An Environmental GIS-based Variable Analysis on SARS-CoV-2 in the City of Recife, Brazil

Arthur Souza

Universidade Federal de Pernambuco

Caroline Maria de Miranda Mota (✉ carolinemota@cidsid.org.br)

Universidade Federal de Pernambuco <https://orcid.org/0000-0001-9617-5726>

Amanda Rosa

Universidade Federal de Pernambuco

Ciro Figueiredo

Universidade Federal Rural do Semi-Arido

Ana Lucia Candeias

Universidade Federal de Pernambuco

Research

Keywords: Covid-19, Spatial Cluster Analysis, Environmental Analysis, GIS

Posted Date: October 23rd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-95748/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Given the increasing rates at which people have been infected by Covid-19 evolving to case-fatality rates on a global scale and the context of there being a world-wide socio-economic crisis, decision-making must be undertaken based on prioritizing effective measures to control and combat the disease since there is a lack of effective drugs.

Method: This paper explores the determinant factors of the COVID-19 pandemic and its impacts on Recife, Pernambuco-Brazil by performing both local and global spatial regression analysis on two types of environmental data-sets. Data were obtained from ten specific days between late April and early July 2020, comprehending the ascending, peaking and descending behaviours of the curve of infections.

Results: This study highlights clusters of the most affected neighbourhoods and their determinant effects. We have observed the increasing phase with hotspots of confirmed cases in a well-developed and heavily densely-populated neighbourhood of Recife city, then evolving for hotspots of case-fatality rates into areas characterised by having a precarious provision of public services and low-income population. The results also help to understand the influence of the age, income, level of education of the population and, additionally, the people's access to public services, on the behaviour of the virus across neighbourhoods.

Conclusion: This study supports government measures against the spread of Covid-19 in heterogeneous cities, evidencing social inequality as a driver for a high incidence of fatal cases of the disease. Understanding the variables which influence the local dynamics of the virus spread becomes vital for identifying the most vulnerable regions for which prevention actions need to be developed.

1. Introduction

The recent identification of a new virus in Wuhan city, China, has been receiving close attention not only due to how quickly it spread but also because of the threat this represents to the public health system. This situation became a concern since the first epicentre of the disease was associated with the intense flow of transportation in China ¹, which led to there being a significant movement of infected agents and, therefore, of SARS-CoV-2 transmitters.

COVID-19 is directly related to infections of the respiratory system ². The most efficient way to combat it is prevention by social distancing and, in extreme cases, lock-down ordered by governmental entities. These actions have presented positive results ³. Previous studies have demonstrated the relevance of quarantine and have done so by estimating the epidemic doubling time and the basic reproductive number of COVID-19 ⁴, by analysing preventative strategies found in the literature and for which a framework has been developed based on differential equations ⁵. This involves conducting a factor analysis such as the number of susceptible people, infected people not entering quarantine, infected

people in quarantine and confirmed cases ⁶ and by applying ARIMA to predict the rate of increase in the number of cases of COVID-19 ⁷ .

Given the dynamic behaviour of the propagation of Covid-19, time and mobility factors have singular importance ⁸ due to trend variation arising from gradually implemented policies to control the spread of the virus. In the current pandemic scenario, visual tools have been used in interactive dashboards ⁹. This approach seems valuable for visualising cases in real-time and can help by prompting insights into how best to mitigate the spread of the disease. But the understanding of determinant factors is also important because this provides a more robust methodology for the combat against future infections. Thus, combined with statistical methods, spatial analysis helps improve understanding how social, demographic and even spatial factors influence the propensity of a neighbourhood for Covid-19 cases. The study is focused on Recife city, located in the state of Pernambuco, Brazil.

The paper is divided into six sections. Section 2 presents a brief review of statistical methods and their applications in the context of public health. In Sect. 3 the methodological steps are described. Section 4 discusses the implementation of GIS-based statistical analysis. In Sect. 5 the results of the study are discussed. Conclusions are presented and suggestions made on approaches that future studies on this topic might take, in Sect. 6.

2. Brief Literature Review

Statistical modelling can be significant in decision-making support in a health context ¹⁰ since it is able to supply a prior understanding of the real problem situation such as labelling the most relevant variables of an event. In view of the pandemic, raw data related to the number of confirmed cases and deaths should be processed in order to capture how the disease has propagated and its overall impacts. This enables an overview of the most affected areas to be obtained, from which their environmental characteristics will be investigated. That information could help to draw up efficient policies to mitigate the causes of Covid-19.

In situations in which an infectious disease is spreading, the use of differential equations has a vast application with regard to the dynamics of how a disease contaminates people over a given period of time with a view to estimating and predicting the number of cases ^{11,12}. In addition to a mathematical approach, a literature review by ¹³ points out that open data-set and artificial intelligence are used as additional tools to analyse the spread of Covid-19.

Mathematical models have been used as a tool to gather information about trends even among countries ^{14,15} for which an extension of the classical SIRD model is used including relevant aspects for decision-making such as parameters for sanitary policies (distancing, border control, reported cases and the elapse of time between doing the exam and obtaining the result) ¹⁴ .

