to the discussion, we incorporated other databases in this study. Data referring to the Municipal Human Development Index (MHDI) were collected from the United Nations Development Programme (UNDP),[4] through UNDP Strategic Plan, 2018–2021. Remembering that HDI is based on three basic features: long and healthy life, education and a decent standard of living.[22] The Gross Domestic Product (GDP) is a measure of the sum of all goods and services found within a territory in some period. The GDP makes it possible to monitor the economic activity and economic differences allowing to analyse its impact on the pandemic control.[23] The website that provides GDP data for all Brazilian municipalities is the Brazilian Institute of Geography and Statistics (IBGE).[5]

Aerosol data (tiny solid and liquid particles suspended in the atmosphere) were also observed. Examples of aerosols include windblown dust, volcanic ash, smoke from fires, and industrial pollution, that can affect climate and people's health.[10] The NASA Earth Observations (NEO)[6] provides data to the sensor with a moderate resolution image spectroradiometer (MODIS) that is aboard NASA's Terra and Aqua satellites and is used to monitor the optical thickness of the aerosol on a given day or over a few days.

Clustering

This work uses an unsupervised machine learning method of clustering called K-means.[24–27] This method uses the data parameters as dimensions in a Euclidean space, where using a k number of seeds, the subjects are grouped given their distance to these seeds. After each algorithm iteration, the seed position is recalculated based on the centroid position of its components. The algorithm ends when no significant change in position occurs.[20,28] K-means allows us to add points to central or centroid nodes, which are determined as the interest groups in our study.

To determine the ideal number of groups to partition a data set, the elbow method [29] was used. The elbow method consists of defining the number of K-groups so that the total variance within the group is minimized through the total square sum within the group using the Within Cluster Sum of Squares (WCSS). In this case, the quality of the cluster is linked to the average distance between the burn points and its centroid for the number of clusters formed.[24,30]

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

[1] https://queimadas.dgi.inpe.br/queimadas/bdqueimadas

[2] https://datasus.saude.gov.br/informacoes-de-saude-tabnet/

[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_M_AER_OD

where x_j and y_j are the geographic coordinates (latitude and longitude) and n is the number of fire occurrences.

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

where k is the number of clusters and x_j is the point in space of a given object (coordinate of fire occurrences) and c_i is the centroid of the cluster and n is the number of fire occurrences.

Post-Processing

To show the impact of fires on respiratory health, two groups were considered: groups of cities with the highest number of fires and the group of cities with the lowest number of fires in the State of Pará. After identifying these groups using the K-means algorithm, we carried a comparison between hospitalization rate for general respiratory diseases (HRD) between the years 2015 to 2019. The hospitalization rate is given by

$$\textit{Hospitalization rate} = \frac{\textit{Totalof monthly hospitalization cases} \times 100,000}{\textit{Total population}}. \quad (3)$$

Time Series Analyses. To take into account the serial correlation of the time series and to verify how much the fire occurrences influence the HRD, an ARIMAX model was built. This model is an extension of the ARIMA (Auto-regressive Integrated Moving Average) model allowing the inclusion of exogenous variables. The usual procedures of identification, estimation and validation are applied. Details on the ARIMAX can be obtained in.[31]

When analysing time series to identify trends using successive differentiation (observation in the moment t minus the observation in the moment $t-1$) is possible to induce a constant average and, consequently, the stationarity. Let F_t and H_t be the time series of fire occurrences and HRD after ∇^d differences to induce stationarity if needed. Thus the model $ARIMAX(p, d, q)$ can be written as:

$$H_t = \gamma_1 F_{t-\ell} + \alpha_1 H_{t-1} + \alpha_2 H_{t-2} + \cdots + \alpha_p H_{t-p} + \quad (4)$$

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

where ϵ_t is the white noise, and p and q are the autoregressive and moving average orders, respectively. These orders are identified from auto-correlation functions. Indeed, the cross-correlation can be evaluated to identify the lag ℓ of the fire variable to explain the HRD. To evaluate the contribution of the independent variable F, the Akaike Information Criterion (AIC) and likelihood rate test are taken into account. The usual residual assumptions of serial correlation, normality, and constant variance were checked after the model estimation.

SARS-Cov-2 Analyses. With the analysis of the HRD potentiated by fires, the analysis of deaths by SARS-Cov-2 by place of residence was performed. The SARS-Cov-2 data were extracted from the OPENDATASUS website for the year of 2020.

We considered the mortality rate for each cluster per 100 thousand inhabitants are given by

$$\text{Mortality rate} = \frac{\textit{Number of } \boldsymbol{SARS-Cov-2\ deaths} \times 100,000}{\textit{Total population}}, \quad (5)$$

instead of the case fatality rate (CFR) (percentage) or lethality rate (ratio) that require the number of SARS-Cov-2 positive cases which are difficult to estimate due to the lack of a broad testing in the population. The CFR data may be biased, since in Brazil COVID-19 reported cases are highly dependent on the quality and quantity of tests applied in each region, testing policies, underreporting of cases, and even duplicate cases due to reinfections.

To compute the mortality rate ratio (MRR) we compare the risk of mortality between the two clusters, one including the cities with the highest number of fires (HF) and the other with the lowest number of fires (LF) in the State of Pará. We aggregated the data in quarter periods and used the following expression:

$$\widehat{MRR} = \frac{\textit{Mortality rate in cluster HF}}{\textit{Mortality rate in cluster LF}} = \frac{Deaths_{HF}/Population_{HF}}{Deaths_{LF}/Population_{LF}}. \quad (6)$$

To evaluate if the estimated $\widehat{MRR}$ indicates risk or not, the confidence interval (CI) need to be computed. Like the risk ratio, the MRR is not normally distributed, but its natural log is. With the standard deviation (SD) of the log rate ratio, $\widehat{SD}[ln(\widehat{IR})] = \left(\frac{1}{DeathsHF} + \frac{1}{DeathsLF}\right)^2$, it is possible to compute the 95%CI of MRR as:

$$e^{ln(\widehat{MRR}) \pm 1.96\widehat{SD}[ln(\widehat{MRR})]}. \quad (7)$$

Role of the funding source

No additional funding source was required for this study.

Results

Fire occurrences in Brazil

Fig. 2 shows the spatial distribution of fires registered in the INPE database in all Brazilian states of cumulative occurrences from 2015 to 2019.

