Spatial analysis of the COVID-19 distribution pattern
in São Paulo State, Brazil

Análise espacial do padrão de distribuição do COVID-19
no Estado de São Paulo, Brasil

Abstract At the end of 2019, the outbreak of COVID-19 was reported in Wuhan, China. The outbreak spread quickly to several countries, becoming a public health emergency of international interest. Without a vaccine or antiviral drugs, control measures are necessary to understand the evolution of cases. Here, we report through spatial analysis the spatial pattern of the COVID-19 outbreak. The study site was the State of São Paulo, Brazil, where the first case of the disease was confirmed. We applied the Kernel Density to generate surfaces that indicate where there is higher density of cases and, consequently, greater risk of confirming new cases. The spatial pattern of COVID-19 pandemic could be observed in São Paulo State, in which its metropolitan region stood out with the greatest cases, being classified as a hotspot. In addition, the main highways and airports that connect the capital to the cities with the highest population density were classified as medium density areas by the Kernel Density method. It indicates a gradual expansion from the capital to the interior. Therefore, spatial analyses are fundamental to understand the spread of the virus and its association with other spatial data can be essential to guide control measures.

Key words Coronavirus, Respiratory disease, Pandemic, Kernel density

Resumo No final de 2019, o surto de COVID-19 foi relatado em Wuhan, China. O surto se espalhou rapidamente para vários países, tornando-se uma emergência de saúde pública de interesse internacional. Sem uma vacina ou medicamentos antivirais, medidas de controle são necessárias para entender a evolução dos casos. Neste estudo, relatamos por análise espacial o padrão espacial do surto do COVID-19. Nosso local de estudo foi no estado de São Paulo, Brasil, onde o primeiro caso da doença foi confirmado. Aplicamos o método “Kernel Density” para gerar superfícies que indicam onde há maior densidade de casos e, consequentemente, maior risco de confirmação de novos casos. O padrão espacial da pandemia de COVID-19 foi observado no estado de São Paulo, em que a região metropolitana do estado foi a que apresentou a maior quantidade de casos, sendo classificada como um “hot spot”. Além disso, as principais rodovias e aeroportos que conectam a capital às cidades com maior densidade populacional foram classificadas como áreas de média densidade pelo método “Kernel Density”. Isso indica uma expansão gradual da capital para o interior. Portanto, as análises espaciais são fundamentais para entender a disseminação do vírus e sua associação com outros dados espaciais pode ser essencial para orientar as medidas de controle.

Palavras-chave Coronavirus, Doença respiratória, Pandemia, Densidade de Kernel
Introduction

At the end of 2019, the abrupt outbreak of the coronavirus disease (COVID-19), (caused by Coronavirus 2 of the Severe Acute Respiratory Syndrome - SARS-CoV-2) in Wuhan City, China, required mitigation and containment measures that brought to a halt the most populous country in the world1-3. COVID-19 was characterized as an acute respiratory disease that may turn into pneumonia with symptoms such as fever, cough and dyspnea. It has an approximate fatality rate of 2–3%3 and a high potential for contagion, which has led its incidence to increase exponentially.

COVID-19 widespread transmission was recognized by the World Health Organization (WHO) as a pandemic. Pandemics are spreads of infectious disease that lead to a rapid raise in the number of deaths. They have been around for centuries and historically reported in different parts of the world, including the recent Ebola spread in West Africa in 2015. The acceleration of the urbanization and interconnectedness of communities contribute significantly to the spread of diseases4,5. COVID-19 was considered to have been initially controlled in China. Nonetheless, the pandemic situation is still severe and grim in the other parts of the world1. By the date of 18th May, 2020, there were more than four million confirmed COVID-19 cases worldwide.