Spatial analysis has been applied in the Covid-19 context, as noted in an exploratory review developed by ¹⁶. They placed research papers into five categories: data mining, environmental variables, health and social geography, spatiotemporal analysis and web-based mapping. ¹⁶ highlighted the importance of matching distinct explanatory variables, investigating spatial and temporal dimensions of the spread of Covid-19, discussing the geographical impact of the disease on decision-making and predicting its evolution.

Autocorrelation and cluster analysis regarding cumulative cases of COVID-19 in China have been investigated by means of Global Moran's I and Getis-Ord General G ¹⁷. In ¹⁸ Local Moran's I is also used for mapping clusters, along with a logistic regression model to predict the growth of the infection curve and a SEIR model to calculate the rate of spread.

In ¹⁹ Global Moran's I, Local Moran's I and Getis-Ord General G statistics are used to characterise spatial autocorrelation regarding confirmed cases of coronavirus at the municipal and county levels. They also examined relationships among cases and described factors related to the population, the economy and the height of the area above seal-level using Spearman's rank correlation.

Contemplating a wide variety of aspects, ²⁰ compared five different local and global spatial regression models in order to explore the relationship between 35 environmental, socio-economic, topographic and demographic variables and the spread of Covid-19 in the United States at a county level. Among them, a small set of four determinants could explain the incidence of the disease: average household income, income inequality, percentage of nurse practitioners and percentage of black women. Considering the regression models compared, the one based on Multiscale Geographically Weighted Regression (MGWR) had the best performance.

Our study combines both statistical and spatial analysis in order to comprehend the behaviour of the outbreak of coronavirus in the neighbourhoods of the city of Recife, which is the capital of the state of Pernambuco, situated in the Northeast region of Brazil. A general context of the municipality is obtained by comparing the curve of the number of confirmed cases with governmental interventions. Hotspots of reported cases and case-fatality are investigated in different moments of the spread of Covid-19 as a way of identifying distinctive patterns in the most affected areas. Spatial regression analysis is used in order to seek to single out which local characteristics among socio-economic factors and the presence of essential services best explain where cases were concentrated.

3. Data And Method

This study aims to develop a spatial evaluation that is concerned with how coronavirus infection rates behave across Recife neighbourhoods by identifying what places are more susceptible to its spread according to local characteristics. For the sake of achieving consistent results, we combined a statistical approach with GIS-based spatially dedicated methods.

We investigated Recife's overall situation regarding confirmed cases over time since the first ones were confirmed in March 2020. A 7-day moving average was applied to raw data, thereby seeking to reduce sudden daily variability due to reporting biases such as the effects of weekends and public holidays, the lack of tests and the delay in recording cases and deaths ²¹.

In order to comprehend disparities in the spread of the virus on a smaller scale, neighbourhoods were explored using a spatial cluster analysis, which have been used to analyse other diseases' spread^{22,23}. Data were obtained from ten specific days between late April and early July as a way of representing the ascending, peaking and descending behaviours of the curve of infections. Local Moran's I ²⁴ was implemented for cumulative confirmed cases and the case-fatality rate, which means the total number of deaths divided by the total number of cases, so that statistically significant hotspots, coldspots and outliers could be identified. Socio-economic factors representing those clusters were extracted from the 2010 Brazilian Census ²⁵ and submitted to quartile analysis and Spearman's rank correlation tests as a means to clarify patterns.

Our study also aims to identify the most relevant factors to explain Covid-19 cases per neighbourhood. Therefore we applied a regression analysis using Covid-19 data gathered from the same ten days used for the cluster analysis as a dependent variable, while previously investigated factors formed the initial set of determinants. Georeferenced environmental elements representing non-stop services during pandemic were tested for Pearson's product-moment correlation so that they could also be added as explanatory variables. Initially, essential services and socio-economic factors were treated separately, thus assuring that each database would be explored in depth. An Ordinary Least Squares (OLS) regression was run twice for each day considered, discarding correlated variables and using a selection method to reduce both sets of determinants.

Additionally, an analysis was performed aggregating both data-sets, socio-economic and georeferenced facilities, as a way of understanding their likely synergistic effect on predicting the number of Covid-19 cases. In this study, the use of multi-source data is possible since we have complementary full data-sets for populations ²⁶, on which a one-to-one linkage procedure was applied for complete observations. Otherwise, in situations of sample-based data-sets, it must be observed if a set of norms and sample conditions are satisfied, as discussed in ²⁷.

Geographically Weighted Regression (GWR) was handled to relax the OLS assumption of estimating global values for regression parameters, thereby allowing relationships among variables to vary over space and to be determined for each location ²⁸. OLS and GWR statistical performances were compared, taking account of the reduced set of determinants found by using data acquired from July 3rd, 2020. Then we distinguished how GWR outputs explain the spread of Covid-19 in each neighbourhood and also elucidate how every relevant factor impacts on the prominence of hotspots for new cases.

4. Results

Loading [MathJax]/jax/output/CommonHTML/jax.js

The city of Recife has a considerable prominence in Brazilian context since it is the second-most densely populated city, with an estimated population of 1.55 million people in an area of 218 Km². Recife is the capital of the third most populated state in the Northeast region, with the highest GDP per capita in the Northeast region ²⁹. So we aimed to explore how Covid-19 cases in Recife neighbourhoods relate to a set of socio-economic and demographic characteristics and the provision of essential services. Data on the disease were obtained from the municipality's Department of Health ³⁰. An evolution of cases over time was plotted on Fig. 1, including relevant actions taken by the State Government and City Council.