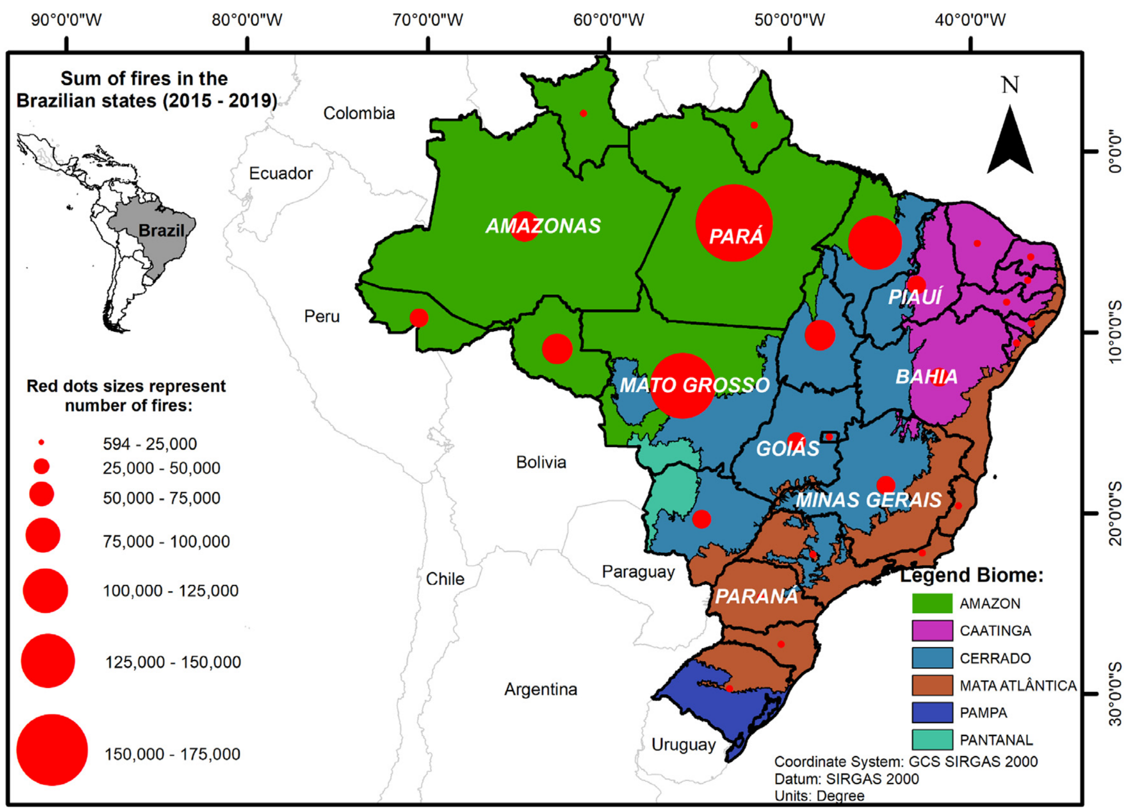


Figure 2. Spatial distribution of the sum of the number of fires (red dots sizes) in the last 5 years (2015–2019) before the pandemic in Brazil states. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The State of Pará, located in the Brazilian Amazon, has the highest incidence of fires in the last 5 years, with more than 170,000 fire occurrences. Pará owns 20% of the Amazon forest in its territory and 24% of the fires reported in Brazil occur in this state. Thus, we selected this state to perform a more detailed analysis, since there is a need to understand these values within the State of Pará through a spatial distribution.

Finding clusters with similar fire occurrences

To understand whether fires occur more frequently in some group of cities in the State of Pará, the k-means algorithm was implemented together with the elbow method. Fig. 3 illustrates the results of the elbow method indicating an ideal number of clusters equal to ten, and Fig. 4 show the quantity of fires for each cluster of cities identified for the State of Pará.

It's possible to highlight in Fig. 4 the clusters 1 and 4 that had the highest number of fires and the lower number of fires respectively. Cluster 1 is formed by five cities with a population estimated to be 268,831 inhabitants in the year of 2020: Altamira, Itaituba, Jacareacanga, Novo Progresso, and Trairão; cluster 4 is formed by twelve cities: Alenquer, Almeirim, Belterra, Curuá, Faro, Juruti, Mojuí dos Campos, Monte Alegre, Óbidos, Oriximiná, Santarém, and Terra Santa, with an estimated population total of 710,867 people.

Observing Human Development Index (HDI) and Gross Domestic Product (GDP) on the clusters

After the identification of the groups of cities with the highest and lowest rate of fires in the State of Pará (cluster 1 and 4, respectively), we observed the Human Development Index (HDI) and Gross Domestic Product (GDP) of both clusters. In Fig. 5 the distribution of HDI and GNP in the State of Pará.

The data in Fig. 5 indicate that the cluster 1 has the highest HDI and GDP in Pará, although the demographic density differs between them. Both clusters have large areas with high HDI and GDP compared to other clusters of the State of Pará, contributing to a fair comparison between them.

Time Series Analysis to investigate fire occurrences influence on HRD

Before building the time series model, the cross-correlation function was investigated to identify the lag time between fire occurrences peaks and respiratory diseases

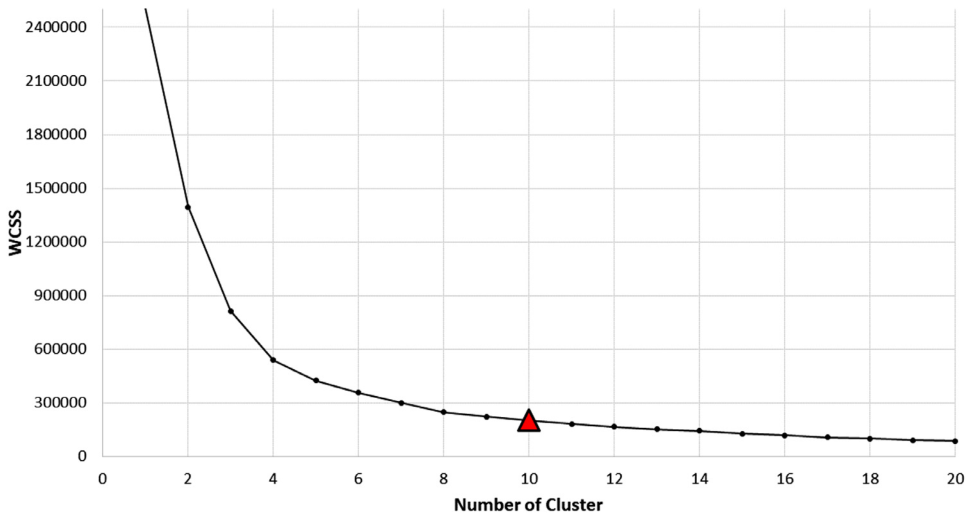


Figure 3. Elbow graphics with the Within Cluster Sum of Squares (WCSS) for State of Pará and fire occurrences for cluster quantities from 1 to 20. The red triangle indicate the chosen cluster number where WCSS is not decreasing so much. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