In Brazil, the Ministry of Health confirmed the first case on 25th February 2020, which represented the first confirmed case in Latin America6. From then on, a series of preventive measures began, such as social distance, the use of masks, among other measures, drastically influencing everyone’s work and living environments. Dubious and false information about factors related to virus transmission, incubation period, geographic reach, number of infected, and the actual mortality rate has led to the aggravation of the situation7. By 13th May 2020, in Brazil there were more than 293.000 confirmed cases.

Spatial analysis enables implementation of health programs that include several municipalities or regions of a State, thus playing an important role in public health diagnosis and planning. Kernel density estimation (KDE) is an important method for mapping spatial patterns of point events and several studies have been proposed in ecology11-13, criminology14, public health and epidemiology15,16, and among other fields. Previous studies have already described the spatial pattern of SARS17,18. However, few literature survey has been found using KDE for these purposes. Lai et al.19 used GIS technology and geostatistical methods such as KDE to analyze the patterns of disease spread of SARS outbreak in Hong Kong. The authors mentioned that its use can offer a scientifically rigorous and quantitative method for identification of unusual disease patterns.

COVID-19 has become a major public health challenge in all countries and its behavior and impacts are still unknown. Therefore, investigating its spread pattern is fundamental aiming to guide the next steps towards overcoming this crisis. Given that São Paulo is the most populous State of Brazil, the objective of this paper is to spatialize the confirmed cases of COVID-19 across the State and relate to demographic data using the KDE method to indicate areas of greatest risk of spread. Specifically, our goals are: (I) geographically spatialize the cities that confirmed positive cases of COVID-19; (ii) spatially the number of cases in each city; (iii) analyze the demographic density of each city. Finally, we crossed-check all information with data from the main highways to identify a pattern in the development of the disease.

Material and methods

Study area

The State of São Paulo is located in the southeast of Brazil (Figure 1), being the most populous State of the country, with an area of approximately 248,173 km², which corresponds to 2.9% of the Brazilian territory. Its capital is the City of São Paulo, the largest urban concentration in the country. The São Paulo administrative structure is composed of 646 municipalities and the population density is heterogeneous over the State20.

The estimated population in 2019 for the State of São Paulo is 45,919,049 inhabitants20. Moreover, São Paulo has the main airport of Brazil, the São Paulo-Guarulhos International...
Airport, with non-stop passenger flights scheduled to 103 destinations in 30 countries, and 52 domestic flights, connecting not only with major cities in Latin America but also with direct flights to North America, Europe, Africa and the Middle East (Dubai).

This first documented case in Brazil was a 61 years-old man, who traveled from 9 to 20th February, 2020, to Lombardy, northern Italy, where a significant outbreak was ongoing. He arrived home on 21st February, 2020, and was attended at the Hospital Albert Einstein in São Paulo, Brazil. By 13th May 2020, it was the state with the highest incidence of COVID-19.

Data description

The data used were acquired from the collaborative platform Brasil.io (https://brasil.io/home/), which is dedicated to compiling all information, in real-time, using the bulletins from the State health departments. The platform was chosen because it has the advantage of delivering these data with the geocode of each city, making it easier to insert the data into a GIS. We selected the date of 13th May, 2020 to start performing the data analysis. On this date, 51,097 confirmed cases were registered in 433 municipalities. Also on this data, Brazil overcame 13,000 deaths confirmed by COVID-19. Data from 40 cases were excluded for not having information on the place of residence.

The data were analyzed and processed in GIS environment (QGIS version 3.8), with meshes and vector points that consider the municipalities with confirmed cases, as well as the polygons of the municipal limits. QGIS is an Open Source Geographic Information System for analysis. The digital mesh of the 645 municipalities was obtained from the Brazilian Institute of Geography and Statistics (IBGE). We geocoded all COVID-19 cases and matched them to the city-level layers of polygon and point by using QGIS 3.8 software. We also set up all layers for the Coordinate Reference System in the SIRGAS 2000 / UTM zone 22S.