It was noted that these authorities took an early decision to close facilities when first cases were reported, such that only supermarkets, grocery stores, bakeries, pharmacies, gas stations and pet shops were allowed to open ³¹. As daily infections were increasing, the state governor issued an edict obliging the population to wear masks in public places. However, as fines were not imposed on shoppers but on the owners of the commercial facilities they entered, adherence to the measure became more dependent on the willingness of the general public to wear masks and intense supervision at the entrances to commercial premises.

Current measures had not been enough to flatten the increase in cases of infection, so in mid-May a 15-day strict quarantine was established in five municipalities in Pernambuco, including Recife. People were only allowed to leave their homes to seek essential services for which they had to show proof and vehicles were only allowed on roads according to a rota system based on the final number of the number plate. The spread of the virus peaked in late May and the cases seem to have been stabilised at a low level for now (July 8th, 2020), even though restrictions have been gradually relaxed and places in which people congregate such as shopping malls and commercial premises have been reopened.

4.1 Spatial cluster analysis

Hotspots with regard to the total number of reported cases per neighbourhood were generated by means of Moran's I. We noted in Fig. 2 that cases of the disease at first were concentrated in the South Zone of the city, specifically in the well-developed and heavily densely-populated neighbourhood called Boa Viagem. As cases increased, other hotspots of High-High and High-Low types were found in the South, the North and the West Zones of the city. A quartile analysis was developed using 15 census indicators ²⁵ concerning residents' average income; government support for piped water, sewage disposal, electricity, and garbage collection; home ownership; population density and five age groups which were considered in absolute numbers. Results revealed social disparities in these areas compared to Boa Viagem, since in terms of access to sanitation, garbage collection, income and literacy, these other hotspots are at least 50% below the indices for these variables compared to all other neighbourhoods in Recife.

Case-fatality hotspots due to Covid-19 were also examined as plotted and shown in Fig. 3, highlighting areas of high case-fatality rate that are surrounded by areas of the same pattern, so-called high-high clusters. A different spatial pattern arises when comparing to the clusters of cases. They kept mostly

stable throughout time so that hotspots usually formed in the North and Southwest zones, while cold spots were seen in the North and West zones.

Although Boa Viagem has been a relevant cluster for reported cases since the first date analysed, this has not occurred for case-fatality. This neighbourhood presents a high number of cases, but the case-fatality rate has been growing in a lower proportion. The first death in Boa Viagem was confirmed on April 10th after 63 cases had been recorded, which represented 15% of the Covid-19 cases in Recife at that time. Boa Viagem has led the number of deaths in the city since April 27th until July, 8th. Although when analysing the relationship between cases and deaths, the case-fatality rate, this neighbourhood did not become the hotspot in any of the transit periods, i.e, Boa Viagem neighbourhood has concentrated the highest number of cases but it has had a low number of deaths per cases. On July 3th, the neighbourhood reached its maximum case-fatality rate, 18%, which, however, is considerably lower than the hotspots of case-fatality, for which the percentages were between 34% and 54%.

The results from quartile analysis show that hotspots of case-fatality rate (the high-high clusters), usually present areas with similar environmental characteristics to those in hotspots of confirmed cases, disregarding Boa Viagem. For instance, most of these areas are characterised by having a precarious public service provision and low-income population. The opposite characteristics were verified when analysing the low-low clusters. It has been observed that neighborhoods in low-low clusters (of both case and case-fatality rates) have fewer residents per household than 75% of all other neighborhoods.

Spearman correlation tests were applied to case-fatality rate and socio-economic factors using data from July 3rd, 2020 as a way of finding monotonic relationships. Significant results were obtained considering a p-value less than 0.05. Regarding population aspects, a negative connection was identified with income ($\rho = -0.51$), literacy rate ($\rho = -0.44$) and percentage of people over 60 years old ($\rho = -0.32$), while residents per household presented a positive one ($\rho = 0.41$). Analysing governmental support showed that access to the sewage system ($\rho = -0.47$) and garbage collection ($\rho = -0.27$) were both negatively related to case-fatality.

Aggregating previous results indicate that some places tend to suffer from fewer deaths due to Covid-19 when residents in such places have a level of income and literacy that is well above average for Recife and because the number of residents per household is lower than elsewhere in Recife. These better-off areas also have more access to public services. So the residents of these privileged areas had the infrastructure and access to goods and services (public and private) that were needed to isolate themselves in relative safety and to respect quarantine restrictions. However, in contrast to what was expected, the results also showed that the percentage of the elderly population was highest in low-low clusters of case-fatality (when comparing the distribution of elderly people over the studied areas). In other words, although the elderly are more prone to catching severe forms of Covid-19 ³², the incidence of these was less in clusters with a high concentration of elderly people.

4.2 Analysis of determinant factors

Local determinants for Covid-19 were explored in greater depth by using spatial regression analysis. An initial set of 15 explanatory variables was compiled from census indicators, as explained above, such as income, literacy and the provision of public services.