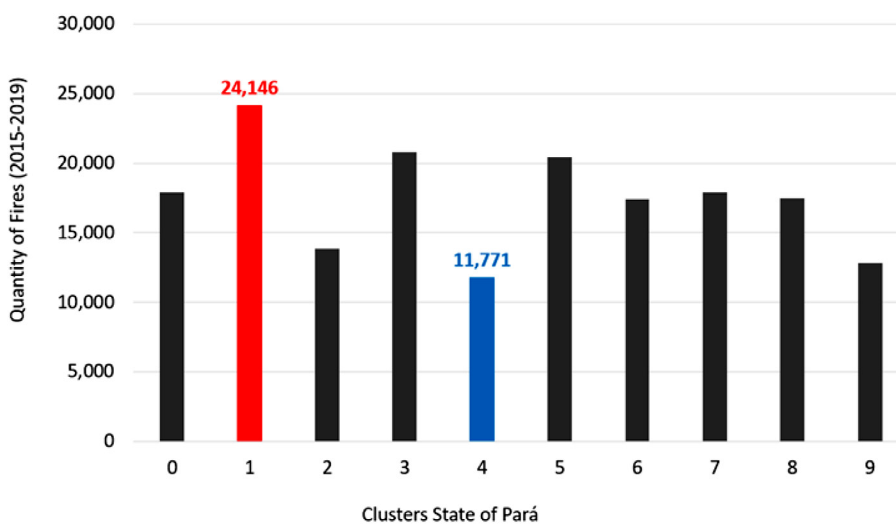


Figure 4. Graphic with the quantity of fires distribution by cluster in the period from 2015 to 2019. Cluster 1 (red) higher frequency of fires and Cluster 4 (blue) lower frequency of fires. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

peaks. We identified a lag of 6 months that can be visualized in Figure 6, alongside the HRD data and fire occurrences considering the clusters 1 and 4.

In Fig 7, the HRD, as well as the ARIMAX models for clusters 1 and 4 are presented. The estimated parameters are in Table 1. The variable related to aerosols could also be included in the model, but because it is highly correlated with fire occurrences, it was not added to avoid multicollinearity.

We can see in Fig. 7 the proposed models (red dashed lines) explain well the HDR data. From Table 1 we confirmed the fire occurrences lagged in about 6 months influence increasing the HRD in both clusters (positive estimates). Observing Fig. 7, only the HRD time series in C1 presented a trend (of decreasing) requiring differentiation. One difference ($d = 1$) was enough to induce stationarity. Because of that, the HRD average was not estimated (too close to zero after differentiation) in C1. For C4, the estimated HDR average during the investigated period was 2.85 (Table 1) and the moving average parameter represents the temporal correlation of the HRD time series.

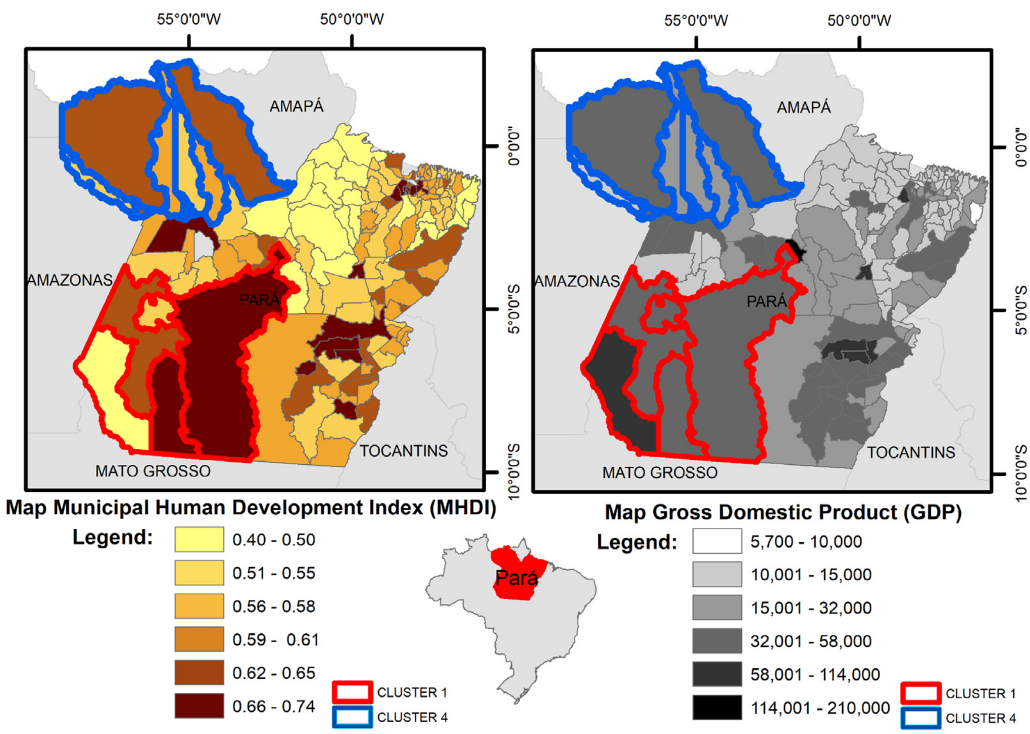


Figure 5. HDI and GDP of the cities of the State of Pará.

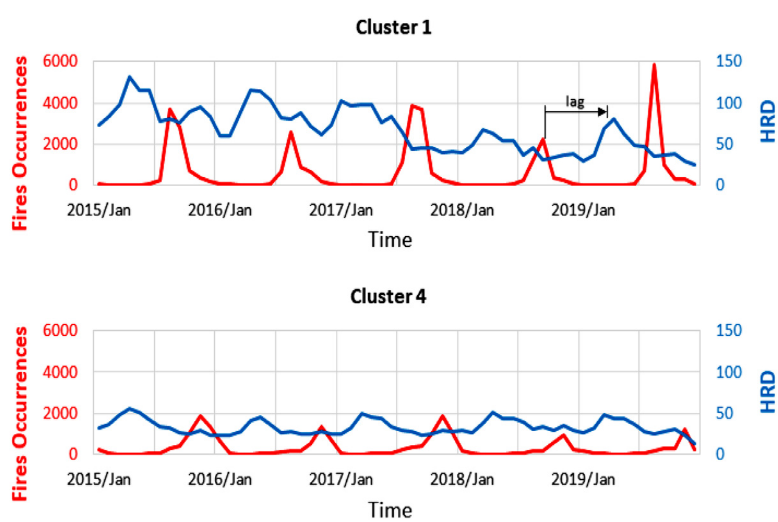


Figure 6. HRD and fire occurrences time series for clusters 1 and 4.