It should be noted that the data is constantly updated. On the portal, it is possible to obtain
the number of cases per municipality. This is an ecological study in São Paulo State, using spatial analysis as data processing technique. Municipalities were used as the geographical units of analysis. This study did not require submission to the Ethics and Research Committee, as it uses public access data, available on the Brasil.io website.

Demographic and road data were also used to complement the analysis. In fact, we used data from the main highways that cross the State of São Paulo. The highways were acquired through shapefile files downloaded from the National Department of Infrastructure Transport (DNIT)\textsuperscript{21}. The demographic data were obtained from the platform DATAGEO – Sistema Ambiental Paulista. We used the most recent demographic data available, which belongs to the year of 2018.

### Kernel density

To explore the spatial distribution pattern of COVID-19 cases on the State level, we applied Kernel Density Estimation (KDE). This methodology is based on points, so it is necessary to link information to a specific point in the space\textsuperscript{22}. KDE is a non-parametric method that uses local information defined by windows (also called kernels) to estimate densities of specified features at given locations. According to Silverman\textsuperscript{23}, the kernel density estimator can be defined as:

\[
\hat{f}(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left\{\frac{x - X_i}{h}\right\} (1)
\]

where: \( h \) is the bandwidth or smoothing parameter, \( K \) is the Kernel and \( f(x) \) is the estimator of the probability density function \( f \). The \( X_i \) are the \( n \) sample values (objective function value of the solutions generated by the \( n \) iterations of the stochastic search algorithm. Thus, the Kernel estimator depends on bandwidth (\( h \)) and Kernel density (\( K \)).

The surface value is the highest at the point location and reduces with increasing distance from the point, reaching zero at the search radius distance from the point. Only a circular neighborhood is possible. The volume under the surface equals the Population field value for the point, or one if NONE is specified\textsuperscript{24}. Using KDE, we could identify the type and the spatial distribution pattern, including significant hot spots, medium spots, and cold spots. Here, the categorization was separated in five categories i.e. Very low, low, medium, high and very high based on their associated case densities.

### Results

#### Current status

In the 646 municipalities of São Paulo State, 51,057 cases of COVID-19 were reported on 10th May, 2020. The number of cases included in the study is concentrated in the metropolitan region of São Paulo, which includes the capital (Table 1). The São Paulo capital has the highest number of reported cases (30,402 people detected with COVID-19), representing approximately 60% of all cases. More than half (61%) of the municipalities had a number of cases less than or equal to 10. Considering the raw data, 213 municipalities did not report cases of COVID-19.

#### Spatial distribution COVID-19

The spatial distribution of COVID-19 confirmed cases in the cities of the São Paulo state is shown in Figure 2. The largest concentration of cities that confirmed positive cases of COVID-19 is located in the Metropolitan Region of São Paulo (MRSP), indicated by warm colors (red and orange).

With the spatialization of confirmed cases of COVID-19 across the state, it is possible to notice a decrease in the concentration of cities that confirmed positive cases is indicated by the colors red to green, that is, the cold colors represent the areas of lowest concentration. It means that in addition to a greater concentration of cities in the metropolitan region with confirmed positive cases, there was probably a gradual spread from the capital to the interior of the State. It is also possible to view some areas of medium concentration of cities with confirmed cases, there was probably a gradual spread from the capital to the interior of the State. It is also possible to view some areas of medium concentration of cities with confirmed cases in the northern and central regions of the State. We realized that these points characterized by medium density have, in short, small cities and at least one medium-size city nearby, which may indicate a pole in the region and possibly a hotspot due to the inflows and outflows of people working in the city and later returning to their respective cities. In addition, we plot the main airports in the state and realize that the vast majority are either in medium density areas or in the metropolitan region, which is responsible for the largest flow of people through airports.

The Kernel map with the spatial distribution of the cases allowed us to indicate the areas where there is a higher concentration of cases as well as where areas of medium and low concentration. In addition to the fact that the Kernel maps indi-
icated the areas with the highest occurrence density, they also demonstrated the probability of risk of exposure to the virus. In other words, hot areas on the map showed where there is the greatest probability of confirming a new case, whereas cold areas there may be a delay to confirm.