A second evaluation set was compiled of places that are typically conducive to attracting crowds of people and that operated even during strict quarantine, and so could have become centers of Coronavirus infections. They were called essential services and six of them were evaluated by means of Pearson correlation tests: bakeries, banks, bus terminals, grocery stores/supermarkets, official lottery establishments which also act as the agent of a public bank, pharmacies. Grocery stores represent a range from small shops to large supermarkets which sell food and general items for domestic use.

Results indicated that all places considered have significant and positive correlations to confirmed cases of Covid-10 for a p-value of 0.05. The Pearson index varied between 0.23 and 0.88, bakeries ($p = 0.88$), pharmacies ($p = 0.86$) and grocery stores ($p = 0.71$) being the top three highest. Thus these six factors formed a second set of determinants.

4.2.1 Analysis of socio-economic versus essential services factors

Multiple regression models were built for different days between April and July, taking the total reported cases as the independent variable and using the OLS method. They were processed twice - for essential services and for socio-economic factors. In each model, highly correlated variables were removed when they had a variance inflation factor (VIF) greater than 7.5. The remaining determinants were submitted to the stepwise method based on statistics that use the Akaike Information Criterion (AIC), thereby seeking to reduce them to a non-redundant set³³, which can sufficiently explain the spread of SARS-CoV-2 in Recife's neighbourhoods and from neighbourhood to neighbourhood. Results are shown in Table 1.

Table 1
The performance of OLS models over time using separate datasets of determinants

Date	GM*	Essential services	Adj. R ²	Socioeconomic factors	Adj. R ²
April 16th	1	Bakeries, bus terminals	0.7793	Age 0 to 9, age over 60, resident per household, income, literacy, rented home	0.8778
April 23rd	1; 2	Bakeries, grocery stores, banks	0.7558	Age over 60, residents per household, income, literacy, piped water, garbage collection, owned home, sewage system	0.8597
May 3rd	1; 2	Bakeries, grocery stores, pharmacies	0.8016	Age over 60, residents per household, income, rented home	0.8939
May 12nd	1; 2	Bakeries, grocery stores, banks, pharmacies	0.8272	Age over 60, piped water	0.9086
May 19th	1; 2; 3	Bakeries, grocery stores, banks, pharmacies	0.8415	Age 0 to 9, age over 60, piped water, garbage collection	0.9279
May 27th	1; 2; 3	Bakeries, grocery stores, banks, pharmacies	0.8583	Age 0 to 9, age over 60, piped water, garbage collection	0.9398
June 3rd	2; 4	Bakeries, grocery stores, banks, pharmacies	0.8583	Age 0 to 9, age over 60, piped water, garbage collection	0.9436
June 12nd	2; 4	Bakeries, grocery stores, banks, pharmacies	0.8631	Age 0 to 9, age over 60, piped water, garbage collection	0.9477
June 24th	2; 4; 5; 6	Bakeries, grocery stores, banks, pharmacies	0.8639	Age 0 to 9, age over 60, piped water, garbage collection, residents per household	0.9509
July 3rd	2; 4; 5; 6	Bakeries, grocery stores, banks, pharmacies	0.8659	Age 0 to 9, age over 60, piped water, garbage collection, residents per household	0.9521
<p>* Government measures (GM) in 2020:</p> <div style="display: flex; flex-wrap: wrap;"> <div style="width: 50%;"> <p>1. Closing of non-essential commercial activities</p> <p>2. Mandatory use of masks</p> <p>3. Lockdown</p> </div> <div style="width: 50%;"> <p>4. Reopening of building supply stores</p> <p>5. Reopening of beauty salons and suburban retailers</p> <p>6. Reopening of malls and places of worship</p> </div> </div>					

The column called GM, government measures, indicates which decisions were being imposed on each

gradually reduce or enable people and traffic to be on the

streets during the various stages of the pandemic to date.

All regression models were significant ($p < 2.2E-16$) and most independent variables had a positive correlation to the number of Covid-19 cases, except for 'residents per household', 'people aged from 10 to 19', 'bus terminals' and 'banks'. In both sets of determinants, the adjusted determination coefficient R^2 mostly increased along with the cumulative total of confirmed cases, which reasserts the connection of the set of original factors with the variance in data. Its value stabilised from early June which is likely due to the scenario of the slow growth of cases in Recife.

The final, relevant, explanatory variables with regard only to essential services remained stable from May 12th, while the main socio-economic factors that explained the results started repeating themselves after May 19th. Although the local government has since implemented relaxation measures as seen in Fig. 1, these findings indicate their importance for predicting cases. When determinants were evaluated individually, it was noted that 'bakeries' could explain 77% of the variability in cases, and 'total residents' nearly 90%.

The subset of determinants obtained with the most recent data was considered for further analysis by applying the spatial Geographically Weighted Regression (GWR): number of bakeries, number of grocery stores, number of banks and number of pharmacies. Unfortunately, the socio-economic factors model could not meet GWR specifications due to the range in the values of the variables not varying sufficiently across space. The cumulative confirmed cases from July 3rd, 2020 were taken as the dependent variable. A R^2 of 0.903 was obtained considering 62 neighbours for every neighbourhood based on inverse distance. GWR residuals were tested for spatial autocorrelation by means of a Moran's I test, which resulted in an index of -0.002 ($p = 0.81$). This implies that there is significant evidence that residuals are randomly distributed, so the model is adequate.