		Estimate	SE	p-value
ARIMAX(0,1,0) for HRD in C1	Lagged Fire Occurrences	0.0005	0.0002	0.0189
ARIMAX(0,0,1) for HRD in C4	Lagged Fire Occurrences	0.0009	0.0003	<0.0001
	β_1	0.5811	0.1119	<0.0001
	HDR average	2.8545	0.1338	<0.0001

***Table 1*: Estimated parameters, Standard Errors (SE), and p-values of ARIMAX models for HRD time series of C1 and C4 in the State of Pará. As for cluster 4, the HRD time series was not differentiated $(\mathbf{d} = \mathbf{0})$ and it has moving average coefficients $(\beta 1)$, a constant term (HRD average) needs to be included in this model.**

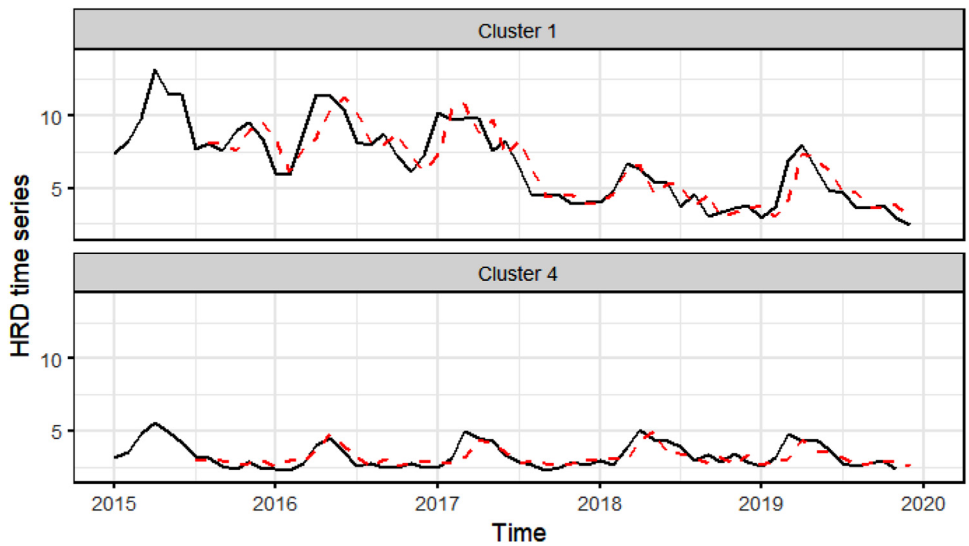


Figure 7. Hospitalization rate time series for general respiratory diseases (HRD) (black line) and ARIMAX models (red dashed lines) in clusters 1 and 4. The HRD in the y axis represents the hospitalization rate in 10 thousands hospitalizations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Although not presented, we also compared the proposed models with similar ones removing the lagged fire occurrences variable. The results were worse. Besides the visual improvement of the proposed models, statistical measures (AIC and likelihood rate test) also proved this fact, corroborating again the influence of fires on HRD.

Association of fires with SARS-Cov-2 mortality and respiratory diseases

To verify the impact of SARS-Cov-2 in families residing in clusters 1 and 4, we elaborated maps related to the behavior of these two regions during the year 2020. In Fig. 8, map 1 represents the distribution of air pollutants in the State of Pará during the year 2020 (clusters 1 and 4 are highlighted). It was found that cluster 1 presents higher values of air pollutants, which is in line with the pattern of annual burnings in this region.

In map 2 of Fig. 8 the HRD in the year of 2020 across the State of Pará are represented showing a reduction in the number of HDR for the year of 2020 (186.21 in cluster 1 and 300.9 in cluster 4) when compared to the average HDR for 2015–2019 (806.78 in cluster 1 and 392.71 cluster 4).

To evaluate if the SARS-Cov-2 pandemic affects the number of cases of respiratory diseases in the regions with fire occurrences, we evaluated the mortality rates of COVID-19 in the State of Pará that are shown in Fig. 9, where we can see that the highest rates are observed in cluster 1. To see the historical pattern of mortality for both respiratory diseases and COVID-19, Fig. 10 shows these rates for the quarters of the years from 2015 to 2020 (top and middle), and also the COVID-19 mortality rates per quarter in the year 2020 (bottom). The fire occurrences are also shown in this figure for the other years' quarters. We can see that the mortality rate for respiratory diseases is higher in cluster 1 where there are more fire occurrences independently of the period of the year. Regarding the COVID-19 mortality, the rate is also larger in cluster 1 than in cluster 4, mainly in the third quarter when the fire occurrences are higher than in the other quarters. Thus, it allows us to raise a possible vulnerability to respiratory complications in cluster 1.

Considering the data presented in Fig. 10, we computed the MRRs (Eq. 6) as well as their CI (Eq. 7). For the mortality rates of respiratory diseases in the period from 2015 to 2019 (before pandemic), we presented the MRRs in the Fig. 11(a). In 2020, the MRRs can be saw separately for the mortality of respiratory diseases and COVID-19 in Figures 11(b) and 11(c), respectively.

Fig. 11 a shows that the mortality rate ratio was twice greater for all quarters except for quarter 2 during 2015 −2019. Figure 11 b shows that the mortality rates from respiratory disease were not statistically different between the clusters throughout the year 2020. It is the same for Figure 11 c except that the mortality from COVID-19 was 1.8 times greater in cluster 1 in quarter 3.

Fig. 10 showed this slight reduction. On the other hand, considering the mortality rate for COVID-19, the MRR comes again to 1.8 (Fig. 11(c)) in the third quarter period, when most of the res occur, indicating that the risk of mortality in the third quarter was about 80% higher in cluster 1 than in cluster 4.