There is a clear pattern of the spread of the virus by the State. A high density of cases occurs in the metropolitan region (red color) and gradually there is a lower concentration of cases in the countryside. We have noticed that in cities close to the State borders there is a lower concentration of confirmed cases of COVID-19, which appears in green color. This result may indicate that cities close to the borders are more disconnected from the metropolitan region as well as the

Table 1. Descriptive analysis with median, minimum and maximum values of COVID-19 in the São Paulo State, on May 10, 2020.

<table>
<thead>
<tr>
<th></th>
<th>Min of cases</th>
<th>Median of cases</th>
<th>Max of cases</th>
<th>County Count</th>
<th>Sum of confirmed cases/municipalities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interior</td>
<td>1</td>
<td>4</td>
<td>1069</td>
<td>394</td>
<td>9390</td>
</tr>
<tr>
<td>MRSP (including capital)*</td>
<td>7</td>
<td>199,0</td>
<td>30402</td>
<td>39</td>
<td>41667</td>
</tr>
<tr>
<td>MRSP (excluding capital)*</td>
<td>7</td>
<td>191,5</td>
<td>1390</td>
<td>38</td>
<td>11265</td>
</tr>
<tr>
<td>Capital</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>30402</td>
</tr>
<tr>
<td>Total</td>
<td>1</td>
<td>5</td>
<td>30402</td>
<td>433</td>
<td>51057</td>
</tr>
</tbody>
</table>

*Metropolitan Region of São Paulo.

Figure 2. Kernel maps of the São Paulo State - Brazil. a) cities with positive cases of COVID-19 and main highways; b) population density and main highways; c) number of confirmed cases per municipality and main highways.
capital where the disease started. We confirm this fact after intersecting the main highways and airports present in the state. Thus, the spread of the virus can be highly associated with the mobility of people by these means of transport. The main highways and airports in the State are mostly in areas of medium to high density of cases and practically all connect to the metropolitan region and capital of the State, the epicenter of Brazil.

Discussion

COVID-19 is an infectious disease caused by a new virus and characterized by a series of symptoms that can vary from person to person, as well as from country to country. Limited knowledge of this new disease led to the sudden outbreak and rapid development of a pandemic. In this study, we demonstrated the spatial distribution of COVID-19 together with the density of infected cities, associating with population data and the main highways, factors that must be considered to adequately model the disease.

We sought to analyze the spatial distribution of COVID-19 cases in the municipalities of the State of São Paulo relating to the geographic density of the cities through the analysis of Kernel density and we were able to identify that there is a uniform distribution among the regions of the State. It is noteworthy that the metropolitan region of São Paulo, composed of the highest population densities in the State, is contained in the area with very high density of cases, thus being an area of greater probability of new cases.

The first countries that suffer with COVID-19 have developed several mathematical models to model its outbreak, mostly focused on forecasting the number of cases and assessing the capacity of country-level healthcare systems to manage disease burden. Large-scale pandemic prevention and control decisions and actions depend on data support. The development and application of spatial analysis tools will undoubtedly help to quickly identify the spatio-temporal process of pandemic development, prevention and control measures and the resulting effectiveness.

A similar study developed by Chen et al. reported that the population that emigrated from Wuhan was the main source of infection in other cities and provinces and that some cities with a low number of cases showed a rapid increase of cases. Additionally, the authors reported that due to the Spring Festival’s next wave of return, it is crucial to understand risk trends in different regions to ensure preparedness at the individual and organizational levels and prevent further outbreaks.

Early on, the spatial distribution of COVID-19 cases in China was explained well by human mobility data. After the implementation of control measures, this correlation dropped and growth rates became negative in most locations, although shifts in the demographics of reported cases were still indicative of local chains of transmission outside of Wuhan. This fact corroborates and is evidenced in our study, where most of the main highways and airports are in medium and high density areas, and only a few in low density areas.