Note that there is an improvement from the OLS results (global analysis) to those of the GWR (local analysis) since adjusted R^2 increased from 0.866 to 0.903, and also AICc statistics reduced from 935.99 to 916.19. On analysing GWR outputs, Fig. 4 illustrates the spatial distribution of local R^2 throughout neighbourhoods, which were allocated to five categories by natural breaks. Hotspots for reported cases found on July 3rd, 2020 were highlighted on the map as a way of better comprehending the behavior of the model in these more affected areas. Since values were generally high (at least 0.737), the performance of the models was highest in areas in which most of the hotspots are located.

Figure 5 reveals the impact of selected determinants on predicting cases across neighbourhoods. Categories were built by natural breaks in the range of the coefficients of the variables obtained from the GWR results. Regarding the same hotspots for cases previously mentioned, note that the Southern hotspots were more impacted by the presence of bakeries and pharmacies, while banks and grocery stores had a negative influence on predicting cases. The Western cluster is best described by local banks, grocery stores and pharmacies. The strongest influence on the incidence of disease in the Northern hotspot was from banks and grocery stores.

4.2.3 Analysis of the factors of socio-economic and essential services

New assessments, by means of regression analysis, were made based on a final set of 21 determinants. As this was conducted beforehand for separate data-sets, at first the same ten dates were considered to obtain the total number of reported cases of Covid-19 and to test them as response variables using multiple regression models based on the Ordinary Least Squares method. A smaller and significant set of explanatory variables was identified for each day by excluding correlated variables and using the stepwise method as applied previously. Findings were summarised in Table 2 and, although determinants were analysed together, the final sets were shown separately to clarify patterns.

Table 2
The performance of OLS models over time using combined datasets of determinants

Date	GM*	Essential services	Socioeconomic factors	Adj. R ²
April 16th	1	Bakeries, grocery stores, lotteries, bus terminals	Owned home, income	0.8117
April 23rd	1; 2	Bakeries, grocery stores, banks	Age 0 to 9, owned home, sewage system	0.8159
May 3rd	1; 2	Bakeries, grocery stores	Age 0 to 9, income	0.8331
May 12nd	1; 2	Bakeries, grocery stores, lotteries, bus terminals	Age 0 to 9, owned home, income, literacy	0.8709
May 19th	1; 2; 3	Bakeries, grocery stores, lotteries	Age 0 to 9, income	0.8958
May 27th	1; 2; 3	Bakeries, grocery stores, lotteries	Age 0 to 9, income	0.9100
June 3rd	2; 4	Bakeries, grocery stores, lotteries	Age 0 to 9, income	0.9162
June 12nd	2; 4	Bakeries, grocery stores, lotteries	Age 0 to 9, owned home, literacy	0.9215
June 24th	2; 4; 5; 6	Bakeries, grocery stores, lotteries	Age 0 to 9, owned home, literacy	0.9255
July 3rd	2; 4; 5; 6	Bakeries, grocery stores, pharmacies	Age 0 to 9, income	0.9260
* Government measures (GM) in 2020: <div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> 1. Closing of non-essential commercial activities 2. Mandatory use of masks 3. Lockdown </div> <div style="width: 45%;"> 4. Reopening of building supply stores 5. Reopening of beauty salons and suburban retailers 6. Reopening of malls and places of worship </div> </div>				

Every designed model was found significant ($p < 2.2E-16$) for forecasting Covid-19 cases, and all determinants contributed positively to their predictions, except for bus terminals. An increasing tendency until early June for the adjusted determination coefficient R^2 followed by a stabilisation pattern was verified as happened previously for separate databases. From late April, the final subset of determinants could be split between commercial facilities and residents' socio-economic attributes. Some of them are frequently repeated, such as bakeries, grocery stores, income and people from 0 to 9 years old, which indicates their importance for predicting cases.

Data from the last day explored, July 3rd, 2020, were kept for a subsequent evaluation using a spatial regression approach by means of GWR. So the following set of relevant explanatory variables was considered for predicting reported cases: number of grocery stores, number of pharmacies, number of bakeries, the average income of residents and the total number of residents aged 0 to 9 years old. Results indicated a R^2 of 0.960 considering 54 neighbours for every neighbourhood based on inverse distance, which represents an improvement on the essential services-only analysis conducted for separate databases. The statistical significance of this analysis was validated by a Moran's I test applied to GWR residuals, since an index of -0.028 ($p = 0.63$) implied they were randomly distributed.

As noted for the analysis of separate databases, an improvement occurred from OLS to GWR results since adjusted R^2 increased from 0.926 to 0.944, and also AICc statistics reduced from 881.42 to 873.78. GWR concerns a local prediction to elucidate spatial variations all over the region of interest²⁸, so the distribution of local R^2 in each Recife neighbourhood is illustrated in Fig. 6. SWith a view to initiating a more in-depth exploration of areas where Covid-19 infections are concentrated, hotspots for cases identified on July 3rd, 2020 were highlighted. Values were found remarkably high since the minimum one explains 82,8% of reported cases, mainly where hotspots were detected.