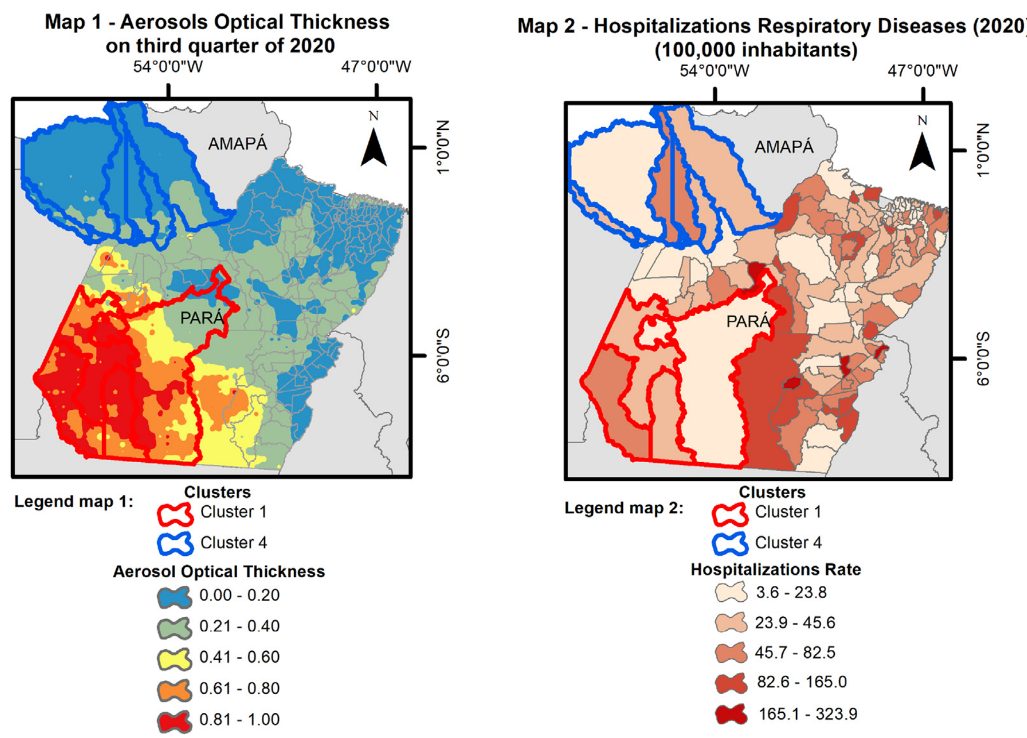


Figure 8. Map 1 represents the optical thickness (less than 0.1 indicates a crystalline sky with high visibility; 1 indicates the presence of dense aerosols). Map 2 represents the HRD in 2020.

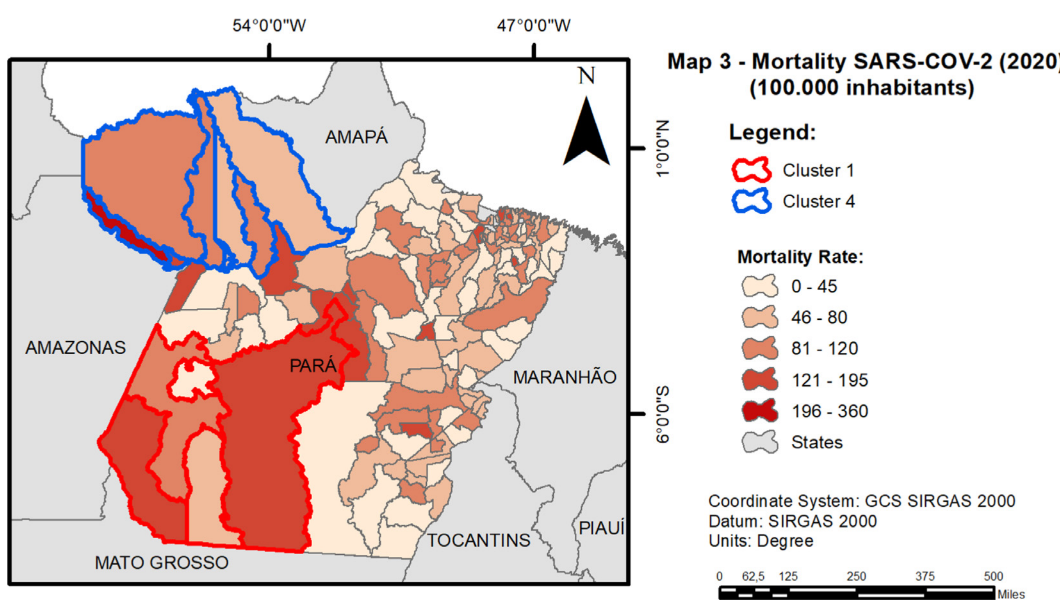


Figure 9. Mortality Rate for SARS-Cov-2 in the year 2020 in the State of Pará.

Discussion and final remarks

The overlap between res and the SARS-Cov-2 related pandemic has shown the urgent need to determine the relationship between fires and respiratory diseases and ultimately the relationship with COVID-19 respiratory complications. This is fundamental to guide public