In a similar study, Vaz and Nascimento identified through spatial analysis techniques that the greater social inequality is also concentrated in the metropolitan region of São Paulo. This factor may be relevant for the authorities and managers of the municipalities since people with less financial resources and a deficiency in the public health system can be a hotspot for the collapse. Joventino et al. suggested that monthly family income equal to or less than a minimum wage is a risk factor for helminthiasis (a parasitic disease), in addition to that a higher income allows better living conditions, such as basic sanitation, better habitability and easier access to personal and home hygiene products, and better means to treat water. The same can be inferred for COVID-19, since minimal hygiene conditions are required for the prevention, as well as the use of gel alcohol, individual masks, and water for cleaning in general.

To date, researchers around the world are running out of time to find an effective method that can stop this new virus, mainly because there are still no specific drugs or treatment protocols for COVID-19. The best way to protect and slow transmission must be well informed about the current COVID-19 virus. Monitoring active ties using GIS spatial analysis is very important to control such as a COVID-19 virus spreading problem. At a time when the number of COVID-19 cases is constantly increasing in Brazil, our findings highlight the high potential for the introduction of new cases in several cities in São Paulo, especially in larger cities close to the capital. The accurate identification of places where clusters of local transmission might first ignite is critical to better coordinate preparedness, readiness and response actions.
population, and there may be a different pattern than the one presented here, since there may be a relatively important proportion of asymptomatic people with the virus. However, this research opens new horizons for the development of future studies with more robust databases. Furthermore, new studies could assess the time-space evolution of the distribution pattern by the kernel density method. Thus, a specific pattern can be identified and useful to face new outbreaks of new diseases. The still incipient studies carried out in relation to the spatial distribution of COVID-19 in Brazil make it difficult to perform the proper comparisons. Nonetheless, some considerations are important in the interpretation of the data found. First, the definition of a confirmed case is based on laboratory criteria. This fact is intrinsically a potential underreporting of cases for both the municipality and the national territory.

This work can provide valuable information to support government monitoring and predict the dissemination of viruses. The WHO announced that the outbreak of Novel Coronavirus disease was a pandemic and reiterated its call for immediate action by governments to step up their response to diagnose, identify, and mitigate spread to save lives. Given the situation, it is essential that effective measures are taken to avoid the risk of continuous outbreaks and the possibility of a local outbreak, especially in smaller cities which are limited in resources and structure. Research that proposes to develop effective methods to monitor and provide early and timely detection of the disease should be carried out and used by local managers to combat the virus.

Conclusions

COVID-19 has become a big challenge to public health around the world. Hence, its spread pattern needs to be investigated in order to support the governments in overcoming this pandemic. Since São Paulo is the most populous state of Brazil, this paper explored the KDE method to spatialize the confirmed cases of COVID-19 across the State taking into account demographic data to indicate the greatest areas with risk of spread. The GIS spatial analysis method revealed a spatial pattern as well as the COVID-19 critical points in São Paulo State on 13th May 2020. The use of spatial analysis tools such as KDE to monitor the evolution of cases can become an essential component in specific programs for taking measures to reduce the public health risk.

Collaborations

FE Rex and CAS Borges jointly participated in preparing the project, obtaining the data, analyzing the data and writing the article, PS Käfer participated in the data analysis and writing of the article as well as in the review. All authors approved the final version to be published.

Acknowledgments

The authors are thankful to the Instituto Brasileiro de Geografia e Estatística (IBGE) for providing the data used in this research. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brazil and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). The authors declare no conflict of interest.

References


24. ESIr. Arcgis Desktop help. (s.d.).


Article submitted 22/05/2020
Approved 01/06/2020
Final version submitted 03/06/2020

This is an Open Access article distributed under the terms of the Creative Commons Attribution License