Again a further exploration was made of the July 3rd, 2020 data, but now seeking to clarify how every relevant contribution of a determinant to spatial regression modelling could influence a prediction of the number of cases. Figure 7 classifies their coefficients on regression local equations into five categories. When the previously mentioned hotspots are analysed, it is noted that the Southern ones were impacted most by the presence of bakeries and their residents' average income. The Western cluster is best described by local pharmacies and grocery stores. The strongest influence on the incidence of disease in the Northern hotspot came from bakeries, pharmacies and people aged from 0 to 9.

5. Discussion

The statistical analysis of Covid-19 cases contributes to elucidating the factors which delay the propagation of the disease. Thus, public governmental entities can react in a preventative way by paying particular attention to gathering data from susceptible neighbourhoods with a view to understanding the paths of transmission and to avoiding the merging of new focus zones.

Recife started applying deterrent measures immediately after the first patients were confirmed in order to reduce contagion, thereby preventing a collapse in the provision of hospital care. Although the infection curve has not flattened as expected, the public authorities' initiative to tighten the quarantine has had convincing results, since the peak of contamination was reached when quarantine was still in force since when the tendency has been for the number of new cases to fall and stabilise.

It seems clear that the propagation of a disease is connected to a spatiotemporal perspective. Building knowledge from local characteristics becomes an important aspect to consider when constructing models since such characteristics indicate more accurately how variables influence infections. This spurs

the need to explore the characteristics of neighbourhoods in Recife as a way of finding differential patterns, which help to implement effective policies to combat and mitigate disease rates for specific places.

As to reported cases, hotspots were first verified in a wealthy and densely-populated neighbourhood. In contrast, since then, other hotspots with worse socio-economic conditions have emerged. This situation could have happened because people in the Boa Viagem hotspot had the resources to travel more frequently to other countries and/or to the Brazilian metropolises of Rio and São Paulo and thus were the first to be infected and on their return to Recife they quickly spread the disease to those around them. However, people who live in less privileged places consequently have less infrastructure in their neighbourhoods and homes which would enable them to meet the requirements for following advice on social isolation and personal care. This includes not having enough money to buy preventive health supplies, a lack of constant access to piped water and not having the option of working from home, which is aggravated by Covid-19 social impacts³⁴.

Evaluation of the determinant factors of Covid-19 cases according to a database helped distinguish the influence of characteristics of the population and people's access to public services, and the flow of people to essential services facilities where concentrations are likely to be formed. Results indicate that the presence of banks is negatively related to the variability of cases, even though these places have led to long queues being formed so that people can receive emergency financial aid from the government ³⁵.

Regression analysis of the set of relevant socio-economic factors found that the number of residents per household, rate of literate people and homes with access to piped water have a negative impact on predicting cases in Recife. It can be interpreted that the virus is likely to spread in public places, not necessarily at home, but further analysis about sanitary conditions also have to be considered. Using the variables of literacy and access to piped water emphasises how socio-economic inequalities may affect the spread of coronavirus since the areas that tend to report the most cases are those in which the population tends to be less informed by trustworthy sources about good habits to avoid the virus and people have fewer resources to keep their environment properly clean.

The average income of residents was statistically significant and positively related to the prediction of cases as found by ²⁰. When only socio-economic factors database was considered, income was only highlighted in the ascending phase of the infections curve, which may indicate a tendency for wealthier regions to be the most affected at that time; regarding both databases, it became more frequent. However, it was also verified that areas with the least average income have had higher case-fatality rates, which likely correlate to governmental infrastructure support and access to the public health system.

The age groups highlighted by means of OLS reveal that places with a large number of children (between 0 and 9 years old) and seniors (over 60 years old) tend to present the largest number of reported cases. Studies affirm that most asymptomatic cases of Covid-19 are verified in children ³⁶, so there is a higher chance of their not being submitted to tests. Moreover, schools and daycare centers have been closed in

Recife since mid-March ³¹, which stimulates those younger groups to stay at home and then it is likely that they will spread the virus to their relatives. But recent studies from China show that children have a lower incidence of coronavirus and are less prone than other groups to being infected by it ³⁷, and so a more in-depth exploration needs to be carried in order to recognise if this pattern also happens in Brazil with regard to interventions and the population structure, even though our results imply the opposite. On the other hand, there should be a focus on the elderly since ³⁸ specified that an increase in coronavirus infection among elderly people had a direct correlation with the risk of infections among other age groups. Therefore, tightening social distancing for the elderly and other measures, such as analyzing spatial accessibility and healthcare resources³⁹, to reduce the risks they face could positively affect the whole of society.

6. Conclusion

This study has significant importance for supporting actions to mitigate the spread of Covid-19 in Recife since the objective is to explore the socio-economic characteristics and the flow of people during this outbreak and how they impact, negatively or positively, rates of infection and unfortunately of deaths.

By using correlation analysis and multiple regressions, it is possible to understand the influence of the age, income and level of education of the population on the behaviour of the virus across neighbourhoods. It sets precedents in a more in-depth structural problem, social inequality, since the lowest case-fatality rate due to Covid-19 was found in areas in which the population has the best living conditions. It has been observed that, although some wealthy neighbourhoods have presented a high number of cases, they have shown a low case-fatality rate, which might be explained by their having access to private health care of a better quality and the full range of public services.