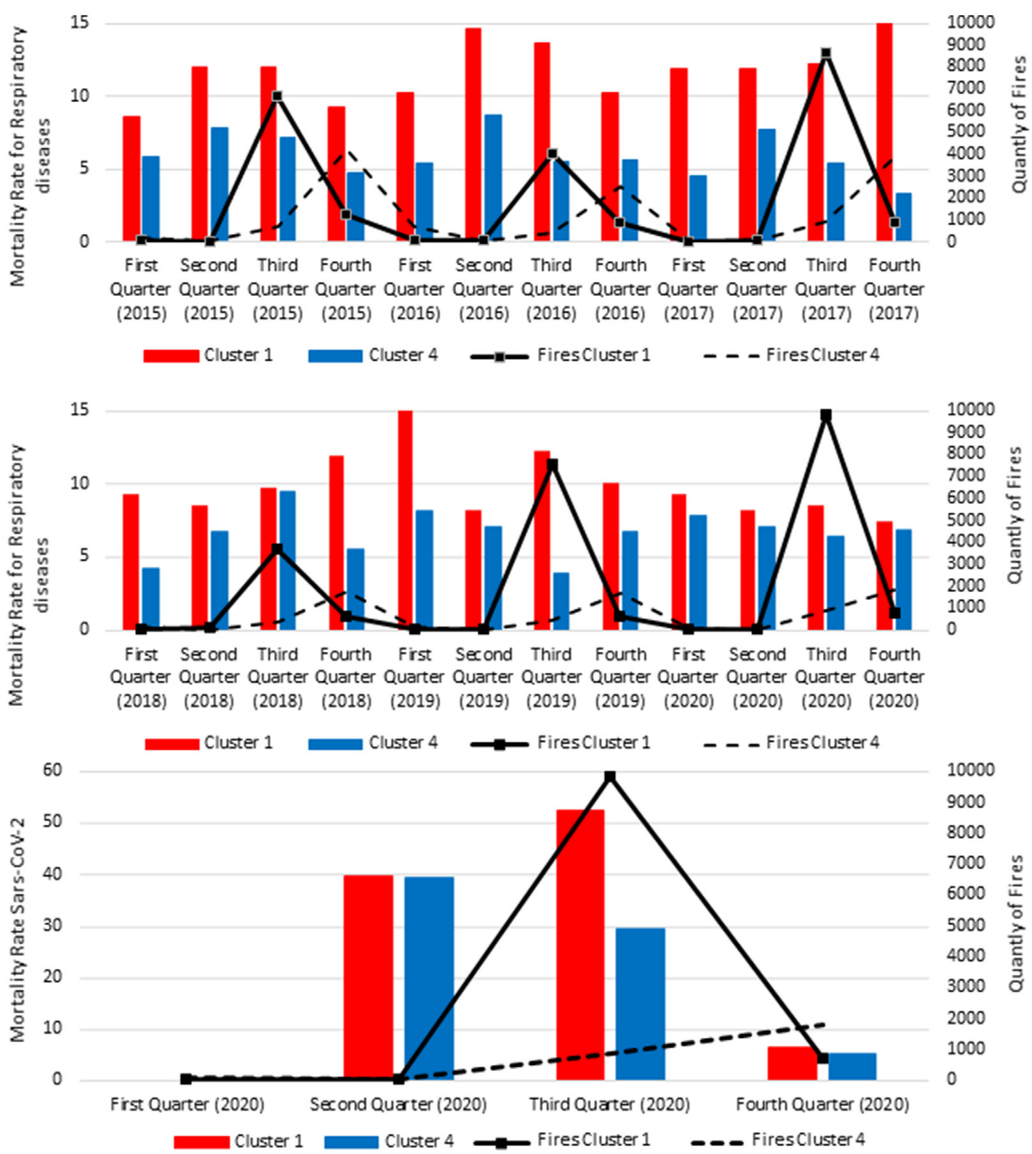
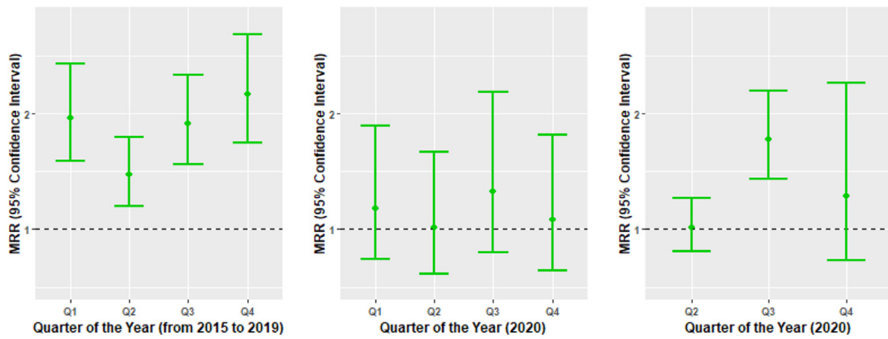


Figure 10. Mortality Rate time series for Respiratory disease (Top and Middle) and Mortality Rate SARS-CoV-2 (Bottom). Blue and Red bars refers to Cluster 1 and 4, respectively. Black continuous and dashed lines refers to Fire occurrences in clusters 1 and 4, respectively.

health actions in different locations considering the vulnerability of the specific population. The State of Para is the second largest state in Brazil.[32]

As of 2015, the intensification of the El Nio phenomenon was identified, causing the warming of waters on the surface of the Pacific, leading to the suppression of rains in eastern Amazonia and an increased risk of burning, mainly in the States of Maranhão, Mato Grosso and Pará[33] causing droughts and forest fires. Besides that, 2020 saw the highest occurrences of forest fires in the Amazon territory in the last ten years, affecting the respiratory health of the population, exacerbating the population's vulnerability.[34].

The occurrence of major fires in Brazilian and worldwide has drawn attention to the environmental problem, but measures to prevent and control their impacts are still insufficient.[35] The burning of biomass produces gaseous pollutants and fine particulate matter, which produce detrimental effects on the respiratory system.[36] Among the major components of wildfire smoke that can affect ambient air quality, is important to mention the fine particles PM2.5 and PM10 (particles with a diameter of less than 2.5 and 10 μm respectively), whose effects are expected to further increase when their concentrations get above air quality standards.[37].



(a) MRR for Respiratory diseases in the quarters of the years 2015-2019

(b) MRR for Respiratory diseases excluding deaths from COVID-19in the quarters of the year 2020

(c) MRR for COVID-19 in the 3 last quarters of the year 2020

Figure 11. MRR and the 95% CI.

The identification of the populations at risk for respiratory complications is of fundamental importance for the control of SARS-Cov-2 pandemic and future pandemicse.[38] Epidemiological studies relating to air pollutants and deaths from SARS-Cov-2 infection are receiving scientific attention. Methods on how the variables highlighted for the respiratory system can exacerbate the symptoms of COVID-19 and increase the risk of coverage,[39] allowing public and environmental agencies to take preventive measures.[40] The coronavirus (SARS-Cov-2) is generating a higher mortality rate and deaths mainly in people with comorbidities. For example,[41] outdoor air pollution causes significant morbidity and mortality. It can affect the respiratory system (exacerbation of asthma and chronic obstructive pulmonary disease) and the cardiovascular system (clearing arrhythmia, heart failure, and strokes).

Pollution generated by forest fires generates particles known as aerosols and long-term exposure to these particles increases the severity of outcomes associated with hospitalizations and mortality associated with COVID-19 as also identified by.[42] The global pandemic of SARS-Cov-2 has been linked to infections and deaths among people in polluted environments. The aerosols inhaled by people who have respiratory diseases have shown significance with the increase in hospitalizations according to,[43], and the forest fires greatly contribute to this factor. Moreover, several studies have reported the SARS-Cov-2 virus has a greater impact on people who already have some comorbidity, e.g. asthma, which is more prevalent in polluted regions.[44]

The analysis of free and public data of studies relating environmental aspects, hospitalizations and current topic related to SARS-Cov-2 allows the elaboration of an approach to study potential associations between historical exposure to atmospheric pollution and the increased vulnerability to deaths by COVID-19. One of the limitations in the development of studies is the lack of more specific details of hospitalizations for respiratory diseases and data on pre-existing vulnerabilities in deaths from SARS-Cov-2. The regionalized analysis proposed in this study does not guarantee the replicability of the results for large representative populations, or even for other cities in the State of Pará. Our study uses space and time approaches to analyze events regularly applied in many areas of research. Using our study as an example, we summarize the unmeasured limitations and confounding factors for future research.