Given that models are statistically significant, they are useful tools not only to reveal the most important factors in the spread of Covid-19, but also to aid re-modelling urban planning by giving greater emphasis to social problems and how people flow through and between public spaces. Therefore, broader positive impacts could be felt by the local population on other issues such as violence which are not directly related to disease.

Further analysis could explore spatial units that are smaller than neighbourhoods, thereby seeking to capture variations in the behaviour of the disease in order to better understand the pandemic context and take appropriate actions, which could include distinguishing between mild and severe cases. More recent data should be used to represent socio-economic characteristics. As the last Brazilian Census occurred in 2010, we assumed constant patterns in the years since them to the date of our study. Obtaining the previous health status of infected people towards comorbidities could help explain local susceptibility to severe forms of Covid-19. The contribution of this paper is our model but for it to have full effect in these difficult times which may see a resurgence of the pandemic, it is important that it comes to the attention of the appropriate authorities and that they adapt it to their local circumstances as suggested above.

Declarations

ETHICS APPROVAL AND CONSENT TO PARTICIPATE:

"Not applicable"

CONSENT FOR PUBLICATION:

"Not applicable"

AVAILABILITY OF DATA AND MATERIALS:

The datasets analysed during the current study are available in the SEPLAG (Secretaria de Planejamento e Gestão de Pernambuco) repository, https://dados.seplag.pe.gov.br/apps/corona_dados.html.

COMPETING INTERESTS:

"The authors declare that they have no competing interests"

FUNDING:

This research was partially support by the Universidade Federal de Pernambuco (UFPE- 06/2020), and by the Conselho Nacional de Desenvolvimento Cientifico e Tecnológico (CNPq).

AUTHORS' CONTRIBUTIONS:

All researchers contributed significantly to the study. Dr Caroline Mota designed the study, drafted the paper, interpreted the analysis results and revised the paper. Mr Arthur de Souza organised the databases, analysed the data and made the figures and tables. Ms Amanda Rosa reviewed the literature, analysed the data and made the figures and tables. Dr Ciro Figueiredo reviewed the literature, organised the databases and revised the paper. Dr Ana Lucia Candeias interpreted the analysis results and revised the paper.

ACKNOWLEDGEMENTS:

Authors acknowledge the financial support received by the UFPE, the SEPLAG for making the data available in their repository and Roderick Kay for his professional support in language editing services.

References

1. Singhal T. A Review of Coronavirus Disease-2019 (COVID-19). *Indian J Pediatr.* 2020;87(4):281-286. doi:10.1007/s12098-020-03263-6
2. Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N*

Loading [MathJax]/jax/output/CommonHTML/jax.js

3. Chinazzi M, Davis JT, Ajelli M, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* (80-). 2020;368(6489):395-400.
doi:10.1126/science.aba9757
4. Li Q, Guan X, Wu P, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *N Engl J Med*. 2020.
5. Adhikari SP, Meng S, Wu YJ, et al. Epidemiology, causes, clinical manifestation and diagnosis, prevention and control of coronavirus disease (COVID-19) during the early outbreak period: A scoping review. *Infect Dis Poverty*. 2020;9(1). doi:10.1186/s40249-020-00646-x
6. Zhao S, Chen H. Modeling the epidemic dynamics and control of COVID-19 outbreak in China. *Quant Biol*. 2020;8(1):11-19. doi:10.1007/s40484-020-0199-0
7. Singh RK, Rani M, Bhagavathula AS, et al. Prediction of the COVID-19 pandemic for the top 15 affected countries: Advanced autoregressive integrated moving average (ARIMA) model. *JMIR public Heal Surveill*. 2020;6(2):e19115.
8. Linka K, Peirlinck M, Sahli Costabal F, Kuhl E. Outbreak dynamics of COVID-19 in Europe and the effect of travel restrictions. *Comput Methods Biomech Biomed Engin*. 2020:1-8.
doi:10.1080/10255842.2020.1759560
9. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis*. 2020;20(5):533-534.
10. Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S. Helping doctors and patients make sense of health statistics. *Psychol Sci public Interes*. 2007;8(2):53-96.
11. Rajagopal K, Hasanzadeh N, Parastesh F, Hamarash II, Jafari S, Hussain I. A fractional-order model for the novel coronavirus (COVID-19) outbreak. *Nonlinear Dyn*. 2020:1-8.
12. Ndairou F, Area I, Nieto JJ, Torres DFM. Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan. *Chaos, Solitons & Fractals*. 2020:109846.
13. Mohamadou Y, Halidou A, Kapen PT. A review of mathematical modeling, artificial intelligence and datasets used in the study, prediction and management of COVID-19. *Appl Intell*. 2020:1-13.
14. Reis RF, de Melo Quintela B, de Oliveira Campos J, et al. Characterization of the COVID-19 pandemic and the impact of uncertainties, mitigation strategies, and underreporting of cases in South Korea, Italy, and Brazil. *Chaos, Solitons & Fractals*. 2020:109888.
15. Pongkitivanichkul C, Samart D, Tangphati T, et al. Estimating the size of COVID-19 epidemic outbreak. *Phys Scr*. 2020.
16. Franch-Pardo I, Napoletano BM, Rosete-Verges F, Billa L. Spatial analysis and GIS in the study of COVID-19. A review. *Sci Total Environ*. 2020:140033.
17. Li H, Li H, Ding Z, et al. Spatial statistical analysis of Coronavirus Disease 2019 (Covid-19) in China. *Geospat Health*. 2020;15(1).
18. Huang R, Liu M, Ding Y. Spatial-temporal distribution of COVID-19 in China and its prediction: A data-driven modeling analysis. *J Infect Dev Ctries*. 2020;14(03):246-253.

19. Xiong Y, Wang Y, Chen F, Zhu M. Spatial Statistics and Influencing Factors of the COVID-19 Epidemic at Both Prefecture and County Levels in Hubei Province, China. *Int J Environ Res Public Health*. 2020;17(11):3903.
20. Mollalo A, Vahedi B, Rivera KM. GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. *Sci Total Environ*. 2020:138884.
21. Buckingham-Jeffery E, Morbey R, House T, Elliot AJ, Harcourt S, Smith GE. Correcting for day of the week and public holiday effects: improving a national daily syndromic surveillance service for detecting public health threats. *BMC Public Health*. 2017;17(1):1-9.
22. Kulldorff M, Huang L, Konty K. A scan statistic for continuous data based on the normal probability model. *Int J Health Geogr*. 2009;8(1):58. doi:10.1186/1476-072X-8-58
23. Francis SS, Enders C, Hyde R, et al. Spatial-Temporal Cluster Analysis of Childhood Cancer in California. *Epidemiology*. 2020:214-223. doi:10.1097/EDE.0000000000001121
24. Anselin L. Local indicators of spatial association - LISA. *Geogr Anal*. 1995;27(2):93-115.
25. IBGE. Censo 2010. <https://censo2010.ibge.gov.br/resultados.html>. Published 2016. Accessed February 21, 2018.
26. de Waal T, van Delden A, Scholtus S. Multi-source Statistics: Basic Situations and Methods. *Int Stat Rev*. 2020;88(1):203-228.
27. Vantaggi B. Statistical matching of multiple sources: A look through coherence. *Int J Approx Reason*. 2008;49(3):701-711.
28. Brunsdon C, Fotheringham AS, Charlton ME. Geographically weighted regression: a method for exploring spatial nonstationarity. *Geogr Anal*. 1996;28(4):281-298. doi:10.1111/j.1538-4632.1996.tb00936.x
29. IBGE. Cidades e Estados. 2020. <https://www.ibge.gov.br/cidades-e-estados/pe.html>. Accessed August 19, 2020.
30. SEPLAG. COVID-19 no Mundo, no Brasil e em Pernambuco. <https://dados.seplag.pe.gov.br/apps/corona.html#painel>. Published 2020. Accessed May 27, 2020.
31. PHS. Pernambuco contra a Covid-19. 2020. <https://www.pecontracoronavirus.pe.gov.br/>. Accessed July 8, 2020.
32. Verity R, Okell LC, Dorigatti I, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect Dis*. 2020;20:669-677.
33. Simsek S, Dag A, Tiaht T, Oztekin A. A Bayesian Belief Network-based probabilistic mechanism to determine patient no-show risk categories. *Omega*. June 2020:102296. doi:10.1016/j.omega.2020.102296
34. Bambra C, Riordan R, Ford J, Matthews F. The COVID-19 pandemic and health inequalities. *J Epidemiol Community Health*. June 2020:jech-2020-214401. doi:10.1136/jech-2020-214401
35. News N. Agências da Caixa voltam a registrar longas filas em todo o Brasil. 2020. <https://g1.globo.com/jornal-nacional/noticia/2020/07/24/agencias-da-caixa-voltam-a-registrar->

longas-filas-em-todo-o-brasil.ghhtml. Accessed August 6, 2020.

36. Kronbichler A, Kresse D, Yoon S, Lee KH, Effenberger M, Shin J Il. Asymptomatic patients as a source of COVID-19 infections: A systematic review and meta-analysis. *Int J Infect Dis.* 2020;98:180-186. doi:10.1016/j.ijid.2020.06.052
37. Lyu J, Miao T, Dong J, Cao R, Li Y, Chen Q. Reflection on lower rates of COVID-19 in children: Does childhood immunizations offer unexpected protection? *Med Hypotheses.* 2020;143:109842. doi.org/10.1016/j.mehy.2020.109842. Accessed August 21, 2020.
38. Yu X. Risk Interactions of Coronavirus Infection across Age Groups after the Peak of COVID-19 Epidemic. *Int J Environ Res Public Heal.* 2020;17(14):5246.
39. Jeon-Young Kang, Alexander Michels, Fangzheng Lyu, Shaohua Wang, Nelson Agbodo, Vincent L. Freeman and Shaowen Wang. Rapidly measuring spatial accessibility of COVID-19 healthcare resources: a case study of Illinois, USA *Int J Health Geogr (2020) 19:36.* doi.org/10.1186/s12942-020-00229-x

Figures

Figure 1

Covid-19 cases evolution in Recife

Figure 2

Spatial distribution of the cases in Recife. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 3

Spatial distribution of case-fatality in Recife. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 4

Performance of local R^2 across the neighbourhoods. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 5

Effects of determinants on prediction of Covid-19 cases. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 6

Performance of local R^2 across the neighbourhoods. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 7

Effects of determinants on prediction of Covid-19 cases. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.