Among the main limitations in working with public health data concerning deaths from SARS-Cov-2, practically in real-time, are the restrictions on data and individual risk factors such as age, race, comorbidities, and data on smoking for example. Considering the spatial scale, it was not possible to carry out in this work analyzes at the individual level, which can affect the data association between group of cities. Additionally, the data on COVID-19 diagnosis or case-fatality rate lack accuracy due to the irregular testing policies or more precise information available in some cases, which can make the findings in this systematically biased. Despite this, our work brings very important results contributing to a better understanding of the behavior of respiratory illnesses due to SARS-Cov-2 in areas with a high incidence of burns, highlighting the importance of specific care in the more vulnerable areas.

Regarding the application of machine learning techniques for knowledge discovery in a large amount of data, our proposed method combining the k-means algorithm, the elbow method, and time series analysis successfully outlined and corroborated with the previous hypothesis that the population exposed to pollution from forest fires were more susceptible hospitalizations and deaths during the SARS-COV-2 pandemic. Previous works attempted to make this correlation or to suggest machine learn techniques to extract such information like in the work of[45] that only suggest several techniques for clusterization to obtain COVID-19 related insights, the work of[46] that carried a space-time serious analysis over forest fires data, and the work of[47] that employed the k-means algorithm to cluster and analyse air pollution data. However, none of these works brought insights regarding respiratory diseases and the SARS-COV-2, which highlight the importance of our findings regarding insights to public policy related to health management and environment.

With this study, we can see that in the State of Pará there was an overlap of situations related to respiratory diseases. Populations residing in places with environmental problems tend to have greater impacts on the emergence of new diseases. Knowing the populations is of fundamental importance for the application of control and prevention policies. Applying to COVID-19 all the responsibility for the high mortality rate may not be the most correct way to interpret the impacts of the pandemic, as there are potentializing factors, in this case, the high rate of hospitalizations for respiratory diseases associated with the practice of burning.

Editor note

The Lancet Group takes a neutral position with respect to territorial claims in published maps and institutional affiliations.

Contributors

Lucas Schroeder - conceptualisation, data curation, formal analysis, investigation, methodology, project administration, resources, software, supervision, validation, visualisation, writing original draft, and writing, review & editing.

Eniuce Menezes de Souza, Clévia Rosset, Mauricio Veronez and Ademir Marques Jnior - data curation, formal analysis, investigation, supervision, validation, visualisation, and writing, review & editing.

Juliano Boquett, Diego Brum, Vinicius Francisco Rofatto, Luiz Gonzaga Jr. and Marcelo Zagonel de Oliveira - review & editing.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data sharing statement

The data used in this work as detailed in the ”Methods” section is made available publicly online. Figures and tables generated in this paper are available for public use given proper citation. Specific data used in data processing and analysis may be available by request.

Acknowledgements

Lucas Schroeder thanks Coordenaão de Aperfeioamento de Pessoal de Nível SuperiorBrasil (CAPES) for the financial support of the MSc scholarship.

References

1 Carrero C, Fearnside M, do Valle DR, Alves S. 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* 2020;**66**:966–84. https://doi.org/10.1007/s00267-020-01354-w.
2 Brando PM, Soares-Filho B, Rodrigues L, Assunção A, Morton D, Tuchschneider D, et al. The gathering firestorm in southern amazonia. *Sci Adv* 2020;**6**(2). https://doi.org/10.1126/sciadv.aay1632.
3 Escobar H. Amazon fires clearly linked to deforestation, scientists say. *Science* 2019;**365**(6456).853−853
4 Smith LT, Aragao LE, Sabel CE, Nakaya T. Drought impacts on children's respiratory health in the brazilian amazon. *Sci Rep* 2014;**4**(1):1–8.
5 Libonati R, Pereira J, Da Camara C, Peres L, Oom D, Rodrigues J, et al. Twenty-first century droughts have not increasingly exacerbated fire season severity in the brazilian amazon. *Sci Rep* 2021;**11** (1):1–13.
6 Ribeiro I, Andreoli R, Kayano M, Sousa T, Medeiros A, Godoi R, et al. Biomass burning and carbon monoxide patterns in brazil during the extreme drought years of 2005, 2010, and 2015. *Environ Pollut* 2018;**243**:1008–14. https://doi.org/10.1016/j.envpol.2018.09.022.
7 Cano-Crespo A, Traxl D, Thonicke K. Spatio-temporal patterns of extreme fires in amazonian forests. *The European Physical Journal Special Topics* 2021: 1–12.
8 National institute for space research burns - INPE. *Queimadas* 2020.
9 Butt EW, Conibear L, Knote C, Spracklen DV. Large air quality and public health impacts due to amazonian deforestation fires in 2019. *GeoHealth* 2021;**5**(7). https://doi.org/10.1029/2021GH000429.
10 Machado-Silva F, Libonati R, Melo de Lima TF, Bittencourt Peixoto R, de Almeida Frana JR, de Avelar Figueiredo Mafra Magalhães M, et al. Drought and fires influence the respiratory diseases hospitalizations in the amazon. *Ecol Indic* 2020;**109**:105817. https://doi.org/10.1016/j.ecolind.2019.105817.
11 Alves L. Amazon fires coincide with increased respiratory illnesses in indigenous populations. *The Lancet Respir Med* 2020;**8**:e84. https://doi.org/10.1016/S2213-2600(20)30421-5.
12 Pierce CA, Preston-Hurlburt P, Dai Y, Aschner CB, Cheshenko N, Galen B, et al. Immune responses to sars-cov-2 infection in hospitalized pediatric and adult patients. *Sci Transl Med* 2020;**12**(564). https://doi.org/10.1126/scitranslmed.abd5487.
13 Xu Z, Shi L, Wang Y, Zhang J, Huang L, Zhang C, Liu S, Zhao P, Liu H, Zhu L, Tai Y, Bai C, Gao T, Song J, Xia P, Dong J, Zhao J, Fu-S, Wang. Pathological findings of covid-19 associated with acute respiratory distress syndrome. *The Lancet* 2020;**8**:420–2. https://doi.org/10.1016/S2213-2600(20)30076-X.
14 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*, 2020. https://doi.org/10.1186/s41256-020-00167-y.
15 Notificaes de síndrome gripal(OPENDATASUS)2021https://opendatasus.saude.gov.br/dataset/casos-nacionais
16 Fayyad UM, Haussler D, Stolorz PE. Kdd for science data analysis: Issues and examples. KDD; 199650–6.
17 Grady NW. Kdd meets big data. 2016 IEEE International Conference on Big Data (Big Data). IEEE; 2016. p. 1603–8.
18 Chang WL, Grady N, et al. Nist big data interoperability framework: volume 1, big data. *definitions* 2015.
19 Limonati R, Camara CD, Setzer A, Morelli F, Melchiori A. An algorithm for burned area detection in the brazilian cerrado using 4 m modis imagery. *Remote Sens* 2015;**7**:15782–803.
20 Schroeder L, Veronez M, Souza E, Brum D, Jr LG, Rofatto V. Respiratory diseases, malaria and leishmaniasis: temporal and spatial

association with fire occurrences from knowledge discovery and data mining. *Int J Environ Res Public Health* 2020;**17**(10):3718–42. https://doi.org/10.3390/ijerph17103718.
21 Information technology department of the public health care system-sus. *(DATASUS)* 2021.
22 Nations United. Human Development Report 2020. *Online: https://www.undp.org/content/undp/en/home.html//* 2020.
23 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. *Nat Med* 2020;**26**(5):640–2. https://doi.org/10.1038/s41591-020-0863-y.
24 Han J, Kamber MPJ. Data mining: concepts and techniques. 3rd ed. Waltham,USA: Elsevier; 2011.
25 Corts S Da C; Porcaro RMLS. Minerao de dados funcionalidades, técnicas e abordagenss. *Pontifícia Universidade Catlica do Rio de Janeiro* 2002.
26 Likas A, Vlassis N, Verbeek JJ. The global k-means clustering algorithm. *Pattern Recognit* 2003;**36**(2):451–61.
27 Hartigan JA, Wong MA. Algorithm as 136: a k-means clustering algorithm. *Journal of the royal statistical society series c (applied statistics)* 1979;**28**(1):100–8.
28 T. Hastie and R. Tibshirani and J. Friedman. The elements of statistical learning: data mining, inference, and prediction. *Yale University Press - 2009* 2009.
29 Kodinariya TM, Makwana PR. Review on determining number of cluster in k-means clustering. *International Journal* 2013;**1**(6):90–5.
30 Kumar V, Steinbach M, Tan P-N. Introduction to data mining. 1st ed Addison Wesley; 2009.
31 Box G, Jenkins G, Reisel G. Time series analysis-wiley series in probability andStatistics (Ber). 2008.
32 Instituto brasileiro de geografia estatistica. *IBGE* - 2020. Accessed on: 10.10.2020.
33 Jiménez-Muñoz JC, Mattar C, Barichivich J, Santamaría-Artigas A, Takahashi K, Malhi Y, et al. Record-breaking warming and extreme drought in the amazon rainforest during the course of el niño 2015–2016. *Sci Rep* 2016;**6**(1):1–7.
34 Junior CHLS, Pessôa ACM, Carvalho NS, Reis JBC, Anderson LO, Aragão LEOC. The brazilian amazon deforestation rate in 2020 is the greatest of the decade. *Nature Ecology & Evolution* 2020;**5**(2):144–5. https://doi.org/10.1038/s41559-020-01368-x.
35 Karen B, W MS. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *The Lancet* 2012;**380**(4836):37–43. https://doi.org/10.1016/S0140-6736(12)60240-2.
36 do Carmo C, Hacon S, Longo K, Freitas S, Ignotti E, Leon AD, Artaxo P. Associao entre material particulado de queimadas e doenas respiratrias na região sul da amazônia brasileira. *Rev Panam Salud Publica* 2010;**27**:10–6.
37 Youssouf H, Liousse C, Roblou L, Assamoi E, Salonen R, Maesano C, Banerjee S, Annesi-Maesano I. Non-accidental health impacts of wildfire smoke. *International Journal of Environmental Research and Public Healtht* 2014;**11**:11772–804.
38 Solimini A, Filipponi F, Fegatelli DA, Caputo B, Marco CMD, Spagnoli A, Vestri AR. A global association between covid19 cases and airborne particulate matter at regional level. *Nature* 2021;**11**. https://doi.org/10.1038/s41598-021-85751-z.
39 Aveyard P, Gao M, Lindson N, Hartmann-Boyce J, Watkinson P, Young D, Coupland CAC, Tan PS, Clift AK, Harrison D, Gould DW, Pavord ID, Hippisley-Cox J. Association between pre-existing respiratory disease and its treatment, and severe covid-19: a population cohort study. *The Lancet* 2021. https://doi.org/10.1016/S2213-2600(21)00095-3.
40 Wu X, Nethery RC, Sabath MB, Braun D, Dominici F. Air pollution and covid-19 mortality in the united states: strengths and limitations of an ecological regression analysis. *Sci Adv* 2020;**6**(45). https://doi.org/10.1126/sciadv.abd4049.
41 Abelsohn A, Stieb D. Health effects of outdoor air pollution: approach to counseling patients using the air quality health index. *Can Fam Physician* 2011;**57**(8):881–7.
42 Wu X, Nethery RC, Sabath MB, Braun D, Dominici F. Air pollution and covid-19 mortality in the united states: strengths and limitations of an ecological regression analysis. *Sci Adv* 2020;**6**(45). https://doi.org/10.1126/sciadv.abd4049.
43 Fennelly K. Particle sizes of infectious aerosols: implications for infection control. *The Lancet* 2020;**8**(9):914–24. https://doi.org/10.1016/S2213-2600(20)30323-4.
44 Hu B, Guo H, Zhou P, Shi ZL. Characteristics of sars-cov-2 and covid-19. *Nature* 2021;**19**:141–54. https://doi.org/10.1038/s41579-020-00459-7.
45 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* 2020;**6**(2):58–68.
46 Parente J, Pereira MG, Tonini M. Space-time clustering analysis of wildfires: the influence of dataset characteristics, fire prevention policy decisions, weather and climate. *Sci Total Environ* 2016;**559**:151–65. https://doi.org/10.1016/j.scitotenv.2016.03.129.
47 Shafi J, Waheed A. K-means clustering analysing abrupt changes in air quality. 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA); 202026–30. https://doi.org/10.1109/ICECA49313.2020.9297493.